Selection and replication of oligonucleotides in early evolution: towards long functional sequences and motifs

Alexandra Kühnlein

Dissertation

München 2021

Selection and replication of oligonucleotides in early evolution: towards long functional sequences and motifs

Dissertation

zur Erlangung des Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Alexandra Kühnlein
aus Erlangen

München, April 2021

Erstgutachter: Prof. Dieter Braun
Zweitgutachter: Prof. Hannes Mutschler
Eingereicht am: 13.04.2021
Tag der mündlichen Prüfung: 02.06.2021

Zusammenfassung

Das Leben ist ein statisches Nicht-Gleichgewicht. Um dieses Nicht-Gleichgewicht aufrechtzuerhalten, sind moderne biologische Prozesse auf eine hochkomplexe enzymatische Maschinerie angewiesen. Zur Entstehung des Lebens, in der damaligen Abwesenheit komplexer Biomoleküle, die an das moderne Leben erinnern, musste das Nicht-Gleichgewicht durch physikalische Mittel hergestellt und aufrechterhalten werden. Eine interessante Nicht-Gleichgewichts Umgebung, die auf der frühen Erde häufig vorkam, sind thermische Flüsse durch wassergefüllte Poren, wie zum Beispiel unter Wasser in vulkanischen Gesteinen oder in der Nähe von hydrothermalen Schloten.

Im ersten Teil dieser Arbeit wird die Entwicklung mikrofluidischer Modellsysteme für solch geheizete, wassergefüllte Poren dargelegt. In einem thermischen Gradienten erfahren Moleküle Konvektion und Thermophorese, was zu einer längen- und ladungsselektiven Anreicherung führt. Dies kann das Natrium-Magnesium-Verhältnis von geologisch plausiblen Lösungen umkehren und Ribozym-Aktivität ermöglichen. Durch die zusätzliche Einführung einer Gas-Wasser-Grenzfläche in eine beheizte Pore werden weitere Effekte wie Verdampfung, Rekondensation und Kapillarströmungen induziert. Diese können sequenzselektive Phasenübergänge von DNA und RNA auslösen und induzieren sowohl Fusion als auch Spaltung von Koazervaten.

Die *RNA-Welt* Hypothese geht davon aus, dass das Leben mit katalytisch aktiven RNA-Strängen, genannt Ribozyme, begann, die grundlegende biologische Funktionen ausführen konnten. Diese Ribozyme sind hochspezifische Sequenzen mit komplizierten Sekundärstrukturen. Allerdings waren Oligonukleotide, die durch eine präbiotische Polymerisationschemie erzeugt wurden, höchstwahrscheinlich zufällig in ihrer Sequenz. Da zufällige Oligonukleotid-Pools, alleine wegen des riesigen Sequenzraums, nur sehr selten funktionelle Motive enthalten ist bisher unklar, wie Ribozym-ähnliche Motive ausreichend angereichert wurden, damit eine *RNA-Welt* beginnen konnte.

Im zweiten Teil dieser Arbeit wird gezeigt, wie ein Wärmefluss über eine wassergefüllte Pore als präbiotischer Filter wirken kann. Er kann Sequenzen auf Basis ihres Phänotyps selektieren und reichert in zufälligen Oligonukleotid-Pools Sequenzen mit niedrigerer freier Energie an, die kompaktere Faltungen bilden können, ein Motiv, das mit einem größeren funktionellen Potenzial in Verbindung gebracht wurde. Dieser Effekt basiert auf der Änderung der thermophoretischen Mobilität, die mit dem räumlichen Kollaps kompakter Faltungen einhergeht.

Im dritten Teil wird ein Mechanismus zur frühen Replikation vorgestellt, der allein auf

Hybridisierung beruht. Neben der Selektion ist die Replikation eine weitere zentrale Eigenschaft des Lebens, die sich früh herausgebildet haben muss. Ein Pool von kreuz-komplementären DNA-Strängen, inspiriert von transfer RNA (tRNA), kann sich selbst zu einem Replikator zusammensetzen. Ein Templat-Strang kann in einer kreuz-katalytischen Reaktion durch thermische Oszillationen mit einer Genauigkeit von 85-90 % exponentiell vervielfältigt werden. Der vorgestellte Replikator ist unabhängig von einer spezifischen präbiotischen Chemie und zeigt stattdessen einen generellen, physikalischen Replikationsmechanismus auf. Darüber hinaus werfen die Ergebnisse die Frage auf, ob die hier gezeigte Verbindung von molekularer Replikation und tRNA eine engere evolutionäre Verbindung der beiden impliziert als bisher angenommen wurde.

Abstract

Life is a non-equilibrium steady state. To maintain this non-equilibrium, modern biology relies on a highly complex enzymatic machinery. At the emergence of life, in the absence of any complex biomolecules reminiscent of modern life, the non-equilibrium had to be provided and maintained by physical means. An interesting non-equilibrium setting that was abundant on the early Earth are heat fluxes across water-filled pores, which could for example have occurred in submerged volcanic rocks or close to hydrothermal vents.

In the first part of this thesis, the development of microfluidic model systems for such heated, water-filled pores is laid out. In a thermal gradient, molecules are exposed to convection and thermophoresis, which results in length- and charge-selective accumulation. This can reverse the sodium to magnesium ratio of geologically plausible ion solutions and enable ribozyme activity. By introducing a gas-water interface to a heated pore, evaporation, recondensation and capillary flows are additionally induced. This can trigger the sequence selective phase-transition of DNA and RNA and causes both fusion and fission of coacervates.

The *RNA world* hypothesis presumes that life started with catalytically active RNAs, called ribozymes, that were able to perform basic biological functions. Those ribozymes are highly specific sequences with intricate secondary structures. However, oligonucleotides generated by a prebiotic polymerization chemistry were most likely random in sequence. As random oligonucleotide pools are scarce in functional motifs, simply due to the large sequence space, it is yet unclear how ribozyme-like motifs were sufficiently enriched to launch an *RNA world*.

The second part of this thesis shows how a heat flow across a water-filled pore can act as a prebiotic filter. It can select sequences based on their phenotype and enriches sequences with lower free energy that form more compact folds, a motif associated with greater functional potential, in random oligonucleotide pools. This effect is based on the change in thermophoretic mobility due to the spatial collapse of compact folds.

In the third part, a mechanism for early replication solely based on hybridization is introduced. Besides selection, replication is another property central to life, that had to be established early on. A set of cross-complementary DNA strands, inspired from transfer RNA (tRNA), can self-assemble into a replicator. A template can be exponentially amplified in a cross-catalytic manner driven by temperature oscillations with a fidelity of 85-90 %. The presented replicator is independent of a specific prebiotic chemistry and instead shows a general physical replication mechanism. Furthermore, the results raise the intriguing question, if the shown link between molecular replication and tRNA implies a closer evolutionary connection of the two than previously assumed.

Contents

1. Introduction

The question of how life emerged and how we came to be has puzzled humankind for centuries. Although, science will never be able to give a definite answer to how life emerged historically it can gather convincing evidence for plausible pathways.

Presumably, life on Earth dates back 3.7-3.8 billion years [44, 125, 135]. Over the last decades, astrobiology has provided us with a better understanding of the chemical repertoire available on the early Earth. Basic organic compounds [72,190], amino acids [22] as well as ribose and other sugars [118] can be found in protoplanetary disks. Furthermore, meteorites and comets have been shown to contain basic organic molecules [23], amino acids, activated phosphates and nucleobases [20]. Using this inventory, the synthesis of organic polymer compounds such as ribonucleotides [150] and precursors for peptides and lipids [143] has recently been demonstrated under prebiotically plausible conditions.

When pursuing such a bottom-up approach in reconstructing life, especially the transition from 'dead' to 'living' matter is of interest. However, it already proves difficult to find a consensus on what exactly 'living' means [13]. The to date most compelling and widely-accepted definition was coined by Gerald Joyce and Carl Sagan and has also been adopted by NASA: '*Life is a self-sustaining chemical system capable of Darwinian evolution*'.

The requirement that life is self-sustaining implies, that it is able to harness energy from external sources to keep itself from away from equilibrium, which is tantamount to death [142]. To be able to do so, the environment itself must constantly be out of thermodynamic equilibrium. This provides boundary conditions for the circumstances under which life could have emerged. Many potentially suitable habitats for early life on Earth have been discussed. Famously, Darwin contemplated the possibility that life might have emerged in a 'warm little pond' [19]. Because the availability of cyclic evaporation and dehydration is essential for polymerization reactions, which require the exclusion of water [34, 131], the 'warm little pond' scenario still enjoys great popularity [146]. Another promising scenario are hydrothermal vents [113, 173]: large chimneys, created by fissures in the seafloor through which hydrothermal fluid discharges, often found in close proximity to volcanic activity. They can provide gradients in temperature [10], pH and ion concentration [12, 113]. Both scenarios exhibit unique physical non-equilibrium conditions. Naturally, it is difficult to pin down the exact scenario or location in which life might have emerged and each scenario has its advantages and disadvantages. A proposal consolidating the above mentioned warm little ponds and a hydrothermal landscape has been put forward [181] and would, for all we know, have been an excellent starting point for life.

The second integral part of the definition of life is the notion of Darwinian evolution, which dictates how we think about life today. Undergoing Darwinian evolution entails the ability to self-replicate and evolve through random mutation to attain a selective advantage in a given environment. This already plays out on a molecular level and governs early evolution. A famous example is known as Spiegelman's monster, where a replicase was used to reproduce and evolve a viral genome. In the absence of biological selection pressure, shorter strands that replicated faster had a selective advantage and thus out-competed longer strands that contain more information [122]. During the emergence of life, this loss of information had to be averted. Here, the loop to suitable environments closes: Physical non-equilibria, provided by the environment, can pose a selection pressure on evolving systems that direct evolution towards higher complexity.

Thermal gradients are such a non-equilibrium environment and were abundant on the early Earth [10]. They have previously been shown to promote prebiotically important processes by convective cycling and thermophoresis, which cause molecules to experience temperature cycling and lead to locally enhanced concentrations of dilute molecules. Thereby, a thermal non-equilibrium can enhance polymerization of oligonucleotides [116], enable length-selective replication [95] and provide a habitat for RNA-based replication [166].

In Chapter 2 of this work an improved experimental model system for a thermal non-equilibrium is introduced. In that connection, I also present results showing that a thermal gradient can trigger sequence-selective phase transition of oligonucleotides, induce fission and fusion of coacervates as well as bring geologically plausible ion solutions within the realm of operation of ribozymes.

The work presented in Chapters 3 and 4 focuses on the molecular origins of life and addresses two of the key aspects of Darwinian evolution: selection and replication. There, I want to leave the much discussed issue of prebiotically plausible polymerization largely aside and build on the premise that, at some point during the emergence of life, oligonucleotides of decent length were available. Keeping in mind the definition of life, two main questions arise:

(i) How were functional motifs selected from the presumably random pool of oligonucleotides, that had a vast sequence space, much too big to screen one by one?

(ii) Given a pool enriched in phenotypic motifs, under which conditions and through which mechanism can functional motifs achieve self-replication?

The first question is approached in Chapter 3. Random pools of DNA and RNA were exposed to a thermal non-equilibrium. Native PAGE analysis and Next-generation sequencing revealed that the thermal gradient is able to selectively enrich phenotypic motifs of lower free energy, that have been associated with larger functional potential. The observed effect is small but consistent and the prospect of amplifying the selection by refeeding is exciting.

The second question is addressed in Chapter 4. A set of cross-complementary oligonucleotides of one distinct phenotype, inspired by transfer RNA (tRNA), can selectively self-assemble into a replicator. Temperature oscillations then drive autonomous, exponential self-replication of binary information, encoded in the molecular assembly. Because the replication is based on hybridization only, it is independent of a specific prebiotic chemistry.

2. Microfluidic chambers: development, setups and physical effects

Summary

Microfluidic chambers represent an ideal experimental model system for hydrothermal pores, that are an excellent environment to kick-start Darwinian evolution at the emergence of life [3]. In this chapter, I will provide an overview of microfluidic chambers and emphasize related experimental results. To highlight the advancements of the current versions of microfluidic chambers, I will first recapitulate the outdated versions and then give an overview of the current state. I will then in detail describe the main types of microfluidic chambers, gas-water interface chambers and thermogravitational chambers and discuss the physical effects at play.

At the current stage, three different types of microfluidic chambers can be constructed in a controlled and reproducible manner:

- **Gas-water interface chambers.** In gas-water interface chambers, a temperature gradient ($\Delta T < 60$ K) is applied across a microfluidic chamber with gas inclusions. Typically, one chamber contains many gas-water interfaces, which allows experimental multi-plexing and increases the total reaction surface. Sequence-selective hydrogelation of oligonucleotides (Section 2.4.3) as well as coacervate fusion and fission (Section 2.4.4) can be observed in gas-water interface chambers. Details of design and assembly are shown in Figure 2.5. The physical effects relevant in gas-water interface chambers are described in Section 2.4.1 and illustrated in Figure 2.4.

- **Thermogravitational chambers.** In thermogravitational chambers, a temperature gradient ($\Delta T < 80$ K) is applied across a water-filled microfluidic chamber. Accumulation of salts (Section 2.5.4) as well as the selection of oligonucleotide motifs based on their phenotype (Chapter 3) can be observed. Details of design and assembly are shown in Figure 2.11. The physical effects relevant in thermogravitational chambers are described in Sections 2.5.1 and 2.5.2.

- **Incubation chambers.** In incubation chambers, no temperature gradient is applied. Instead, experiments (up to ten at a time) can be performed under isothermal conditions with full optical readout. Peltier elements are directly attached to the backside of

the chamber, which allows full thermal control and permits heating and cooling rates of up to 2 K/s. In essence, experiments in incubation chambers are comparable with experiments in thermocylers, but they have the advantage of full concurrent optical readout. Incubation chambers are used to monitor self-assembly and subsequent sedimentation for a tRNA-inspired replicator in Chapter 4. Details of design and assembly are shown in Figure 2.14.

2.1. PLA printed chambers

After moving away from capillary chambers with laser induced heating as described in [116, 129], the temperature gradient was going to be established perpendicular to the direction of gravity to naturally achieve convection and thus similar effects as observed previously. The temperature gradient was installed by using a waterbath (Julabo, Germany) to cool the traps from one side (back) and heater cartridges for 3D printers (12 V, 40 W) were used for heating from the other side (front). Full optical control of the chamber while maintaining good heat conductivity (42 W/mK) was ensured by using a sapphire as window (14x30 mm, 5 mm thick) into the chamber. The sapphire was equipped with four holes (diameter 1.8 mm), two at the top and two at the bottom (Figure 2.2c).

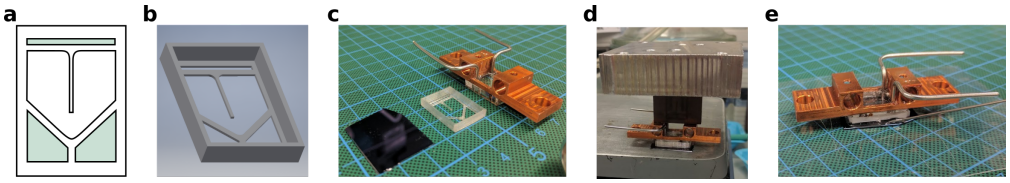

Figure 2.1: Chamber design and assembly of a PLA printed chamber. a, 2D design of the chamber geometry. **b,** 3D chamber geometry as displayed in Autodesk Inventor [84]. **c,** Components of the chamber assembly prior to melting. From left to right: silicon wafer, PLA printed chamber and already assembled top part consisting of a sapphire, metal tubing fit into the sapphire top holes and the copper bracket with holes for heating cartridges. **d,** Melting procedure of the PLA chambers. The components from panel c are sandwiched, put on a heat plate at $\sim$ 150 °C and weighed down. Glass beads are used as spacers during melting in the areas highlighted in green in panel a. **e,** Assembled chamber. The bottom holes of the sapphire are closed with PDMS or epoxy glue after melting. The assembled trap will be put on a copper back plate and is ready if tight.

3D printing of the chambers promises full flexibility in chamber design and quick iterations. The desired chamber design is drawn in 2D (Figure 2.1a) and extruded into a 3D object (Figure 2.1b) using Autodesk Inventor and exported as '.stl' file. The geometry of the chambers was 3D printed from polylactic acid (PLA), usually transparent or white PLA filament (filamentworld, Germany), using FELIX printers, generation 2 & 3 (FELIX printers, Netherlands). An example of a PLA printed chamber geometry is shown in Figure 2.1c. Metal tubings are polished, inserted and adhered to the top holes of the sapphire. The copper bracket is fit onto the sapphire and thermally connected using thermal conduction paste (10 W/mK, Keratherm® KP 92, KERAFOL, Germany). The 3D printed geometry is fit inbetween the sapphire and a silicon wafer, which constitutes the back of the microfluidic chamber. It is chemically inert and has good optical properties. The assembled sandwich is then placed on a heating plate at $\sim$ 150 °C and weighed down from the top (Figure 2.1d). Upon melting of PLA, the chamber melts onto wafer and sapphire and creates a permanent, air-tight connection. Glass beads (105-150 µm, Polysciences, Germany) are used to define the height during melting. The ready assembled chamber (Figure 2.1e) is then thermally

connected to a copper back plate using liquid metal (80 W/mK, Liquid Pro Flüssigmetall, Coollaboratory, Germany) and screwed down with Teflon screws to not create a thermal bridge.

This generation of microfluidic chambers was used by Keil *et al.* to show that a thermal gradient induces a pH gradient [90].

Problems with PLA printed chambers were mainly, but not limited to, poor printing results of the 3D extrusion printing method, insufficient height definition and high breakage rate of sapphires due to fine-grained dust or dirt in the sapphire holes in combination with the thermal expansion of the metal tubings.

2.2. Resin printed chambers

In the next major iteration, chambers were printed with a UV-curable resin (Photocentric 3D Daylight Resin, flexible, color: amber). A laminated LCD display, covered with resin, illuminated and thereby cured the inverted chamber design (Figure 2.2a) onto a silicon wafer (Figure 2.2b). The height of the chambers was defined by metal spacers, available in thicknesses of 50-250 µm, placed between display and silicon wafer. After the complete curing of the resin structure, sapphires were placed on top of the printed chamber and curable resin was again applied around the rim of the saphhire to close the gap between silicon wafer and sapphire. This made the chamber air- and water-tight (Figure 2.2c). Micofluidic access to the chamber was provided through silicone rubber tubings, which were fixated in the top holes of the sapphires using instant adhesive (Loctite, Henkel). The wafer was attached to the copper back plate and thermally connected using liquid metal. A copper bracket was thermally connected to the sapphire using thermal conduction paste and used to press the sapphire onto the copper back plate (Figure 2.2d). Again, heater cartridges were inserted into the preformed sockets of the copper bracket to provide heating from the front side.

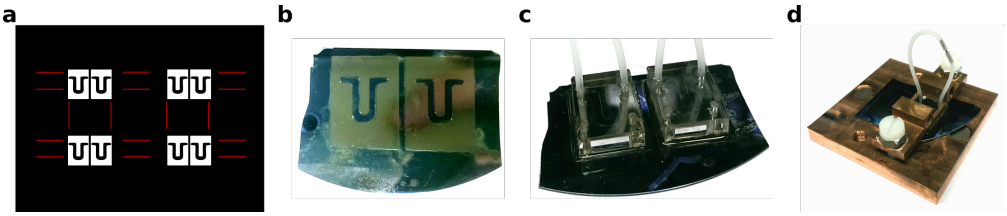

Figure 2.2: Chamber design and assembly of a resin printed chamber. **a,** Inverted chamber design as illuminated onto the LCD-screen. Everything displayed in white is cured onto the silicon wafer. The red lines indicate the position of the metal spacers (50-250 µm) for height definition. **b,** A printed chamber structure on a silicon wafer. **c,** Sapphires equipped with tubings in the top holes are placed on the wafer with the printed chamber structures. The rim around the sapphire and additional holes in the sapphire are closed with resin and the structure is cured again and subsequently tested for tightness. **d,** The ready assembled chamber screwed onto the copper back plate.

Problems with Resin printed chambers were mainly, but not limited to inadequate tightness at the sapphire rim and the in- and outlets, frequent breaking of the silicon wafer due to the necessary strain exerted by the copper frame, uneven temperature profile due to selective heating at only two spots and lack of chemical inertness.

2.3. FEP sheet chambers

In the next and current generation of chambers, the trap geometry is cut from an FEP (teflon) sheet. The additional advances compared to the resin printed chambers, which are described in the following, mainly apply to thermogravitational chambers. However, also when aiming for gas-water interfaces FEP sheets are now used to define the chamber geometry. Then, a simpler setup as described in Section 2.2 is used, only that the resin printed chamber geometry is substituted by a FEP printed chamber geometry.

Four main components (images are shown in Figure 2.3) were improved or redesigned during the last major iteration of chambers:

- **FEP sheet printed chamber geometry.** The chamber geometry was no longer 3D printed but instead cut from a chemically inert FEP sheet (Holscot, Netherlands) of desired thickness (25-1000 µm) using an industrial plotter (CE6000-40 Plus, Graphtec, Germany).

- **Heater.** The heater was redesigned to a new shape, which provides almost perfectly uniform heating of the top surface, while not disturbing the visual readout through the front sapphire. The new heater is shown in Figure 2.3e. Heater cartridges (24 V, 160 W) can be inserted into shafts in y-direction on the left and right side of the heater. A temperature sensor is directly connected to the heater for temperature regulation.

- **Sapphires.** The new chamber assembly contains two (instead of only one) sapphires. The silicon wafer was eliminated from the assembly and the two sapphires now enclose the FEP sheet from top and bottom and thus form the chamber. The sizing of the sapphire as well as the position of the holes on the sapphire was modified. The sapphire length was increased from 14x30 mm to 22x60 mm, which allows for better trapping selectivity in a thermogravitational setting (Section 2.5.2). The four holes in the sapphire are now all located in the upper part of the sapphire (Figure 2.3e). Generally, to maximize the heat transfer across the whole assembly, the sapphires should as slim as possible. However, in order to compensate for the grid-like shape of the heater (Figure 2.3e, Figure 2.12a), a thicker sapphire (2 mm) is chosen for the hot side to better distribute the heat generated by the heater. The sapphire on the cold side is slimmer (0.5 mm) and equipped with four holes for micofluidic access.

- **Tubing access.** Microfluidic access to the chamber is provided by FEP tubings (ID 0.18-1.0 mm, KAP 100.969, Techlab, Germany), which are screwed into the aluminum back plate using fittings and ferrules. Since the sample enters and leaves the chamber now

on the cold instead of the hot side, uncontrolled evaporation and bubble formation is reduced.

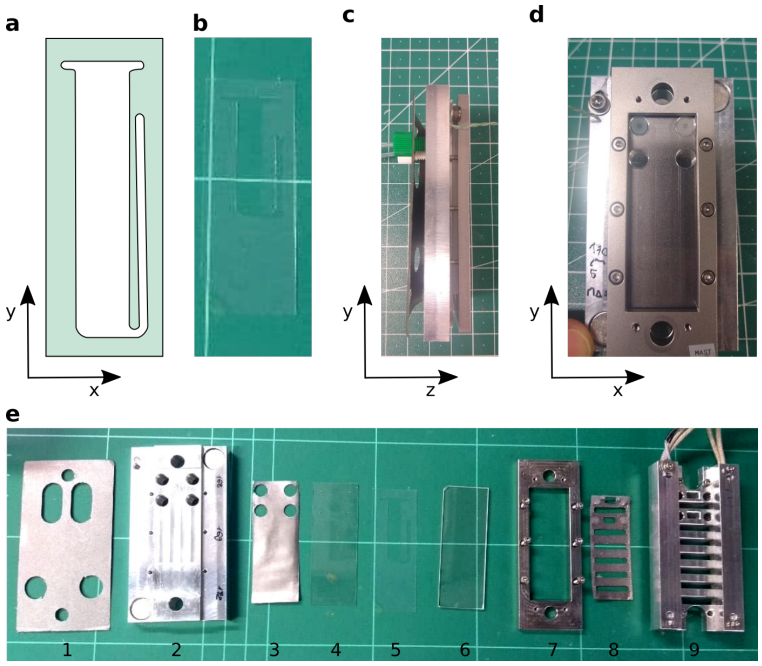

Figure 2.3: Chamber design and assembly of an FEP sheet chamber. **a,** Design of a chamber, with one inlet on the left and two outlets, top and bottom on the right. The plotter cuts along solid lines and the area of the FEP sheet is marked in green. **b,** A chamber geometry cut from a 170 µm FEP sheet. **c,** Lateral and **d,** frontal view of an assembled chamber. The in- and outlet tubings are connected from the back side and a temperature sensor (top left in panel d) is attached for temperature monitoring and regulation. **e,** Chamber components (for a thermogravitational chamber) in correct order from left to right prior to assembly. (1) Thick graphite foil for thermal connection to cryostat-cooled aluminum block, (2) aluminum back plate, (3) thin graphite foil, (4) 0.5 mm sapphire with holes, (5) cut 170 µm FEP sheet, (6) 2 mm sapphire, (7) steel frame, (8) thick graphite foil, (9) heater with attached temperature sensor (top left) and heating cartidges (inserted from top).

A typical trap geometry (for thermogravitational trapping) and an assembled FEP sheet chamber are displayed in Figure 2.3. The cut out FEP sheet (5), as displayed also in Figure 2.3b, is fit between the two sapphires of different thickness (4,6) and, using a thin graphite foil (3), (EYGS091203DP, 25 µm, 1600 W/mK, Panasonic, Japan) for thermal connection, put on top of the aluminum back plate (2). This sandwich is then fixed by a steel frame (7) using six torque controlled steel screws. Then, the heater (9) can be attached with an intermediate layer of thick graphite foil (8), (EYGS0811ZLGH, 200 µm, 400 W/mK, Panasonic, Japan) again using torque-controlled screws. Heater cartridges for 3D printing are used (24 V, 160 W), which are connected to a solid state relais and controlled by Arduino

boards. Cooling is provided by a waterbath (DYNEO DD, Julabo, Germany) or cryostat (Grant TXF200-R5, UK), that cools the aluminum block to which the aluminum back plate (2) is attached by magnets and torque-controlled teflon screws and thermally connected using a thick graphite foil (1). The indicated numbers relate to the annotations in Figure 2.3e.

The tubings are cut to the desired length such that they can accommodate the sample volume. For microfluidic connection, tubings are screwed to the back side of the aluminum back plate (Figure 2.3c) with ferrules (Ferrule VBM 100.632, Techlab, Germany) and fittings (Nut, Delrin, flangeless, VBM 100. 823, VBM 100.824, VBM 100.825, VBM 100.826, VBM 100.827, VBM 100.828, Techlab, Germany) additionally using connectors to connect multiple tubings (Verbinder zöllig, UP P-702-01, Techlab, Germany). After the tubings are attached, the chamber is flushed with low-viscosity flourinated oil (3M™Novec™7500 Engineered Fluid) to test for tightness and flush out gas inclusions.

Although, the height should be relatively well defined by the thickness of the FEP sheet, it is again measured after chamber assembly at three positions along the y-axis using a confocal micrometer (CL-3000 series with CL-P015, Keyence, Japan). This is done to guarantee homogeneous thickness of the chamber. If the thickness differs by more than approximately 5 % across the length of the chamber, it is reassembled.

2.4. Gas-water interface chambers

Gas-water interfaces in a thermal gradient represent an attractive non-equilibrium environment when considering processes involved in the emergence of life. Simple heat flows [127] as well as recondensation [1] are non-equilibrium processes found frequently on the early Earth. The potential relevance of bubbles in an early Earth setting has been hinted at before [104]. Gas-water interfaces lead to the adsorption of DNA [51] and lipid monolayers [6] and can aid the formation of peptide bonds [66]. Furthermore, it was recently demonstrated that DNA can be involved in the formation of rigid fibrous aggregates at gas-water interfaces [27]. In this section, the effects of gas-water interfaces are studied in connection with thermal gradients.

2.4.1. General mechanisms at heated gas-water interfaces

The phenomenon of a liquid droplet evaporating and depositing a characteristic pattern of dried molecules is known as the 'coffee-ring' effect [40]. During the evaporation of the liquid droplet, capillary flows increasingly replenish the material at the rim of the droplet until evaporation is complete. A filled gas-water interface chamber can be described as an inverted coffee ring effect. Because of the additionally applied temperature gradient across the gas-water interface, several physical effects are happening simultaneously [128].

- **Convection in the liquid phase.** Due to gravity and the temperature dependence of the density, the liquid phase experiences convection. Close to the interface, convective

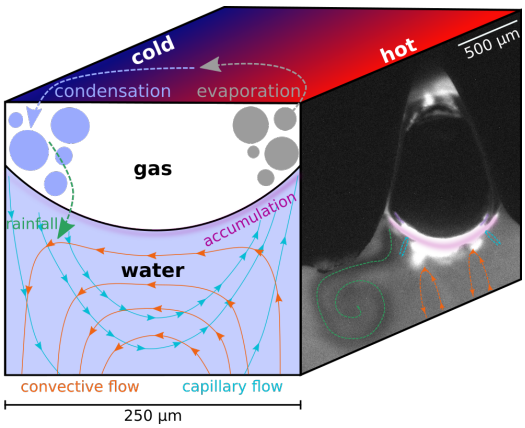

Figure 2.4: Sketch of the relevant effects at a gas-water interface and the corresponding fluorescence image. The sketch shows effects relevant in gas-water interface chambers. The temperature gradient drives a constant cycle of evaporation on the hot side and condensation on the cold side. The drops condensating on the cold side can fall back into the water phase (rainfall, green). This leads to periodic rise and fall of the water level inducing drying-rehydration cycles at the interface. In solution, the temperature gradient causes convective flow (from hot to cold, orange) as well as capillary flow (from cold to hot, cyan). Molecules accumulate at the interface, due to the capillary flow especially at the hot side. These effects are also indicated in the fluorescence image on the right face of the cube, showing 36 nt DNA at 10 µM in 1xPBS.

flow moves material from the hot to the cold side, which is indicated by orange arrows in Figure 2.4.

- **Evaporation-recondensation cycles.** At the hot chamber wall, water from the liquid phase evaporates, enters the gas phase and recondensates at the cold chamber wall. As more and more water gathers there, occasional 'rainfall' occurs, i.e. a recondensation droplet falls back into solution. Those droplets contain pure water and momentarily cause a drop in concentration (dark trace in Figure 2.4, fluorescence image). It has previously been shown that 'rainfall' events provide a local perturbation in salt concentration that lead to melting of oligonucleotides [80]. Furthermore, the evaporation-recondensation cycles cause fluctuations in the water level, which can lead to periodic drying and rehydration of molecules specifically at the rim of the interface (dark purple in fluorescence image in Figure 2.4).

- **Capillary flow at the interface.** Capillary flows are pushing molecules along the meniscus towards the hot side, where evaporation takes place. In Figure 2.4 the capillary flow is indicated by blue arrows. The capillary flow is stronger than the convective flow and flow velocities are higher close to the hot chamber wall [128].

- **Marangoni flow at the interface.** Marangoni flows occur due to the temperature and concentration dependence of surface tension. The Marangoni flow has the same direction as convection but is typically much weaker. In systems with modified surface

properties, Marangoni flow can be strong and suppress the capillary flow towards the hot side [78].

Together, those effects lead to an accumulation of molecules at the gas-water interface, as clearly visible in the fluorescence image of Figure 2.4. Detailed numeric simulations and bead-tracking measurements to determine the flow patterns can be found in [128].

2.4.2. Setup

The experimental setup for gas-water interface chambers is illustrated in Figure 2.5. The chamber geometry as exemplary shown in Figure 2.5a, can be adapted according to specific experimental requirements. If the design contains triangular cavities (Figure 2.5a), filling of the chamber creates gas inclusions. In the subsequently shown experiments, the chamber geometries are either printed with UV-curable resin or cut from an FEP sheet. In both cases, the z-thickness was 250 µm. The fluorescence images are taken from the hot side, showing an image plane perpendicular to the direction of the temperature gradient as indicated in Figure 2.4.

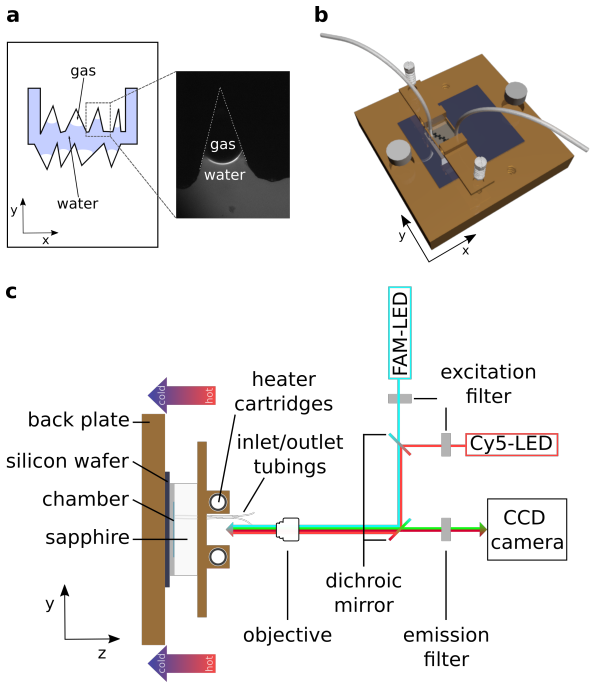

Figure 2.5: Overview of the experimental setup of gas-water interface chambers. **a,** Typical design of a gas-water interface chamber. The geometry is either printed with UV-curable resin or cut from an FEP sheet. The triangular cavities result in gas inclusions. The thickness in z-dimension typically 250 µm. **b,** Visualization of an assembled gas-water interface chamber. **c,** Lateral sketch of the gas-water interface chamber including the optical setup which allows fluorescence imaging and is equipped with two optical channels.

2.4.3. Formation of hydrogels from DNA and RNA at gas-water interfaces

It was previously demonstrated that a temperature gradient across a water-filled microfluidic compartment can evoke a sequence-selective phase transition to hydrogels [129]. Expanding the experimental system to temperature gradients including gas-water interfaces revealed that oligonucleotides experience strong accumulation at the interface as described in Section 2.4.1.

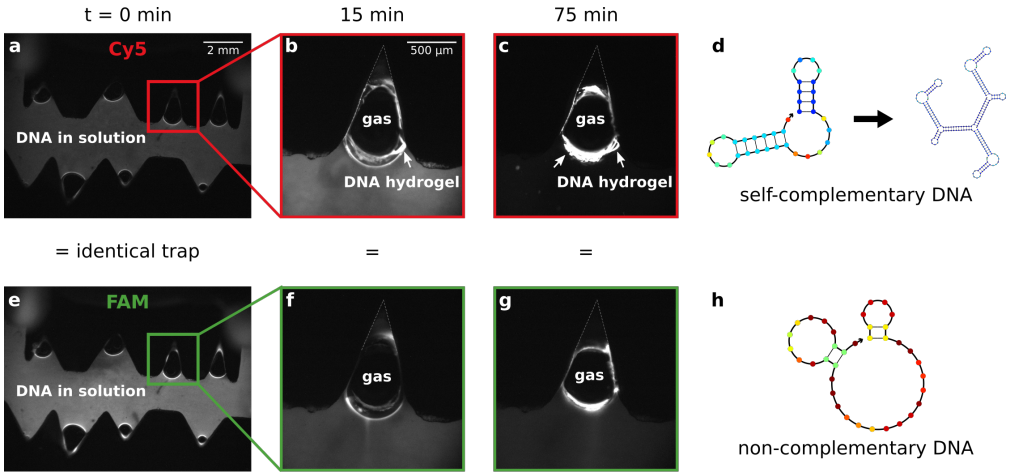

Figure 2.6: Sequence selective hydrogel formation of DNA. **a-c,** Self-complementary (Cy5-labeled, red) and **e-g,** non-complementary (FAM-labeled, green) DNA single strands (36 nt) were mixed at 10 µM in 1x PBS and flushed into a gas-water interface chamber. Fluorescence microscopy showed that after switching on a temperature gradient ($T_{cold} = 10\,°C$, $T_{hot} = 40\,°C$, $\Delta T = 30\,K$) the DNA which is initially in solution (a,e) starts to accumulate at the interface (b,f). Only for the self-complementary sequence this accumulation also leads to the formation of DNA hydrogels (b,c). The hydrogel forms within minutes and after approximately one hour all DNA is adsorbed at the interface. The initially higher fluorescence signal at the gas-water interface is due to slight surface adsorption effects. **d, h,** Predicted equilibrium secondary structures of both strands calculated using NUPACK [199]. The self-complementary strand (a-d) can branch out and form a network.

When self-complementary and non-complementary strands were present in the same gas-water interface chamber, fluorescence microscopy revealed, as shown in Figure 2.6, that both

This section was published by Morasch *et al.* [130] in *Nature chemistry* under the title "Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules" and is adapted and reprinted in part with permission from Nature publishing group.
Authors: M. Morasch, J. Liu, C.F. Dirscherl, A. Ianeselli, **A. Kühnlein**, K. Le Vay, Ph. Schwintek, S. Islam, M.K. Corpinot, B. Scheu, D.B. Dingwell, P. Schwille, H. Mutschler, M.W. Powner, C.B. Mast, D. Braun.
Contributions: M.M., J.L., C.F.D., A.K., A.I., and Ph.S. performed the experiments, M.M., J.L., K.L.V., S.I., B.S., D.B.D., H.M., P.S., M.W.P., C.B.M., and D.B. conceived and designed the experiments, M.M., J.L., K.L.V., S.I., M.K.C., H.M., M.W.P., and D.B. analyzed the data, M.M., J.L., and D.B. wrote the paper.

strands accumulate at the interface. The strands were present at a concentration of 10 µM each in 1x PBS and carried a covalently attached Cy5- and FAM-label respectively. The phase transition to a hydrogel occurred selectively only for self-complementary sequences that can form larger structures by hybridization (Figure 2.6, Cy5 (a-c) vs. FAM (e-g) channel). This is supported by the observation that the hydrogel melted if it was caught in a convection stream and transported to the hot side [130]. If the hydrogel grew large enough, it broke apart, detached partially from the gas-water interface and floated around in solution (slightly visible in Figure 2.6c).

Figure 2.7 shows that hydrogel formation occurs across a wide range of parameters. The self-complementary sequence from Figure 2.6 also leads to hydrogel formation in RNA (Figure 2.7a). Furthermore, DNA strands of various length and sequence also exhibit hydrogel forming properties (Figure 2.7b-d). Notably, even when parts of the sequence are random (Figure 2.7d) hydrogel formation at the interface can be observed.

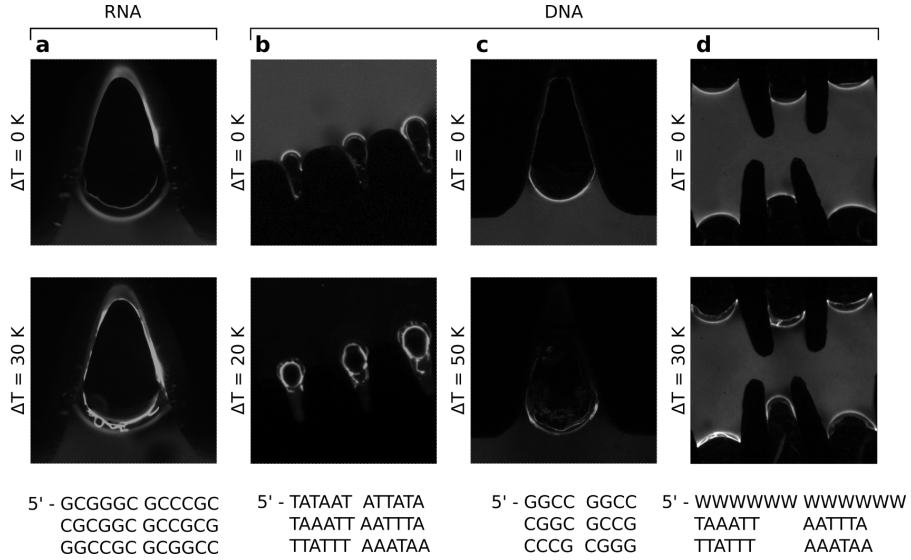

Figure 2.7: Hydrogel formation of different DNA and RNA sequences. Hydrogel formation in a temperature gradient (ΔT indicated for each experiment) is a prevalent property of at least partially self-complementary oligomers. The chamber design was different in every experiment, which highlights the universality of gas-water interfaces and its effects. **a,** Fully self-complementary RNA single strands (36 nt) at 10 µM in 1x PBS, identical sequence as shown in Figure 2.6a-c for DNA. **b,** Fully self-complementary DNA single strands (36 nt) with AT only at 10 µM in 1x PBS. **c,** Fully self-complementary DNA single strands (24 nt) with GC only at 10 µM in 10 mM $MgCl_2$ and 1x PBS. **d,** Partially (2/3) self-complementary, partially (1/3) random DNA single strands (36 nt) at 10 µM and 1x PBS. In the sequence, W indicates a random base {A,T}, i.e. the sequence space is $2^{12} = 4096$ different sequences.

In this study by Morasch *et al.* the effects of gas-water interfaces on prebiotic molecules were examined in a holistic approach. In addition to the sequence selective hydrogelation of oligonucleotides, it could be demonstrated that gas-water interfaces direct (i) accumulation of functional oligonucleotides increasing their catalytic activity, (ii) phosphorylation of

monomers due to dry-wet cycling at the interface, (iii) accumulation and subsequent crystallization of RNA precursors (RAO), (iv) encapsulation of oligonucleotides in vesicles.

2.4.4. Coacervate fusion and fission at gas-water interfaces

Compartmentalization was one of the key steps during the emergence of life. Only in a compartmentalized system an evolutionary advantage can be maintained. The role of liquid-liquid phase transitions in biological systems has recently been acknowledged [16] and gave them a renaissance after their initial recognition almost one hundred years ago [36, 138]. Coacervates are a liquid-liquid phase transition [149] and have been proposed as potential protocells and a possible pathway to early compartmentalization [100, 148, 149]. They are liquid microdroplets that form spontaneously when two oppositely charged polymers like polypeptides and oligonucleotides are mixed. After coacervation, the initial suspension is decomposed into a polymer-dense phase (coacervates) and a dilute supernatant phase. Formation, organization and properties of liquid-liquid phase transitions and coacervates specifically have recently been reviewed [4, 9, 133]. Due to their ability to length-selectively recruit RNA [48], spatially localize and accumulate molecules [93, 107, 132] and support biochemical reactions [148, 184] coacervates are especially interesting in an origin-of-life context.

After we had observed the general relevance of gas-water interfaces for various prebiotic systems [130], as described in Section 2.4.3, we became interested in the effects of a heated gas-water interface on coacervates. A suspension of CM-Dex:PDDA (Carboxymethyl-Dextran:Polydiallyldimethylammonium chloride) coacervates was mixed and inserted into a gas-water interface chamber. Polymers were present at 20 mM total concentration in a 6:1 ratio in 0.1 M Na^+ bicine buffer, pH 8.5. The phase separation between the dense, fluorescently labeled polymer-rich phase and the dilute phase was clearly visible as shown in Figure 2.8a. After switching on the thermal gradient ($\Delta T = 30$ K), the coacervate droplets accumulated rapidly (within two minutes) at the interface. Initially, they were around 10 µm in size but experienced strong growth and reached a size of more than 150 µm within ten minutes after switching on the temperature gradient (Figure 2.8c). A precise size distribution over time for different coacervate compositions can be found in Figure 3e, of [81]. Figure 2.8 clearly shows that a heated gas-water interface prompts accumulation and catalyzes the fusion of coacervates.

Interestingly, apart from fusion, fission events of coacervates were also be observed at the heated gas-water interface. This is due to the different flows directly at the interface (Figure 2.4), which generate shear forces within the coacervates tearing them apart. Moreover, fission or fragmentation also happened as consequence of a rainfall event, when condensed

---

This section has been submitted for publication by Ianeselli *et al.* [81] to *Nature chemistry* under the title "Coacervate protocells assembled at the gas-water interface of heated rock pores coalesce, divide and select RNA" and is currently in resubmission.
Authors: A. Ianeselli, D. Tetiker, J.A.C. Stein, **A. Kühnlein**, C.B. Mast, D. Braun and T-Y D. Tang.
Contributions: A.I. and D.T. performed the experiments and analyzed the data. A.I., D.T., J.A.C.S., A.K., C. M., D.B. and T-Y D.T. conceived and designed the experiments. A.I, D.T., D.B and T-Y D.T wrote the manuscript.

water falls back into solution and divides a coacervate at the interface.

Sole convection as well as an isothermal gas-water interface proved insufficient to induce fusion or fission of coacervates, indicating that the observed features are a unique property of coacervates at heated gas-water interfaces. These results allude to the important role coacervates might have played as an early protocell. The ability to grow and subsequently divide into two daughter 'cells' is particularly relevant, because a categorical property of modern biological compartments, i.e. cells, is their autonomous growth and division, which permits the propagation of information to daughter cells.

Due to its prebiotic relevance, RNA was also studied as a coacervate component additionally to the modified sugars (CM-Dex,-), peptides (Plys,+) and synthetic polymers (PDDA,+). A 51 nt single strand RNA was introduced into the system as a third component in addition to CM-Dex and Plys during coacervation. This encouraged competition with CM-Dex for the incorporation of the negatively charged polymer. Fluorescence microscopy showed initial co-localization of RNA and out-competition of CM-Dex within the coacervate droplets due to higher affinity of RNA to Plys. At the heated gas-water interface, the RNA-rich coacervate droplets were accumulated and enriched in CM-Dex. The gas-water interface chamber can thus maintain and select multiple populations of different phenotype and sustain energetically unfavorable conformations.

The study by Ianeselli *et al.* has important implications regarding the relevance of thermal fluxes and gas inclusions for an evolutionary scenario including coacervate droplets as protocells.

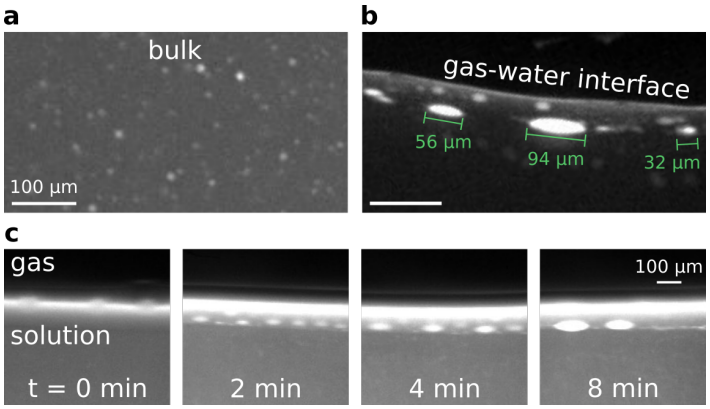

Figure 2.8: Coacervate accumulation and fusion at gas-water interfaces. a, Bulk fluorescence microscopy image directly after loading the chamber with coacervate suspension made from 20 mM CM-Dex:PDDA (6:1 ratio) in 0.1 M Na^+ bicine buffer, pH 8.5. Initially, the coacervates were smaller than 10 µm in size. **b,** In a thermal gradient ($\Delta T = 30$ K) the coacervates accumulated at the gas-water interface and started growing, resulting in differently sized coacervates. **c,** Time-lapse of fluorescence microscopy images after switching on the thermal gradient ($\Delta T = 30$ K). Within only 8 minutes almost all initial coacervates at the interface have fused to larger coacervates. Figure adapted from [81].

2.5. Thermogravitational chambers

2.5.1. Thermophoresis

Thermophoresis is the directed movement of molecules along a temperature gradient. Their drift velocity v depends linearly on the temperature gradient ∇T and is characterized by the thermophoretic mobility D_T.

$$v = -D_T \nabla T \tag{2.1}$$

In equilibrium, the concentration gradient created by the thermophoretic movement is counteracted by diffusion. Therefore, the flow density is given by

$$j = j_{D_T} + j_D = -D_T \nabla T(x,y) \cdot c(x,y,t) - D\nabla c(x,y,t), \tag{2.2}$$

with the diffusion coefficient D and concentration gradient ∇c. The Soret coefficient $S_T = D_T/D$ gives a measure for the relative strength of thermophoresis [17, 49]. In steady state ($j = 0$), the concentration of a molecule with initial concentration c_0 in a temperature gradient between T and T_0, with cold temperature T_0 is then

$$\frac{c}{c_0} = \exp[S_T \cdot (T - T_0)]. \tag{2.3}$$

The precise mechanism of thermophoresis is still subject of active research. Currently, it is believed that the effect is mainly driven by the temperature dependence of the electrical double layer of molecules. In addition, in salt solution, thermophoresis of the small ions themselves establishes an electrical field that causes electrophoretic movement of other charged species, such as oligonucleotides.

2.5.2. Thermogravitational trapping

The mechanism that leads to accumulation of molecules was first described in the early 20th century [29]: the superposition of thermophoresis and convection. While Clusius and Dickel described this for gases, it also works in a water filled pore as shown in Figure 2.9. Dissolved biomolecules such as DNA or RNA typically have a positive S_T, which causes them to move from the hot to the cold side of a pore (horizontal arrows). At the same time, the thermal expansion of water induces a convection roll (vertical arrows). The superposition of these two effects leads to the accumulation of dissolved molecules in the cold bottom corner of the pore. Figure 2.9 shows a temperature gradient (ΔT = 30 K) and the resulting concentration profile over time calculated using finite element software COMSOL [177].

The flow density from Equation 2.2 is then extended by a convective term with the convection velocity v_c for a pore of thickness ϑ and tilting angle α

$$j = j_{D_T} + j_D + j_c = -D_T \nabla T(x,y) \cdot c(x,y,t) - D\nabla c(x,y,t) + \mathbf{v}(\alpha,\vartheta) \cdot c\,. \tag{2.4}$$

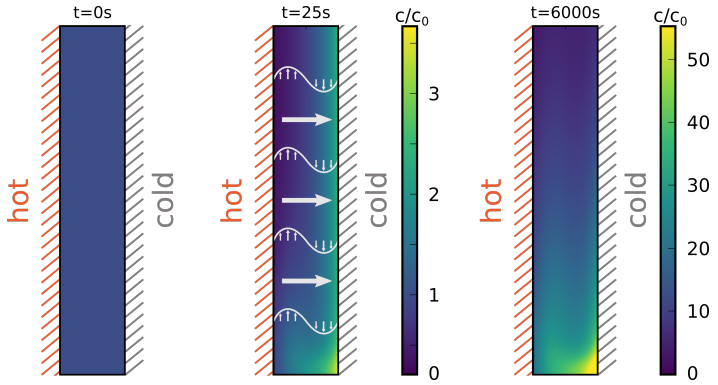

Figure 2.9: Simulation of thermal trapping of DNA using COMSOL. A temperature gradient of $\Delta T = 30$ K across a pore simultaneously leads to convection (center, vertical arrows) and thermophoresis (center, horizontal arrows). The superposition of both effects leads to a relative concentration increase at the cold bottom corner of the pore.

The change in concentration over time can be derived analytically using the continuity equation. A 2-dimensional approach is chosen, with the thermal gradient along the x-axis and gravitation along the y-axis. The temperature gradient is assumed to be linear across the width of the pore ϑ and ∇T can thus be written as $\Delta T/\vartheta$.

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) + D \cdot S_T \cdot \nabla \cdot (\nabla T \cdot c) - \nabla \cdot (\mathbf{v}(\alpha, \vartheta) \cdot c) \tag{2.5}$$

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) + D \cdot S_T \cdot \nabla T \cdot \nabla c - \mathbf{v}(\alpha, \vartheta) \cdot \nabla c \tag{2.6}$$

For the convective velocity in a vertical pore ($\alpha = 90°$) Debye derived a differential equation [38]. The turning points in the velocity profile are neglected, such that the convective flow is only dependent on x, $v(x)$.

$$\frac{\partial^3 v}{\partial x^3} = \frac{-\beta \cdot \rho \cdot g \cdot \Delta T}{\mu \cdot \vartheta}, \tag{2.7}$$

where g is the gravitational acceleration, μ the dynamic viscosity, ρ the density and β the volumetric expansion coefficient of the fluid. With a no-slip boundary condition, integration of Equation 2.7 yields the convective flow between to plates

$$v(x) = \frac{-\beta \cdot \rho \cdot g \cdot \Delta T \cdot \vartheta^2}{6\mu} \cdot \left(\frac{x}{2\vartheta} - \frac{3x^2}{2\vartheta^2} + \frac{x^3}{\vartheta^3} \right) = -v_0 \cdot \left(\frac{x}{2\vartheta} - \frac{3x^2}{2\vartheta^2} + \frac{x^3}{\vartheta^3} \right). \tag{2.8}$$

The velocity profile can then be plugged into Equation 2.6. To solve the resulting differential equation for the concentration profiles several non-dimensional variable are introduced: $q = v_0\vartheta/D$, $p = S_T \Delta T$, $\eta = y/\vartheta$, $\xi = x/\vartheta$ and $\tau = Dt/\vartheta^2$. Equation 2.6 then reads

$$\frac{\partial c}{\partial \tau} = D\left(\frac{\partial^2 c}{\partial \xi^2} + \frac{\partial^2 c}{\partial \eta^2}\right) + p \cdot \frac{\partial c}{\partial \xi} + q \cdot f(\xi)\frac{\partial c}{\partial \eta}, \text{ with } f(\xi) = \frac{\xi}{2} - \frac{3\xi^2}{2} + \xi^3 . \tag{2.9}$$

Using a separation of variables ansatz, $c(\xi,\eta) = U(\xi)V(\eta)$, and the assumptions that V is an exponential in η, $V(\eta) = \exp(\alpha\eta)$ and $p \ll 1$, a steady state solution for the concentration profile can be found

$$c(\xi,\eta) = \left(1 + p\left(-\xi + \frac{84p \cdot q}{10080 + q^2}\left(\frac{\xi^5}{20} - \frac{\xi^4}{8} + \frac{\xi^3}{12}\right)\right)\right) \cdot \exp\left(\frac{84p \cdot q}{10080 + q^2}\eta\right). \tag{2.10}$$

This yields a spatial concentration gradient in a pore with height h and width w of

$$\frac{c_{bot}}{c_{top}} = \exp(0.42 \cdot S_T \cdot \Delta T \cdot h \cdot \vartheta_{opt}^{-1}). \tag{2.11}$$

Equation 2.11 underlines the exponential nature of thermal trapping. Since the Soret coefficient is highly sensitive to small changes in charge, surface or size, this results in a robust, selective trapping mechanism. It has been shown that a temperature gradient across a closed water filled pore can exponentially accumulate oligonucleotides by several orders of magnitude [10] and create a pH gradient over two units by the spatial separation of buffering molecules and amino acids [90].

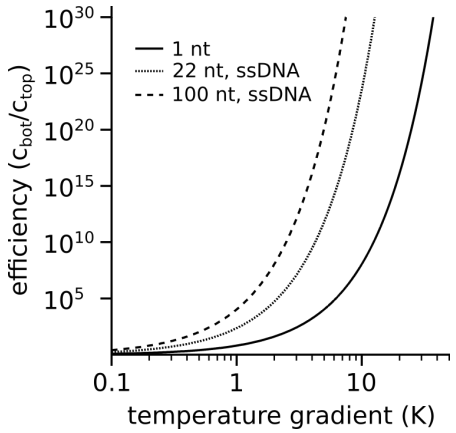

Figure 2.10: Exponential accumulation of different lengths of ssDNA. The accumulation of DNA depends exponentially on the temperature gradient ΔT and the aspect ratio $h \cdot \vartheta^{-1}$ of the chamber. The accumulation is calculated according to Equation 2.11 for a thermogravitational chamber of $h = 50$ mm and $\vartheta = 170$ µm. The strong accumulation of up to 30 orders of magnitude is a steady state effect and in this extend experimentally unrealistic. The Soret coefficients are taken from [10].

Expanding the system to an open system with a feeding flow allows many more opportunities. Kreysing *et al.* demonstrated that an open pore system can enable replication and positive length selection [95].

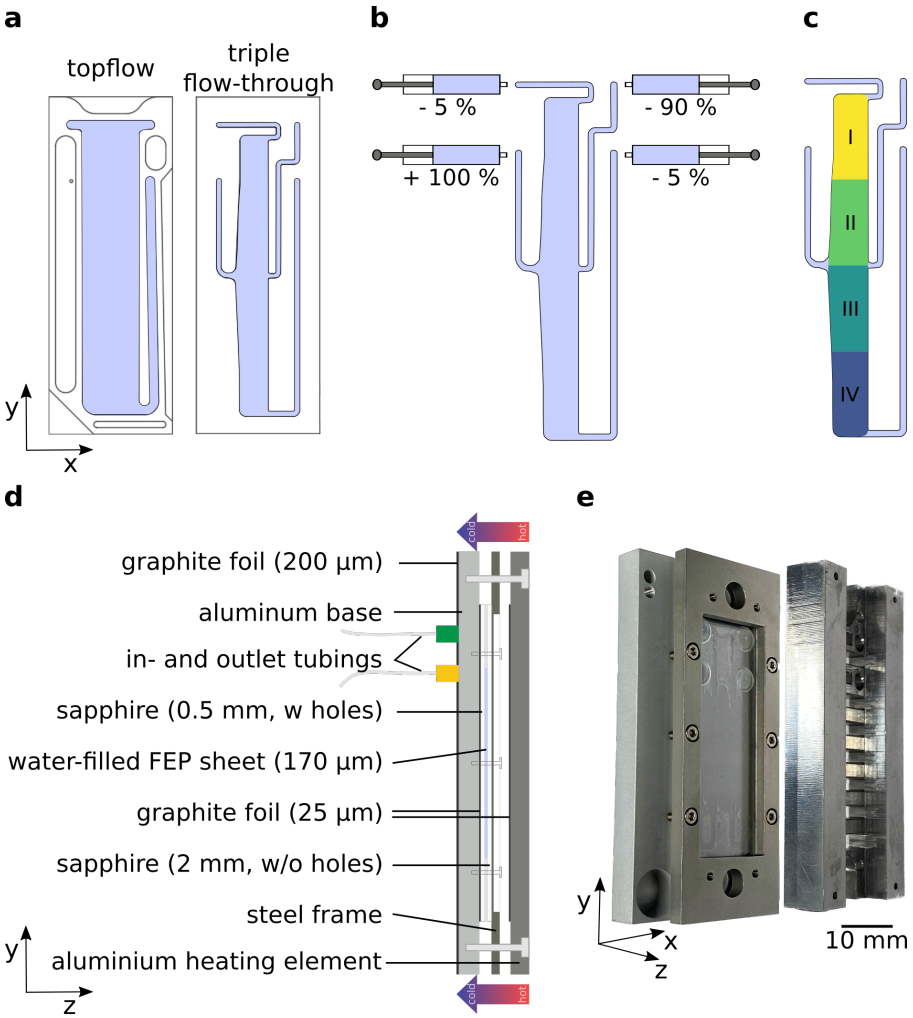

Figure 2.11: Overview of design, methodology and assembly of a current thermogravitational chamber. a, Typical designs for thermogravitational chambers. Fluid flow is from left to right with different number of outlets. The chamber geometry is typically cut from a 170 µm FEP sheet, which results in a chamber volume of $\sim$ 45 µl. **b,** In- and outlets of the chamber are connected to syringe pumps to steer in- and outflow. The net flow has to be zero to not create pressure in the chamber. Flow velocities between 0.5 - 1000 nl/s are possible. **c,** By freeze extraction, i.e. shock-freezing the chamber before disassembly, the main part of the chamber (sometimes also the in- and outlet channels) can be extracted as an ice sheet and cut into four separate sections, which can be analyzed independently. **d,** Lateral sketch of a thermogravitational chamber. The chamber is assembled from individual components as shown in Figure 2.3. **e,** Picture of the assembled chamber with the heating element to be attached. Heating cartridges are inserted into shafts in the heating element in y-direction.

2.5.3. Current version of thermogravitational chambers

For thermogravitational experiments, the chamber components are assembled as shown in Figure 2.3. There are two major types of chamber designs: 'topflow' chambers and 'triple flow-through' chambers, Figure 2.11a. In both cases, the sample is flushed into the chamber through an inlet on the left side and then exposed to convection and thermophoresis in the main chamber. Depending on the chamber design, the sample enters the chamber at the top (topflow) or the center (triple flow-through) and is removed through up to three outlets.

Generally, thermogravitational chambers can be operated in two modes, either in steady mode without flow or with a constant flow-through of sample. In the first case, the sample is loaded into the chamber, the tubings are closed with end caps and the total concentration inside the chamber is conserved. In flow-through chambers, a constant inflow of new sample is provided and accumulation can cause an uptake of material inside the chamber. Syringe pumps (Low Pressure Syringe Pump neMESYS 290N with Quadruple Syringe Holder Low-Pressure, Cetoni, Germany) are connected to in- and outlets (Figure 2.9b) and the respective flow speeds at the outlets can be adjusted, thereby modulating the flow inside the chamber. Since the flow pattern in the chamber is governed by laminar flow, the relative flow speeds translate directly to the spatial distributions od the extracted volume fractions. The syringe pumps can be equipped with syringes holding 100-1000 µl (Göhler-HPLC Syringes, Germany) and can move the pusher with 1 nm/s - 6 mm/s.

For thermogravitational trapping, usually no optical readout is used. However, it is in principle possible and can be done if necessary. Before and after the experiment, the chamber is checked for air bubbles by eye. To quantify the results of thermogravitational trapping, the samples are extracted from the different outlets and can be analyzed separately and compared to the inlet or stock sample. Moreover, freeze extraction allows the extraction of up to four sections of the main chamber as indicated in Figure 2.11c. To do so, the heater is disconnected after the experiment and the chamber assembly is immediately shock-frozen at -80 °C. After approximately twenty minutes at -80 °C, the chamber is disassembled on ice and the frozen ice sheet is thawed sequentially, separating the different sections. The time after switching off the heaters and shock-freezing should be minimized to avoid mixing by diffusion.

There are two main conditions to be considered regarding heat conductivity of an optimal thermogravitational chamber. Firstly, it should have a uniform temperature distribution in the x-y-plane, perpendicular to the direction of the temperature gradient. In previous versions (Section 2.2), this constituted a problem. Secondly, thermal conductivity in z-direction along the temperature gradient should be good to ensure that the majority of the temperature difference between hot and cold side is applied to the sample in a linear temperature drop. In order to investigate the thermal properties of thermogravitational chambers as shown in Figure 2.11, the finite-element software COMSOL [177] was used to simulate the heat distribution in a CAD-created 3D model, which is displayed in Figure 2.12. A temperature gradient of $\Delta T = 71$ K between $T_{cold} = 10$ °C and $T_{hot} = 81$ °C was simulated. The thermal conductivity was assumed to be 0.62 W/mK for water (sample), 44.5 W/mK for the steel frame, 237 W/mK for the aluminum back plate, 35 W/mK for the sapphire, 1600 W/mK for

the thin (25 µm) graphite foil, 400 W/mK for the thick (200 µm) graphite foil and 0.2 W/mK for the FEP sheet.

The simulated temperature distribution shows indeed, that the temperature is uniformly distributed along the x-y-plane (Figure 2.12b,c) and that the temperature drop across the water-filled part of the chamber is steep and linear (Figure 2.12d).

During the experiment, the temperature can be checked using an IR camera (ShotPRO Wärmebildkamera, EAN: 0859356006217, Seek Thermal, USA) from the hot side and a temperature sensor on the cold side (GTH 1170, Greisinger, Germany and B&B Thermo-Technik 06001301-10, Germany).

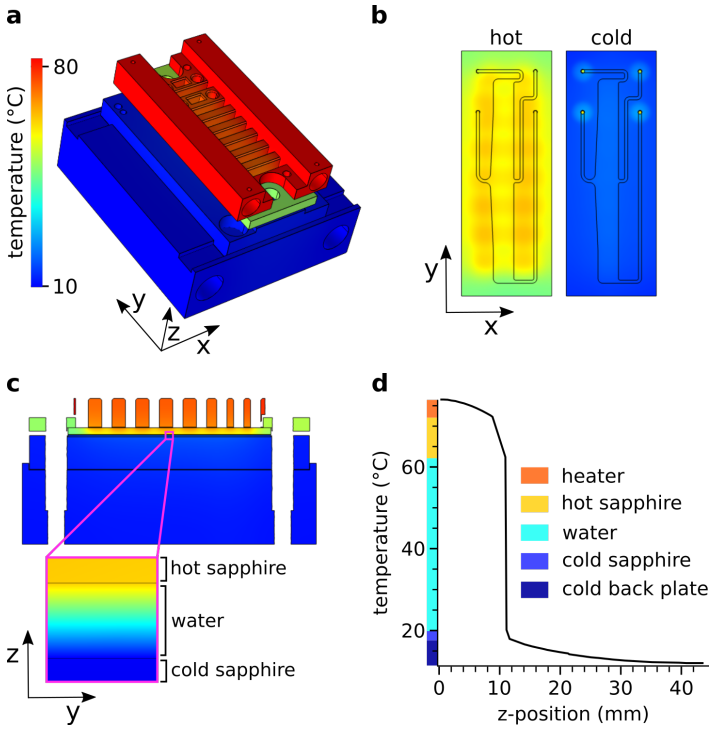

Figure 2.12: Temperature distribution across a thermogravitational chamber. a, Visualization of the temperature distribution across an assembled chamber, including the heater. The temperature distribution was simulated using the finite-element software COMSOL. A heater temperature of $T_{hot} = 81$ °C and backside temperature of $T_{cold} = 10$ °C were assumed. **b,** Temperature projection on the hot (left) and cold (right) sapphires, where they touch the sample. This yields the effective boundary temperatures of the temperature gradient the sample is exposed to. **c,** Projection of temperature distribution along the z-y plane on the complete assembly. It shows that the heat profile along the y-axis is uniform. The cut-out zooms in on the chamber area, between the two sapphires. **d,** One-dimensional temperature profile plotted across the z-dimension of the chamber assembly (averaging over y). It is evident the heat conduction of the assembly is very good, such that the temperature drop across the sample-filled part of the assembly is steep and the majority of the temperature gradient is applied there.

2.5.4. Accumulation of salts

Catalytically active RNAs are assumed to have played a central role at the emergence of life. It is well known that magnesium (or manganese) ions are indispensable for the catalytic activity of such RNAs and direct the folding of their active site [35, 101]. Interestingly, it is difficult to substitute magnesium by other divalent ions such as calcium or manganese [41, 62]. Experimentally, many ribozymes actually require magnesium concentrations as high as 200 mM $MgCl_2$ [8, 76]. Furthermore, some ribozymes are inhibited by high concentrations of monovalent ions, such as sodium [8, 59, 196].

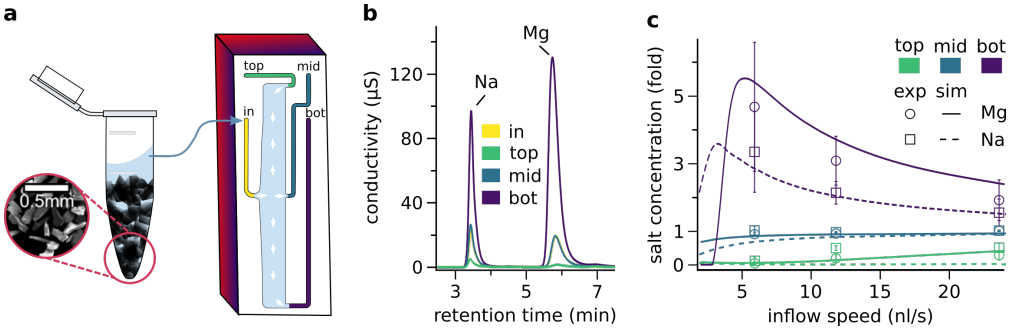

Figure 2.13: Magnesium accumulation in a thermogravitational chamber after leaching from basalt. **a,** Incubating crushed and sieved basaltic rock and glass (inlet: SEM image) in water at elevated temperature of 60 °C leads to the release of ions into solution over the course of hours. A solution mimicking the supernatant of the leached sample is inserted (yellow) into a thermogravitational chamber of 170 µm thickness with $\Delta T = 40$ K across. A triple flow-trough design is used and samples are extracted at top (green), mid (blue) and bottom (purple). **b,** In- and outlet samples are analyzed by ion chromatography. Sodium and magnesium ions can be separated and quantified. The inlet trace shows that the basalt leaching produces a surplus of sodium ions, with a Na:Mg ratio ≥ 1. By thermogravitational trapping this ratio can be shifted leading to an excess of magnesium ions in the bottom outlet (Na:Mg ≤ 1) and a corresponding decrease in the top outlet. **c,** Thermogravitational trapping of a solution mimicking the leached basalt sample was performed for different inflow speeds. The salt concentration relative to the inlet is shown. The error bars correspond to $N = 3$. Slower flow rates cause better uptake by convection and enhance selectivity between the ions. The simulation shows a numerical model including thermal convection, thermophoresis and fluid flow using finite element software. The Soret coefficients and diffusive mobilities were taken from [147, 183]. Experimental data and simulation are in good agreement. Figure adapted from [117].

This section has been accepted for publication in *Nature chemistry* by Matreux *et al.* [117] under the title "Heat flows in early rocks naturally optimize salt compositions for ribozymes" and is currently in resubmission. Authors: T. Matreux*, K. LeVay*, A. Schmid, P. Aikkila, L. Belohlavek, Z. Caliskanoglu, E. Salibi, **A. Kühnlein**, C. Springsklee, B. Scheu, D. B. Dingwell, D. Braun, H. Mutschler, C. B. Mast.
Contributions: T.M., K.L.V., A.S., P.A., L.B., Z.C., E.S., C.S., C.B.M. performed the experiments. T.M., K.L.V., A.S., L.B., Z.C., E.S., A.K., C.S., B.S., D.B.D., D.B., H.M., C.B.M. conceived and designed the experiments. T.M., K.L.V., P.A., L.B., Z.C., E.S., A.K., C.S., B.S., H.M., C.B.M. analyzed the data. T.M., K.L.V., A.S., P.A., L.B., Z.C., A.K., B.S., D.B., H.M., C.B.M. wrote the paper. All authors discussed the results and commented on the manuscript.

Ionic solutions representative of early Earth conditions can be attained from incubating basaltic rock or glass at slightly elevated temperature of 60 °C (Figure 2.13a). This leads to leaching of ions such as sodium and magnesium into solution. After a leaching time of 36 hours, concentrations of up to 200 µM Mg^{2+} and 1 mM Na^{+} were reached. This constitutes a Na:Mg ratio $\sim$ 5, which inhibits ribozyme activity (Figure 4 in [117]). Ion concentrations were quantified with ion chromatography.

To test whether thermal gradients across water-filled pores can shift the Na:Mg ratio and thereby create a viable environment for prebiotically relevant RNAs, ionic solutions mimicking the supernatant from the leaching experiments were prepared and flushed through a thermogravitational chamber using a triple flow-through design with three outlets. These experiments showed that the thermogravitational trapping can greatly enhance the magnesium concentration relative to the sodium concentration at the bottom of the chamber (Figure 2.13b). Studying different flow speeds revealed that low flow speeds lead to generally increased material uptake but also better relative separation. These findings were confirmed by simulating the effects of convection, thermophoresis and fluid flow using finite element software COMSOL [177] and using literature values for Soret coefficient and diffusive mobilities [147, 183].

The study by Matreux *et al.* also looked at the impact of different Na:Mg ratios on ribozyme activity for self-replication and RNA extension using a (modified) R3C ribozyme. It could be experimentally confirmed that in bulk experiments (thermocycler) high relative sodium concentrations (up to 1000x) indeed inhibit ribozyme activity. An up to ten-fold excess of sodium could be tolerated by the ribozyme at high total magnesium concentration (4 mM). As expected, thermogravitational chambers enhance ribozyme activity due to magnesium accumulation and the ribozyme commences its activity at lower total magnesium concentrations compared with bulk experiments.

An extrapolation from the experimental of two concatenated thermogravitational chambers showed that a cascade of up to 30 thermogravitational chambers can turn an Na:Mg ratio of 1000, which is realistic for carbonate lakes [185], to 0.1. This would completely alter the reaction dynamics for any involved ribozymes, suddenly providing highly favorable conditions for ribozyme activity.

The results by Matreux *et al.* underline the prebiotical relevance of thermal non-equilibria, modeled by thermogravitational chambers, as they can establish a viable environment for ribozyme activity from initially plausible geological conditions.

2.6. Incubation chambers

In contrast to gas-water interface chambers and thermogravitational chambers, incubation chambers are not operated with a temperature gradient across the chamber. They do not intend to induce specific physical effects at the interface or in solution, but are merely an aid to better observe effects happening under thermal incubation or thermal cycling. Peltier elements (QuickCool QC-31-1.4-3.7AS, purchased from Conrad Electronics, Germany) are directly connected to the backside of the chamber providing full thermal control and are reg-

ulated by a custom PID regulation implemented in LabVIEW (National instruments). Heating and cooling rates of up to 2 K/s can be reached.

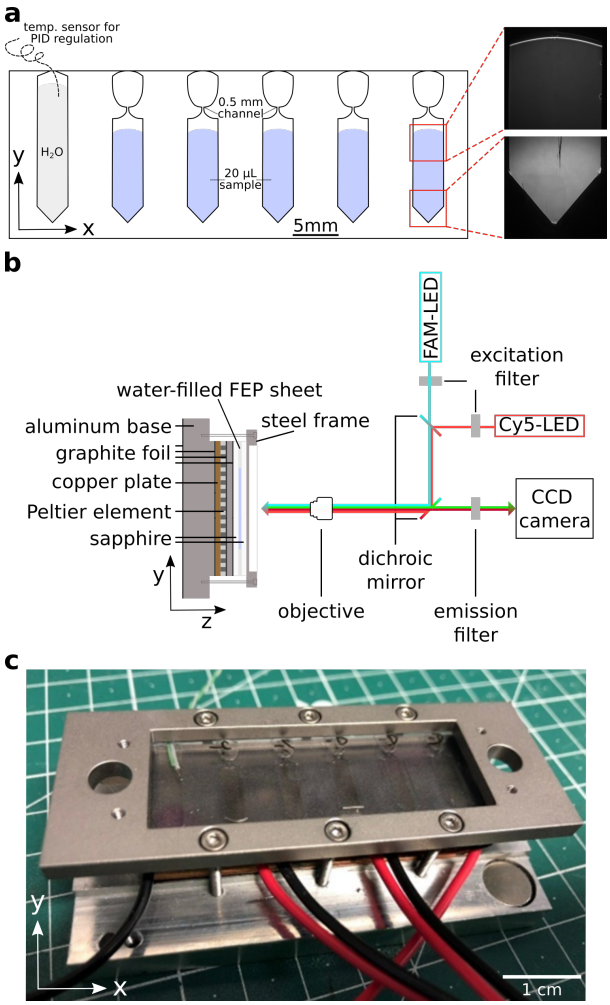

Figure 2.14: Overview of a microfluidic incubation chamber and setup. a, An incubation chamber typically contains five independent pockets, which can be accessed through a 0.5 mm channel from the top. After loading the sample, the pockets are sealed with Parafilm. An additional pocket contains a temperature sensor for the PID regulation. On the right, exemplary fluorescence images from top and bottom of the chamber are shown containing 20 µM of 84 nt DNA in 150 mM NaCl, 20 mM $MgCl_2$ and 20 mM Tris, pH 8. **b,** Lateral sketch of the chamber assembly and the optical setup. The chamber geometry (0.5 mm FEP sheet) is fitted between two plane sapphires. Three Peltier elements (in x-direction) are attached to the backside of the chamber which allows quick heating and cooling. This sandwich is then screwed onto an aluminum base using a steel frame. The sapphire allowed full visual access to the sample. The assembled chamber was attached to a motorized stage for full optical coverage of the chamber. **c,** Image of an assembled chamber. Figure adapted from [98].

The typical design and setup of an incubation chamber is displayed in Figure 2.14. The chamber geometry is cut from a thick (0.5 mm or 1 mm) FEP sheet and can contain up to nine sample-filled pockets. One additional water-filled pocket contains the temperature sensor (B&B Thermo-Technik 06001301-10, Germany) for the PID regulation. The cut chamber is fitted between two plane sapphires (2 mm thick) and assembled as shown in Figure 2.14b. To ensure good heat conduction, graphite foils are inserted wherever possible. The sample can be loaded through a 0.5 mm channel at the top of each pocket, which is sealed with Parafilm after loading. The chamber is then mounted onto a custom-built motorized stage. During the experiment the aluminum base of the chamber is connected to a water bath (CORIO CD, Julabo, Germany) to increase the thermal reservoir and ensure that the heat generated by the Peltier elements can dissipate quickly.

Due to imperfect heat conduction, small temperature gradients ($\Delta T \sim 5$ K) across the chamber can arise, especially for incubation temperatures that differ more than 20 K from room temperature. This is evident from condensation droplets visible at the top of the chamber. However, these temperature gradients are small compared to the temperature gradients created in gas-water interface chambers ($\Delta T \geq 20$ K), especially keeping in mind the different chamber thicknesses, and can thus be neglected.

3. Thermally driven selection of oligonucleotide motifs

Summary

RNA can not only store genetic information, but also assume catalytically active conformations by self-hybridization, and is therefore central to the *RNA world* hypothesis. However, it is yet unclear how catalytic function arose de novo from a random pool of oligonucleotides synthesized by a prebiotic chemistry, given the vast sequence space and resulting low concentration of specific functional sequences. Our experiments show, that heat fluxes across water-filled pores, which were abundant on the early Earth, can act as a prebiotic filter. Starting from a random pool, a thermal gradient is able to selectively filter oligonucleotide motifs based on their phenotype. The mechanism is driven by the positive size-selection of thermophoresis, which leads to a divergent effect for poorly and amply hybridized strands. Therefore, strands with lower free energy that form compact secondary structures, a motif that has previously been associated with greater functional potential, can be enriched. Although the observed shift in mean free energy is small, it is consistent. Especially the prospect of amplifying the effect by repeated selection cycles is exciting.

3.1. Introduction

Enzymatic function dominates modern life. The replication of the genetic information stored in DNA and its translation into proteins is catalyzed by proteins themselves. The discovery that not only proteins but also nucleic acids have catalytic properties revolutionized the origin of life research and also paved the way for the wide-spread acceptance of the *RNA world* hypothesis [61,70,96], which presumes that early on genetic information was encoded in RNA that self-replicated in the absence of catalytically active proteins. Naturally occurring ribozymes have been shown to support reactions such as (self-) cleavage and (self-) splicing [45, 47]. Furthermore, in vitro selection experiments have been used to identify ribozymes that carry out a wide range of biological function [26, 101, 160].

The ensuing path from the *RNA world* to modern life has already been drafted by Darwin: As soon as catalytic function has emerged in biopolymers such as RNA or peptides, recombination and mutation will improve and evolve their function and fidelity. However, it is yet unclear how catalytic function arose de novo [73, 182]. Keeping in mind that the ribozymes we know today are typically 40-150 nt long [45, 47], this quest can be divided into two aspects: (i) how did oligonucleotides of sufficient length form? And, (ii) how were those strands that contained functional motifs selected from a (semi-) random pool? This chapter will be dedicated to the latter question.

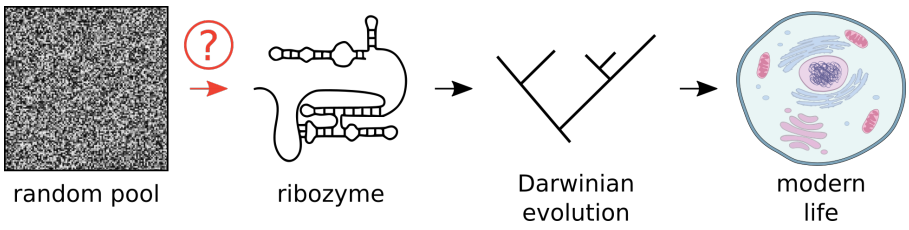

Figure 3.1: Evolutionary road from a random pool to modern life. Prebiotic polymerization produced a pool of oligomers of random sequence. To kick-start evolution, there had to be a process in play, which was able to select functional sequences or sequence motifs from the random pool. Those functional motifs would later undergo Darwinian evolution and evolve to what we today know as modern life.

Though the prebiotic polymerization and ligation of oligonucleotides is yet to be settled, it is fair to assume that it produced a dilute pool of (semi-) random strands [73,86]. It should be noted here, that alone due to mass constraints, the sequence space must have been limited early on. Already a pool of completely random 100 nt strands exceeds the mass of the Earth by several orders of magnitude [87,124]. Nevertheless, even if prebiotic polymerization and ligation chemistries had a considerable sequence bias, as it has been suggested [124], the resulting sequence space would still have been vast. Since the subset of catalytically active strands in a random pool is very small [54] and their concentration would have been very low [73], to launch an *RNA world* with substantial ribozyme activity one or more physical processes must have exerted a selection pressure on the (semi-) random sequence pool of oligonucleotides to enrich it in functional motifs.

Non-equilibrium conditions are crucial for the emergence of living entities but also for modern life itself. Today, the metabolism is responsible for keeping up a chemical disequilibrium. But in the absence of a metabolism, the emergence of life must have been dictated by physical non-equilibria, that provided a flux of free energy to allow further development. One simple and abundant non-equilibrium setting on the early Earth are heat fluxes across elongated, water-filled pores as illustrated in Figure 3.2. Such submerged pores could have occurred on early Earth in volcanic rock, rock fissures or in inclusions of mud or clay, where the temperature gradient was provided by the spatial vicinity to a warm volcanic outflow or hydrothermal system, which released heat into a cold ocean [10, 120]. As described in detail in Sections 2.5.1 and 2.5.2, a thermal gradient causes thermophoresis of dilute molecules, which combined with convection leads to local accumulation. The strength of a molecule's movement in the thermal gradient is determined by its Soret coefficient, which is highly sensitive to small changes in size or charge. While this microenvironment selects for longer lengths, there is no guarantee that longer strands are more functional. Therefore, there also has to be a mechanism to select for function in order to allow particular nucleotide sequences or motifs to become dominant over others of similar length [25].

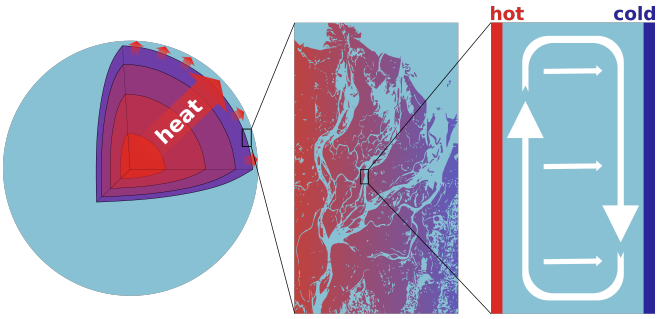

Figure 3.2: Illustration of heat fluxes on early Earth. Heat fluxes across water-filled pores were abundant on the early Earth, for example in submerged volcanic rocks. Convection and thermophoresis occur as described in Sections 2.5.1 and 2.5.2 and lead to length selective, exponential accumulation of dilute molecules.

In ribozymes, like in proteins, functionality is reached by folding into secondary structures by self-hybridization [5, 164, 195]. It is worth noting, that catalytic abilities in oligonucleotides are not limited to RNA but have also been observed in DNA [101, 167, 172]. Oligonucleotides have a wide repertoire of possible secondary structures, including stem loops, bulges, internal loops and pseudoknots, which are illustrated in Figure 3.3. Ribozymes show intricate secondary structures with a combination of the above mentioned elements as depicted in Figure 3.1.

Although functional sequences are sparse in random pools [54], computational analysis has shown that in a random pool of RNA single strands $\geq$ 25 nt only a small fraction ($\leq$ 15 %) is unfolded and most strands are present in single or multiple stem loop structures [18]. Furthermore, it has been shown that while such structural motifs are generally abundant due to the large sequence space, their frequency is heavily biased, i.e. the majority of sequences occurs within a few structural motifs [109, 176]. Interestingly, this bias is

more pronounced for longer sequences and in an RNA pool of 55 nt for example, 50 % of genotypes are represented in only 0.1 % of phenotypes [43].

A recent experimental study associated sequences that assume highly compact secondary structures with the ability to converge faster towards better functionality in in vitro selection starting from a random oligonucleotide pool [28].

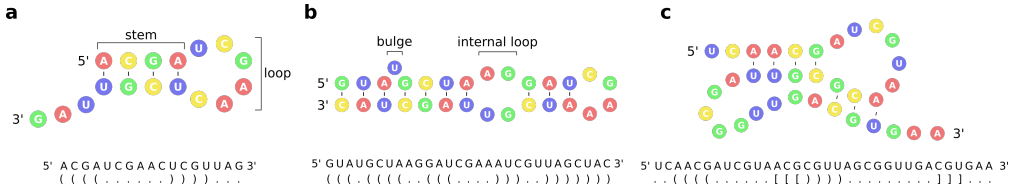

Figure 3.3: Sketch of RNA secondary structures. a, Stem loops, **b,** bulges and internal loops and **c,** pseudoknots are frequent RNA secondary structures that can for example occur in ribozymes. Generally, those secondary structures also appear in DNA. The most stable loops are 6-7 nt long in RNA and 4-5 nt long in DNA [187]. Structures that contain a mixture of elements from panels a and b are typically referred to as hairpins. Noted below each sketch are the sequence and corresponding dot-bracket notation.

Folding of oligonucleotides in general and of evolved structures like ribozymes in particular causes them to spatially collapse. This alters the Soret coefficient and thus their behavior in a thermal gradient. Considering the different secondary structures that can occur in a random oligonucleotide pool it is important to understand how exactly potential folding impacts the thermophoretic behavior of oligonucleotides.

The behavior of single-stranded versus double-stranded oligonucleotides was investigated by Reichl *et al.* and Maeda *et al.* in two controversial studies [111, 159]. While Reichl *et al.* found an increased Soret coefficient for single-stranded DNA and RNA, compared to double strands of the same length, the data from Maeda *et al.* suggests that DNA and RNA single strands show no thermophoretic mobility at all. Only when base pairing occurs, like in stem loops, strands start to accumulate. As the authors themselves noted, this would be especially promising for thermally driven selection of oligonucleotides towards functional motifs. Wienken *et al.* provided yet other data, which suggests that the Soret coefficient of DNA double strands is higher than for DNA single strands [193]. They also studied the Soret coefficient for partially hybridized DNA duplexes, i.e. duplexes with overhangs on both sides of the double-stranded section and could show that the Soret coefficient of the duplex depends on length and sequence of the overhangs. Furthermore, they demonstrated that changes in secondary structure, when a strand forms a stem loop, also affect the thermophoretic behavior.

The above mentioned studies also compared the behavior of DNA and RNA. Data from Reichl *et al.* suggests that DNA and RNA of identical length behave comparably and their Soret coefficient as well as effective electric charge are within the same order of magnitude [159]. Despite the controversial results, these studies agree over the fact that DNA and RNA single strands, that exhibit self-binding have a positive Soret coefficient and that the extend to which they are double-stranded alters the strength of their thermophoretic mobility.

In this chapter, I will show that prebiotically abundant heat flows across rock pores on early Earth could have selectively enriched a random oligonucleotide pool with sequences that form compact folds, a phenotypic motif that has been associated with larger functional potential [28]. The thermal non-equilibrium acts as a prebiotic filter and can select sequences based on their phenotype due to the associated change in thermophoretic mobility. Such selectivity of sequences with lower free energy is observed in static as well as dynamic settings for random pools of DNA and RNA over a wide range of parameters and thus appears to be a robust effect. Especially the prospect of several connected rock pores through which a random pool would experience a successively increased selection is promising.

3.2. Determination and mapping of genotype and phenotype

To measure whether thermogravitational trapping is able to select sequences based on their phenotype from a random pool, several analysis methods can be used. First and foremost, samples are sequenced before and after exposure to the thermal gradient using Next-generation sequencing (NGS), which determines their genotype. Subsequently, computational tools allow to obtain the phenotype from the sequenced genotype. Additionally, native polyacrylamide gel electrophoresis (PAGE) can be used to resolve different secondary structures in a gel matrix.

3.2.1. DNA and RNA sequencing

Recently, scalable, low-cost sequencing technologies were developed, which are typically subsumed under the term Next-generation sequencing (NGS) technologies [63, 92]. Due to their high level of parallelization, a large amount of reads can be generated from a single sequencing run.

The sequencing data in this thesis was generated using the Illumina technology. It is based on sequencing by synthesis, in which the single-base addition of fluorescently labeled bases is used for sequence determination. Prior to sequencing, a library of the oligonucleotide sample has to be prepared. This step is specific to the sequencing technology and includes sample amplification and indexing. The main steps during library preparation and the structure of the prepared library strands are shown in Figure 3.4a.

The highly diluted library is then flushed into the flow cell of the Illumina machine, where the adapters hybridize onto a solid surface. The bound strands are amplified through bridge amplification, creating clusters of strands with identical sequence in close spatial proximity. The reverse complement strands, which are created during clonal amplification, are then cleaved and washed off. Together with a primer, fluorescently labeled monomers are flushed into the flow cell, where a polymerase starts adding single bases to the forward strands. After every cycle of monomer addition, the clusters are excited by a light source and a characteristic fluorescent signal is emitted, which allows to identify the incorporated base.

In this way, the template sequence for each cluster is read out. The HiSeq 2500 machine from Illumina generates 3-6 billion reads per flow cell [82]. The sequencing process in the flow cell is illustrated in detail in Figure 3.4b.

Internal data processing in the sequencing machine generates FASTQ files, which contain the read sequence for each cluster and a corresponding quality string. For data analysis the sequence data is demultiplexed, i.e. each read is assigned to a certain sample, according to the indices introduced during the library preparation.

3.2.2. RNAfold

A very powerful collection of tools for oligonucleotide secondary structure prediction is provided by the ViennaRNA package [74, 108], which contains the program RNAfold. RNAfold takes a single RNA or DNA sequence as input and uses an algorithm originally proposed by Zuker and Stiegler [204] to minimize the Gibbs free energy of a sequence's secondary structure. Additionally, it can calculate the partition function, the base pairing matrix and the minimum free energy for the thermodynamic ensemble [69]. As an output it provides the minimum free energy and the corresponding secondary structure in dot-bracket notation, where a dot · represents an unpaired base and brackets () represent a pair of Watson-Crick paired bases as illustrated in Figure 3.3.

3.2.3. Native PAGE of random oligonucleotides

Polyacrylamide gel electrophoresis (PAGE) is a technique commonly used in biochemistry to determine the length (SDS-PAGE) or composition (native PAGE) of oligonucleotide samples. The electrophoretic mobility of oligonucleotides depends on their length, conformation and charge.

The behavior of random RNA strands in comparison to known functional RNA sequences and permutations of those functional sequences in native PAGE has been studied by Schultes *et al.* [169]. It is important to understand that the authors did not look at fully random pools but instead probed strands, which had one defined but random sequence. Their data suggests strands that are permuted from a functional sequence assume many coexisting folds, the randomly generated sequences simply collapse with increasing magnesium concentration. It has later been suggested that the observed collapse is due to unspecific aggregation [28].

Recently, Chizzolini *et al.* [28] investigated the behavior of truly random RNA pools in native PAGE and found that structural differences within a random pool can be detected. Furthermore, the authors were able to show that native PAGE itself can be used as a means to select sequences of higher phenotypic potential from a random pool. A pool, that had been pre-selected for compact secondary structures, yielded more potent sequences in subsequent in vitro selection experiments. While the authors themselves acknowledge the relevance of their findings to considerations on the emergence of functional sequences, their focus lay on improving the functional potential of starting pools for in vitro evolution and they did thus not provide a prebiotically realistic scenario for such phenotypic selection.

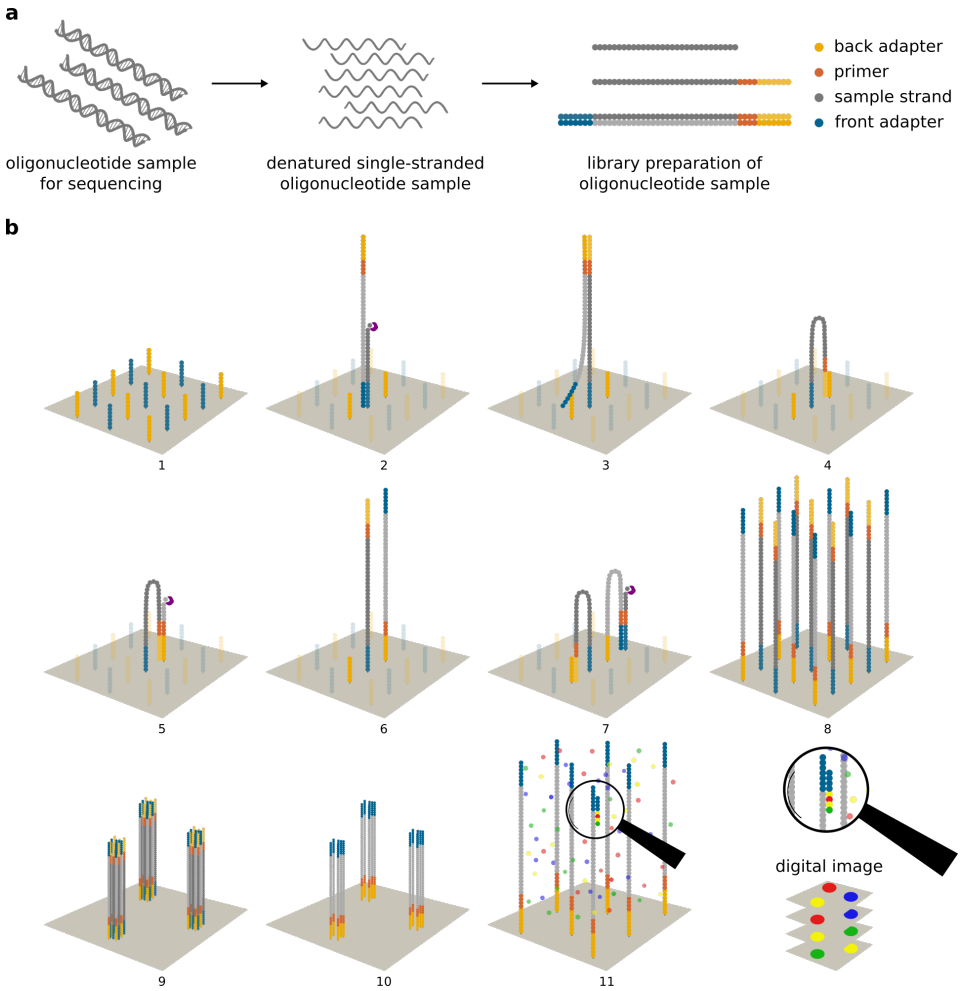

Figure 3.4: Illustration of library preparation and sequencing by synthesis. **a,** The sample must be quantified and denatured. During library preparation, primers as well as front and back adapters are added. The adapters are specific to the sequencing machine but also include sample specific barcodes that later allow sample identification. **b,** Illumina technology of sequencing by synthesis. The sample is loaded onto a flowcell (1) that contains adapters complementary to the library. The sample strands bind to the flowcell and their complement is polymerized (2). In a denaturing step (3) the original strand is washed away. Using bridge amplification (4-7) the strands are amplified. First, the strand folds over and hybridizes to another adapter on the flow cell (4) from where a polymerase generates its complement (5). After denaturing (6) the forward and reverse strands are attached to the flow cell. Refolding onto other adapters and repeated strand amplification (7) generates a cluster of identical sequence within close spatial proximity (8). Many clusters of different sequence are generated across the flow cell (9). The unwanted reverse strands are washed away (10) and the sequencing by synthesis starts. Primers, enzymes and labeled terminated bases are flushed into the flow cell and attached to the forward strand (11). Now, sequencing proceeds base-by-base, with illumination after each base incorporation. Illumination generates a digital image sequence in which each cluster lights up in a different color depending on the incorporated base.

3.3. Outline of the parameter space

Heat fluxes across elongated pores on the early Earth can experimentally be mimicked by thermogravitational chambers, which are described in detail in Section 2.5.

Naturally, the parameter space to study the potential impact of thermogravitational trapping on the sequence space of a random pool is vast. In the following, an overview of the parameters, their influence and a reasonable range in which they can be employed is given. The first four points (1.-4.) are strictly related to the physical properties of the thermogravitational chambers, whereas the latter four (5.-8.) apply to the loaded oligonucleotide sample and its handling.

1. **Chamber dimensions.** Chamber height (y) and chamber thickness (z) are parameters relevant for thermal trapping (Section 2.5.2) and can experimentally easily be modified. The chamber height typically ranges from 30-50 mm. The possible chamber thicknesses are discrete due to the availability of FEP foils and can be 25, 50, 75, 100, 127, 150, 170, 200 or 250 µm.

2. **Static or flow-through mode.** Chambers can be operated in static or flow-through mode. For the first, total concentration is conserved, while the latter is more dynamic and an increase in total concentration in the pore can occur. Furthermore, the operation mode also affects the sample extraction after the experiment. In static mode, the sample must be retrieved by freeze extraction (described in Section 2.5.3 and Figure 2.11c), whereas in flow-through mode, the material from the outlets is collected during the experiment and freeze extraction can be considered as an additional extraction method.

3. **Temperature gradient.** The temperature gradient ΔT determines the strength of thermophoresis (Section 2.5.1) and can range from 0-80 K. However, not only the absolute temperature gradient but also values of T_{cold} and T_{hot} can be relevant. T_{cold} can range from 0-50 °C, T_{hot} from 10-100 °C.

4. **Inflow speed.** In a flow-through system the syringe pumps are connected to in- and outlets (Figure 2.3) and can generate a wide range of inflow speeds (nl/s). Since the syringe pump only determines the velocity with which the piston moves, the flow speed depends on the volume of the syringe.

5. **Oligonucleotide sample: concentration, length, DNA or RNA.** The oligonucleotide length as well as the base composition affect the sequence space of the sample. Oligonucleotide samples can be either DNA or RNA, which alters the folding energy landscape.

6. **Magnesium concentration.** Magnesium ions stabilize oligonucleotides, affect their melting temperature and enhance folding into secondary and tertiary structures [123]. To compensate the effect of magnesium with monovalent ions, a 5-fold concentration excess has to be used [141]. The initial magnesium concentration ranges typically from 0-10 mM, which is low and prebiotically plausible [37]. It should be considered

that Magnesium ions can accumulate in a temperature gradient up to 10-fold (Section 2.5.4).

7. **Buffer conditions.** Potential buffers are PBS, Tris or MOPS. Generally, it should be paid attention to whether the buffer molecule itself accumulates and thereby creates a pH gradient [90]. PBS was chosen as buffer because it is the most physiological and prebiotically plausible buffer. To prevent phosphate precipitation, which occurs at elevated temperatures ($>$ 5 min at $>$ 90 °C), the concentration was reduced to 0.1x PBS and exposure to temperatures $>$ 90 °C was minimized.

8. **Temperature protocol prior to trapping.** The sample handling prior to loading into the thermogravitational chamber is relevant as it can alter the conformation of oligonucleotides: while quenching (q) reduces the formation of secondary structures, annealing (a) ramps can promote folding of secondary and tertiary structures. A comparison by native PAGE of various temperature protocols is shown in Figure 3.14.

3.4. Finite element simulation of thermogravitational trapping

In order to first explore the system, finite element simulations using COMSOL [177] were performed. A simple flow-through chamber with one inlet and one outlet was used as a geometry (Figure 3.5a). At time point t = 0, the whole chamber is filled with water. Boundary conditions at the in- and outlet created a continuous feeding flow containing a predefined molecules species (DNA) across the top of the chamber. The DNA concentrations were determined over the course of 36 hours at top and bottom of the chamber by averaging over volumes indicated in Figure 3.5a. In an attempt to cover the parameter space outlined in Section 3.3 as good as possible, a parameter sweep was calculated for inflow speed (2-50 nl/s), DNA length (1-100 nt) and temperature gradient ΔT (20, 40 K). This was repeated for different chamber geometries permuting chamber height (32.25, 52.25 mm) and thickness (170, 250 µm). The parameter values were chosen considering experimental constraints. Diffusion coefficient and Soret coefficient for DNA were taken from the literature [116].

The trapping efficiency, i.e. concentration at the bottom versus concentration at the top, is a measure for how well a certain species is accumulated in the chamber. With regard to phenoytype-selective trapping, it is especially interesting how efficiencies of different lengths compare. In a simplified model, we can assume that the change in thermophoretic mobility between two strands, one with a poorly folded secondary structures and one with an exceptionally well folded secondary structure corresponds to the change between a strand and one half its length. Thus, a perfectly self-complementary 100 nt strand would have similar thermophoretic properties as a 50 nt unfolded strand.

In Figure 3.5c-f the trapping efficiency is characterized within the covered parameter space. As expected, the trapping efficiency increases over time. Because the chamber is initially filled with water it takes up to a few hours until the material, which is absorbed into

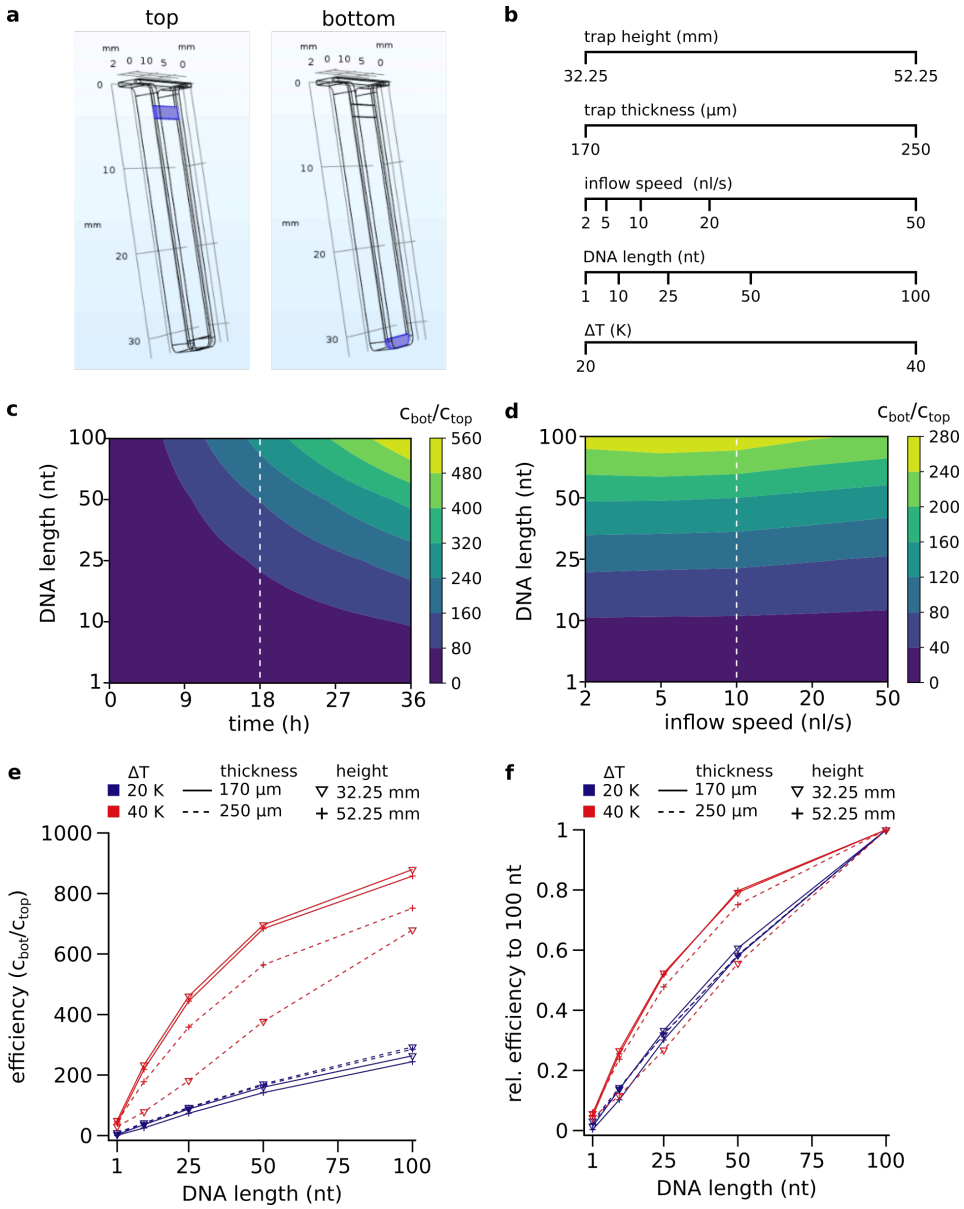

Figure 3.5: Finite element simulation of thermogravitational trapping in a flow-through system. **a,** A chamber design with a simple feeding flow over the top of the chamber was simulated with a time resolution of 20 mins for 36 hours over **b,** a wide range of experimentally feasible parameters. At time point t = 0 the chamber is filled with water. The molecule concentration is measured by averaging over the indicated volumes. **c,** 2-dimensional contour plots of trapping efficiency (c_{bot}/c_{top}) for different DNA lengths versus time at inflow speed 10 nl/s and **d,** versus inflow speeds after 18 hours. Note that the axes for DNA length and inflow speed are not linear. Dashed lines indicate the fixed values for time and inflow speed respectively, which also apply for panels e and f. **e,** Trapping efficiency plotted against DNA lengths for two temperatures (blue/red), chamber thicknesses (solid/dashed lines) and chamber heights (triangle/cross). **f,** To quantify the sensitivity for separation the trapping efficiency from panel e is normalized by the value for 100 nt.

into the chamber reaches the concentration of the feeding flow. Thus, experiments should ideally run for at least several hours. The comparison of different inflow speeds shows that it should be low for longer molecules and high for shorter molecules. Generally, slow inflow speeds are experimentally inconvenient since they limit the material that can be collected from the outlet and complicate downstream analysis.

After fixing time and inflow speed to 18 hours and 10 nl/s respectively, the trapping efficiency was analyzed for different temperature gradients and chamber dimensions as displayed in Figure 3.5e. From Equation 2.11 it can be expected, that a higher temperature gradient as well as a greater chamber aspect ratio lead to more efficient trapping. Within the simulated parameter space, the effect of a temperature change is pronounced, while the effect of the aspect ratio is almost negligible.

The normalized trapping efficiency in Figure 3.5f indicates how well different lengths can be separated from a 100 nt strand. Note that counter-intuitively, a lower value means better separation. Although a higher temperature gradient yields higher absolute trapping efficiencies, better selectivity between lengths is provided by a lower temperature gradient. The only exception is the case of the lowest aspect ratio chamber, again hinting that to achieve selectivity strong absolute accumulation is not necessarily desired.

In conclusion, experimental timescales should exceed a few hours and rather have an intermediate temperature gradient or a low chamber aspect ratio to allow for best selectivity.

3.5. Experimental characterization

The finite element simulations gave some insight into the effects and kinetics of the parameter space but do not reflect all experimental aspects. Therefore, it was important to also experimentally characterize the accumulation of oligonucleotides in different settings.

In order to first verify that DNA and RNA strands accumulate in the chamber as expected, experiments in static mode, i.e. without flow, were performed. From freeze extraction, four samples were obtained, which are labeled in roman numbers from top-most (I) to bottom-most (IV). The concentration after 18 hours across the chamber relative to the stock is plotted in Figure 3.6 for a 60 nt RNA and a 140 nt DNA sample for $\Delta T = 15$ K and a magnesium concentration of 1 mM. As expected, the strands are accumulated at the bottom and depleted at the top, with stronger accumulation of the longer oligonucleotide strands. Interestingly, the depletion is strongest in part II, not part I as might be naively expected.

Subsequently, the more interesting, dynamic flow-through system was investigated in larger experimental parameter screens. Here, the topflow design with a feeding flow across the top was used. Considering the results from the simulation, the chamber dimensions were fixed to a thickness of 170 µm and a height of 52.25 mm. From each flow-through experiment, four samples were obtained: inlet (i), outlet (o), pore (p) and stock (s). A thorough description of the sample extraction procedure is given in Section 3.8.2 and it is illustrated in Figure 3.7a. Flow-through experiments were performed with both DNA and RNA over a wide range of parameters (Figure 3.7b-f), varying $MgCl_2$ concentration, temperature gradient ΔT and inflow speed. In the experiments, as in the simulation, the chamber is initially

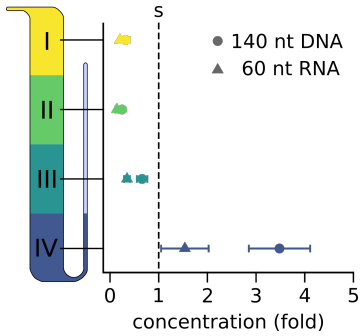

Figure 3.6: Accumulation of DNA and RNA. Concentration increase for 140 nt DNA and 60 nt RNA random pools in a thermogravitational chamber with a temperature gradient $\Delta T = 15K$ (20 °C-35 °C) relative to stock sample. Samples are mixed at 4 µM oligonucleotide concentration, 1 mM $MgCl_2$ and 0.1x PBS and are quenched before loading into the chamber. After a runtime of 18 hours, the chamber is shock-frosted and the separate sections are extracted as indicated. Error bars correspond to N = 3 independent experiments.

filled with water. For all experiments, the concentration of the inlet (i), outlet (o) and pore (p) samples is plotted relative to the corresponding stock (s) concentration.

All experiments in Figure 3.7 ran for 3-4 hours, with the notable exception of experiment shown in panel f, which ran for 20 hours and had a chamber height of only 32.25 mm. Evidently, the timescales in panels b-e were too short to achieve effective accumulation, but some material has been taken up into the chamber. However, in the case of longer timescales and a shorter chamber as in panel f, both factors that favor stronger accumulation, a ten-fold increase of concentration in the chamber was achieved.

The data shows that DNA and RNA strands of the same length behave comparable (Figure 3.7b), which supports the notion that Soret coefficient is similar [159]. In accordance with the simulation, there is a slightly better uptake of oligonucleotides into the chamber for higher temperature gradients (Figure 3.7c) and lower inflow speeds (Figure 3.7d).

Experiments in which the magnesium concentration is varied as shown in Figure 3.7e,f are very instructive. While thermal trapping of a 60 nt DNA pool over the timescale of 4 hours (τ_{short}, panel e) shows almost no magnesium dependent effect, a strong dependence is observed for a 140 nt DNA pool over the timescale of 20 hours (τ_{long}, panel f). For τ_{long}, no or very low magnesium concentrations lead to much better (up to 10-fold) accumulation of DNA. A magnesium concentration of 8 mM limits or rather, since material is still absorbed into the chamber, slows down DNA accumulation. The evident decrease of DNA concentration in the pore with increasing magnesium concentration (inset Figure 3.7f) can readily be understood considering that higher magnesium concentrations result in a lower Debye length, which in turn results in a lower Soret coefficient [158].

Furthermore, when comparing panels e and f, it becomes clear how important timescales are in thermal trapping. Depending on whether an experiment ran for 4 or 20 hours, the results differ strongly. Although, also other parameters such as DNA length, inflow speed

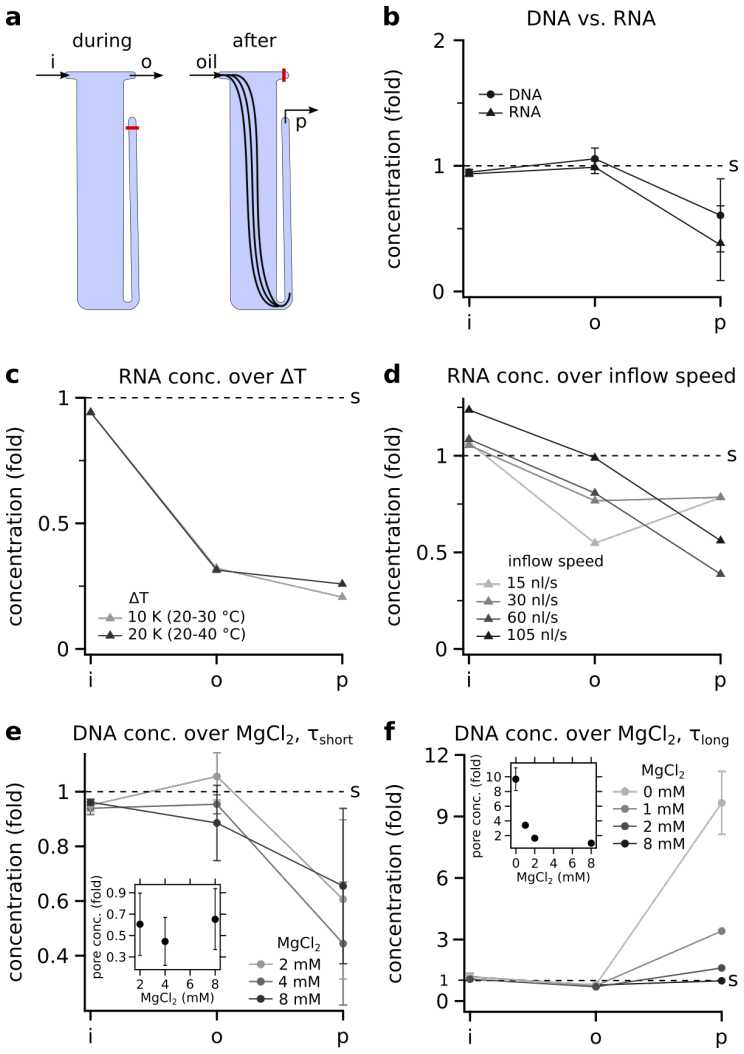

Figure 3.7: Concentrations in flow-through experiments relative to stock. a, Illustration of sample collection for a flow-through experiment. During the experiment, sample is continuously collected from the outlet (o). After the run, sample still retained in the chamber (p) is recovered through the bottom outlet. Oligonucleotides are at 4 µM in 0.1x PBS in all experiments. Note that the x-axis does not correspond to the trap height and the lines are merely a guide to the eye. Concentrations are plotted relative to stock (dashed line) in flow-through chambers for **b,** 60 nt DNA and RNA at 2 mM $MgCl_2$, 20 K (10-30 °C), 60 nl/s after 4 hours, **c,** 60 nt RNA at various ΔT, 2 mM $MgCl_2$, 30 nl/s after 2.5 hours, **d,** 60 nt RNA at various inflow speeds, 2 mM $MgCl_2$, 20 K (40-60 °C) after 4 hours, **e,** 60 nt DNA at various $MgCl_2$ concentrations, 20 K (10-30 °C), 60 nl/s, 52.25 mm chamber on a short timescale τ_{short}, after 4 hours. The inset shows pore versus $MgCl_2$ concentration. **f,** 140 nt DNA at various $MgCl_2$ concentrations, 20 K (30-50 °C), 30 nl/s, 32.25 mm chamber a long timescale, τ_{long}, after 20 hours. The inset shows pore versus $MgCl_2$ concentration. Error bars correspond to N = 2 independent experiments.

and chamber height are different, they all make the difference if only more distinct, since they all influence the trapping in the same way as a longer timescale. For τ_{short} (panel e) the pore concentration is still below the stock and the inlet concentration, which is because the chamber is initially water-filled. However, for τ_{long} (panel f) this effect is eliminated and an up to ten-fold concentration increase in the pore sample is observed. From this point on, the concentration in the pore will keep increasing until eventually, the same amount of material enters and leaves the chamber by the flow-through. Therefore, it has to be differentiated between two timescales: τ_{short}, which is still strongly influenced by the initial conditions and τ_{long}, which is en route to a steady state.

There is good agreement between simulation and experimental data and general trends could be confirmed. The inherently different readout methods however, pose difficulties in directly drawing conclusions from simulation data. In the experiments, the pore sample (p) averages the concentration over a large fraction of sample retained in the chamber, while the simulation provides a much better spatial resolution. Therefore, the experiments are less sensitive to small parameter changes and need longer timescales to achieve robust trapping. If timescales are short, intermediate effects might dominate the properties of the system before it moves towards its steady state.

3.6. Phenotype-selective trapping

In the following, the phenotype of different samples after exposure to a thermal gradient is analyzed for some of the experiments for which the concentration after trapping was characterized in Figures 3.6 and 3.7. All experiments were conducted at 4 µM oligonucleotide concentration in 0.1x PBS and each oligonucleotide sample consisted of a fully random pool of a given length.

3.6.1. Mean free energies of in silico generated random pools

First, DNA and RNA random pools of different lengths were generated in silico to determine the mean free energy of different pools. RNAfold was used to calculate the minimum free energy secondary structures and the corresponding minimum free energy dG for the synthetic pools. Figure 3.8 shows the minimum free energy distributions (Gibbs free energy) of the random pools as well as the mean free energy of each pool. The minimum free energy calculated by RNAfold is a measure for how well any given sequence can fold onto itself and is therefore also sometimes referred to as the folding energy. Generally, RNA exhibits lower free energies than DNA and the free energy decreases with increasing length.

For the experiments, three random pools were chosen: a 60 nt RNA pool, a 60 nt DNA pool and a 140 nt DNA pool. The length of 60 nt is well in the range of known ribozymes and phenotype-selective effects on this length scale are therefore interesting. The 140 nt DNA pool is chosen to match the folding energy of the 60 nt RNA pool. Considering that the sequence space of a 60 nt random pool is $4^{60} \simeq 10^{36}$, which corresponds to $\simeq 10^{13}$ mol,

it is clear that in the experiments the sequence space is undersampled by many orders of magnitude.

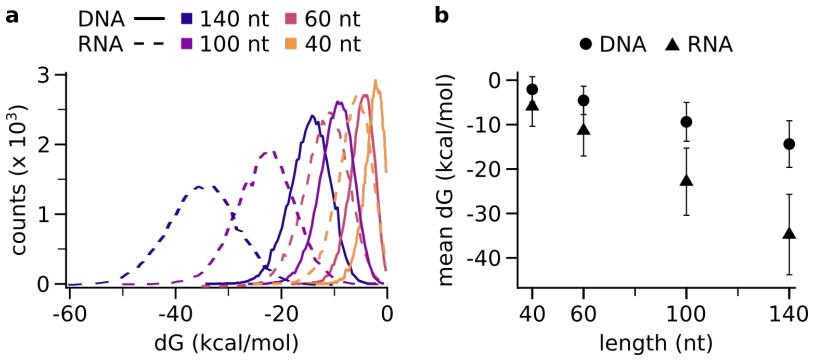

Figure 3.8: Mean free energies for synthetically generated random pools. a, DNA (solid lines) and RNA (dashed lines) pools of random sequence were generated synthetically for different lengths. RNAfold was used to calculate the minimum free energy secondary structure for randomly picked strands from each pool. The distribution of the minimum free energy of each pool is shown. **b,** Mean values for the minimum free energy are given for each pool. The mean is extracted from the distributions in panel a by fitting a Gaussian function. Error bars indicate the fitted Gaussian width σ.

3.6.2. Determination of phenotype from native PAGE

Native PAGE as well as Next-generation sequencing (NGS) are used to study the phenotype of samples after thermal trapping. Whenever native PAGE analysis is used, denaturing gels were run in parallel to exclude that hydrolysis or other effects influenced the sample composition. To be able to correlate native PAGE and NGS results, an untreated random pool of 60 nt DNA strands was first analyzed using native PAGE.

The gel, which is shown in Figure 3.9a, displayed several distinct bands. Thus, the random pool consisted of several fractions, which migrated through the gel at different speeds, indicating different degrees of compactness. When those fractions were recovered from the gel individually and sequenced, a shift in the mean free energy between the fractions was observed (Figure 3.9b). The slowest migrating fraction was found to have the highest mean free energy, while the fastest migrating fraction had the lowest. Therefore, the migration speed in native PAGE and the folding energy of an oligonucleotide strand are directly correlated. Strands that exhibit a more compact fold with a lower mean free energy migrate faster in native PAGE.

To be able to compare the migration speed and thus determine a relative change in phenotype between samples extracted from the thermal trap, images from native PAGE had to be analyzed quantitatively. How normalized intensity traces are obtained from native PAGE images is shown in Figure 3.10. First, the lane intensities from native PAGE are plotted using a custom LabVIEW tool (see Section 3.8.4). Because the same amount of oligonu-

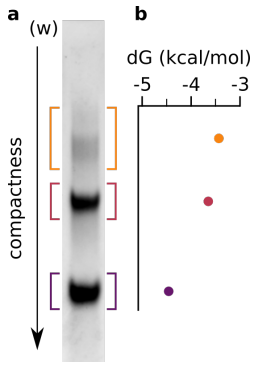

Figure 3.9: Correlation of native PAGE and mean free energy. **a,** Native PAGE (8 % acrylamide, 10 mM $MgCl_2$ and 1x TB buffer) of an untreated random pool of 60 nt DNA strands in 10 mM $MgCl_2$ and 0.1x PBS shows that the pool consists of different fractions, which migrate at different speeds from the well (w) through the gel. DNA strands from the three fractions (as indicated) are recovered separately using a gel recovery kit and are sequenced subsequently. **b,** Calculating the mean free energy from the NGS data shows that strands which, migrate faster have a lower mean free energy, i.e. are more likely to fold onto themselves. Therefore, a mobility shift in native PAGE analysis corresponds to a change in mean free energy and thus secondary structure.

cleotide sample is loaded in each lane, the intensity traces are internally normalized to 1. Then, the main band of the gel is determined, which divides the gel into three sections: the main band itself and two sections that migrate slower or faster relative to the main band. This allows the coarse division into secondary structure motifs that are less or more compact respectively. Integration over those sections, normalized by a reference will later provide the relative frequency of those coarsely classified secondary structure motifs.

3.6.3. Selection for compact folds by static thermal trapping

In static thermogravitational trapping a clear phenotype-based selection from a random oligonucleotide pools can be observed. A 60 nt RNA pool and a 140 nt DNA pool, both containing 1 mM $MgCl_2$, were exposed to a temperature gradient of $\Delta T = 15$ K. The stock sample was split into two fractions and incubated at the boundary temperatures of the thermal gradient (20 °C and 35 °C) for the duration of the experiment. After thermal trapping, samples were retrieved by freeze extraction yielding samples I-IV (from top to bottom) for each chamber. The concentration profile of the 60 nt RNA pool and 140 nt DNA pool after trapping is given in Figure 3.6 and shows an up to 2-fold concentration increase of the 60 nt RNA and an up to 4-fold concentration increase of the 140 nt DNA in the bottom-most fraction (IV).

The extracted samples were analyzed using native PAGE and NGS with regard to their phenotype. The results are shown in Figure 3.11, panels a and b for the 60 nt RNA pool and 140 nt DNA pool respectively. All extracted samples as well as the corresponding stock

were first diluted to the same concentration, then analyzed in native PAGE (Figure 3.11, (i)) and subsequently subjected to quantitative analysis as illustrated in Figure 3.10. There, a coarse subdivision into more and less compact secondary structure motifs, that run faster or slower than the main band respectively, is carried out. Integrating over those sections and normalizing by the stock sample yields the relative frequency of a certain motif and thus indicates the structural change relative to the stock (Figure 3.11, (ii)). In addition, the extracted samples are also sequenced. The mean free energy of each sample is calculated from the NGS results using RNAfold (Figure 3.11, (iii)).

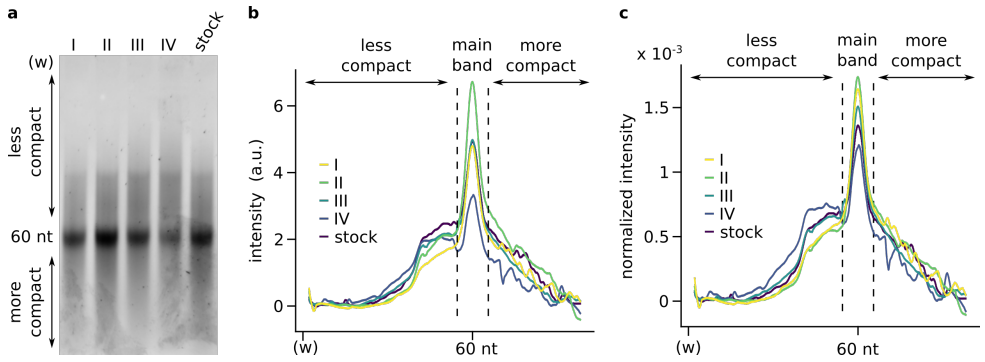

Figure 3.10: Quantitative analysis of native PAGE. The quantification of native PAGE is shown exemplary for **a,** a gel of a 60 nt random RNA sample (identical to Figure 3.11a) after thermal trapping. **b,** A custom LabVIEW tool is used to plot the intensities from each gel lane against a uniform x-axis. In the process, well (w) and main band of all traces are aligned. The custom LabVIEW tool is described in more detail in Section 3.8.4. Subsequently, position and width of the main band are determined, as indicated by the dashed lines. This divides the gel into three sections: the main band itself and the two sections left and right of it (top and bottom in the gel image of panel a), which run slower or faster respectively and thus contain less or more compact oligonucleotides. **c,** Because the same amount of oligonucleotide sample is loaded in each lane, each trace is internally normalized to 1, to quantitatively compare different samples from the same gel run.

While it is only marginally recognizable by eye from the native PAGE images, subpanels (i) in Figure 3.11, the quantitative analysis, subpanels (ii), shows clearly that for both, the 60 nt RNA and the 140 nt DNA pool, there is an intensity increase in the less compact section towards the bottom of the chamber. Correspondingly, at the top of the chamber, there is an intensity increase in the more compact section. This implies that thermal trapping causes the oligonucleotide strands of a random pool to spatially separate in the chamber depending on their compactness. The resulting distribution contains more compact folds at the top and less compact folds at the bottom of the chamber.

To understand this effect it is instructive to revisit the considerations on thermophoretic mobility. While all strands have the same length, those that exhibit less compact folds have a larger hydrodynamic radius and are thus trapped better. In turn, those strands that exhibit more compact folds do not trap equally well and become over-represented in the top part of the chamber.

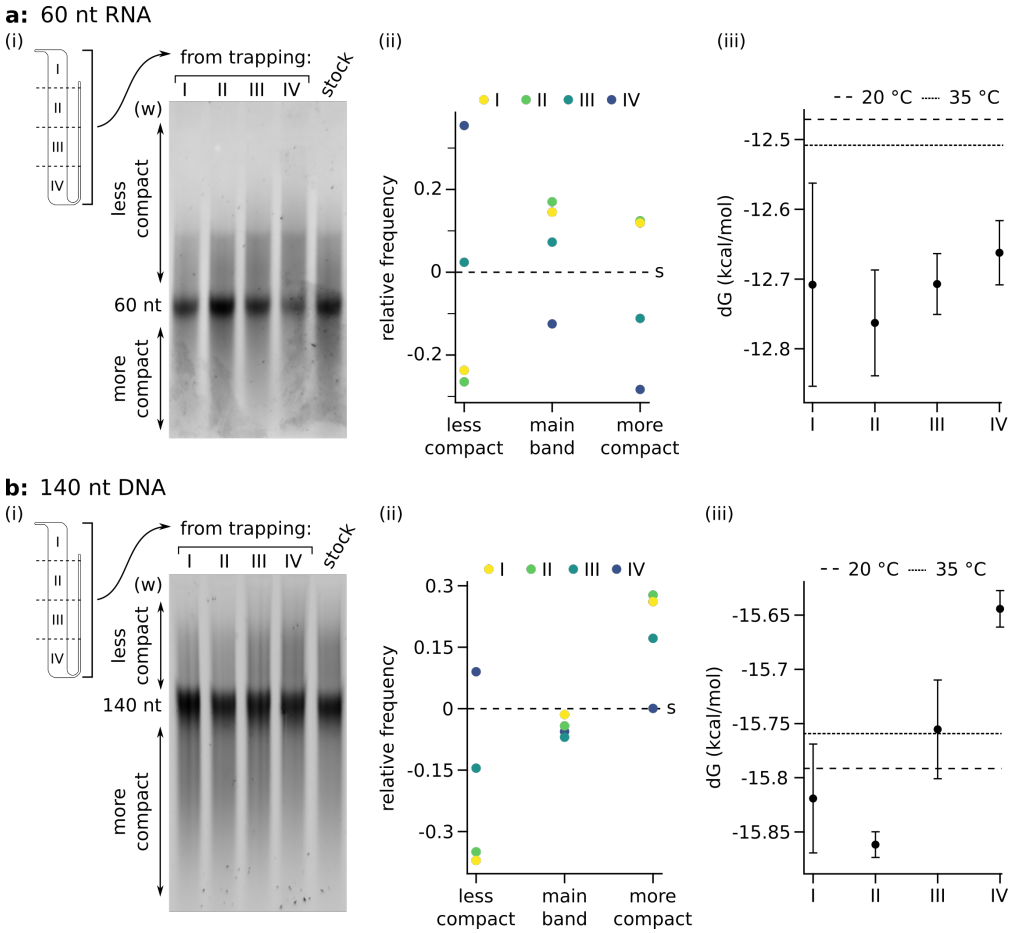

Figure 3.11: Phenotype-selective enrichment of random DNA and RNA pools by static thermogravitational trapping. a, A random 60 nt RNA and **b,** a random 140 nt DNA pool, both containing 1 mM $MgCl_2$, were exposed to a temperature gradient of $\Delta T = 15$ K in a static thermogravitational chamber for 18 hours. The concentration profile after freeze extraction (yielding fractions I to IV, from top to bottom) is shown in Figure 3.6. (i) Samples from thermal trapping of the respective oligonucleotide pool and the corresponding stock are analyzed using native PAGE. (ii) The intensity in native PAGE is integrated within the less compact, main and more compact sections (as indicated in (i) and Figure 3.10) and normalized against the stock samples (dashed lines). This yields the relative frequency of the different secondary structure motifs in each section relative to the stock. (iii) The mean free energies of the extracted samples are calculated from NGS data. The dashed and dotted lines indicate the stock samples, which are incubated at the boundary temperatures of 20 °C and 35 °C respectively. Native PAGE analysis shows the for both oligonucleotide samples, the top-most (I) sample is enriched in more compact secondary structures and accordingly depleted in less compact ones. Correspondingly, the bottom-most sample (IV) exhibits opposed behavior. Sequences with higher free energy are found at the bottom of the chamber and the mean free energy generally increases within the chamber from top to bottom, except for the top-most sample (I). In both, panel a and b, the signal from native PAGE and NGS data are concurrent. Error bars indicate N = 3 independent experiments.

This phenotype-based selection is confirmed by the NGS results shown in Figure 3.11, subpanels (iii). Leaving aside the top-most fraction for the moment, the mean free energy distinctly increases towards the bottom of the chamber. This again suggests, that strands at the bottom fold poorly and thus feature more bulky secondary structures compared to strands further at the top, that are more compact. The top-most fraction (I) diverges from this trend because due to diffusive pull, fresh sample is continuously drawn in from the sample-filled tubing, which is connected at the very top. This leads to an effective 'dilution' with stock sample in the top-most fraction (I). This effect is also visible in the concentration profile shown in Figure 3.6, where the top-most fraction (I) is not depleted as strongly as the fraction second to top (II). The volume stored in the tubing is at most 10 % of the trap volume, so the effect is small and should disappear on very long timescales.

3.6.4. Dynamic selection in flow-through trapping

Random DNA pools were also subjected to thermogravitational trapping in a flow-through system. Since the feeding flow continuously provides fresh material that can be absorbed into the chamber or pore, this promises to provide an even more potent selection than observed in static thermogravitational trapping.

Flow-through experiments were performed using a 60 nt DNA pool with 8 mM $MgCl_2$ and a 140 nt DNA pool with 2 mM $MgCl_2$. The pools were exposed to a thermal gradient of $\Delta T = 20$ K for 4 and 20 hours respectively. The stock samples were incubated at room temperature for the duration of the experiment. During and after trapping, inlet (i), outlet (o) and pore (p) samples were extracted from the chamber as illustrated in Figure 3.7a. Their concentrations relative to the corresponding stock are shown in Figure 3.7e,f for the 60 nt pool and the 140 nt pool respectively.

The analysis with regard to phenotype is again performed with native PAGE and NGS, following the identical workflow as in the static case (Section 3.6.3). The results are shown in Figure 3.12, panel a for the 60 nt DNA pool after trapping for 4 hours, τ_{short} , and panel b for the 140 nt DNA pool after trapping for 20 hours, τ_{long}.

In contrast to the static experiments, the flow-through experiments exhibit a more differentiated picture. When comparing panels a and b of Figure 3.12 it is easy to see that the two experiments show opposing outcomes. For the 140 nt DNA pool (panel b) more compact secondary structures are less efficiently trapped and thus enriched in the outlet sample, just as seen in the static case (Figure 3.11). However, for the 60 nt DNA pool (panel a) the opposite behavior is observed: more compact secondary structures are more frequently found in the pore sample and are underrepresented in the outlet sample. For both cases, the signal from native PAGE and NGS are in agreement.

To understand why those two flow-through experiments deliver such opposite outcomes, it is instructive to return to the consideration of timescales. An effect of the oligonucleotide length can be excluded, since in the static case oligonucleotide pools of different lengths behaved analogously (Figure 3.11). It was discussed before, that in flow-through trapping the experimental timescales are relevant especially with regard to the influence of the initial conditions. Due to the initially water-filled chamber, the concentration in the pore stays be-

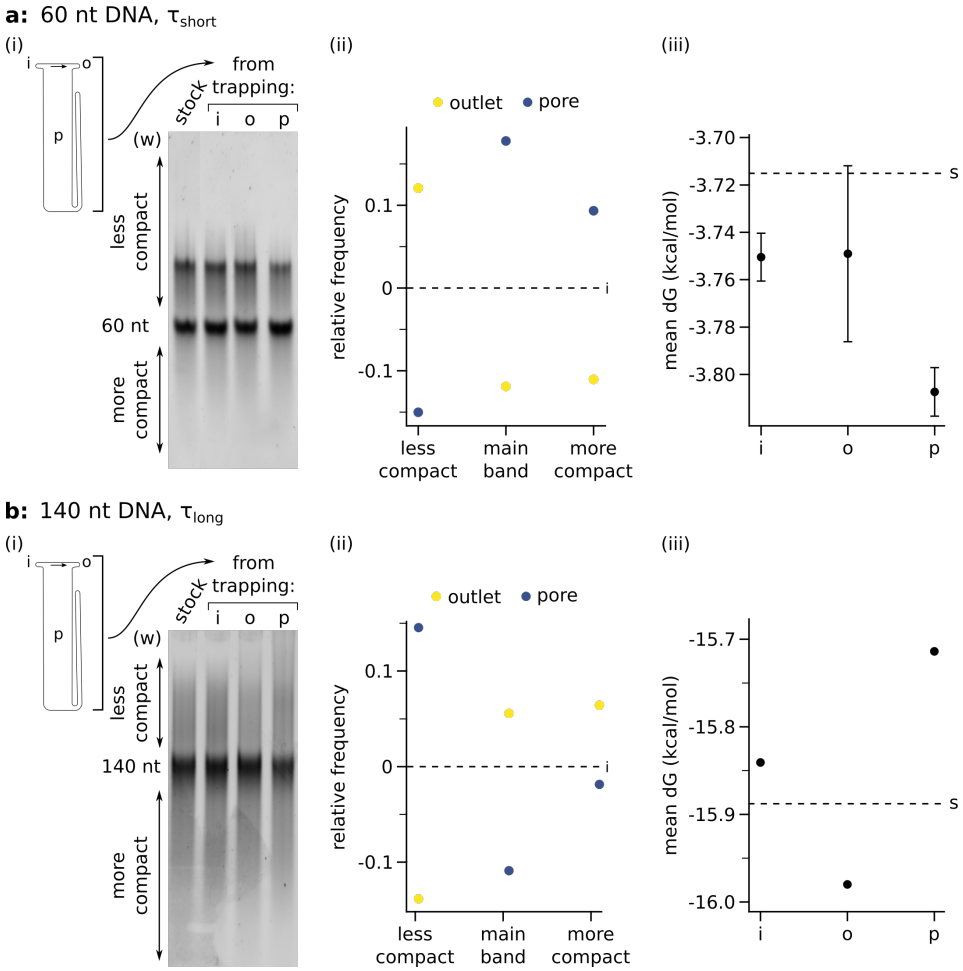

Figure 3.12: Phenotype-selective enrichment of random DNA pools by flow-through thermogravitational trapping. Random DNA pools of **a,** 60 nt with 8 mM $MgCl_2$ and **b,** 140 nt with 2 mM $MgCl_2$ were exposed to a thermal gradient of $\Delta T = 20$ K in a flow-through chamber. Inlet (i), outlet (o) and pore (p) samples are collected after 4 and 20 hours of thermal trapping respectively. Their concentration relative to the stock is shown in Figure 3.7e,f for the 60 nt pool and the 140 nt pool respectively. (i) Samples from thermal trapping of the respective DNA pool and the corresponding stock are analyzed using native PAGE. (ii) Frequency of secondary structure motifs in outlet and pore relative to the inlet (dashed line) as calculated from native PAGE (i). (iii) The mean free energies of the extracted samples are calculated from NGS data. The dashed lines indicate the mean free energies of the stock samples. In panel a, native PAGE analysis shows that the outlet sample is depleted in more compact secondary structures and enriched in less compact ones. The opposite behavior is observed for the pore sample. This is confirmed by the NGS data, which shows that the pore sample contains sequences with the lowest free energy. In panel b, the outlet sample is depleted in less compact secondary structures and enriched in more compact ones. Again, the correspondingly opposite behavior is observed for the pore sample. This is confirmed by the NGS data, which shows that the outlet sample contains sequences with lower free energy that form more compact folds compared to the pore sample. Error bars correspond to $N = 2$ independent experiments.

low those of inlet and stock on short timescales. This initial phase of populating the chamber with oligonucleotide strands happens, firstly, by uptake in the convective flow and secondly, by diffusion. While convection acts on all molecules equally, diffusion is size dependent and smaller molecules spread further in the same amount of time. Therefore, strands that exhibit a more compact secondary structure will initially populate the chamber overproportionally during a diffusion dominated timescale.

In summary, thermogravitational trapping in a flow-through system introduces a phenotype-based bias into random oligonucleotide pools. We observe a dynamic effect, where the diffusion-driven enrichment of compact oligonucleotides prevails on short time-scales and is later out-competed by the stronger thermogravitational accumulation of less compact oligonucleotides. This turnaround underlines the potential of phenotype-selective trapping in flow-through systems.

3.7. Discussion

The prebiotic synthesis of RNA molecules has quite recently experienced such considerable advances [150, 151] that pessimism regarding the prospects of experimental support for the *RNA world* hypothesis [87, 140] has predominantly been dispelled. However, it still lies largely in the dark, how functional sequences such as ribozymes emerged from prebiotically generated oligonucleotides, given that such specific sequences would have been present at extremely low concentrations in a random pool [73, 180].

In this chapter, I approached this question and presented a mechanism that can act as a prebiotic filter on random oligonucleotide pools. A thermal gradient can pose a physical selection pressure on such pools and is able to select for or against certain oligonucleotide motifs based on the compactness of their phenotype. Such compact phenotypes are often a signature of ribozymes [137,169] and have been associated with greater functional potential in in vitro evolution [28]. Even though the observed shift in mean free energy induced by the thermal gradient is relatively small, up to ½ k_BT, it emerges consistently for both DNA and RNA of different lengths. Thus, it is a promising starting point for further experiments to explore phenotypic enhancement of random oligonucleotide pools.

Especially the flow-through system still needs better experimental characterization. So far, only a scarce data set is available. Often, the flow leads to experimental complications, insufficient amounts of sample are recovered from the pore or one of the three downstream analyses (NanoDrop, native PAGE, NGS) fails, which renders the experiment useless for further evaluation. Notably, the investigation of RNA in a flow-through system as well as the examination of the influence of $MgCl_2$ concentrations would be of interest.

In this work, Next-generation sequencing was used to determine the mean free energy of a given pool, which was in turn used as a measure for the compactness of a secondary structure, largely following the approach of Chizzolini *et al.* [28]. The underlying rationale was that a better fold implies a lower free energy and vice versa. However, here it is necessary to reinterpret the term 'better' for the presented selection mechanism. Following the hypothesis that the selection pressure is exerted on the basis of a molecule's size it would

be a more suitable approach to use the NGS data to calculate the hydrodynamic radius of the most common secondary structure for a given sequence and compare the mean radii of different samples. In this context, the importance of branching structures ensuing from partial self-complementarity for the correct determination of the hydrodynamic radius has been pointed out [15, 58]. For future analysis, an approach such as proposed by Yoffe *et al.* [198] could be used. The authors approximate an oligonucleotide secondary structure by its 'maximum ladder distance', adding up all double-stranded sections and treating them as a stiff rod, while either ignoring or attributing very little size to floppy, single-stranded sections. Such an effective contour length for each secondary structure could be calculated from the dot-bracket notation provided by RNAfold.

Another exciting prospect for future work is the amplification of selection by reinsertion of a previously enriched pool. In comparison to simply running the experiment on longer timescales, reinsertion has higher potential for increased selectivity. Generally, in thermogravitational trapping reinsertion is technically challenging but would be feasible especially for the static case considering that a concentration increase of approximately four-fold is reached in the bottom-most quarter of the chamber (Figure 3.6). This is a technique also commonly used in in vitro evolution, where several rounds of selection are performed to yield improved functionality. Considering the results from Chizzolini *et al.*, who used six selection cycles, it is reasonable to believe that reinsertion will indeed enhance the overall phenotypic selectivity of thermogravitational trapping. After a few rounds of selection, a more thorough analysis of secondary structure motifs as well as their classification and a comparison with known ribozyme motifs could be undertaken.

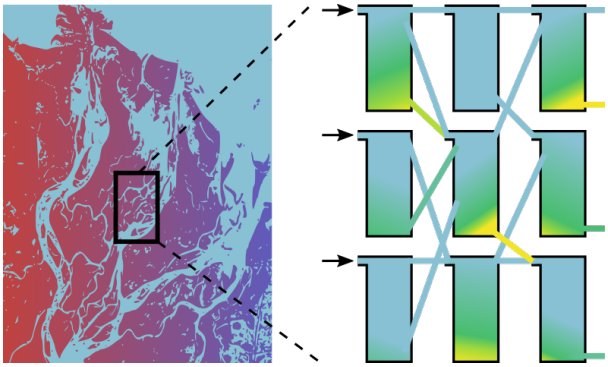

Figure 3.13: Illustration of a network of interconnected pores. Rock fissures on early Earth, e.g. in submerged volcanic rocks, did not only include a singular pore but consisted of a whole network of many interconnected pores. Large scale convection induces drift flows that can transport material from one pore to another. Such a network of interconnected pores represents a plausible scenario for repeated cycles of sample reinsertion. Figure adapted from [117].

In a prebiotic context, one could even imagine a whole network of rock pores, which are interconnected by branching channels as illustrated in Figure 3.13. A drift flow through the network, e.g. induced by large scale convection, would automatically lead to reinsertion of

material from one pore to another. Because the number of reinsertion cycles is experimentally limited to a low digit number, simulating such a network is an appealing idea. The relative concentration increase for different hydrodynamic radii could be calculated using finite element simulations, additionally sweeping over other parameters such as temperature gradient or inflow speed (similar to Figure 3.5). An algorithm for reinsertion could then randomly pick parameters for the thermal gradient, inflow speed or also the number and position of in- and outlets through which the material is transferred and calculate the concentration distribution for different hydrodynamic radii after exposure to a whole network of pores.

The idea that oligonucleotide secondary structures might be more or at least as important as their precise sequence is not entirely new. It is well known that ribozymes assume intricate secondary structures and rely on native folds to execute their function. However, the same holds true for non-coding RNAs such as transfer RNA (tRNA) or ribosomal subunits. There, it has been shown that in order to maintain their function a conservation in phenotype is often more substantial than a conservation of genotype [14, 68, 201]. This provides an interesting perspective and raises the question whether or to what extent specific ribozyme phenotypes, known from naturally occurring ribozymes, could execute catalytic functions independent of their genotype. This is especially interesting given the potential of the presented mechanism to promote the emergence of alike, compact phenotypes, irrespective of their genotype in a prebiotically plausible setting.

3.8. Materials and methods

3.8.1. Sample preparation

All experiments were performed at an oligonucleotide concentration of 4 µM in 0.1x PBS (13.7 mM NaCl, 0.27 mM KCl, 1 mM Na_2HPO_4, 0.18 mM KH_2PO_4) at varying magnesium concentrations. RNA samples contained a 5'-phosphate to allow subsequent sequencing. Oligonucleotides were purchased from Biomers, Germany and stock concentrations were quantified by UV-absorption measurements using a NanoDrop™ (ThermoFisher, Germany).

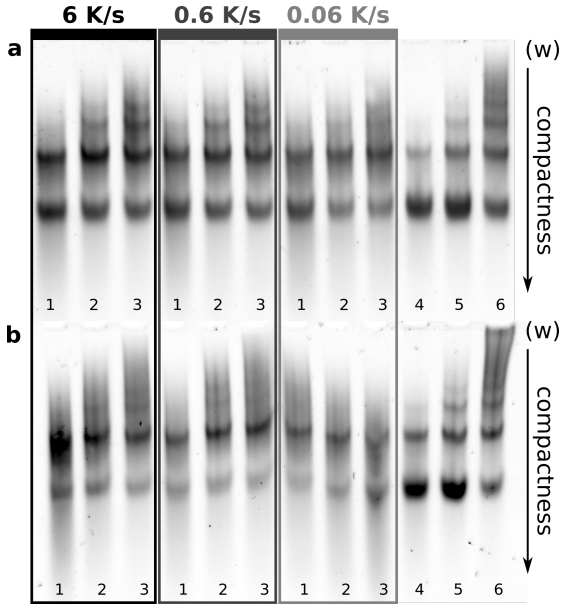

Figure 3.14: Comparison of different temperature protocols for sample preparation by native PAGE. A random 60 nt RNA sample was mixed at 20 µM with **a,** 2 mM $MgCl_2$ and **b,** 5 mM $MgCl_2$ and subsequently exposed to different temperature protocols. All samples, except (5), were first heated for 2 minutes and then cooled quickly to 70 °C. Then, the samples were incubated for 18 hours at (1) 60 °C, (2) 50 °C and (3) 40 °C with different ramp speeds (as indicated) to reach the final incubation temperature. Lanes (4) were quenched on ice after the initial heating step. Lanes (5) and (6) were incubated at room temperature (∼ 25 °C) for 18 hours without (5) and with (6) the initial heating step.

The impact of different temperature protocols during sample preparation was tested for a 60 nt RNA random pool under different magnesium concentrations. The samples were initially heated to 80 °C for two minutes to ensure complete melting. Then, they were either quenched on ice or annealed to their incubation temperature. The results from native PAGE, shown in Figure 3.14, confirmed that the chosen temperature protocol largely influences secondary and tertiary structures in the sample. Interestingly, slow annealing did not result in larger secondary structures and simple incubation at room temperature showed similar

results as the more elaborate annealing protocols. However, in order to always provide uniform starting conditions, all experiments were conducted with a freshly quenched sample (93 °C heating for DNA). Though, in flow-through experiments runtimes can be as high as 20 hours, during which the sample is effectively incubated at room temperature in the inlet tubing.

3.8.2. Sample extraction

After the trapping experiment, samples are extracted from the chamber. To simplify sample extraction, two tubings, a short one coming directly from the chamber and a much longer one containing more than 95 % of the sample, are concatenated. The inlet sample (i = inlet) is collected from the leftover sample in the inlet tubing. In a topflow chamber, the outlet sample (o = outlet) is recovered from the sample collected in the outlet tubing. If the chamber content shall be extracted by freeze extraction, the short tubings, which are directly connected to the chamber, are closed off and the chamber is shock-frozen to -80 °C. Otherwise, low viscosity oil can be used to recover the remaining volume from within the chamber (p = pore) through the bottom outlet. To do so, all other outlets are closed and oil is carefully flushed in through the inlet. During the whole extraction process, extreme caution must be exercised such that never more than one in- or outlet is open to avoid additional flows. The sample extraction is illustrated in Figure 3.7a. The stock sample (s = stock) is stored at room temperature for the duration of the trapping experiment. After sample extraction, each sample is split into three to parts to determine the concentration and proceed with native PAGE and NGS analysis.

3.8.3. Gel electrophoresis

Native polyacrylamide gel electrophoresis (PAGE) was performed with 0.75 mm thick gels containing 8 % acrylamide (acrylamide:bisacrylamide 37.5:1), 10mM $MgCl_2$ and 1x TB buffer. Samples were mixed 5:1 with 6x loading dye (4 g sucrose, 25 mg Orange G (New England Biolabs, Germany) in 10 ml volume). Gels were run in 1x TB buffer in a two-step protocol with electric field strengths of first 50 V/cm for 5 minutes and then 300 V/cm for 28 minutes at 8-14 °C.

Denaturing polyacrylamide gel electrophoresis (PAGE) was performed with 0.75 mm thick gels containing 15 % acrylamide and 50 % urea. Samples were mixed 1:1 with 2x loading dye (9.5 ml formamide, 0.5 ml glycerol, 1 µl EDTA (0.5 M), 100 µl Orange G (New England Biolabs, Germany)). Gels were run in 1x TBE buffer in a two-step protocol with electric field strengths of first 50 V/cm for 5 minutes and then 300 V/cm for 28 minutes at 45-50 °C.

Gels were loaded with 3-5 µL of 0.05-0.25 µM oligonucleotide sample and empty wells were filled with the same volume of 1x loading dye. All samples run within one gel were equimolar. Gels were post-stained in 50 ml 1x TBE buffer with 2x SYBR Gold (Invitrogen, Germany) for 5 minutes, followed by two washing steps in 1x TBE buffer. Imaging was performed using a Bio-Rad ChemiDoc MP System (Bio-Rad, Germany).

3.8.4. Native PAGE quantification

A custom LabVIEW tool was used for quantification of native PAGE from the unprocessed gel image. By manually adjusting four points (ROI), a number of signal (black) and background (red) lanes are projected onto the gel image as displayed in Figure 3.15. Lanes that do not contain sample can be faded out (dashed lines). For each lane, the average intensity between the black lines is plotted along the y-axis. The resulting intensity plot is then background subtracted using the average of the background intensity from the two adjacent background lanes. Data and background are smoothed with a rolling average algorithm over 8 pixels. Furthermore, the peaks in the intensity plots are aligned at the well and at the main band. To simplify further analysis, all intensity plots from one gel are plotted against a uniform axis.

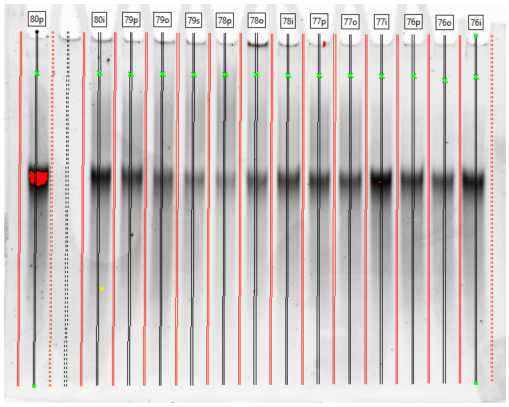

Figure 3.15: Image from custom LabVIEW tool. The unprocessed gel image is loaded into a custom LabVIEW tool and annotated. Lines marking the signal (black) and background (red) areas are projected onto the gel image. Dashed lines indicate lanes that did not contain sample.

3.8.5. Recovery from native PAGE

Sections of a native PAGE gel were recovered using the ZR small-RNA PAGE Recovery Kit (Zymo Research, Germany). Native PAGE and imaging were first performed as usual. Under constant imaging control, sections of interest were cut out from the gel using a scalpel. The sample from the gel sections was then recovered according to the manufacturer's protocol.

3.8.6. Library preparation and Illumina sequencing

Prior to library preparation, sample concentration was determined by UV-absorption. The required sample input concentration for library preparation kit is usually given in ng/µl. However, because the concentration of enzymes in the kit is optimized for a certain number of

5'/3'-ends, the important measure for the sample input is not mass per volume but molarity, which should for example be ∼ 20 nM for the Accel-NGS® 1S Plus DNA Library Kit.

The library preparation for DNA samples was performed using the Accel-NGS® 1S Plus DNA Library Kit (Swift Biosciences, USA). The protocol provided by the manufacturer is scaled down to one quarter of enzyme, buffer and index volumes and filled up with water.

The library preparation for RNA samples was performed using the NEBNext® Multiplex Small RNA Library Prep Kit for Illumina® according to the manufacturer's protocol. After library preparation, the library was purified using bead purification (AMPure XP, Beckman Coulter, Germany) with 1.2x capture ratio.

After library preparation, capillary gel electrophoresis is performed for quality control. The samples are loaded onto a High-sensitivity DNA Bioanalyzer chip (Agilent, Germany) and run according to the protocol provided by the manufacturer. The main criterion for the quality check is whether there is material of the desired length (typically original length plus 120 nt) and whether it exceeds the amount of primer-dimers and other side products.

Illumina HiSeq 2500 and NextSeq 500 machines were used for sequencing of DNA and RNA samples respectively.

3.8.7. Sequence analysis

Reads from the sequenced library were demultiplexed and quality-filtered. During quality-filtering, a minimal quality value of 20 with an accuracy of 90 % was required. Subsequently, cutadapt [112] was used to trim the reads to the expected length.

RNAfold (Section 3.2.2) was used to calculate the minimum free energy secondary structure and the corresponding dG for each sequence within each sample. The dG distributions for each sample were then fitted with a Gaussian function to find mean and width of the distribution.

4. tRNA-like sequences can assemble into a replicator

Summary

Can replication and translation emerge in a single mechanism via self-assembly? The key molecule, transfer RNA (tRNA), is one of the most ancient molecules and contains the genetic code. Our experiments show how a pool of oligonucleotides, adapted with minor mutations from tRNA, spontaneously formed molecular assemblies and replicated information autonomously using only reversible hybridization under temperature oscillations. The pool of cross-complementary hairpins self-selected by agglomeration and sedimentation. The metastable DNA hairpins bound to a template and then interconnected by hybridization. Temperature oscillations separated replicates from their templates and drove an exponential, cross-catalytic replication. The molecular assembly could encode and replicate binary sequences with a replication fidelity corresponding to 85-90 % per nucleotide. The replication by a self-assembly of tRNA-like sequences suggests that early forms of tRNA could have been involved in molecular replication. This would link the evolution of translation to a mechanism of molecular replication.

This chapter was published by Kühnlein *et al.* [98] in *eLife* and is adapted and reprinted in part with permission from eLife.
Authors: A. Kühnlein*, S.A. Lanzmich*, D. Braun.
Contributions: A.K.: Conceptualization, Data curation, Formal analysis, Visualization, Methodology, Writing - original draft, Writing - review and editing; S.A.L.: Conceptualization, Data curation, Software, Formal analysis, Visualization, Methodology, Writing - original draft; D.B.: Conceptualization, Software, Funding acquisition, Methodology, Writing - original draft, Writing - review and editing.

4.1. Introduction

Modern life is described by the central dogma of molecular biology: genetic information is stored in DNA, which is transcribed into intermediate RNA, from where it is translated into proteins following the genetic code [32, 33]. The replication of DNA is performed by proteins but in turn proteins cannot be made without DNA. This constitutes a well-known chicken-and-egg paradox within the origin of life, that poses the question: what came first? The almost logical answer to this is the assumption that long ago all tasks, which are today distributed across the three dominant biopolymers of life, were united in one molecule species [152]. The *RNA world* hypothesis [61] suggests that this species was RNA, since it can store genetic information but also has catalytic abilities. Such catalytically active RNA molecules, ribozymes, could among other reactions catalyze their own synthesis, which would have made them the central molecules of early life. Especially the discovery that the ribosome is indeed a ribozyme [24, 175] gave rise to the widespread acceptance of the *RNA world* hypothesis, concluding that it likely was the last predecessor of modern life [161]. As described and discussed in Chapter 3, the path to the emergence of functional motifs and eventually ribozymes is so far only partially elucidated.

Independent of the consideration of functional motifs and catalytic activity, at some point during the emergence of life, there must have been a molecule that was able to self-replicate, i.e. to autonomously catalyze the formation of identical copies of itself. Those copies were imperfect, which eventually allowed Darwinian evolution to kick off. Considering that strand lengths of at least 30 nt are necessary to perform self-replication [54,139,182], it remains unclear how oligonucleotides of sufficient length might have formed [140, 161]. The details of the emergence of replication and its (chronological) connection to the emergence of translation are still heavily debated [42,162,170,194]. Nevertheless, there is agreement over some of the key aspects and problems an early template-based, non-enzymatic (self-)replication mechanism must address [144, 161, 170, 180]:

- **Template inhibition.** In a template based replication mechanism, there needs to be a means of strand separation to avoid template inhibition [88, 110, 181]. Template and replicate are complementary and form a stable duplex. To avoid stalling of the replication, both template and replicate need to be made available for the next round of replication. Strand dissociation, commonly referred to as melting, can be achieved by changes in e.g. temperature, salt or pH.

- **Strand reannealing.** Provided a means for strand separation, e.g. temperature, template inhibition can still not be overcome, because the kinetics of reannealing are up to several orders of magnitude larger than the copying rate of the replicator [180].

- **High melting temperature.** Self-replication is only observed in oligonucleotides with strand lengths exceeding 30 nt [54, 139, 182], but the separation of template and replicate becomes increasingly difficult for perfectly complementary strands longer than 30 nt. Especially for RNA this is a problem, because its duplex structure is so stable [161, 180]. Furthermore, high concentrations of divalent ions, which are often

necessary for ribozyme catalyzed polymerization or ligation [8,77] would in a one-pot scenario increase the melting temperature and hinder strand separation.

- **Fast, exponential replication.** Natural population growth undergoes reproduction (growth) and degradation (decay). Exponential growth can compensate for linear degradation and is thus the preferred growth mode for Darwinian evolution [115]. When degradation cannot be implemented chemically, it is experimentally often realized through serial transfer experiments [2]. Interestingly, in a sub-exponential but super-linear growth mode, several replicators can coexist [179]. Only in an exponential growth mode, survival of the fittest is necessarily fulfilled [7, 178].

- **Replication without template.** The template should catalyze and accelerate the product formation [145]. Conversely, if no template is present, the rate of product formation should be low. This avoids unspecific product formation and also ensures a low spontaneous background rate in the presence of a template.

- **Fidelity in copying.** Even though mutation is necessary for Darwinian evolution, early replication must have had a sufficiently high fidelity in copying. The error rate of a replicator is inversely proportional to the length of the genome it can replicate [52, 178]. A replication fidelity of 80 % per base allows only for a genome of length 5 [71, 155] and an increase to 97 % per base could maintain a genome of length 30 [88, 161]. Szathmary [178] estimates that the error rate per nucleotide must have been as low as 10^{-3} to allow life to develop. Most likely, the improvement of fidelity would have been the strongest early selection pressure [161].

- **Energy source.** Through continuous growth and decay, any replicating system is in an out-of-equilibrium steady state. There needs to be a widely available energy source that enables the replication to run autonomously. Overviews of potential energy sources available on the early Earth are given by Miller *et al.* [121], and with a focus on hydrothermal vents by Martin *et al.* [114].

Some examples of non-enzymatic molecular (self-)replication have been reported using nucleic acids [8, 77, 154, 189], peptide-based approaches [103, 163] or organic molecules [157, 165]. However, many of them still suffer from at least one shortcoming regarding the above mentioned requirements for an early self-replicator.

The ability to self-replicate is not the sole requirement for a molecule involved in the origin of life and the importance of emergence of other properties such as compartmentalization or encapsulation and the development of a metabolism can hardly be overstated. But self-replication is such a dominant feature of life that it must have been present early on. Therefore, its emergence is one of the essential quests when trying to understand the origin of life.

In this chapter, I will introduce a mechanism that is able to replicate sequence information through hybridization of oligonucleotides. Replication is accomplished using sequences inspired by tRNA that exhibit a double hairpin secondary structure and is driven by temperature oscillations.

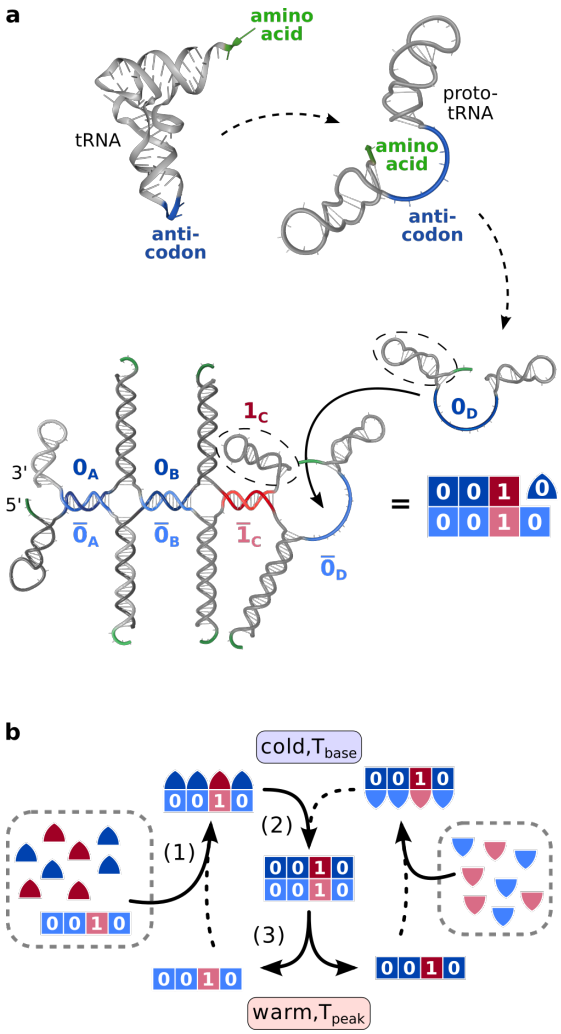

Figure 4.1: Heat-driven replication by hybridization using hairpin structures inspired from transfer RNA. a, Transfer RNA (tRNA) folds into a double hairpin conformation upon very few base substitutions. In that configuration, the 3'-terminal amino acid binding site (green) is close to the anticodon (blue) and a double hairpin structure forms. A set of pairwise complementary double hairpins can encode and replicate sequences of information. A binary code implemented at the position of the anti-codon, the information domain, allows to encode and replicate binary sequences (red vs blue). Each strand (82-84 nt) comprises two hairpin loops (gray) and an interjacent unpaired information domain of 15 nt length (blue/red, here: 0_D). The displayed structure of eight strands shows replication of a template corresponding to the binary code 0010. Note that no covalent linkage is involved in the process. **b,** Replication is driven by temperature oscillations in four steps: (0) The hairpins are activated into their closed conformation by fast cooling indicated by triangles. (1) Strands with matching information domain bind to the template. (2) Fluctuations in a bound strands' hairpins facilitate the hybridization of neighboring strands. (3) Subsequent heating splits replicate from template, while keeping the longer hairpin sequences connected, freeing both as templates for the next cycle. Figure reproduced from [98].

No covalent bonds are formed during replication. Thus, the mechanism is completely reversible and does not produce any waste products. Most importantly, it does therefore not rely on a specific prebiotic chemistry, e.g. using activated monomers or activation agents [39, 50, 154, 203], but is based on a more general physical mechanism. The replication scheme has previously been explored in a limited subset [94]. Now, binary information is encoded in four bits using a set of 16 different strands. Since the replication mechanism is independent of specific sequences or oligonucleotide lengths and merely requires a certain secondary structure, it could easily be extended to much higher complexity and its capability to encode and replicate information is not per se restrained.

4.2. Sequence design and replication mechanism

The replication mechanism is based on hybridization only and driven by temperature oscillations. It can replicate and amplify information, which is encoded in the succession of oligomers. It has previously been shown that upon a few point mutations the secondary structure of tRNA changes from the characteristic cloverleaf structure to a double-hairpin structure, where the former anti-codon in the center is flanked by a double-hairpin on each side [94]. A pool of double hairpins can be designed to be pairwise complementary, such that the 3' hairpin of one strand is complementary to the 5' hairpin of the next and can form a backbone, see Figure 4.1a. The system has a periodic boundary condition, i.e. the double hairpin of the last strand is complementary to the first one. Due to the periodic boundary condition the system size and complexity is not per se limited. Here, a periodicity of four different double hairpins was implemented.

A pool of cross-complementary strands with the desired secondary structure was designed using a computer algorithm [199]. Details of the strand design and what boundary conditions were used to achieve the desired design are given in [99].

The length of a single strand was 82-84 nt, similar to the length of a tRNA molecule [171]. The hairpins were made up of 30-33 nt each and the information encoding domain ('information domain') in the center was 15 nt long. A binary set of information domains was implemented, allowing the system to encode either 0 or 1 (blue/red in Figure 4.1) at each position. The four positions are labeled A-D and complementary information domains are 0 and $\overline{0}$ (dark/light colors in Figure 4.1). The used sequences are given in Table 4.1.

The replication mechanism is schematically shown in Figure 4.1b. Replication proceeds in four steps: (0) By quenching (i.e. fast cooling) the hairpins are brought into their closed conformation with an unpaired information domain in the center flanked by two closed hairpins (strand 0_D in Figure 4.1a). (1) At a moderate temperature T_{base}, strands with matching information domains can bind to a template (0010 in Figure 4.1b). (2) Due to thermal fluctuations at T_{base}, the hairpins alternate between an opened and closed conformation. When two strands are bound to the template at the information domain, those fluctuations allow two adjacent hairpins to hybridize. In this way, the 'hairpin backbone' of the replicator is formed and the succession of information domains is replicated. (3) A short temperature spike to T_{peak} separates template and replicate. Due to the higher melting temperature of the

hairpin backbone compared to the information domain, template and replicate stay intact and can both act as templates for the next round of replication.

The main constraint for choosing the oscillation temperatures T_{base} and T_{peak} is their relationship to the melting temperatures of the hairpins, $T_{m,hairpin}$, and the information domain $T_{m,information\ domain}$ respectively. The replication mechanism as sketched in Figure 4.1b only works, if $T_{base} < T_{m,information\ domain} < T_{peak} < T_{m,hairpin}$. During the templating phase (1,2) information domain bonds must be stable, but must break during the heating step (3) in order to separate template and replicate. The hairpin backbone must always remain stable. This constraint was considered during strand design and detailed melting curves of the distinct domains can be found in [99]. The optimal oscillation temperatures were determined experimentally (Figure 4.9).

The initial quenching step (0) induces a meta-stable state that stores energy [64]. Because of the initial fast cooling all hairpins are closed in free solution. This inhibits the formation of replicates without template. While the binding of adjacent hairpins with the respective template happens within minutes, hairpins without template in free solution connect only on timescales of hours and thus the spontaneous replication rate is very low (Figures 4.9b,c and 4.10a).

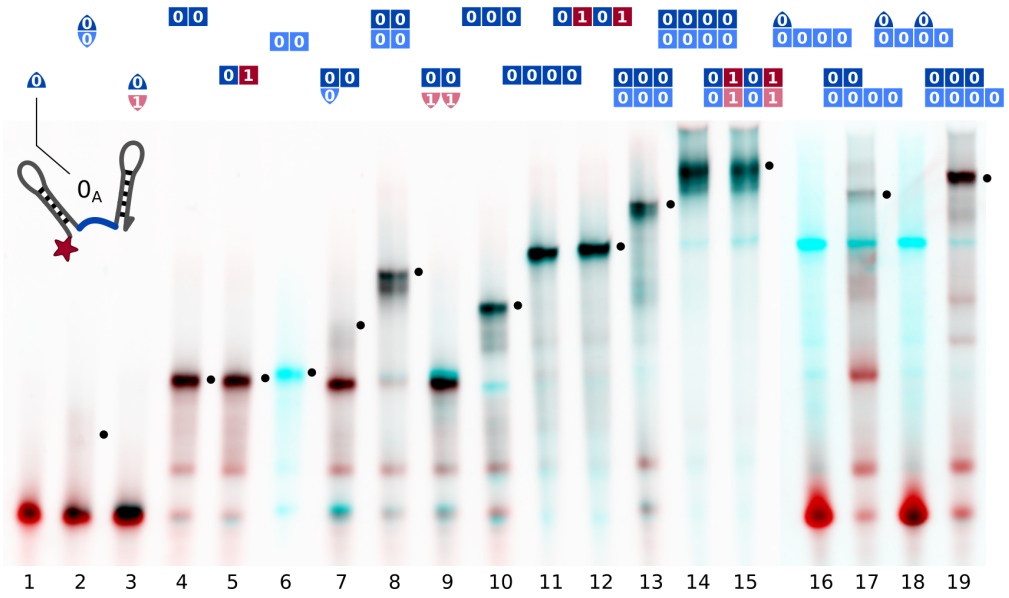

Figure 4.2: Assembly of different subsets of the cross-replicating system observed by native PAGE. Samples contained strands at 200 nM concentration each and were slowly annealed. The different subsets provided for each lane are indicated at the top of each lane. Comparison of different lanes allowed for the attribution of bands to complexes. Complexes incorporating all present strands are marked (•). The red channel shows the intensity 0_A-Cy5, the cyan channel shows SYBR Green I fluorescence. Single information domain bonds (lane 2, 7) break during gel electrophoresis. Figure reproduced from [98].

The replicator was implemented in DNA, not RNA as in the study by Krammer *et al.* [94]. However, since the mechanism relies on hybridization only it can be expected to work equally well in DNA and RNA. An RNA version of this replicator exhibits the identical, desired secondary structure (Figure 4.13), and RNA specific features such as G-U wobble pairs seem to be negligible. When implemented in RNA, oscillation temperatures and salt concentrations would need to be slightly adapted.

4.2.1. Analysis of complex formation

To confirm the predicted conformations and to analyze and quantify the template and product formation, native polyacrylamide gel electrophoresis (PAGE) was used. In contrast to denaturing PAGE (SDS-PAGE), native PAGE is also sensitive to the structure of oligonucleotide molecules or assemblies apart from their charge and length [28, 169].

Different subsets of the replicator were annealed at 200 nM per strand as described in Section 4.6.4 and analyzed with native PAGE (Figure 4.2). The replicator assembled as intended. The different assembly stages could be resolved and all bands could be assigned. By comparing lanes 14 (15), 17 and 19 it is clear that templates with two, three or four bound hairpins can be resolved. Detailed measurements on the friction coefficient of the assemblies can be found in [99].

4.3. Self-assembly of the replicator

4.3.1. Tying self-replication and self-assembly

Self-assembly is a feature that modern biology has mastered to perfection but man-made self-replicating systems are often lacking. There have been few reports of replicators that can self-assemble autonomously from their building blocks [21, 67], even though this clearly is a necessary feature for truly autonomous replication. Therefore, the emergence of a self-replicating system was probably closely tied to its ability to selectively self-assemble. Any replicating system capable of self-assembly would have had a profound selection advantage against others [153]. Eventually, a thermodynamically driven self-assembly would then have transitioned into a kinetically controlled self-assembly to keep the system out of equilibrium [7, 21].

Previously, particularly hairpin motifs have proven to foster self-assembly in oligonucleotide systems [129, 134, 197], making the proposed replicator an excellent contender for unifying self-replication and self-assembly. In the following section, I will show that the self-replicator can indeed self-assemble. While here, it is merely a demonstration that the system can self-assemble in principle, it seems absolutely possible that self-assembly and self-replication of the presented replicator occur in a one-pot scenario, under slightly altered conditions.

4.3.2. Agglomeration and sedimentation

When combining a complete set of eight cross-complementary strands, for example strands 0_{A-D} & $\overline{0}_{A-D}$ (full_0), spontaneous agglomeration and subsequent sedimentation occurred (Figure 4.3). When an incomplete set was provided, i.e. one strand was omitted, agglomeration and sedimentation were inhibited (Figure 4.6). The characteristic time of the sedimentation depended on the oligonucleotide concentration (Figure 4.8).

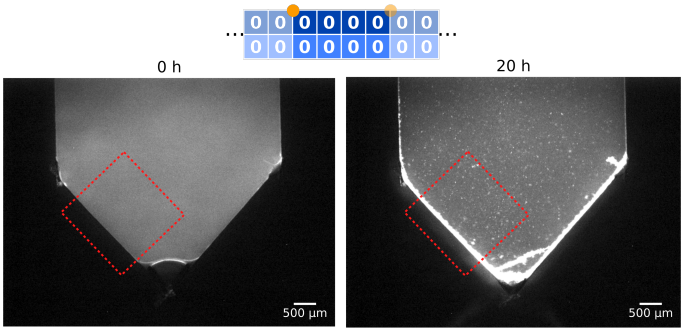

Figure 4.3: Fluorescence microscopy images of spontaneous self-assembly and sedimentation. A full set of cross-complementary strands was mixed at 625 nM per strand (5 µM total) in 150 mM NaCl, 20 mM $MgCl_2$ and 20 mM Tris-HCl pH 8. The sample was loaded in an airtight microfluidic chamber and was initially heated to 95 °C for 10 seconds to ensure an unbound initial state, then rapidly (within ∼ 30 s) cooled to 25 °C, where self-assembly and sedimentation occurred. The first frame after heating shows initially homogeneous fluorescence (left). In the last frame, after 20 hours, the agglomerates have sedimented and cover the bottom of the chamber (right). Figure reproduced from [98].

Agglomeration and sedimentation of the replicator were monitored by fluorescence microscopy in an airtight microfluidic chamber. A custom chamber design was cut from a 500 µm thick FEP foil and fit between two plane sapphires. Three Peltier elements, regulated by a PID controller, were attached at the backside of the chamber. This allowed direct temperature control, with maximum cooling rates of 2 K/s. A sketch of design and setup and a picture of the chamber are shown in Figure 2.14.

The strands were mixed at a total oligonucleotide concentration of 5 µM (i.e. 625 nM per strand for a full set) and inserted into the chamber. One of the strands carried a covalently attached fluorescent label that allowed to monitor the sample. In the chamber, the sample was first heated to 95 °C to separate all molecules and then rapidly cooled (within ∼ 30 seconds) to 25 °C. The initially homogeneous solution (left column Figure 4.3) formed agglomerates, which were tens of micrometers in size and sedimented under gravity within hours (right column Figure 4.3). The formation of such large agglomerates is possible due to the periodic design of the double hairpins (such that 0_D can bind to 0_A). Therefore, a full set of matching hairpins can 'polymerize' freely.

This stands in contrast to the targeted template or complex formation (Figure 4.2), where a complex consists of a maximum of eight strands. For the targeted template and complex

formation, slightly lower oligonucleotide concentrations of 1.6 µM (i.e. 200 nM per strand) were used. Figure 4.8 shows that for a complete set of matching hairpins lower oligonucleotide concentration leads to increased sedimentation time and from 2.5 µM downwards, no agglomeration or sedimentation was detectable. It can be concluded, that the assembly operates quite differently in the two concentration and temperature regimes: (i) For the spontaneous self-assembly the concentration is high (> 2.5 µM) and the temperature is lowered quickly (within seconds) below the melting temperature of the hairpin backbone (<70 °C). (ii) For the assembly of templates and single complexes the concentration is low (<2.5 µM) and the solution is incubated above or around the melting temperature of the hairpin backbone for at least one hour. In case (ii) the hairpins have a long time to equilibrate to their energetic minimum. The quick cooldown in (i) freezes the hairpins (Figure 4.4). If the concentration is high enough, the hairpins can form intermolecular bonds before they get locked in their hybridization state (below 70 °C, Figure 4.4). The slow annealing in (ii) allows the hairpins to equilibrate and find their energetically optimal conformation. The concentration threshold, where the hairpin bonds change from intra- or intermolecular lies thus at around 2.5 µM.

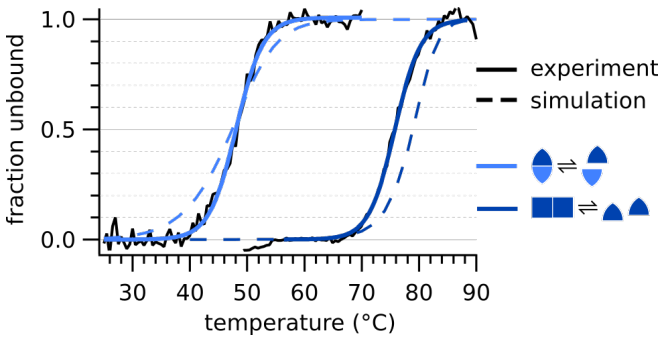

Figure 4.4: Determination of melting temperatures of information domain and hairpin backbone. Melting temperatures of complementary information domains (light blue, strands 0_A+$\overline{0}_A$) and hairpin backbones (dark blue, strands 0_A+0_B). Dashed lines show simulation data from NUPACK. Figure reproduced from [98].

Additionally, steric effects influence the agglomeration. When eight matching hairpins were provided, large agglomerates form, but with only strands of either top or bottom half of the assembly (Figure 4.6, rows 4 and 5) no agglomeration and sedimentation occured. For the top or bottom half, the same arguments regarding the dynamics of hairpin bonds and their 'polymerization' are valid. The only difference is that the assembly has only half the size and is less hindered by steric effects. The half-assembled complex is likely flexible enough to quickly close and form the bond between hairpin D and A, so that no large agglomerates can form. Presumably, also top and bottom half would eventually form agglomerates at higher concentrations.

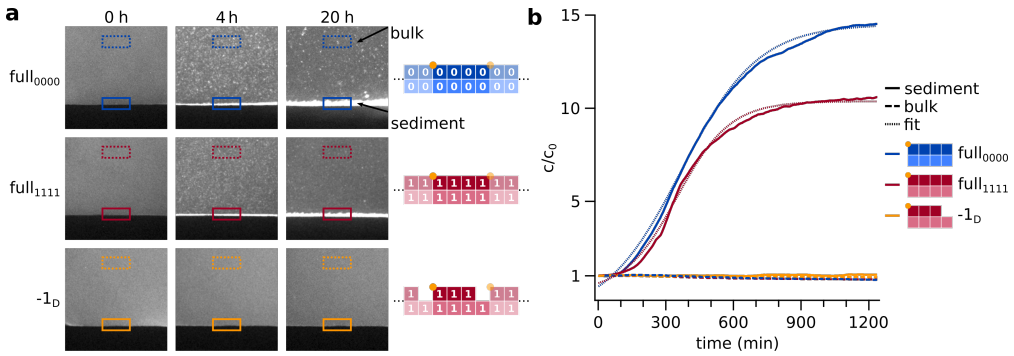

Figure 4.5: Spontaneous self-assembly of complete and incomplete sets. a, Cut-outs from fluorescence microscopy images over time (as indicated in Figure 4.3) for complete sets of matching hairpins with information domains 0, full_{0000} (top) and 1 full_{1111} (middle) as well as for one incomplete set -1_D (bottom). Note that agglomeration and sedimentation only occurred in the presence of all eight matching hairpins, but not when one strand was omitted. For quantification, the bulk and sediment intensities were normalized against the first frame after heating (left column). The total oligonucleotide concentration was 5 µM, about threefold higher than in Figure 4.2 and the following replication experiments. **b,** The relative concentration of sediment and bulk, normalized against the first frame after heating is plotted over the time course of 20 hours. The plot shows the integrated fluorescence intensity over the 2τ width of the intensity profile as described below. Time-traces are plotted for different configurations (same examples as shown in panel a). The complete sets (blue, red) show pronounced concentration increase, i.e. sedimentation. When one strand is omitted (yellow) there is no sediment building. When sedimentation occurs, a slight drop in bulk fluorescence (dashed lines) can be detected. The sedimentation time-traces are fitted with a sigmoid function (dotted line) to determine the saturation concentration $c/c_{0,\text{sat}}$. Figure reproduced from [98].

Agglomeration and sedimentation occurred selectively only when eight matching hairpins were present (Figure 4.5a, top two rows), but was inhibited when one strand was omitted (Figure 4.5a, bottom row). The relative increase of concentration at the sediment was quantified by normalizing the intensity against the first frame after heating. The intensity of the sediment was integrated over the whole flank (Figure 4.3). To determine the width of the sediment, the intensity of the last recorded frame was plotted (along the axis of the box in Figure 4.3 which is perpendicular to the sediment). The intensity was fitted with an exponential function

$$I(x) = y_0 + A \exp\left(\frac{-(x - x_0)}{\tau}\right) \tag{4.1}$$

with a manual offset x_0 to the peak position. The width of the sediment was defined to be 2τ. The integration interval $[x_0{:}x_0 + 2\tau]$ determined from the last frame was applied to the whole image stack for the intensity integration. When no sedimentation could be detected, the average values of x_0 and τ were used. The relative concentrations over time, normalized against the first frame after heating of sediment and bulk, for the three described cases are plotted in Figure 4.5b.

To determine the saturation level $c/c_{0,\mathrm{sat}}$ of the agglomeration, time-traces as presented in Figure 4.5 are fitted with a sigmoid function (dotted line)

$$c/c_0(t) = \mathrm{base} + \frac{\mathrm{max}}{1+\exp\left(\frac{t_{1/2}-t}{\mathrm{rate}}\right)}. \tag{4.2}$$

The saturation level is then $c/c_{0,\mathrm{sat}} = \mathrm{base} + \mathrm{max}$. When no detectable sedimentation occurred, a linear fit was used to determine the final value for $c/c_{0,\mathrm{sat}}$.

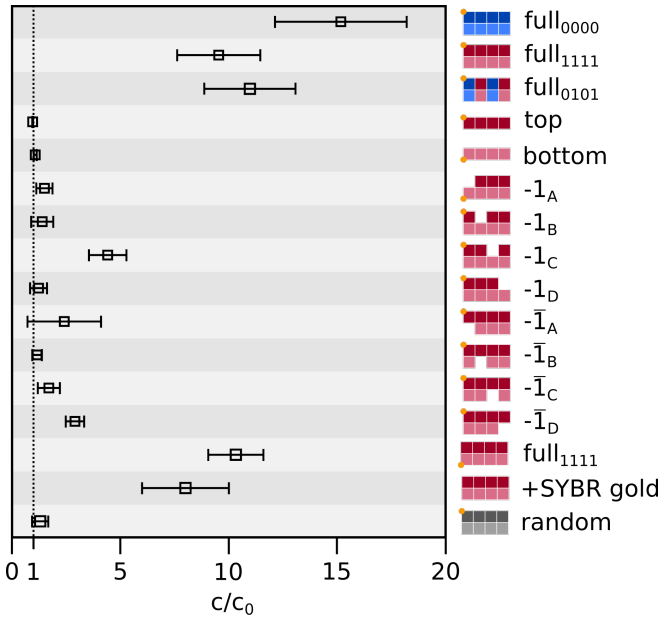

Figure 4.6: Saturation concentrations of the sediment for all configurations. The values for c/c_0 are retrieved from fitting the time traces. Error bars correspond to N $\geq$ 3 independent experiments. For the full set of complementary hairpins, agglomeration and sedimentation is most pronounced (top three rows). When one strand is omitted, mostly no and occasionally weak sedimentation can be detected. Controls show that the agglomeration and sedimentation is independent of the label and its position. Random sequences do not show agglomeration nor sedimentation. Figure reproduced from [98].

Whenever a full set of complementary strands is provided, strong agglomeration and subsequent sedimentation occurred. This was independent of the sequences of information domains (top three rows, Figure 4.6). Similar to the incomplete set shown in Figure 4.5, all permutations of incomplete sets, as well as only the top (0_{A-D}) and bottom ($\overline{0}_{A-D}$) set were measured. The incomplete sets showed no or only a weak signal of agglomeration and sedimentation. Labeling controls with a differently positioned covalent Cy5 label attached to $\overline{1}_A$ and with SYBR gold staining proved that the agglomeration was independent of the fluorescent label and its position. A pool of 84 nt long strands with random sequence showed no signal, excluding unspecific agglomeration.

It is unclear why the saturation levels of c/c_0 for full sets (Figure 4.6, top three rows) are different depending on the incorporated information domain. Also unsettled is the question, why some sets show stronger agglomeration with one omitted strand than others, e.g. -1_C vs. -1_B. The analysis of melting temperature and free energy with NUPACK does not give clues for either case (see Figure 4.13).

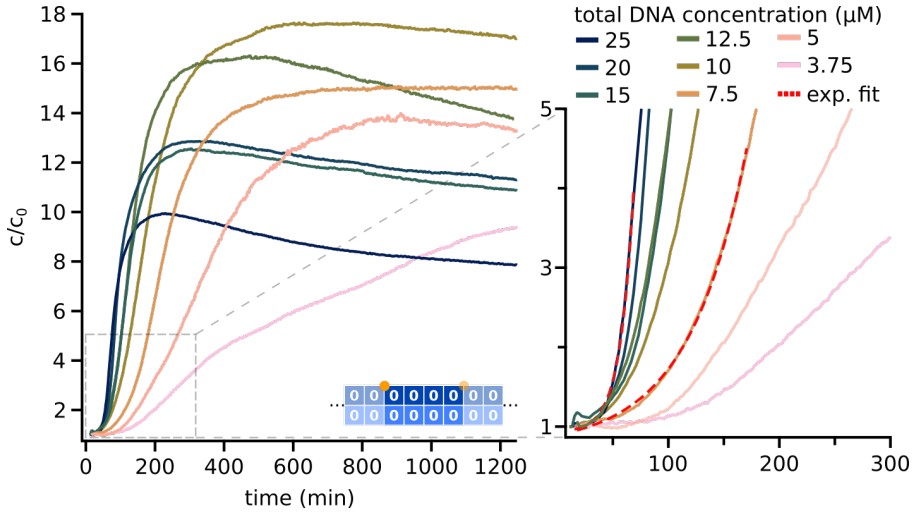

Figure 4.7: Time-traces of sedimentation for different oligonucleotide concentrations of 0000. On the left side, the sedimentation time-traces for different total oligonucleotide concentrations are plotted. At a total oligonucleotide concentration of 2.5 µM, no sedimentation is detectable (not shown here). The drop in intensity after initial sedimentation for high concentrations is due to compacting of the agglomerated phase. On the measured timescale, the level of saturation concentration varies by a factor of two for different oligonucleotide concentrations. It can be hypothesized that they would eventually meet on longer timescales. The zoom on the right, shows the onset of sedimentation. To quantify the sedimentation kinetics, this initial phase is fitted with an exponential function $c/c_0(t) = \alpha \cdot \exp(t/t_{1/2}) + \gamma$. The fit is shown exemplary for two concentrations (red dashed line).

As it can be expected, the agglomeration and sedimentation of the full sets is strongly concentration dependent. Sedimentation time-traces for different oligonucleotide concentrations of the complete set with information domain sequence 0000 (full_{0000}) are shown in Figure 4.7. Interestingly, the saturation concentration is highest for intermediate concentrations. Especially in the movies with high oligonucleotide concentration ≥ 15 µM, it is clearly visible that after the initial sedimentation, the sedimented phase becomes more compact. Due to the depth of the chamber, this results in a decrease in measured intensity. It can be speculated that this compacting might also happen for lower oligonucleotide concentrations but on extended timescales. This would assimilate the saturation concentration for different oligonucleotide concentrations.

To determine the kinetics of the sedimentation, i.e. how fast the agglomerates initially form and sediment, sedimentation onset is fitted with an exponential function as shown in

Figure 4.7. This yields a characteristic sedimentation time, which is lower for higher oligonucleotide concentrations. Figure 4.8 shows the concentration dependence of the characteristic sedimentation time, that can be fitted with a power law. The exponent is close to -1, which suggests that cooperativity does not play a strong role.

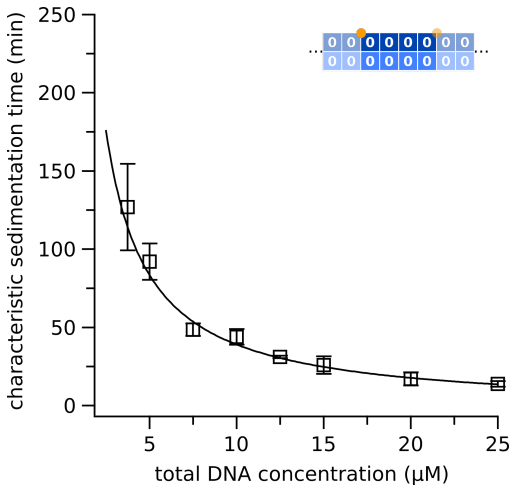

Figure 4.8: Concentration dependence of the sedimentation kinetics for 0000. The characteristic sedimentation time is determined by fitting the initial increase of c/c_0 over time with an exponential function. The error bars correspond to N $\geq$ 3 independent experiments. The concentration dependence can be fitted with a power law function $y(c) = a_1 + a_2 \cdot c^p$, which yields an exponent of $p = -1.06$. The fit is weighed by the displayed error bars. Figure reproduced from [98].

4.4. Exponential cross-catalytic replication

4.4.1. Template-assisted product formation

The replication mechanism is template-based and driven by temperature oscillations. To determine the optimal conditions for replication, template-assisted product formation, step (1) from Figure 4.1b, was studied in isothermal and equimolar conditions. Single strands were provided at 200 nM per strand. Template-assisted product formation was initiated with template concentrations $\overline{c}_0$ between 0 nM and 120 nM (Figure 4.9b). Higher initial template concentrations sped up product formation and produced higher yields.

The potential temperature range for the base temperature T_{base} was given by the melting curve of the information domain (Figure 4.4). Product formation worked better at elevated temperatures but slowed down above 48 °C when the melting temperature of the information domain is exceeded and hybridization between information domains becomes unstable (Figure 4.4). At high temperatures (> 45 °C), because hairpins start to open in solution, the rate of spontaneous, untemplated product formation increased. At optimal conditions,

approximately 80 % of the final product yield was reached within 20 minutes. For amplification experiments, T_{base} was set to 45 °C and each incubation cycle at T_{base} was determined to last 20 minutes.

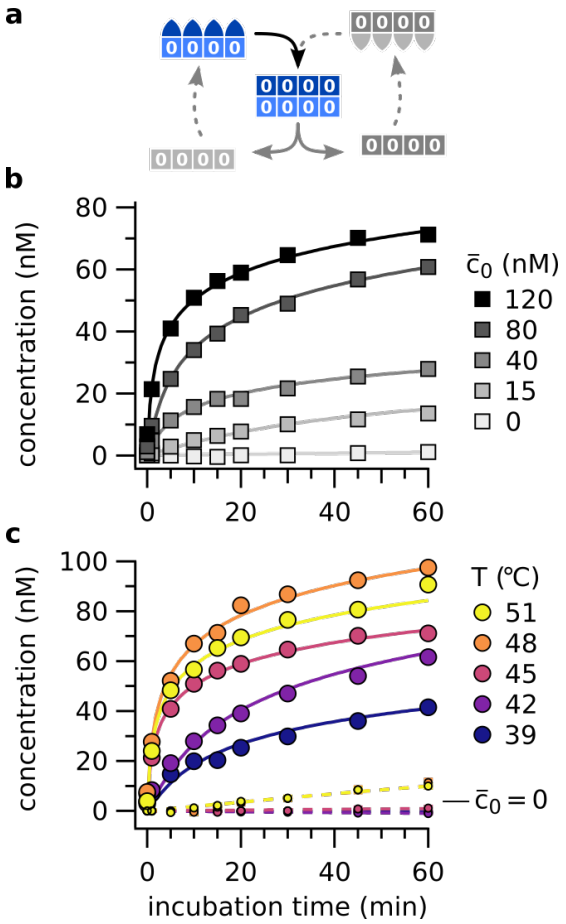

Figure 4.9: Templating kinetics. **a,** Schematic representation of the templating step at constant temperature. For grayed out steps, temperature cycling is necessary. **b,** Isothermal tetramer formation at 45 °C. The template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ is provided at different starting concentration $\bar{c}_0$. Single strands were provided at 200 nM per strand. The product concentration is plotted over time and includes all complexes containing tetramers with labeled strand 0_A. Higher initial template concentration causes faster templating and higher yield. **c,** Templating observed for different temperatures starting with an initial concentration $\bar{c}_0 = 120$ nM of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ and 200 nM per single strand (large circles). Small circles show spontaneous, untemplated product formation ($\bar{c}_0 = 0$), which increases for T > 45 °C. Templated product formation worked best for T = 48 °C. Figure reproduced from [98].

4.4.2. Exponential template amplification

A simple cross-catalytic model for the amplification of template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ with concentration $\overline{c}$ and product $0_A0_B0_C0_D$ with concentration c is given by

$$\frac{d}{dt}c(t) = k \cdot \overline{c}(t) + k_0\,, \quad \frac{d}{dt}\overline{c}(t) = k \cdot c(t) + k_0\,, \tag{4.3}$$

where k is the rate of cross-catalysis and k_0 is the rate of spontaneous, untemplated product formation. This corresponds to a cycle-dependent exponential growth of two interdependent species and does for $c(t) \approx \overline{c}(t)$ simplify to an exponential growth on a per-cycle basis. This model can be solved analytically but does not account for saturation effects when monomers become scarce. Therefore, concentrations of template and product have to be kept well below 200 nM, which is the chosen monomer concentration, or refeeding with monomers has to be considered.

Template amplification was studied under temperature cycling (Figure 4.10). For amplification experiments, the sample was subjected to six temperature cycles with an asymmetric temperature profile. The base temperature was held for 20 minutes, the peak temperature for 1 second. With heat-up and cool-down this summed to approximately 30 seconds off the base temperature. Sequences encoding for information domain 0 (0_{A-D} & $\overline{0}_{A-D}$) were present at 200 nM per strand.

Amplification was tested for different initial template concentrations $\overline{c}_0$ from 0 nM to 45 nM of the template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ as shown in Figure 4.10a, with temperature oscillations between $T_{base} = 45$ °C and $T_{peak} = 67$ °C. The experimental data points were fitted using the simple cross-catalytic model from Equation 4.3 (solid lines). The fit produced rate constants $k = 0.16$ cycle^{-1} and $k_0 = 0.04$ nM·cycle^{-1}.

The initial growth rate was extracted and plotted against the initial template concentration c_0 as depicted in Figure 4.10b. As implicated by Equation 4.3, the growth rate can be simplified to a linear dependence of c_0 and the data showed good agreement with a linear fit (Figure 4.10b).

With regard to the peak temperature, the amplification reaction showed to be robust (Figure 4.10c), as long as the peak temperature was kept below 71 °C. This is in good agreement with the beginning of hairpin melting at around 72 °C (Figure 4.4), which causes the hairpin backbone to become unstable.

The exponential nature of the amplification reactions was also confirmed by serial dilution experiments (Figure 4.10d). Over the course of 15 temperature cycles, a reaction containing the template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ and single strands 0_{A-D} & $\overline{0}_{A-D}$ was diluted with monomers by a factor of ½ every three cycles (black circles). A second reaction (open circles) contained the identical template but only monomers 0_{A-D}, but lacked monomers $\overline{0}_{A-D}$, which are necessary to make copies of the template. Only when all monomers necessary for the cross-catalytic amplification were available, the reaction was able to withstand dilution.

The concentration under dilution could be fitted using Equation 4.3 with the dilution factor as a free parameter (solid lines).

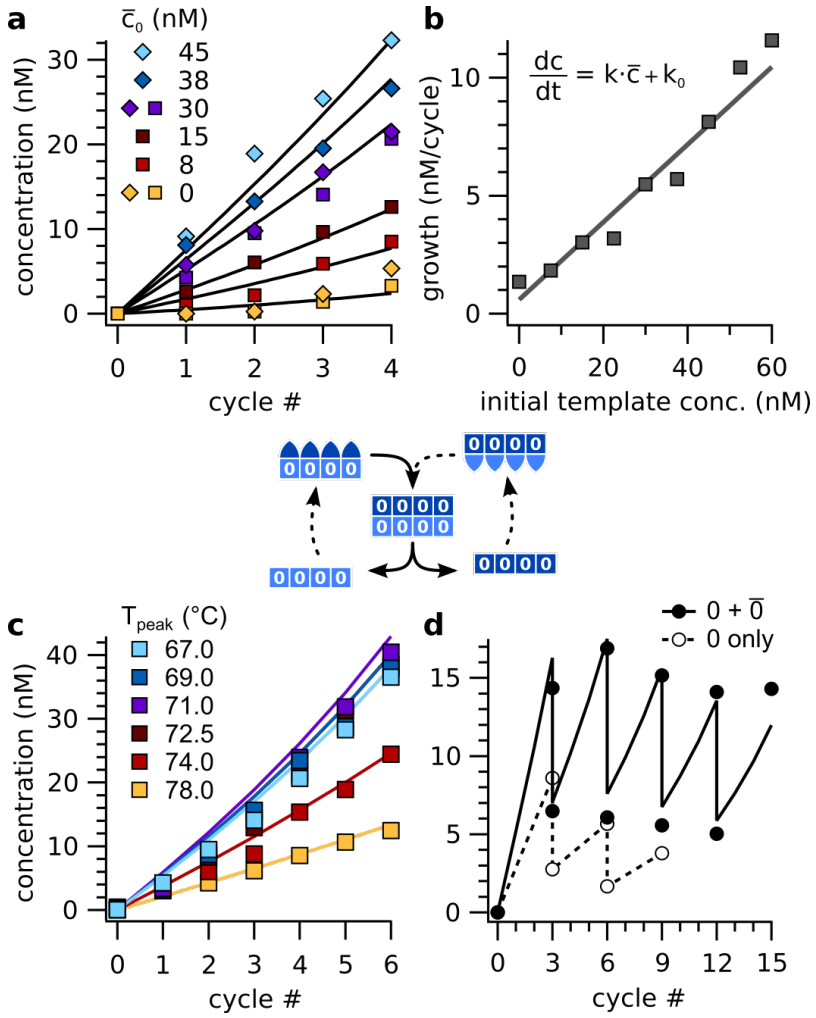

Figure 4.10: Exponential amplification of a template with temperature oscillations. **a,** Amplification of sequence 0000 was monitored for different initial template concentrations $\bar{c}_0$ (0-45 nM) of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$. Temperature was oscillated between T_{base} = 45 °C and T_{peak} = 67 °C. Monomers 0_A, $\bar{0}_A$, 0_B, …, $\bar{0}_D$ were provided at 200 nM each. The concentration of complexes 4:4 after each cycle is shown. The data was fitted using the cross-catalytic model from Equation 4.3. **b,** Initial reaction velocity as a function of initial template concentration $\bar{c}_0$. The data points show good agreement with the line calculated from the fits in panel a. **c,** Amplification with initially 30 nM of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ proceeded well for T_{peak} < 74 °C. At higher temperatures, the hairpin backbone starts to melt. The base temperature was T_{base} = 45 °C. **d,** The reaction containing strands 0_A, $\bar{0}_A$, 0_B, …, $\bar{0}_D$ (black circles) survived successive dilution by a factor of ½ every three cycles at almost constant concentration. A second reaction containing the identical amount of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, but lacking monomers $\bar{0}_{A–D}$, died out (open circles). The solid and dashed lines show the fit with Equation 4.3 and the dilution factor as a fitting parameter. Figure reproduced from [98].

4.4.3. Sequence replication and fidelity

As described in Section 4.2, the replicator can encode binary information by incorporating either information domain 0 or 1 at each position. So far, the amplification reactions only contained a singular set of monomers (0_{A-D} & $\overline{0}_{A-D}$) and did therefore not show the replication of information encoded in the succession of the information domains.

To perform sequence replication, template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ was provided at 15 nM initial concentration with 100 nM each of all 16 monomers (0_A, $\overline{0}_A$, 1_A, ..., $\overline{1}_D$). Temperature oscillated between $T_{base} = 45$ °C and $T_{peak} = 67$ °C. The progress of the replication over six temperature cycles was monitored by native PAGE (Figure 4.11a, left). The product concentration includes all bands containing tetramers and is plotted over time in Figure 4.11b (yellow).

To determine to which rate a template was replicated faithfully, the replication was run with defective sets of monomers, i.e. lacking one or more monomers required to replicate a template faithfully. When template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ was provided, a defective set of monomers lacked for example monomer 0_D, denoted '+++-' whereas the full set of monomers is denoted '++++'. Omitting one or more strands means that the correct product cannot be built and any product from a defective set must contain mutations in the succession of information domains. The product concentration over six temperature cycles for three different defective sets, '+++-', '++–' and '+-+-', is shown in Figure 4.11b, together with the replication reaction containing the complete set of monomers ('++++'). A set with two defects performs comparable to the case when no template is provided (black line).

From the native PAGE analysis (Figure 4.11a), it is clearly visible that for the defective set '+++-' (right side of panel a) mostly complexes $0_A0_B0_C : \overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ of size 3:4 are produced, especially during the first couple of cycles. In contrast, the full set produces almost exclusively complexes of size 4:4 (left side of panel a).

Replication with a single defect ('+++-') produced $\sim$ 40 % yield compared to the complete reaction. Introducing a second defect reduced the yield further to $\sim$ 15-20 %. Since $0.4 \times 0.4 = 0.16 \simeq$ 15-20 %, it can be concluded that the effect of two defects is similar to two independent single defects. The reduction in product concentration for single and double defects is independent of the template sequence (Figure 4.11c). For other templates, the according monomer strands are omitted in the defective sets, i.e. for template $\overline{0}_A\overline{1}_B\overline{0}_C\overline{1}_D$, monomer 1_D is omitted in the defective set '+++-'.

It is instructive to consider how an unfaithful replicate can form. Let's assume the case displayed in blue in Figure 4.11. The replication of sequence 0000 when the template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ is provided with the defective set '+++-', i.e. monomer 0_D is missing. Now, there are two categories of unfaithful replicates that can form:

- **Mutation in position.** A monomer with the correct information domain, but incorrect hairpin sequence (e.g. 0_B) binds to the 'open' information domain on the template, $\overline{0}_D$. The information domain bond is stable at base temperature, but due to the incorrect position (5' hairpin of 0_B cannot hybridize with 3' hairpin of 0_C), the hairpin backbone cannot close. During the next temperature oscillation, the information domain bond will melt and because the backbone did not close, 0_B will be released back into solution as a monomer (probably accompanied by the trimer $0_A0_B0_C$).

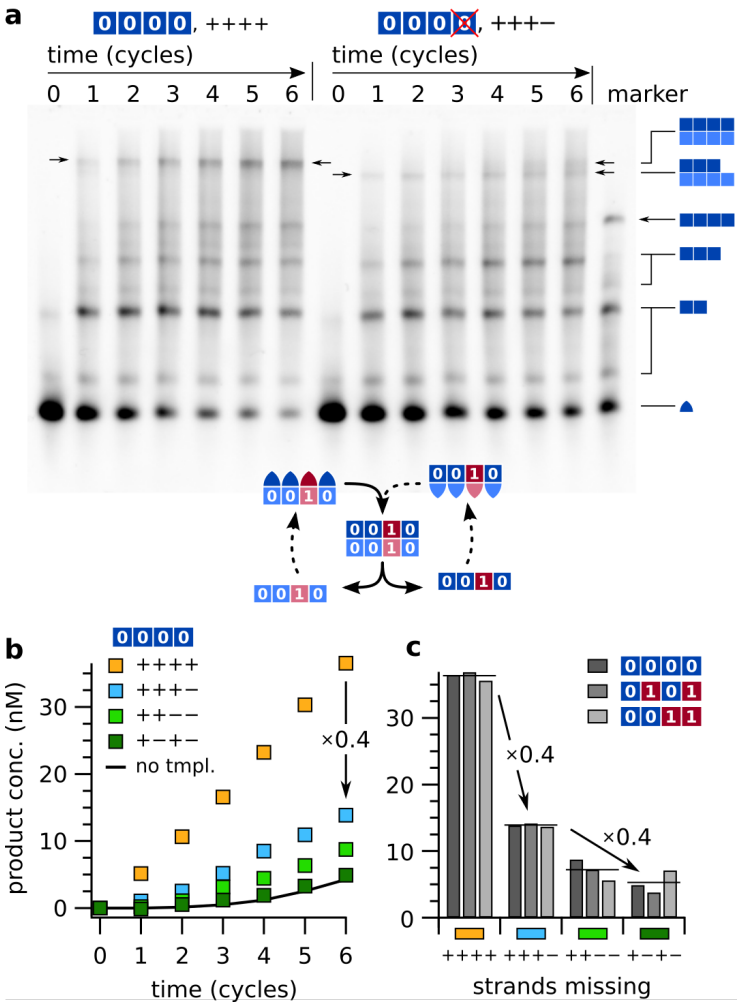

Figure 4.11: Sequence replication with temperature oscillations and fidelity check by forcing mutations from 0 to 1 at different locations. **a,** Replication of sequence $0_A0_B0_C0_D$. Reactions were started with 15 nM initial template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$. Monomer strands were present at 100 nM each. Native PAGE results comparing the complete reaction with all 16 strands, '++++' (left) with a defective reaction lacking strand 0_D, '+++-' (right). The defective set +++- mostly produced 3:4 complexes instead of 4:4 complexes. Schematics on the right indicate position of different complex sizes. The overall yield of tetramer-containing complexes was greatly reduced. As a size reference, the marker lane contained complexes $0_A0_B0_C0_D$, $0_A0_B0_C$, 0_A0_B, and monomers 0_A. **b,** Product concentration over time for the complete sequence network (yellow) and three defective sets with missing strands. Data was integrated by quantitative image analysis from electrophoresis gels using covalent markers on the 0_A-strand counting all product complexes containing tetramers. Mutations of information in the product from 0 to 1 were induced by defective reactions that lacked strands 0_D (+++-), $0_C/0_D$ (++−), and $0_B/0_D$ (+-+-). All reactions were initiated with 15 nM of $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$. The solid line shows data from reaction '++++' without template. **c,** End point comparison of reactions with templates $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ (panels a, b), $\overline{0}_A\overline{1}_B\overline{0}_C\overline{1}_D$, and $\overline{0}_A\overline{0}_B\overline{1}_C\overline{1}_D$ after six cycles for the various intact and defective sets. Horizontal lines indicate averages of the three template sequences for each set. A single missing strand reduced the product yield to about 40 %, two missing strands to 15-20 %. Figure reproduced from [98].

- **Mutation in information domain.** A monomer with the incorrect information domain, but correct hairpin sequences (e.g. 1_D) binds to one of the open hairpins of an incomplete replicate (trimer $0_A0_B0_C$), which is already bound to the template at the information domains. 1_D can either bind with its 5' hairpin to the 3' hairpin of 0_C or with its 3' hairpin to the 5' hairpin of 0_A. Once one of those bonds has formed, likely the other will happen too as they are in such close proximity. Then, a correctly assembled backbone is formed. During the next temperature oscillation, the product $0_A0_B0_C1_D$ will be released from the original template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$. Henceforth, it can act as a template for the unfaithful replication of sequence 0001, which will proceed as exponentially, since all strands for the replication of 0001 are present.

A product yield of $\sim$ 40 % for an unfaithfully replicated template seems high at first glance. However, the mutations were intentionally imposed by not providing all strands necessary to replicate the template. Through the way in which we test for unfaithful replication, the effect of binding competition is neglected. Sticking with the example '+++-' (blue in Figure 4.11b), there is no binding competition for position D since only strand 1_D is present. In the full replication reaction '++++' however, the better matching strand 0_D is also present and binds preferentially. For a similar binding competition in a protein-catalyzed ligation reaction the inferior reaction decreased 7-fold in the presence of competition [186]. It should also be noted that the product concentration in Figure 4.11b,c covers all complexes containing tetramers, i.e. it also includes complexes 2:4 and 3:4 which are replicated faithfully and have no mutation.

From the experiments shown in Figure 4.11b, the replication fidelity can be estimated to be $1 - \frac{14\text{ nm}}{37\text{ nm}} = 62$ %. Due to the above mentioned reasons, this is a lower bound estimate and it can be expected that the replication fidelity in the full reaction is much higher.

It is important to mention that due to the periodic design of the replicator, this analysis of the replication fidelity is complete. The defective sets presented in Figure 4.11 represent all possible defective sets with one or two lacking strands.

4.4.4. Translation of per information domain fidelity to per nucleotide fidelity

It is interesting to compare this replicator in its performance and fidelity to other replicators. Considerations about replicators in an origin of life context, both experimentally [77] and theoretically [52] are typically based on base-by-base replication (i.e. polymerization). In this paragraph, the aim is to calculate a corresponding hypothetical per nucleotide fidelity from the experimentally measured per information domain fidelity.

The deviation in melting temperature due to point mutations in the information domain was used as a criterion to translate between the two replication types. Considering the melting curve of the information domain (Figure 4.4) and the temperature oscillations between $T_{base} = 45$ °C and $T_{peak} = 67$ °C, it was estimated that a drop in melting temperature by more than 10 °C would render the replication unfunctional. It can then be estimated how

many point mutations in the information domain can be tolerated to stay within that melting temperature threshold.

The information domain duplex 0:$\overline{0}$ was compared with an information domain duplex 0:$\overline{0}$*, where $\overline{0}$ and $\overline{0}$* differ by less than K point mutations. Their information domain binding energies and the resulting melting temperatures were analyzed as displayed in Figure 4.12. The cumulative free energy distributions of mutated information domain sequences with one to three point mutations show that the vast majority of information domains with three point mutations have an information domain binding energy of $\Delta G(0{:}\overline{0}^*) \geq$ -12.5 kcal/mol, in comparison to the original binding energy of $\Delta G(0{:}\overline{0})$ = -15.4 kcal/mol (Figure 4.12a). For an information domain with three point mutations, even at the low end of the binding energy distribution, this results in a drop in melting temperature of at least 10 °C. Therefore, the defined melting temperature threshold of $\Delta T_m = 10$ °C translates to an information domain with up to $K = 3$ point mutations.

Within the set temperature threshold, the replicator cannot differentiate between information domain $\overline{0}$ and information domain $\overline{0}$*, if $\overline{0}$ and $\overline{0}$* differ by less than K point mutations. Then, the per information domain fidelity p_K with up to K point mutations across an information domain of length N is given by a cumulative binomial distribution

$$p_K(N) = \sum_{k=0}^{K-1} \binom{N}{k} p^{(N-k)} (1-p)^k , \tag{4.4}$$

where p is the per nucleotide replication fidelity we are looking for.

For $K = 3$ mutations and $N = 15$ nucleotides in the information domain, we experimentally found a per information domain fidelity of $p_3(15) = 0.62$ and can thus calculate the per nucleotide fidelity to be $p = 0.85$. Internal mutations have a much bigger effect on the binding energy and melting temperature than terminal mutations. The free energy distribution for $K = 2$ internal mutations is similar to the distribution for $K = 3$ total mutations [99]. Taking this into account, Equation 4.4 can be recalculated with $K = 2$ and $N = 13$ and gives a per nucleotide fidelity of $p = 0.90$. Therefore, the presented replication mechanism is comparable to a base-by-base replication mechanism with a fidelity of $\simeq$ 85-90 %.

4.5. Discussion

In this chapter, I presented a cross-catalytic replicator that can self-assemble and self-replicate binary information encoded in a succession of oligomers. The replication is driven by temperature oscillations. Because the replication mechanism is based on hybridization only, no covalent bonds are involved or required. The replication is independent of external fuel and does not generate waste. Instead, all bonds are reversible and all products or potential side-products, i.e. incomplete or malformed complexes, can be recycled by a high temperature spike.

The replication proceeded with temperature peaks every 20 minutes and reached saturation limits on the timescale of hours without replenishing monomer building blocks. This

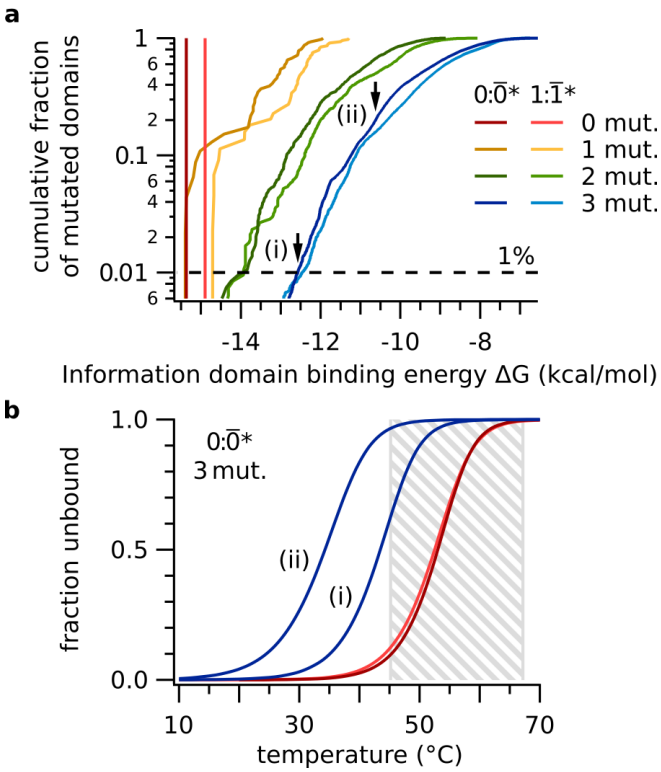

Figure 4.12: Change in binding energy and melting temperature due to point mutations in the information domain. a, Cumulative free energy distributions of mutated information domain duplexes 0:$\bar{0}$* and 1:$\bar{1}$* with up to three point mutations in $\bar{0}$* and $\bar{1}$* (yellow, green, blue). The free energies of the original duplexes 0:$\bar{0}$ (red) and 1:$\bar{1}$ (light red) are also indicated. 99 % of duplexes 0:$\bar{0}$* and 1:$\bar{1}$* with three point mutations have free energies $\Delta G \geq$ -12.5 kcal/mol (dashed line). **b,** Melting curves of information domain duplexes 0:$\bar{0}$ (red), 1:$\bar{1}$ (light red) and two duplexes 0:$\bar{0}$* with three mutations in the information domain, indicated by arrows in panel a. Duplex (i), which is located at the low end of the ΔG distribution, has a melting temperature of $\simeq$ 10 °C below that of 0:$\bar{0}$, whereas duplex (ii) differs by $\simeq$ 20 °C. This decrease in melting temperature destabilizes the binding of the information domain and renders these sequences unfunctional for replication in the temperature oscillation regime between T_{base} = 45 °C and T_{peak} = 67 °C (gray box). Figure reproduced from [98].

speed is similar to other replicators [91, 192], cross-polymerizing ribozymes [77, 200] or auto-catalytic DNA networks [197, 202].

The replicator was able to transfer the information encoded in the specific succession of information domains with a fidelity of at least 62 %. This translates to a 85-90 % replication fidelity per nucleotide. Non-enzymatic template based replicators exhibit fidelity in coping of around 80-90 % [71, 155]. The here presented replicator is therefore comparable to other replication studies. In contrast to these studies, the replication machinery in modern biological systems includes a mechanism for error detection and correction, which increases

the replication fidelity substantially [102].

Generally, the replication fidelity is limited by the background rate of spontaneous bond formation without template. In DNA or RNA hairpin-fueled reactions, this spontaneous product formation is often a problem and is attempted to be reduced by careful design [64, 94, 197]. Here, this background rate was low, at maximum 10 % (Figure 4.9). However, the interaction of strands in solution is strongly concentration dependent and the rate of spontaneous product formation decreases with the overall strand concentration. In a prebiotic setting, very low concentrations are to be expected. This would minimize undesired spontaneous product formation at the expense of slower replication dynamics.

When considering the replicator in an origin of life context, it must be admitted that it seems unlikely that oligonucleotide strands of substantial length, capable of replication, arose spontaneously [87, 161, 170, 180]. Instead, a combination of non-enzymatic polymerization [31, 79, 83, 105] and template-directed copying [46, 55, 191] produced oligonucleotide that could initiate replication. However, oligonucleotide provided by these reactions often suffer from issues with regiospecificity [87, 124]. A heterogeneous backbone of 2'-5'- and 3'-5'-linkages might inhibit chemical downstream reactions, but could also offer a solution to the template inhibition problem [180]. For a purely hybridization based approach however, regiospecificity is not equally critical [56]. Typically, prebiotic polymerization does not produce strands that exceed 10-20 nt in decent yields [139], especially when all four bases are incorporated at an equimolar ratio [126]. Nevertheless, the synthesis of RNA strands longer than 50 nt has been demonstrated, although under very specific conditions [30, 79, 156].

Since the replication mechanism is based on hybridization only, as long as the order of melting temperatures is conserved, this replication mechanism is also expected to work with shorter strands, that might have been produced by an upstream prebiotic polymerization. Indeed, this would even have two advantages. Firstly, a shorter information domain would be more sensitive to mismatches and might already reject single base mismatches, thereby largely increasing the fidelity of replication. Secondly, shorter hairpins would have lower melting temperatures and therefore higher fluctuations, which were here induced artificially by introducing intentional mismatches in the 30-33 nt hairpin sequences. In this context, it is important to mention that a replicator based on shorter strands does not debunk the notion of tRNA-inspired sequences, which could grant replication an important role in the emergence of translation. Already Hopfield [75] proposed that modern tRNA emerged from two primal tRNA strands.

Any known non-enzymatic polymerization reaction would produce an oligonucleotide pool with a wide variety in length and sequence [57, 87, 124, 161]. It has been shown that already simple ligation systems break the symmetry in sequence space and will further a bias in the sequence pool [18, 97, 186]. An additional reduction of the sequence space could have been achieved through sequence-specific phase transition. Indeed, the ability of this replicator to spontaneously agglomerate and sediment in a sequence selective manner was demonstrated. Analogous sequence-specific phase-transitions of DNA into a gel-like phase have been observed with shorter strands of similar secondary structure (triple and quadruple hairpins) [129, 134]. Self-selection could have provided a pathway for self-similar sequences

capable of performing such a replication to emerge from a (semi-) random sequence pool provided by polymerization. Here, strong and fast agglomeration was observed. When starting with an at least semi-random pool, this process would be slowed down enormously and not result in similarly large agglomerates. But even a very shallow temperature gradient would be sufficient to effectively localize strands capable of agglomeration [116] that could form first templates and initiate replication. The biased hydrolysis of single- versus double-stranded oligonucleotides could represent a further selection pressure favoring assembled complexes [136].

The temperature oscillations that drive the replication can for example be realized through convection cycling, with asymmetric temperature profiles on similar timescales [17, 166]. In an early Earth scenario, such convection cycling could have been evoked by a temperature gradient across a rock pore [95, 115, 116]. Depending on the geometry of the pore, the temperature profile can have a wide variety of shapes [10, 89] such that for any variation of the replicator, there likely were plenty of suitable pores. As already described in Section 2.5.2 thermal gradients also lead to the accumulation of molecules [116], which can boost self-assembly, initial template generation and replication activity. In different environments, it is also conceivable that a similar replication mechanism is driven by pH oscillations [90] or thermo-chemical oscillations [11].

As discussed, this replication mechanism must have been preceded by a prebiotic polymerization chemistry and might have coexisted with a non-enzymatic ligation chemistry, that can link 3'-5' ends of oligonucleotides, e.g. in hairpin overhangs [174]. Most likely, the ligation reaction would have been slow, but even though it would be a promising means to permanently fix the replication results by chemically linking adjacent hairpins. It would also have made template and replicate impervious to higher temperature spikes. Another potentially helpful downstream addition are small ribozymatic centers that can induce scission at the onset of a hairpin [35, 188]. This would cleave the hairpins from the assembled information domains creating an assembly that much resembles mRNA when ligated. Due to its rejection of sequences containing too many mutations, the replication mechanism could also have served as a check on polymerization to improve the overall fidelity of the emergence of longer strands.

This replication mechanism is an interplay between oligonucleotide sequence and length, salt concentration and oscillation temperature and time. Here, one specific set of those parameters was studied. This should serve as a proof of principle that a certain set of oligonucleotide strands - for which the boundary conditions have been elaborated in Sections 4.2 - can replicate information according to the proposed mechanism. Considering the vast amount of time and attempts available on early Earth to randomly screen the parameter space, it is certainly plausible that maybe not exactly this but a similar replication once prevailed.

Keeping in mind how much the single oligonucleotide strands of the replicator resemble modern tRNA in structure and sequence, the potential existence of such a replicator also raises the question how intertwined the emergence of translation is with such an ancient replication mechanism. In modern biology, tRNA is the central molecule bridging between the information carrying units (DNA, mRNA) and the executing units (proteins). The func-

tion and structure of tRNA is conserved across the three domains of life (archea, bacteria, eukaryotes), indicating that tRNA is one of the most ancient molecules [152]. Furthermore, almost the complete translation machinery was already present in the last universal common ancestor (LUCA) [60]. One could speculate that self-replication emerged from a proto-tRNA, a simple double hairpin molecule, which later evolved into tRNA and took up its role in protein translation. Already Eigen [53] proposed that proto-tRNAs were 'the first genes' and suggested that primitive tRNA molecules were the earliest components of the translational machinery.

The tRNA-like sequences used in the presented replication mechanism follow conformational considerations of proto-tRNAs [168]. The information domain would evolve to become the anti-codon and is spatially close to the 3' end of an unbound strand, which would later carry the amino acid. Recently, short peptide-RNA hybrids [65, 85] have been shown to form under prebiotic conditions. Equipped with a mechanism to ensure correlation between the 3' end and the anti-codon, such hybrids could have provided the link between proto-tRNAs and the genetic code.

This replication mechanism proposes that predecessors of tRNA had a role in hybridization-based replication, independent of protein translation. This would justify the existence of tRNA-like molecules before the emergence of translation and suggests that the emergence of translation is closely tied to a replicator. However, many questions regarding intermediate evolutionary steps remain unsettled, e.g. how an amino acid could have been attached to and released from the 3' end of a hairpin, or how the anti-codon could have been reduced to three nucleotides, given the instability of a 3 nt hydrogen bond.

4.6. Materials and methods

4.6.1. Strand design and secondary structure prediction

DNA double hairpin sequences were designed using the NUPACK software package [199]. In addition to the secondary structures of the double hairpins, the design algorithm was constrained by all target dimers. Candidate sequences were selected for optimal homogeneity of binding energies and melting temperatures. On the one had, backbone domains connecting consecutive strands (e.g. $0_A0_B0_C$) had to be the most stable bonds in the system, in particular more stable than between a template and a newly formed product complex (e.g. 0_B:$\overline{0}_B$). On the other hand, hairpin melting temperatures had to be low enough to allow for a sufficient degree of thermal fluctuations. To reconcile this with the length of the strands, mismatches were introduced in the hairpin stems.

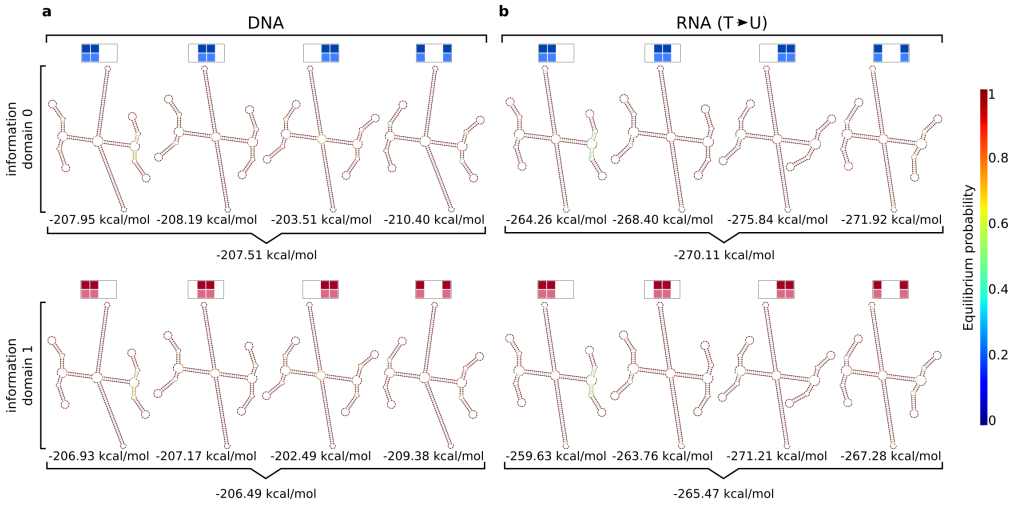

Figure 4.13: Secondary structure predictions and calculated free energies for the replicator in DNA and RNA using NUPACK. All 2:2 complexes of double hairpins are displayed for **a,** DNA and **b,** RNA at 25 °C, 1 M Na^+ and 0.625 µM per strand. In panel b every T in the DNA sequence is substituted by a U. Due to computational limitations, only four strands could be calculated at a time. The predicted structures look as expected, forming the backbone (top and bottom from center) and the information domain bonds (left and right from center), which are flanked by the closed hairpins. The DNA (**a**) and RNA (**b**) implementations are identical. Also, there is no difference in structure or free energy between information domain 0 and information domain 1. The calculated free energies are given below each structure and are averaged over each complete set. Figure reproduced from [98].

Sequences were picked such that the melting temperatures and free energies of potential dimers were as homogeneous as possible. The secondary structure predictions as well as the calculated free energies of all 2:2 complexes are shown in Figure 4.13. It is evident that the

secondary structure is identical in a DNA and an RNA implementation. The sequences of all strands are given in Table 4.1. For more detail on the sequence design, see [99].

4.6.2. Native PAGE and product quantification

DNA complexes were analyzed using native polyacrylamide gel electrophoresis (PAGE) at 5 % acrylamide concentration and 29:1 acrylamide/bisacrylamide ratio (Bio-Rad, Germany). Gels were run at electric fields of 14 V/cm at room temperature. Strands $0_A/1_A$ were covalently labeled with Cy5. Fluorescence intensities were later used to compute strand concentrations. As an additional color channel, strands were stained using SYBR Green I dye (New England Biolabs). Initial complex identification was done by comparing the assembly of different subsets (Figure 4.2). To correctly identify bands in the time-resolved measurements, each gel contained a marker lane. The marker contained strands 0_A (200 nM), 0_B (150 nM), 0_C (50 nM), and 0_D (100 nM), and was prepared using the two-step annealing protocol from 95 °C to 70 °C. The unequal strand concentrations ensured that the sample contained a mixture of mono-, di-, tri- and tetramers.

PAGE gels were imaged in a multi-channel imager (Bio-Rad ChemiDoc MP). Image post processing and data analysis were performed using a custom LabVIEW software. Post processing corrected for inhomogeneous illumination by the LEDs, image rotation, and distortions of the gel lanes, if applicable. Background fluorescence was determined from empty lanes on the gel, albeit generally low in the Cy5 channel. For the determination of reaction yields, the intensities of all gel bands containing strands of the sequence length of interest were integrated. To quantify the tetramer yield, single tetramers were considered as well as their complexes with di- and tri- and tetramers. Single strands separated from their complements during electrophoresis (Figure 4.2, lanes 2, 7, 16, 18).

4.6.3. Template preparation and complex formation

All reactions were performed in 150 mM NaCl, 20 mM $MgCl_2$ and 20 mM Tris-HCl pH 8. DNA oligonucleotides (purchased from Biomers, Germany) were used at 200 nM concentration per strand in reactions containing a fixed-sequence subset of eight strands (e.g. $0/\overline{0}$ only) and 100 nM per strand in reactions containing all 16 different strands. Temperature cycling was performed in a standard PCR cycler (Bio-Rad C1000). Reaction kinetics were obtained by running each reaction for different run times or numbers of cycles in parallel. The products were analyzed using native PAGE. The time between Temperature cycling and PAGE analysis was minimized to exclude artifacts from storage on ice. Template sequences were prepared using a two-step protocol. Annealing from 95 °C to 70 °C within one hour, followed by incubation at 70 °C for 30 minutes. Afterwards, samples were cooled to 2 °C and stored on ice. When assembling complexes containing paired information domains (Figure 4.2), samples were slowly cooled down from 70 °C to 25 °C within 90 minutes before being transferred onto ice. DNA double hairpins were quenched into mono-molecular states by heating to 95 °C and subsequent fast transfer into ice water.

4.6.4. Thermal melting curves

Thermal melting curves were measured using either UV absorbance at 260 nm wavelength in a UV/Vis spectrometer (JASCO V-650, 1 cm optical path length), via quenching of the Cy5 label at the 5'-end of strand 0_A (excitation: 620–650 nm, detection: 675–690 nm), or using fluorescence of the intercalating dye SYBR Green I (excitation: 450–490 nm, detection: 510–530 nm). Fluorescence measurements were performed in a PCR cycler (Bio-Rad C1000). Samples measured via fluorescence were at 200 nM of each strand, those measured via UV absorption contained 1 µM total DNA concentration to improve the signal-to-noise ratio. Before analysis of the melting curves [119], data was corrected for baseline signals from reference samples containing buffer and intercalating dye, if applicable.

4.6.5. Agglomeration and sedimentation experiments

The samples were mixed in the replication buffer (150 mM NaCl, 20 mM $MgCl_2$, 20 mM Tris-HCl pH 8) at a total oligonucleotide concentration of 5 µM, i.e. varying concentration per strand depending on the number of different strands in the configuration (4, 7 or 8). The microfluidic chamber was assembled with a custom cut, 500 µm thick, FEP foil placed between two plane sapphires as shown in Figure 2.14. Three Peltier elements (QuickCool QC-31-1.4-3.7AS, purchased from Conrad Electronics, Germany) were attached to the backside of the chamber to provide full temperature control. The chamber was initially flushed with 3M™Novec™7500 (3M, Germany) to avoid bubble formation. The samples were pipetted into the microfluidic chamber through the 0.5 mm channels using microloader pipette tips (Eppendorf, Germany). The chamber was then sealed with Parafilm and heated to 95°C for 10 seconds to fully separate the strands and cooled rapidly (within 30 s) to 25 °C. Assembly and sedimentation were monitored for 20 hours on a fluorescence microscope (Axiotech Vario, Zeiss, Germany) with two LEDs (490 nm and 625 nm, Thorlabs, Germany) using a 2.5x objective (Fluar, Zeiss, Germany). Over the course of the experiment, also the top of the chamber, as indicated in Figure 2.14a, was monitored to check against evaporation. Prior to image analysis, the image stacks were stabilized using an ImageJ plugin [106].

4.6.6. Sequences

Table 4.1: Sequences of all DNA strands used in Chapter 4. Strand 0_A and 1_A are 5' labeled with Cy5. All other strands have a 5' terminal phosphate. Dashed underlines indicate information domains, hairpin loops are marked by solid underlines.

Name	Sequence (5' to 3')
0_A	P-GCA G CG TTAATTCCCG CGCCTAT CGGGAATGTAACGC AGTGGGTAATAATGA CGATAGCCGTTCG GGAAAAG CGAACGGT ATCG
1_A	P-GCA G CG TTAATTCCCG CGCCTAT CGGGAATGTAACGC AAAAGAAGAGAAAGA CGATAGCCGTTCG GGAAAAG CGAACGGT ATCG
0_B	P-GCA G CGAT ACCGTTCG CTTTTCC CGAACGGCTATCGC AGTGGGTAATAATGA GCG A ACTGTCG GTGCTTG CGACAGT GTCGC
1_B	P-GCA G CGAT ACCGTTCG CTTTTCC CGAACGGCTATCGC AAAAGAAGAGAAAGA GCG A ACTGTCG GTGCTTG CGACAGT GTCGC
0_C	P-GCA G GCGAC ACTGTCG CAAGCAC CGACAGT T CGCC AGTGGGTAATAATGA GCGG TTCCTTGC GGAGTAG GCAAGGAATCCGC
1_C	P-GCA G GCGAC ACTGTCG CAAGCAC CGACAGT T CGCC AAAAGAAGAGAAAGA GCGG TTCCTTGC GGAGTAG GCAAGGAATCCGC
0_D	P-GCA G GCGGATTCCTTGC CTACTCC GCAAGGAATC GCC AGTGGGTAATAATGA CGTTACATTCCCG ATAGGCG CGGGAATTAA CG
1_D	P-GCA G GCGGATTCCTTGC CTACTCC GCAAGGAATC GCC AAAAGAAGAGAAAGA CGTTACATTCCCG ATAGGCG CGGGAATTAA CG
$\overline{0}_A$	P-GCT G CGC ATTAACGCG CTTGTCC CGCGTTAATTGCGC TCATTATTACCCACT CGCT CTCGGCTG TTTTGCC CAGCCGAGCAGCG
$\overline{1}_A$	P-GCT G CGC ATTAACGCG CTTGTCC CGCGTTAATTGCGC TCTTTCTCTTCTTTT CGCT CTCGGCTG TTTTGCC CAGCCGAGCAGCG
$\overline{0}_B$	P-GCT G CGTT GCATTGGC GATCAAA GCCAATGCGAACGC TCATTATTACCCACT CGCAATTAACGCG GGACAAG CGCGTTAAT GCG
$\overline{1}_B$	P-GCT G CGTT GCATTGGC GATCAAA GCCAATGCGAACGC TCTTTCTCTTCTTTT CGCAATTAACGCG GGACAAG CGCGTTAAT GCG
$\overline{0}_C$	P-GCT G GTTGGAGAAGGCG AACAGCA CGCCTTC CCAACC TCATTATTACCCACT CGTTCGCATTGGC TTTGATC GCCAATGCAA CG
$\overline{1}_C$	P-GCT G GTTGGAGAAGGCG AACAGCA CGCCTTC CCAACC TCTTTCTCTTCTTTT CGTTCGCATTGGC TTTGATC GCCAATGCAA CG
$\overline{0}_D$	P-GCT G CGCTGCTCGGCTG GGCAAAA CAGCCGAG AGCGC TCATTATTACCCACT GTTGG GAAGGCG TGCTGTT CGCCTTCTCCAAC
$\overline{1}_D$	P-GCT G CGCTGCTCGGCTG GGCAAAA CAGCCGAG AGCGC TCTTTCTCTTCTTTT GTTGG GAAGGCG TGCTGTT CGCCTTCTCCAAC

References

[1] Yutaka Abe. Physical state of the very early Earth. *Lithos*, 30(3-4):223–235, 1993.

[2] Paul Adamski, Marcel Eleveld, Ankush Sood, Ádám Kun, András Szilágyi, Tamás Czárán, Eörs Szathmáry, and Sijbren Otto. From self-replication to replicator systems en route to de novo life. *Nature Reviews Chemistry*, 4(8):386–403, 2020.

[3] Emil Dandanell Agerschou, Christof B Mast, and Dieter Braun. Emergence of life from trapped nucleotides? Non-equilibrium behavior of oligonucleotides in thermal gradients. *Synlett*, 28(01):56–63, 2017.

[4] Simon Alberti, Amy Gladfelter, and Tanja Mittag. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, 176(3):419–434, 2019.

[5] Elvin A Alemán, Rajan Lamichhane, and David Rueda. Exploring RNA folding one molecule at a time. *Current opinion in chemical biology*, 12(6):647–654, 2008.

[6] Katsuhiko Ariga and Jonathan P Hill. Monolayers at air-water interfaces: from origins-of-life to nanotechnology. *The Chemical Record*, 11(4):199–211, 2011.

[7] Gonen Ashkenasy, Thomas M Hermans, Sijbren Otto, and Annette F Taylor. Systems chemistry. *Chemical Society Reviews*, 46(9):2543–2554, 2017.

[8] James Attwater, Aniela Wochner, and Philipp Holliger. In-ice evolution of RNA polymerase ribozyme activity. *Nature chemistry*, 5(12):1011, 2013.

[9] William M Aumiller Jr, Fatma Pir Cakmak, Bradley W Davis, and Christine D Keating. RNA-based coacervates as a model for membraneless organelles: formation, properties, and interfacial liposome assembly. *Langmuir*, 32(39):10042–10053, 2016.

[10] Philipp Baaske, Franz 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, 2007.

[11] Rowena Ball and John Brindley. Hydrogen peroxide thermochemical oscillator as driver for primordial RNA replication. *Journal of the Royal Society Interface*, 11(95):20131052, 2014.

[12] Laura M Barge, Yeghegis Abedian, Michael J Russell, Ivria J Doloboff, Julyan HE Cartwright, Richard D Kidd, and Isik Kanik. From Chemical Gardens to Fuel Cells: Generation of Electrical Potential and Current Across Self-Assembling Iron Mineral Membranes. *Angewandte Chemie International Edition*, 54(28):8184–8187, 2015.

[13] Steven A Benner. Defining life. *Astrobiology*, 10(10):1021–1030, 2010.

[14] Deeksha Bhartiya and Vinod Scaria. Genomic variations in non-coding RNAs: structure, function and regulation. *Genomics*, 107(2-3):59–68, 2016.

[15] Alexander Borodavka, Surendra W Singaram, Peter G Stockley, William M Gelbart, Avinoam Ben-Shaul, and Roman Tuma. Sizes of long RNA molecules are determined by the branching patterns of their secondary structures. *Biophysical journal*, 111(10):2077–2085, 2016.

[16] Clifford P Brangwynne, Christian R Eckmann, David S Courson, Agata Rybarska, Carsten Hoege, Jöbin Gharakhani, Frank Jülicher, and Anthony A Hyman. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science*, 324(5935):1729–1732, 2009.

[17] Dieter Braun and Albert Libchaber. Trapping of DNA by thermophoretic depletion and convection. *Physical review letters*, 89(18):188103, 2002.

[18] Carlos Briones, Michael Stich, and Susanna C Manrubia. The dawn of the RNA World: toward functional complexity through ligation of random RNA oligomers. *Rna*, 15(5):743–749, 2009.

[19] Bradley Burcar, Matthew Pasek, Maheen Gull, Brian J Cafferty, Francisco Velasco, Nicholas V Hud, and César Menor-Salván. Darwin's warm little pond: a one-pot reaction for prebiotic phosphorylation and the mobilization of phosphate from minerals in a urea-based solvent. *Angewandte Chemie International Edition*, 55(42):13249–13253, 2016.

[20] Aaron S Burton, Jennifer C Stern, Jamie E Elsila, Daniel P Glavin, and Jason P Dworkin. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chemical Society Reviews*, 41(16):5459–5472, 2012.

[21] Jacqui MA Carnall, Christopher A Waudby, Ana M Belenguer, Marc CA Stuart, Jérôme J-P Peyralans, and Sijbren Otto. Mechanosensitive self-replication driven by self-organization. *Science*, 327(5972):1502–1506, 2010.

[22] GM Muñoz Caro, UJ Meierhenrich, WA Schutte, B Barbier, A Arcones Segovia, H Rosenbauer, WH-P Thiemann, A Brack, and JM Greenberg. Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature*, 416(6879):403–406, 2002.

[23] Paola Caselli and Cecilia Ceccarelli. Our astrochemical heritage. *The Astronomy and Astrophysics Review*, 20(1):1–68, 2012.

[24] Thomas R Cech. The ribosome is a ribozyme. *Science*, 289(5481):878–879, 2000.

[25] Irene A Chen. Replicating towards complexity. *Nature chemistry*, 7(3):191–192, 2015.

[26] Xi Chen, Na Li, and Andrew D Ellington. Ribozyme catalysis of metabolism in the RNA world. *Chemistry & biodiversity*, 4(4):633–655, 2007.

[27] NS Chirkov, AV Akentiev, RA Campbell, S-Y Lin, KA Timoshen, PS Vlasov, and BA Noskov. Network Formation of DNA/Polyelectrolyte Fibrous Aggregates Adsorbed at the Water–Air Interface. *Langmuir*, 35(43):13967–13976, 2019.

[28] Fabio Chizzolini, Luiz FM 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, 2019.

[29] Klaus Clusius and Gerhad Dickel. Neues Verfahren zur Gasentmischung und Isotopentrennung. *NW*, 26(33):546–546, 1938.

[30] Giovanna Costanzo, Samanta Pino, Fabiana Ciciriello, and Ernesto Di Mauro. Generation of long RNA chains in water. *Journal of Biological Chemistry*, 284(48):33206–33216, 2009.

[31] Giovanna Costanzo, Raffaele Saladino, Giorgia Botta, Alessandra Giorgi, Anita Scipioni, Samanta Pino, and Ernesto Di Mauro. Generation of RNA Molecules by a Base-Catalysed Click-Like Reaction. *ChemBioChem*, 13(7):999–1008, 2012.

[32] Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[33] Francis HC Crick. The origin of the genetic code. *Journal of molecular biology*, 38(3):367–379, 1968.

[34] Laura Da Silva, Marie-Christine Maurel, and David Deamer. Salt-promoted synthesis of RNA-like molecules in simulated hydrothermal conditions. *Journal of molecular evolution*, 80(2):86–97, 2015.

[35] V Dange, Reuel B Van Atta, and Sidney M Hecht. A Mn2(+) -dependent ribozyme. *Science*, 248(4955):585–588, 1990.

[36] Bungenberg de Jong. Coacervation. In *Proc. Royal Acad. Amsterdam*, volume 32, pages 849–856, 1929.

[37] Cornel EJ De Ronde, Dominic M deR Channer, Kevin Faure, Colin J Bray, and Edward TC Spooner. Fluid chemistry of Archean seafloor hydrothermal vents: Implications for the composition of circa 3.2 Ga seawater. *Geochimica et Cosmochimica Acta*, 61(19):4025–4042, 1997.

[38] Peter Debye. Zur Theorie des Clusiusschen Trennungsverfahrens. *Annalen der Physik*, 428(3-4):284–294, 1939.

[39] 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.

[40] Robert D Deegan, Olgica Bakajin, Todd F Dupont, Greb Huber, Sidney R Nagel, and Thomas A Witten. Capillary flow as the cause of ring stains from dried liquid drops. *Nature*, 389(6653):827–829, 1997.

[41] Natalia A Denesyuk and D Thirumalai. How do metal ions direct ribozyme folding? *Nature chemistry*, 7(10):793–801, 2015.

[42] Massimo Di Giulio. Was it an ancient gene codifying for a hairpin RNA that, by means of direct duplication, gave rise to the primitive tRNA molecule? *Journal of theoretical biology*, 177(1):95–101, 1995.

[43] Kamaludin Dingle, Steffen Schaper, and Ard A Louis. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. *Interface focus*, 5(6):20150053, 2015.

[44] Matthew S Dodd, Dominic Papineau, Tor Grenne, John F Slack, Martin Rittner, Franco Pirajno, Jonathan O'Neil, and Crispin TS Little. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*, 543(7643):60–64, 2017.

[45] Elizabeth A Doherty and Jennifer A Doudna. Ribozyme structures and mechanisms. *Annual review of biophysics and biomolecular structure*, 30(1):457–475, 2001.

[46] NG Dolinnaya, NI Sokolova, OI Gryaznova, and ZA Shabarova. Site-directed modification of DNA duplexes by chemical ligation. *Nucleic acids research*, 16(9):3721–3738, 1988.

[47] Jennifer A Doudna and Thomas R Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 2002.

[48] Björn Drobot, Juan M Iglesias-Artola, Kristian Le Vay, Viktoria Mayr, Mrityunjoy Kar, Moritz Kreysing, Hannes Mutschler, and TY Dora Tang. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nature communications*, 9(1):1–9, 2018.

[49] Stefan Duhr and Dieter Braun. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52):19678–19682, 2006.

[50] 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.

[51] Thomas H Eickbush and Evangelos N Moudrianakis. A Mechanism for the Entrapment of DNA at an Air-Water Interface. *Biophysical journal*, 18(3):275–288, 1977.

[52] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.

[53] Manfred Eigen and Ruthild Winkler-Oswatitsch. Transfer-RNA, an early gene? *Naturwissenschaften*, 68(6):282–292, 1981.

[54] Eric H Ekland, Jack W Szostak, and David P Bartel. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, 269(5222):364–370, 1995.

[55] Aaron E Engelhart, Brian J Cafferty, C Denise Okafor, Michael C Chen, Loren Dean Williams, David G Lynn, and Nicholas V Hud. Nonenzymatic ligation of DNA with a reversible step and a final linkage that can be used in PCR. *ChemBioChem*, 13(8):1121–1124, 2012.

[56] Aaron E Engelhart, Matthew W Powner, and Jack W Szostak. Functional RNAs exhibit tolerance for non-heritable 2'–5' versus 3'–5' backbone heterogeneity. *Nature chemistry*, 5(5):390–394, 2013.

[57] Gözen Ertem, Robert M Hazen, and Jason P Dworkin. Sequence analysis of trimer isomers formed by montmorillonite catalysis in the reaction of binary monomer mixtures. *Astrobiology*, 7(5):715–722, 2007.

[58] Li Tai Fang, William M Gelbart, and Avinoam Ben-Shaul. The size of RNA as an ideal branched polymer. *The Journal of chemical physics*, 135(15):10B616, 2011.

[59] Nina M Fischer, Marcelo D Polêto, Jakob Steuer, and David van der Spoel. Influence of Na+ and Mg2+ ions on RNA structures studied with molecular dynamics simulations. *Nucleic acids research*, 46(10):4872–4882, 2018.

[60] George E Fox. Origin and evolution of the ribosome. *Cold Spring Harbor perspectives in biology*, 2(9):a003483, 2010.

[61] Walter Gilbert. Origin of life: The RNA world. *Nature*, 319(6055):618–618, 1986.

[62] 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.

[63] Sara Goodwin, John D McPherson, and W Richard McCombie. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6):333, 2016.

[64] Simon J Green, Daniel Lubrich, and Andrew J Turberfield. DNA hairpins: fuel for autonomous DNA devices. *Biophysical journal*, 91(8):2966–2975, 2006.

[65] Helmut Griesser, Maren Bechthold, Peter Tremmel, Eric Kervio, and Clemens Richert. Amino Acid-Specific, Ribonucleotide-Promoted Peptide Formation in the Absence of Enzymes. *Angewandte Chemie International Edition*, 56(5):1224–1228, 2017.

[66] Elizabeth C Griffith and Veronica Vaida. In situ observation of peptide bond formation at the water–air interface. *Proceedings of the National Academy of Sciences*, 109(39):15697–15701, 2012.

[67] Saul Griffith, Dan Goldwater, and Joseph M Jacobson. Self-replication from random parts. *nature*, 437(7059):636–636, 2005.

[68] Andreas R Gruber, Stephan H Bernhart, Ivo L Hofacker, and Stefan Washietl. Strategies for measuring evolutionary conservation of RNA secondary structures. *BMC bioinformatics*, 9(1):1–19, 2008.

[69] Andreas R Gruber, Ronny Lorenz, Stephan H Bernhart, Richard Neuböck, and Ivo L Hofacker. The vienna RNA websuite. *Nucleic acids research*, 36(suppl_2):W70–W74, 2008.

[70] 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, 1983.

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

[72] Thomas Henning and Dmitry Semenov. Chemistry in protoplanetary disks. *Chemical Reviews*, 113(12):9016–9042, 2013.

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

[74] Ivo L Hofacker. Vienna RNA secondary structure server. *Nucleic acids research*, 31(13):3429–3431, 2003.

[75] John J Hopfield. Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proceedings of the National Academy of Sciences*, 75(9):4334–4338, 1978.

[76] 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, 2016.

[77] 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, 2016.

[78] Hua Hu and Ronald G Larson. Marangoni effect reverses coffee-ring depositions. *The Journal of Physical Chemistry B*, 110(14):7090–7094, 2006.

[79] Wenhua Huang and James P Ferris. One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. *Journal of the American Chemical Society*, 128(27):8914–8919, 2006.

[80] 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, 2019.

[81] Alan Ianeselli, Damla Tetiker, Julian A C Stein, Alexandra Kühnlein, Dieter Braun, and T.-Y. Dora Tang. Coacervate protocells assembled at the gas-water interface of heated rock pores coalesce, divide and select RNA. (in press), 2021.

[82] Inc. Illumina. Performance specifications for the HiSeq 2500 System, 2021.

[83] Thomas Inoue and Leslie E Orgel. A nonenzymatic RNA polymerase model. *Science*, 219(4586):859–862, 1983.

[84] Autodesk Inventor and Autodesk Inventor Tooling. Autodesk®, 2019.

[85] Mario Jauker, Helmut Griesser, and Clemens Richert. Spontaneous formation of RNA strands, peptidyl RNA, and cofactors. *Angewandte Chemie International Edition*, 54(48):14564–14569, 2015.

[86] Gerald F Joyce. The antiquity of RNA-based evolution. *nature*, 418(6894):214–221, 2002.

[87] Gerald F Joyce and Leslie E Orgel. Prospects for understanding the origin of the RNA world. *Cold Spring Harbor Monograph Series*, 24:1–1, 1993.

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

[89] Lorenz Keil, Michael Hartmann, Simon Lanzmich, and Dieter Braun. Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations. *Physical Chemistry Chemical Physics*, 18(30):20153–20159, 2016.

[90] Lorenz MR 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):1–9, 2017.

[91] Maik Kindermann, Insa Stahl, Malte Reimold, Wolf Matthias Pankau, and Günter von Kiedrowski. Systems Chemistry: Kinetic and Computational Analysis of a Nearly Exponential Organic Replicator. *Angewandte Chemie International Edition*, 44(41):6750–6755, 2005.

[92] Daniel C Koboldt, Karyn Meltz Steinberg, David E Larson, Richard K Wilson, and Elaine R Mardis. The next-generation sequencing revolution and its impact on genomics. *Cell*, 155(1):27–38, 2013.

[93] Shogo Koga, David S Williams, Adam W Perriman, and Stephen Mann. Peptide–nucleotide microdroplets as a step towards a membrane-free protocell model. *Nature chemistry*, 3(9):720, 2011.

[94] Hubert Krammer, Friederike M Möller, and Dieter Braun. Thermal, Autonomous Replicator Made from Transfer RNA. *Physical Review Letters*, 108(23):238104, 2012.

[95] 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, 2015.

[96] Kelly Kruger, Paula J Grabowski, Arthur J Zaug, 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, 1982.

[97] Patrick W Kudella, Alexei V Tkachenko, Annalena 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), 2021.

[98] Alexandra Kühnlein, Simon A Lanzmich, and Dieter Braun. tRNA sequences can assemble into a replicator. *eLife*, 10:e63431, 2021.

[99] Simon A Lanzmich. *Replication in Early Evolution*. PhD thesis, LMU, 2016.

[100] Antonio Lazcano. Historical development of origins research. *Cold Spring Harbor perspectives in biology*, 2(11):a002089, 2010.

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

[102] Kristian Le Vay, Laura Isabel Weise, Kai Libicher, Judita Mascarenhas, and Hannes Mutschler. Templated self-replication in biomimetic systems. *Advanced Biosystems*, 3(6):1800313, 2019.

[103] David H Lee, Juan R Granja, Jose A Martinez, Kay Severin, and M Reza Ghadiri. A self-replicating peptide. *Nature*, 382(6591):525–528, 1996.

[104] Louis Lerman. Potential role of bubbles and droplets in primordial and planetary chemistry: exploration of the liquid-gas interface as a reaction zone for condensation process. *Origins of life*, 16(3-4):201–202, 1986.

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

[106] Kang Li and S Kang. The image stabilizer plugin for ImageJ. *can be found under http://www. cs. cmu. edu/˜ kangli/code/Image_Stabilizer. html*, 2008.

[107] Saskia Lindhoud and Mireille MAE Claessens. Accumulation of small protein molecules in a macroscopic complex coacervate. *Soft Matter*, 12(2):408–413, 2016.

[108] Ronny Lorenz, Stephan H Bernhart, Christian Höner Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. ViennaRNA Package 2.0. *Algorithms for molecular biology*, 6(1):1–14, 2011.

[109] Ard A Louis. Contingency, convergence and hyper-astronomical numbers in biological evolution. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 58:107–116, 2016.

[110] Anatol Luther, Romana Brandsch, and G Von Kiedrowski. Surface-promoted replication and exponential amplification of DNA analogues. *Nature*, 396(6708):245–248, 1998.

[111] 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, 2012.

[112] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12, 2011.

[113] William Martin, John Baross, Deborah Kelley, and Michael J Russell. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*, 6(11):805–814, 2008.

[114] William F Martin, Filipa L Sousa, and Nick Lane. Energy at life's origin. *Science*, 344(6188):1092–1093, 2014.

[115] Christof B Mast and Dieter Braun. Thermal trap for DNA replication. *Physical review letters*, 104(18):188102, 2010.

[116] 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, 2013.

[117] Thomas Matreux, Kristian LeVay, Almuth Schmid, Paula Aikkila, Lukas Belohlavek, Zeynep Caliskanoglu, Elia Salibi, Alexandra Kühnlein, Christina Springsklee, Bettina Scheu, Donald B Dingwell, Dieter Braun, Hannes Mutschler, and Christof B Mast. Heat flows in early rocks naturally optimize salt compositions for ribozymes. (in press), 2021.

[118] Cornelia Meinert, Iuliia Myrgorodska, Pierre De Marcellus, Thomas Buhse, Laurent Nahon, Søren V Hoffmann, Louis Le Sergeant d'Hendecourt, and Uwe J Meierhenrich. Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs. *Science*, 352(6282):208–212, 2016.

[119] Jean-Louis Mergny and Laurent Lacroix. Analysis of thermal melting curves. *Oligonucleotides*, 13(6):515–537, 2003.

[120] Randall E Mielke, Kirtland J Robinson, Lauren M White, Shawn E McGlynn, Kavan McEachern, Rohit Bhartia, Isik Kanik, and Michael J Russell. Iron-sulfide-bearing chimneys as potential catalytic energy traps at life's emergence. *Astrobiology*, 11(10):933–950, 2011.

[121] Stanley L Miller, Harold C Urey, and J Oro. Origin of organic compounds on the primitive earth and in meteorites. *Journal of Molecular Evolution*, 9(1):59–72, 1976.

[122] Donald R Mills, RL Peterson, and Sol Spiegelman. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences*, 58(1):217, 1967.

[123] Vinod K Misra and David E Draper. On the role of magnesium ions in RNA stability. *Biopolymers: Original Research on Biomolecules*, 48(2-3):113–135, 1998.

[124] Shin Miyakawa and James P Ferris. Sequence-and regioselectivity in the montmorillonite-catalyzed synthesis of RNA. *Journal of the American Chemical Society*, 125(27):8202–8208, 2003.

[125] Stephen J Mojzsis, G Arrhenius, KD McKeegan, TM Harrison, AP Nutman, and CRL Friend. Evidence for life on Earth before 3,800 million years ago. *Nature*, 384(6604):55–59, 1996.

[126] Pierre-Alain Monnard, Anastassia Kanavarioti, and David W Deamer. Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *Journal of the American Chemical Society*, 125(45):13734–13740, 2003.

[127] William B Moore and A Alexander G Webb. Heat-pipe earth. *Nature*, 501(7468):501–505, 2013.

[128] Matthias Morasch. *Accumulation, gelation, and crystallization of prebiotic molecules in a thermal gradient and deep UV circular dichroism*. PhD thesis, lmu, 2019.

[129] Matthias Morasch, Dieter Braun, and Christof B Mast. Heat-Flow-Driven Oligonucleotide Gelation Separates Single-Base Differences. *Angewandte Chemie*, 128(23):6788–6791, 2016.

[130] Matthias Morasch, Jonathan Liu, Christina F Dirscherl, Alan Ianeselli, Alexandra Kühnlein, Kristian Le Vay, Philipp Schwintek, Saidul Islam, Mérina K Corpinot, Bettina Scheu, et al. Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nature chemistry*, 11(9):779–788, 2019.

[131] 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.

[132] Gregory A Mountain and Christine D Keating. Formation of multiphase complex coacervates and partitioning of biomolecules within them. *Biomacromolecules*, 21(2):630–640, 2019.

[133] Karina K Nakashima, Mahesh A Vibhute, and Evan Spruijt. Biomolecular chemistry in liquid phase separated compartments. *Frontiers in molecular biosciences*, 6:21, 2019.

[134] Dan T Nguyen and Omar A Saleh. Tuning phase and aging of DNA hydrogels through molecular design. *Soft Matter*, 13(32):5421–5427, 2017.

[135] Allen P Nutman, Vickie C Bennett, Clark RL 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.

[136] Benedikt Obermayer, Hubert Krammer, Dieter Braun, and Ulrich Gerland. Emergence of information transmission in a prebiotic RNA reactor. *Physical review letters*, 107(1):018101, 2011.

[137] Carlos G Oliver, Vladimir Reinharz, and Jérôme Waldispühl. On the emergence of structural complexity in RNA replicators. *RNA*, 25(12):1579–1591, 2019.

[138] Aleksandr Ivanovich Oparin. The origin of life. 1953.

[139] Leslie E Orgel. Polymerization on the rocks: theoretical introduction. *Origins of Life and Evolution of the Biosphere*, 28(3):227–234, 1998.

[140] Leslie E Orgel. Prebiotic chemistry and the origin of the RNA world. *Critical reviews in biochemistry and molecular biology*, 39(2):99–123, 2004.

[141] Richard Owczarzy, Bernardo G Moreira, Yong You, Mark A Behlke, and Joseph A Walder. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry*, 47(19):5336–5353, 2008.

[142] Robert Pascal, Addy Pross, and John D Sutherland. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open biology*, 3(11):130156, 2013.

[143] Bhavesh H Patel, Claudia Percivalle, Dougal J Ritson, Colm D Duffy, and John D Sutherland. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature chemistry*, 7(4):301–307, 2015.

[144] Natasha Paul and Gerald F Joyce. Self-replication. *Current Biology*, 13(2):R46, 2003.

[145] Natasha Paul and Gerald F Joyce. Minimal self-replicating systems. *Current opinion in chemical biology*, 8(6):634–639, 2004.

[146] Ben KD Pearce, Ralph E Pudritz, Dmitry A Semenov, and Thomas K Henning. Origin of the RNA world: The fate of nucleobases in warm little ponds. *Proceedings of the National Academy of Sciences*, 114(43):11327–11332, 2017.

[147] CJ Petit, M-H Hwang, and J-L Lin. The Soret effect in dilute aqueous alkaline earth and nickel chloride solutions at 25° C. *International Journal of Thermophysics*, 7(3):687–697, 1986.

[148] Raghav R Poudyal, Rebecca M Guth-Metzler, Andrew J Veenis, Erica A Frankel, Christine D Keating, and Philip C Bevilacqua. Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nature communications*, 10(1):1–13, 2019.

[149] Raghav R Poudyal, Fatma Pir Cakmak, Christine D Keating, and Philip C Bevilacqua. Physical principles and extant biology reveal roles for RNA-containing membraneless compartments in origins of life chemistry. *Biochemistry*, 57(17):2509–2519, 2018.

[150] Matthew W Powner, Béatrice Gerland, and John D Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244):239–242, 2009.

[151] 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.

[152] 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, 2015.

[153] Addy Pross. The driving force for life's emergence: kinetic and thermodynamic considerations. *Journal of theoretical Biology*, 220(3):393–406, 2003.

[154] 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, 2016.

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

[156] Sudha Rajamani, Alexander Vlassov, Seico Benner, Amy Coombs, Felix Olasagasti, and David Deamer. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins of Life and Evolution of Biospheres*, 38(1):57–74, 2008.

[157] Julius Rebek. Synthetic self-replicating molecules. *Scientific American*, 271(1):48–55, 1994.

[158] 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, 2014.

[159] 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, 2015.

[160] Debra L Robertson and Gerald F Joyce. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, 344(6265):467–468, 1990.

[161] Michael P Robertson and Gerald F Joyce. The origins of the RNA world. *Cold Spring Harbor perspectives in biology*, 4(5):a003608, 2012.

[162] Andrei S Rodin, Eörs Szathmáry, and Sergei N Rodin. On origin of genetic code and tRNA before translation. *Biology Direct*, 6(1):1–24, 2011.

[163] Boris Rubinov, Nathaniel Wagner, Hanna Rapaport, and Gonen Ashkenasy. Self-Replicating Amphiphilic β-Sheet Peptides. *Angewandte Chemie*, 121(36):6811–6814, 2009.

[164] Rick Russell, Ian S Millett, Mark W Tate, Lisa W Kwok, Bradley Nakatani, Sol M Gruner, Simon GJ Mochrie, Vijay Pande, Sebastian Doniach, Daniel Herschlag, et al. Rapid compaction during RNA folding. *Proceedings of the National Academy of Sciences*, 99(7):4266–4271, 2002.

[165] Jan W Sadownik and Douglas Philp. A simple synthetic replicator amplifies itself from a dynamic reagent pool. *Angewandte Chemie*, 120(51):10113–10118, 2008.

[166] Annalena Salditt, Lorenz MR 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, 2020.

[167] Stephen W Santoro and Gerald F Joyce. A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences*, 94(9):4262–4266, 1997.

[168] Paul Schimmel and Barry Henderson. Possible role of aminoacyl-RNA complexes in noncoded peptide synthesis and origin of coded synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 91(24):11283, 1994.

[169] Erik A Schultes, Alexander Spasic, Udayan Mohanty, and David P Bartel. Compact and ordered collapse of randomly generated RNA sequences. *Nature structural & molecular biology*, 12(12):1130–1136, 2005.

[170] Robert Shapiro. A replicator was not involved in the origin of life. *IUBMB life*, 49(3):173–176, 2000.

[171] Stephen Jefferson Sharp, Jerone Schaack, Lyan Cooley, Debroh Johnson Burke, and Dieter Soil. Structure and transcription of eukaryotic tRNA gene. *Critical Reviews in Biochemistry*, 19(2):107–144, 1985.

[172] Scott K Silverman. Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends in biochemical sciences*, 41(7):595–609, 2016.

[173] Victor Sojo, Barry Herschy, Alexandra Whicher, Eloi Camprubi, and Nick Lane. The origin of life in alkaline hydrothermal vents. *Astrobiology*, 16(2):181–197, 2016.

[174] Petr Stadlbauer, Jiří Šponer, Giovanna Costanzo, Ernesto Di Mauro, Samanta Pino, and Judit E Šponer. Tetraloop-like Geometries Could Form the Basis of the Catalytic Activity of the Most Ancient Ribooligonucleotides. *Chemistry–A European Journal*, 21(9):3596–3604, 2015.

[175] Thomas A Steitz and Peter B Moore. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in biochemical sciences*, 28(8):411–418, 2003.

[176] Michael Stich, Carlos Briones, and Susanna C Manrubia. On the structural repertoire of pools of short, random RNA sequences. *Journal of theoretical biology*, 252(4):750–763, 2008.

[177] COMSOL AB Stockholm, Sweden. COMSOL Multiphysics® version 5.4.

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

[179] Eörs Szathmáry and Irina Gladkih. Sub-exponential growth and coexistence of non-enzymatically replicating templates. *Journal of Theoretical Biology*, 138(1):55–58, 1989.

[180] Jack W Szostak. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry*, 3(1):1–14, 2012.

[181] Jack W Szostak. On the origin of life. *MEDICINA (Buenos Aires)*, 76(4):199–203, 2016.

[182] Jack W Szostak and Andrew D Ellington. In vitro selection of functional RNA sequences. *COLD SPRING HARBOR MONOGRAPH SERIES*, 24:511–511, 1993.

[183] Naokata Takeyama and Kimie Nakashima. Proportionality of intrinsic heat of transport to standard entropy of hydration for aqueous ions. *Journal of solution chemistry*, 17(4):305–325, 1988.

[184] T-Y Dora Tang, Dirk van Swaay, Andrew DeMello, JL Ross Anderson, and Stephen Mann. In vitro gene expression within membrane-free coacervate protocells. *Chemical Communications*, 51(57):11429–11432, 2015.

[185] Jonathan D Toner and David C Catling. A carbonate-rich lake solution to the phosphate problem of the origin of life. *Proceedings of the National Academy of Sciences*, 117(2):883–888, 2020.

[186] Shoichi Toyabe and Dieter Braun. Cooperative ligation breaks sequence symmetry and stabilizes early molecular replication. *Physical Review X*, 9(1):011056, 2019.

[187] Gabriele Varani. Exceptionally stable nucleic acid hairpins. *Annual review of biophysics and biomolecular structure*, 24(1):379–404, 1995.

[188] Alexander V Vlassov, Sergei A Kazakov, Brian H Johnston, and Laura F Landweber. The RNA world on ice: a new scenario for the emergence of RNA information. *Journal of molecular evolution*, 61(2):264–273, 2005.

[189] Günter von Kiedrowski. A self-replicating hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, 25(10):932–935, 1986.

[190] Catherine Walsh, Tom J Millar, Hideko Nomura, Eric Herbst, Susanna Widicus Weaver, Yuri Aikawa, Jacob C Laas, and Anton I Vasyunin. Complex organic molecules in protoplanetary disks. *Astronomy & Astrophysics*, 563:A33, 2014.

[191] Travis Walton and Jack W Szostak. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *Journal of the American Chemical Society*, 138(36):11996–12002, 2016.

[192] Bing Wang and Ian O Sutherland. Self-replication in a Diels–Alder reaction. *Chemical Communications*, (16):1495–1496, 1997.

[193] Christoph J Wienken, Philipp Baaske, Stefan Duhr, and Dieter Braun. Thermophoretic melting curves quantify the conformation and stability of RNA and DNA. *Nucleic acids research*, 39(8):e52–e52, 2011.

[194] Yuri I Wolf and Eugene V Koonin. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct*, 2(1):1–25, 2007.

[195] Sarah A Woodson. Compact intermediates in RNA folding. *Annual review of biophysics*, 39:61–77, 2010.

[196] Kun Xi, Feng-Hua Wang, Gui Xiong, Zhong-Liang Zhang, and Zhi-Jie Tan. Competitive binding of Mg2+ and Na+ ions to nucleic acids: from helices to tertiary structures. *Biophysical journal*, 114(8):1776–1790, 2018.

[197] Peng Yin, Harry MT Choi, Colby R Calvert, and Niles A Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451(7176):318–322, 2008.

[198] Aron M Yoffe, Peter Prinsen, Ajaykumar Gopal, Charles M Knobler, William M Gelbart, and Avinoam Ben-Shaul. Predicting the sizes of large RNA molecules. *Proceedings of the National Academy of Sciences*, 105(42):16153–16158, 2008.

[199] 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, 2011.

[200] Hani S Zaher and Peter J Unrau. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *Rna*, 13(7):1017–1026, 2007.

[201] Anna Zampetaki, Andreas Albrecht, and Kathleen Steinhofel. Long non-coding RNA structure and function: is there a link? *Frontiers in physiology*, 9:1201, 2018.

[202] David Yu Zhang, Andrew J Turberfield, Bernard Yurke, and Erik Winfree. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 318(5853):1121–1125, 2007.

[203] 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:e51888, 2019.

[204] Michael Zuker and Patrick Stiegler. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic acids research*, 9(1):133–148, 1981.

Acknowledgements

First and foremost, I want to thank **Dieter** for giving me the opportunity to work with him on truly exciting projects with wonderful people around. It took me a while to recognize, but I really appreciate the vision and passion you have for science and for the origin of life field.

A huge thanks also goes to **Christof** for your great wit and your inability to accept inadequate solutions - without you we'd still be stuck with some gruesome version of the traps covered in heat paste.

The most heartfelt thanks goes to **Annalena**, **Patta** and **Thomas** - and you are in alphabetical order here, because there is no other way to do this right. Thank you for always listening, helping and supporting, for motivation, comfort and appreciation, for early and sometimes also very late coffees and for everything else. Thanks to **Annalena** for being the greatest companion through California, **Patta** for covering up social anxieties and **Thomas** for being the best project buddy - I didn't know it was my lucky day when I taught you how to build traps.

Thanks to **multibunt-Chrissy** for initially and always trusting in me, to **Matzi** and **Zhenya** for a great trip to San Diego and for taking me in, to **Julian** for new perspectives and endless discussions and to **Noel** for your patience and endurance with us and with the traps and for your educational physics questions. Also, a great shout-out to the rest of **the Braunies** for making the last years remarkable.

Thanks to the **real** and the **surreal David**, **Joris**, **Kimbu** and the rest of the QBM and WG crowd for all those amazing trips, evenings and parties - well, back when we still had them. I don't know what Munich would have been without you.

Thank you **Veronika** & **Lukas** for being my homebase away from home, for all the comfort and friendship and for our lovely Christmas tradition.

I also want to thank **Lea**, **Lukas**, **Jonas**, **Marvin** and **Flo** for their joy and passion, which gave us so many unforgettable moments, and for teaching me what team-spirit truly means.

Thank you to my **mom** and **dad**. You have always supported me, believed in me, cheered me on and have been my advocates when I couldn't.

My last and very special thanks goes to **my sister**, who is always there for me with as much love and time as you can have.

A. Associated First Author Publication

A. Kühnlein, S. A. Lanzmich, D. Braun. **tRNA sequences can assemble into a replicator.** *eLife*, 10:e63431, 2021.

RESEARCH ARTICLE

tRNA sequences can assemble into a replicator

Alexandra Kühnlein†, Simon A Lanzmich†, Dieter Braun*

Systems Biophysics, Physics Department, Center for NanoScience, Ludwig-Maximilians-Universität München, Munich, Germany

Abstract Can replication and translation emerge in a single mechanism via self-assembly? The key molecule, transfer RNA (tRNA), is one of the most ancient molecules and contains the genetic code. Our experiments show how a pool of oligonucleotides, adapted with minor mutations from tRNA, spontaneously formed molecular assemblies and replicated information autonomously using only reversible hybridization under thermal oscillations. The pool of cross-complementary hairpins self-selected by agglomeration and sedimentation. The metastable DNA hairpins bound to a template and then interconnected by hybridization. Thermal oscillations separated replicates from their templates and drove an exponential, cross-catalytic replication. The molecular assembly could encode and replicate binary sequences with a replication fidelity corresponding to 85–90 % per nucleotide. The replication by a self-assembly of tRNA-like sequences suggests that early forms of tRNA could have been involved in molecular replication. This would link the evolution of translation to a mechanism of molecular replication.

***For correspondence:** dieter.braun@lmu.de

†These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 18

Received: 24 September 2020
Accepted: 28 January 2021
Published: 02 March 2021

Reviewing editor: Gonen Ashkenasy, Ben-Gurion University, Israel

Introduction

A machine to create replicate of itself is an old dream of engineering (***von Neumann, 1951***). Biological systems have solved this problem long ago at the nanoscale with DNA and RNA. Their replication machinery was optimized to perfection through Darwinian evolution. In modern living systems, the replication of DNA and RNA necessitates the formation of covalent bonds. It requires an interconnected machinery: proteins need to perform base-by-base replication of sequence information, a modern metabolism to supply activated molecules, and tRNA as well as the ribosome to create the required proteins.

This is a complex system to set up in the first place at the emergence of life. The RNA world hypothesis proposes, that early on, the catalytic function of highly defined RNA sequences was used for self-replication (***Horning and Joyce, 2016***; ***Orgel, 2004***; ***Turk et al., 2011***). These ribozymes catalyze the ligation of RNA (***Doudna et al., 1991***; ***Mutschler et al., 2015***; ***Paul and Joyce, 2002***; ***Robertson et al., 2001***; ***Walton et al., 2020***) and the addition of individual bases (***Attwater et al., 2013***; ***Horning and Joyce, 2016***). These very special sequences were engineered using in vitro evolution. It is unclear how autonomous evolution of early life could have reached such levels of sequence complexity.

Here, we focus on how such replication may have been predated by simpler forms of self-replication. Creating a replicator must fulfill a series of requirements. Replication must yield fidelity in copying, be fast, enable exponential replication, be fed by an autonomous energy source, not require complex sequences and should not form too many replicates without the existence of a template.

We show that replication of information can be realized by the reversible hybridization interactions between tRNA-like molecules alone. The proposed mechanism is driven by an external physical non-equilibrium setting, in our case thermal oscillations. Since the process does not involve chemical ligation, it does not rely on a particular non-enzymatic or catalytic ligation chemistry (***Dolinnaya et al., 1988***; ***Engelhart et al., 2012***; ***Patzke et al., 2014***; ***Pino et al., 2011***;

eLife digest The genetic code stored within DNA contains the instructions for manufacturing all the proteins organisms need to develop, grow and survive. This requires molecular machines that 'transcribe' regions of the genetic code into RNA molecules which are then 'translated' into the string of amino acids that form the final protein. However, these molecular machines and other proteins are also needed to replicate and synthesize the sequences stored in DNA. This presents evolutionary biologists with a 'chicken-and-egg' situation: which came first, the DNA sequences needed to manufacture proteins or the proteins needed to transcribe and translate DNA?

Understanding the order in which DNA replication and protein translation evolved is challenging as these processes are tightly intertwined in modern-day species. One theory, known as the 'RNA world hypothesis', suggests that all life on Earth began with a single RNA molecule that was able to make copies of itself, as DNA does today. To investigate this hypothesis, Kühnlein, Lanzmich and Braun studied a molecule called transfer RNA (or tRNA for short) which is responsible for translating RNA into proteins. tRNA is assumed to be one of the earliest evolved molecules in biology. Yet, why it was present in early life forms before it was needed for translation still remained somewhat of a mystery.

To gain a better understanding of tRNA's role early in evolution, Kühnlein, Lanzmich and Braun made small changes to its genetic code and then carried out tests on these tRNA-like sequences. The experiments showed these 'early' forms of tRNA can actually self-assemble into a molecule which is capable of replicating the information stored in its sequence. It suggests early forms of tRNA could have been involved in replication before modern tRNA developed its role in protein translation.

With these experiments, Kühnlein, Lanzmich and Braun have identified a possible evolutionary link between DNA replication and protein translation, suggesting the two processes emerged through one shared pathway: tRNA. This deepens our understanding about the origins of early life, while taking biochemists one step closer to their distant goal of recreating self-replicating molecular machines in the laboratory.

Rohatgi et al., 1996; *Sievers and von Kiedrowski, 1994*; *von Kiedrowski, 1986*) or particular catalytically active sequences, but merely requires sequence complementarity. The advantage of reversible hybridization is the re-usability of educts and products. Moreover, sequence-encoded interactions can self-select by forming agglomerates.

Nature's approach to achieve exponential growth is the usage of cross-catalysis: the replicate of a template serves as a template for the next round of replication. For short replicators under isothermal conditions, the binding between template and replicate has to be weak such that the dissociation of strands happens spontaneously and is not rate limiting (*Paul and Joyce, 2002*; *Sievers and von Kiedrowski, 1994*; *von Kiedrowski, 1986*). For longer replicates, temperature change has successfully been used to separate strands for replication catalyzed by thermostable proteins (*Barany, 1991*; *Saiki et al., 1985*). For catalytic RNA, elevated salt concentrations disfavor strand separation by temperature and catalyze hydrolysis (*Horning and Joyce, 2016*). In an interesting alternative to strand separation by temperature, Schulman et al. used moderate shear flows to separate DNA tile assemblies (*Schulman et al., 2012*).

Apart from nucleotide-based replicators, very interesting replication systems using non-covalent interactions have been developed with non-biological compounds (*Bottero et al., 2016*; *Sadownik and Philp, 2008*; *Tjivikua et al., 1990*), peptide-based approaches (*Altay et al., 2017*; *Bourbo et al., 2011*; *Carnall et al., 2010*; *Lee et al., 1996*; *Rubinov et al., 2012*), and peptide nucleic acids (*Ura et al., 2009*). We also want to point to several instructive reviews about the state-of-the-art systems chemistry regarding self-replication (*Adamski et al., 2020*; *Ashkenasy et al., 2017*; *Kosikova and Philp, 2017*).

In the past, metastable hairpin states have been prepared in a physically separated manner. The reaction was then triggered by mixing. For example, the mixing of hairpins with a trigger sequence has been shown to form long concatemers (*Dirks and Pierce, 2004*). With a similar logic, mixing a low entropy combination of molecules was used to create entropically driven DNA machines,

including exponentially amplifying assemblies (**Zhang et al., 2007**). These reactions run downwards into the binding equilibrium. However, the preparation of the initial low entropy state required human intervention or a unique flow setting for mixing.

Sequence design

We designed a set of cooperatively replicating DNA strands using the program package NUPACK (**Zadeh et al., 2011**). The sequences are designed to have self-complementary double hairpins and are pairwise complementary within the molecule pool, such that the 3′ hairpin of one strand is complementary to the 5′ hairpin of the next. Their structure resembles the secondary structure of proto-tRNAs proposed by stereochemical theories (**Figure 1a**), comprising two hairpin loops that surround the anticodon with a few neighboring bases (**Krammer et al., 2012**). The lengths of 82–84 nt of the double hairpins are that of average tRNA molecules (**Sharp et al., 1985**), with stem loops consisting of 30–33 nt and the information-encoding interjacent domains of 15 nt. As the replication mechanism is based on hybridization only, it is expected to perform equally well for DNA and RNA. Here, we implemented the system with DNA and not RNA as done previously (**Krammer et al., 2012**). Both, in the design and the implementation we did not see significant differences between the two versions. Because of the simpler and more inexpensive synthesis of the 82–84 nt long sequences we now implemented the replicator in DNA. Due to short heating times and moderate magnesium concentrations, we estimate that an RNA version could survive for days if not weeks (**Li and Breaker, 1999**; **Mariani et al., 2018**). The most critical step regarding the RNA stability would be the initial temperature spike to 95 ˚C, which remains unchanged from our previous study (**Krammer et al., 2012**) and did not prove critical. We also show that an RNA version behaves structurally identical to the implemented DNA version (**Figure 1—figure supplement 1**).

Replication mechanism

The replication mechanism is a template-based replication, where instead of single nucleotides, information is encoded by a succession of oligomers. The domain, at the location of the anticodon in tRNA, is the template sequence and thus contains the information to be replicated. We therefore term it information domain. The goal is to replicate the succession of information domains.

To allow longer replicates, we chose the resulting meta-sequences to be periodic with a periodicity of four different hairpins. This makes the minimal cyclic meta-sequence large enough to keep the information domains accessible even in cyclic configuration. The information domains feature a binary system and contain sequences marked by '0' and '1' (blue/red). For replication, two sets of strands replicate strings of codons in a cross-catalytic manner (**Figure 1b**), using complementary information domains (light/dark colors).

The replication is driven by thermal oscillations and operates in four steps (**Figure 1b**): (0) Fast cooling within seconds brings the strands to their activated state with both hairpins closed. (1) At the base temperature, activated strands with complementary information domains can bind to an already assembled template. (2) Thermal fluctuations cause open-close fluctuations of the hairpins. When strands are already bound to a template at the information domain, those fluctuations permit adjacent complementary hairpins of different strands to bind. In this way, the succession of information domains is replicated. (3) Subsequent heating splits the newly formed replicate from the template at the information domains. Due to their higher melting temperatures, the backbone of hairpin strands remains stable. Both, replicate and template, are available for a new replication round. This makes both the replicate and the template replication cross-catalytic in a subsequent step. Later, high temperatures spikes can unbind and recycle all molecules for new rounds of replication.

Because of the initial fast cooling, all hairpins are closed in free solution. This inhibits the formation of replicates without template. While the binding of adjacent hairpins with template happens within minutes, hairpins in free solution connect without template only on timescales slower than hours and thus give false positives at a very low rate.

The basic principle of this replication mechanism was previously explored by Krammer et al. using a set of four hairpins using half a tRNA sequence (36 nt) that amplified into dimers (**Krammer et al., 2012**). This amplification could not encode information and suffered from a high rate (>50 %) of unspecific amplification without template (Figure 4 therein). Here, in contrast, we demonstrate exponential amplification, and the replicator can now encode sequence information '0' and '1' with four

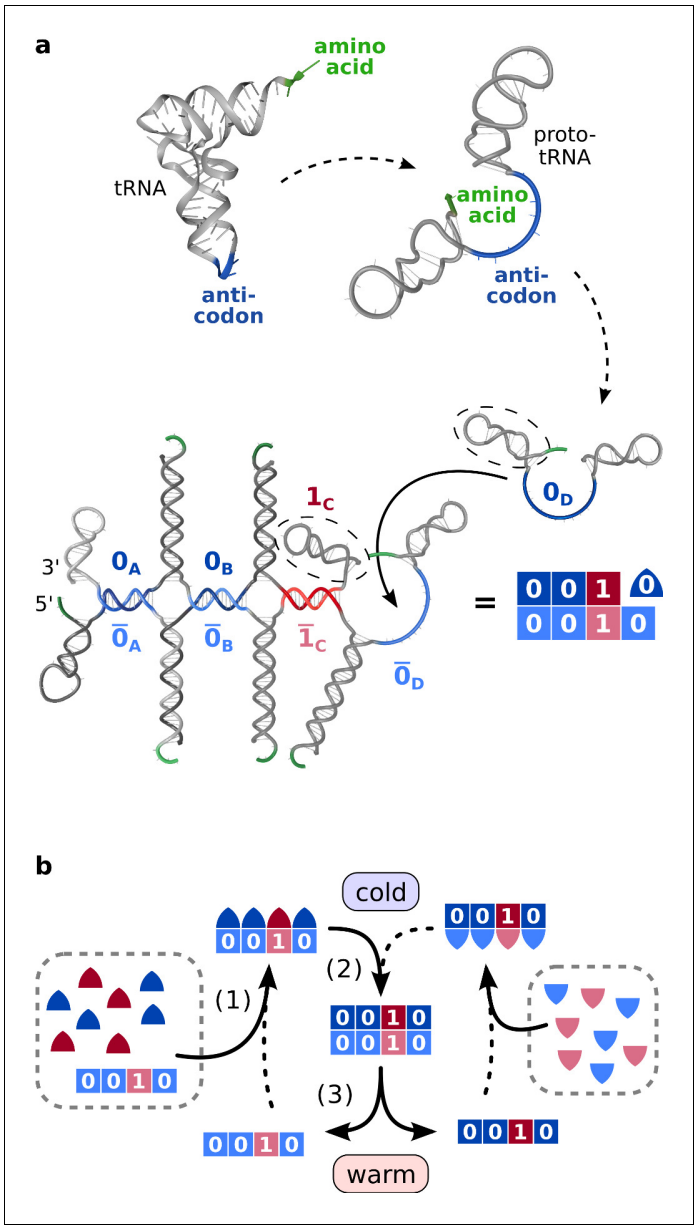

Figure 1. Heat-driven replication by hybridization using hairpin structures inspired from transfer RNA. (**a**) Transfer RNA folds into a double-hairpin conformation upon very few base substitutions. In that configuration, the 3'-terminal amino acid binding site (green) is close to the anticodon (blue) and a double hairpin structure forms. A set of pairwise complementary double hairpins can encode and replicate sequences of information. A binary code implemented in the position of the anti-codon, the information domain, allows to encode and replicate binary sequences (red vs blue). Each strand (82-84 nt) comprises two hairpin loops (gray) and an interjacent unpaired information domain of 15 nt length (blue/red, here: 0_D). The displayed structure of eight strands shows replication of a template corresponding to the binary code 0010. Note, that no covalent linkage is involved in the process. (**b**) Replication is driven by thermal oscillations in four steps: (0) The hairpins are activated into their closed conformation by fast cooling indicated by triangles. (1) Strands with matching information domain bind to the template. (2) Fluctuations in the bound strands' hairpins facilitate the hybridization of neighboring strands. (3) Subsequent heating splits replica from template, while keeping the longer hairpin sequences connected, freeing both as templates for the next cycle.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Secondary structure predictions and free energy calculations for the replicator in DNA and RNA using NUPACK.

bits. Moreover, the strands making up the new replicator are double hairpins with the sequence structure and length of tRNA. The replicator now shows a significantly decreased unspecific amplification without template of approximately 10 % (Figure 5a).

Results

Analysis of molecule conformations

Native polyacrylamide gel electrophoresis (PAGE) showed that the double hairpins assembled as intended (***Figure 2***). Comparing different subsets of strands allowed to identify all gel bands.

All complexes were formed at concentrations of 200 nM of each strand and could be resolved despite their branched tertiary structure. Friction coefficients of complexes of two to four strands were 1.6–1.8-fold higher than for linear dsDNA, and 2.4-fold higher for larger complexes (4:4 configuration, ca. 660 nt, ***Figure 2—figure supplement 1***). This agrees with the branched structure of the suggested strand assembly geometry (***Figure 1a***). Partially assembled complexes of two or three strands bound to a four-strand template could be resolved (***Figure 6—figure supplement 1***). Complexes containing single bound information domains were not stable during electrophoresis (***Figure 2***, lanes 2, 7 and ***Figure 6—figure supplement 1***). This allowed to differentiate fully assembled complexes from those where individual strands are bound to a template but have not formed

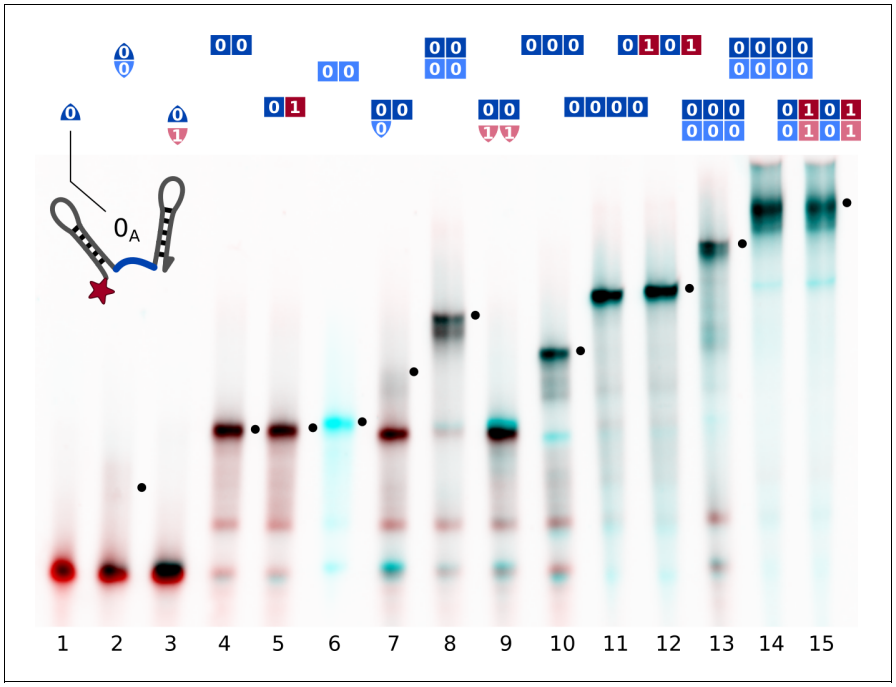

Figure 2. Assembly of different subsets of the cross-replicating system of strands observed by native gel electrophoresis. Samples contained strands at 200 nM concentration each and were slowly annealed as described in Materials and methods. Lane contents are indicated at the top of each lane. Comparison of different lanes allowed for the attribution of bands to complexes. Complexes incorporating all present strands are marked (•). The red channel shows the intensity 0_A-Cy5, the cyan channel shows SYBR Green I fluorescence. Single information domain bonds (lane 2, 7) break during gel electrophoresis.

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Source data for assembly of different subsets of the cross-replicating system of strands observed by native gel electrophoresis.

Figure supplement 1. Gel mobilities of different complexes compared to linear dsDNA.

Figure supplement 1—source data 1. Source data for gel mobilities of different complexes compared to linear dsDNA.

backbone duplexes. Covalent end labels and two reference lanes on each gel were used to quantify concentrations from gel intensities using image analysis as described in Materials and methods.

Selection by agglomeration and sedimentation

For a replicator to be autonomous, there must be a mechanism in place to select, assemble and (re-) accumulate its molecular components purely at one location. We argue that DNA hydrogels could offer such a solution. While DNA often, also in our case, assembles into agglomerates, DNA hydrogels have been shown to be able to form fluid phases if gaps of single bases were added to create flexible linkers between molecules (***Nguyen and Saleh, 2017***).

We combined eight matching hairpin sequences of design as introduced in ***Figure 1*** at moderately elevated concentrations and cooled the system to only 25 ˚C after separating the molecules at 95 ˚C (***Figure 3***). We found the spontaneous formation of agglomerates that were large enough to sediment under gravity. The initial homogeneous fluorescence turned into micrometer-sized grains and sedimented within hours. The fluorescence was provided by a covalently attached label to either strand 0_A or 1_A. Since the double hairpins have a periodic boundary condition, they can create large assemblies (***Figure 3a***).

It is evident from ***Figure 3—video 1*** that the sedimentation was very selective. When only seven of the eight matching hairpins were present, sedimentation was much weaker and, in most cases, undetectable (***Figure 3b,c***). For the full system, the sedimentation kinetics showed to be strongly concentration dependent (***Figure 3—figure supplement 1b***). Analogous experiments with random sequences (random pool of 84 nt strands) at equal concentration did not show agglomeration nor sedimentation (***Figure 3—figure supplement 1c***). We have previously found that similar hairpin molecules provided the shortest sequences capable of forming agglomerates (***Morasch et al., 2016***).

The above results suggest that agglomeration could serve as an efficient way to assemble matching hairpins from much less structured and selected sequences in an autonomous way. After the molecules have been assembled as sedimented agglomerates, a convection flow can carry the large assemblies into regions of warmer temperatures, where the molecules would be disassembled by heat and activated for replication with a cooling step. Similar recycling behavior is seen in thermal gradient traps (***Morasch et al., 2016***), which were also found to enhance the molecular assembly (***Mast et al., 2013***) with characteristics that can match the above scenario.

Templating kinetics

Hybridization between stems of neighboring hairpins (***Figure 1b***, step 2) was catalyzed by the presence of already assembled complexes $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, confirming its role as a template. Assembly kinetics at 45 ˚C were recorded in reactions containing 200 nM of each strand for a range of template concentrations. At 120 nM template concentration, 40 % yield was achieved within 10 min (***Figure 4b***, black line). The untemplated, spontaneous reaction proceeded significantly slower (1.4 % yield, light gray line).

Assembly rates showed a strong dependence on incubation temperature (***Figure 4c***). At 39 ˚C, the reaction proceeded significantly slower than at 42 ˚C or 45 ˚C. This is because the hairpins are predominantly in closed configuration and cannot bind to neighboring molecules in the assembly. Binding between complementary information domains still occurs, but the formation of bonds between neighboring strands becomes rate limiting. Above the melting temperature of the information domain (48 ˚C) (see ***Figure 4—figure supplement 1***), template-directed assembly becomes slower. However, the slower kinetics of template-directed product formation are partially superposed by the spontaneous product formation lacking an initial template (***Figure 4c***, small circles), which becomes an additional reaction channel due to the now open hairpins.

Exponential amplification

As intermediate step toward replication, we studied amplification reactions under thermal oscillations (***Figure 5***). The amplification reactions only contained strands encoding for information domain '0', that is 0_A, $\bar{0}_A$, 0_B, $\bar{0}_B$, …, $\bar{0}_D$. The strands were subjected to thermal oscillations between T_{base} = 45 ˚C and T_{peak} = 67 ˚C. The lower temperature was held for 20 min, the upper for one second with temperature ramps amounting to 20±1 s in each full cycle. This asymmetric shape of the temperature cycle accords with differences in kinetics of the elongation step and the melting of the information

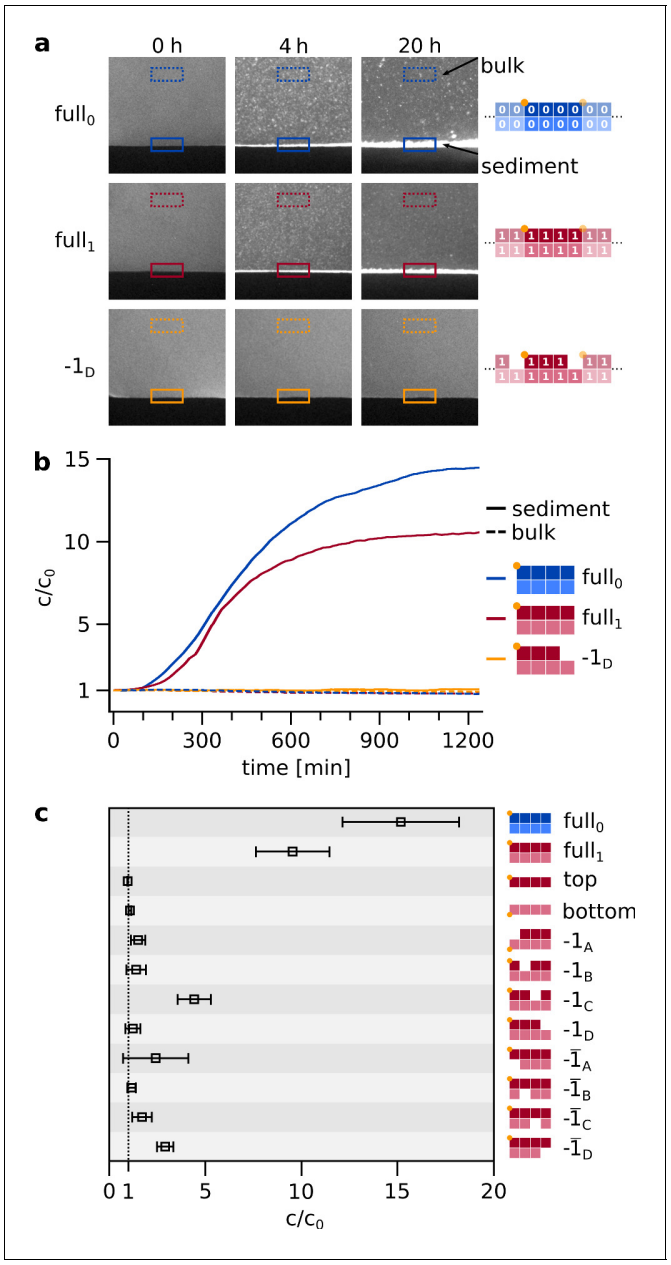

Figure 3. Spontaneous self-assembly and sedimentation of matching hairpins. (**a**) In a simple, sealed microfluidic chamber (***Figure 3—figure supplement 2***), the hairpin strands can self-assemble into agglomerates and sediment on a timescale of hours. The sample was initially heated to 95 ˚C for 10 s to ensure an unbound initial state, then rapidly (within 30 s) cooled to 25 ˚C, where self-assembly and sedimentation occured. Note, that agglomeration and sedimentation only occured if all eight matching hairpins were provided (top two rows) but not in the case of a knockout (-1_D, bottom row). For quantification, the bulk and sediment intensities were normalized by the first frame after heating. Samples contained strands at total concentration of 5 µM, about threefold higher than in ***Figure 2*** and the following replication experiments. (**b**) Time traces of concentration increase for sediment and bulk of different configurations, same examples as shown in **a**. The time traces of all further knockout permutations are shown in ***Figure 3—figure supplement 1b***. (**c**) Final concentration increase of sediment, relative to first frame after heating, for all configurations. The final values ($N \geq 3$) for c/c_0 are retrieved from fitting the time traces. For the full set of complementary hairpins, self-assembly and sedimentation is most pronounced.
The online version of this article includes the following video, source data, and figure supplement(s) for figure 3:

Source data 1. Source data for spontaneous self-assembly and sedimentation of matching hairpins.
Figure supplement 1. Extended data on self-assembly and sedimentation.

Figure 3 continued on next page

Figure 3 continued

Figure supplement 1—source data 1. Source data for extended data on self-assembly and sedimentation.
Figure supplement 2. Sketch of microfluidic chamber.
Figure 3—video 1. Sedimentation of DNA agglomerates.
https://elifesciences.org/articles/63431#fig3video1

domain. It is typical for trajectories in thermal convection settings with local heating (***Braun et al., 2003***).

The growth of molecular assemblies with different initial concentrations of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ revealed an almost linear dependence of the reaction velocity on the initial amount of template (***Figure 5a, b***). This confirms the exponential nature of the replication. The cross-catalytic replication kinetics can be described by a simplistic model that only considers the concentrations $c(t)$ of the template $0_A0_B0_C0_D$ and its complement $\bar{c}(t)$ of $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$:

$$\frac{d}{dt}c(t) = k\cdot\bar{c}(t) + k_0\,,\ \frac{d}{dt}\bar{c}(t) = k\cdot c(t) + k_0 \qquad (1)$$

Here, k is the rate of cross-catalysis and k_0 the spontaneous formation rate. For $c(t)\approx\bar{c}(t)$, the model corresponds to simple exponential growth on a per-cycle basis. The model can be solved in closed form but does not account for saturation effects from the depletion of monomers. Therefore, it is not valid for concentrations similar to the total concentration of each strand. Fitting the model to the amplification reactions with 0–45 nM of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ revealed rate constants of k = 0.16 cycle^{-1} and k_0 = 0.4 nM cycle^{-1} (***Figure 5b***). Amplification was robust with regard to the peak temperature of the oscillations. For T_{peak} below 74 ˚C, the reaction remained almost unaffected (***Figure 5c***). Above, the temperature is too close to the melting transitions of the hairpin-hairpin duplexes, ranging from 76 to 79 ˚C (***Figure 4—figure supplement 1***).

The ability to withstand consecutive dilutions is characteristic for exponentially growing replicators and was tested for in serial transfer experiments. Strands encoding for '0' (i.e. 0_A, $\bar{0}_A$, 0_B, etc.) were thermally cycled with 30 nM of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$. After three cycles each, samples were diluted one to one with buffer containing all eight strands as monomers at 200 nM each (***Figure 5d***). This high frequency of dilutions prevented the reaction from transitioning into the saturating regime. The cross-catalytic model was fitted to the data with the dilution factor as single free parameter, that was found

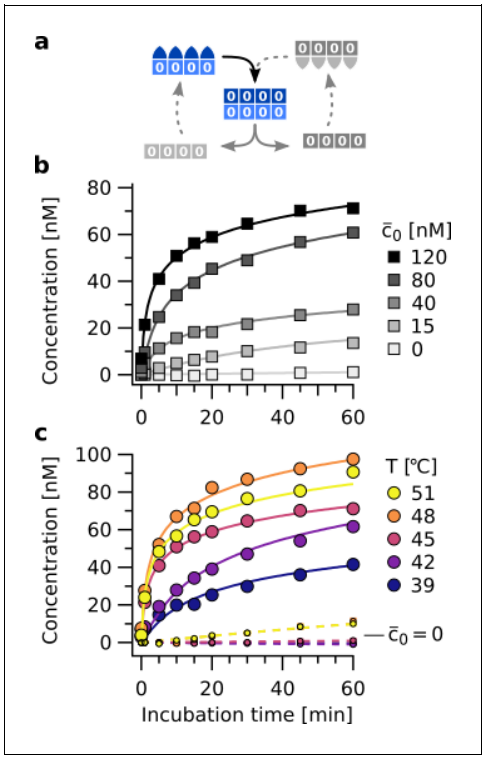

Figure 4. Isothermal template assisted product formation. (**a**) Schematic representation of the templating step at constant temperature. (**b**) Kinetics of tetramer formation at 45 ˚C with different starting concentrations of template ($\bar{c}_0$). Data includes concentrations of all complexes containing tetramers. (**c**) Templating observed over a broad temperature range. Large circles show data for reactions at $\bar{c}_0 = 120$ nM of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, small circles show the spontaneous formation ($\bar{c}_0 = 0$). The latter increases at T > 45 ˚C. Above 48 ˚C, binding of monomers to the template gets weaker, slowing down the rate of template assisted formation. This is consistent with the melting temperatures of the information domains (see ***Figure 4—figure supplement 1***).
The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Source data for determination of thermal oscillation temperatures.
Figure supplement 1. Determination of thermal oscillation temperatures.
Figure supplement 1—source data 1. Source data for isothermal template assisted product formation.

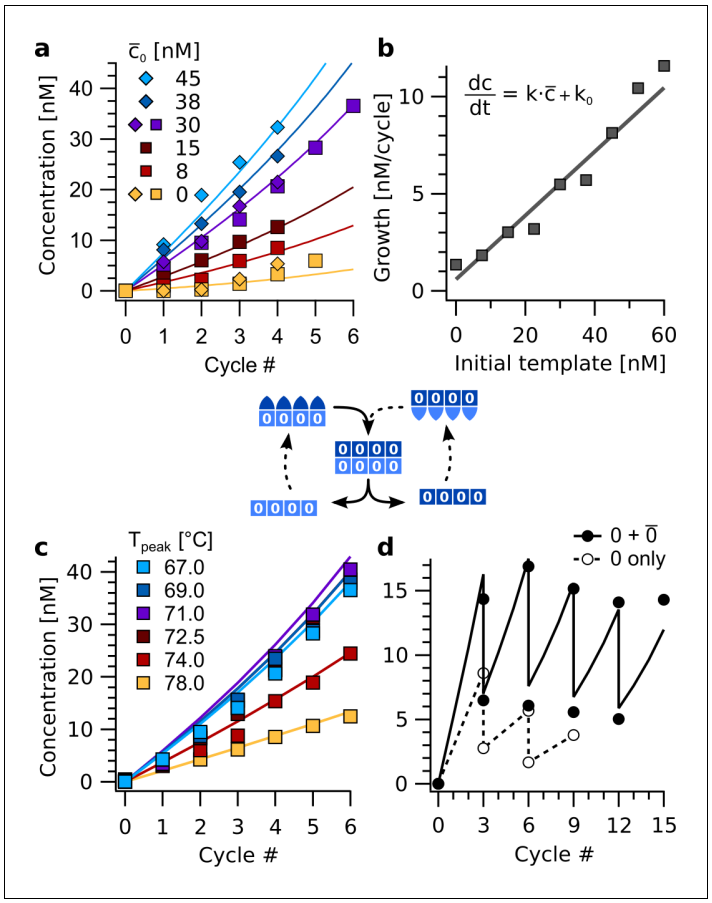

Figure 5. Exponential amplification of a restricted sequence subset with thermal oscillations. (**a**) Amplification time traces for concentration c for sequence 0000 during the first four to six cycles (T_{peak} = 67 ˚C) for template ($\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$) concentrations $\bar{c}_0$ from 0 to 45 nM. The data was fitted using the cross-catalytic model from equation (1). Strands 0_A, $\bar{0}_A$, 0_B, …, $\bar{0}_D$ were used at 200 nM concentration each. Data points show concentrations of complexes 4:4. (**b**) Initial reaction velocity as a function of initial template concentration $\bar{c}_0$. The data points show good agreement with the line calculated from the fits in panel **a**. (**c**) Amplification proceeded for peak temperatures below 74 ˚C. Above, backbone duplexes start to melt, and the complexes are no longer stable. The base temperature was 45 ˚C, reactions initially contained 30 nM of complex $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ as template. (**d**) Serial transfer experiment. The reaction containing strands 0_A, $\bar{0}_A$, 0_B, …, $\bar{0}_D$ (black circles) survived successive dilution by a factor of 1/2 every three cycles at almost constant concentration. In contrast, a reaction with the same amount of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, but lacking monomers $\bar{0}_{A-D}$, fades out (open circles). The solid line shows the model from ***Equation 1***.

The online version of this article includes the following source data for figure 5:

Source data 1. Source data for exponential amplification of a restricted sequence subset with thermal oscillations.

to be 0.43. The difference from the theoretical value of 0.50 was likely due to strands sticking to the reaction vessels before dilution. As a control, a reaction with the same initial concentration of template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, but without monomers $\bar{0}_A$, $\bar{0}_B$, $\bar{0}_C$, $\bar{0}_D$, was subjected to the same protocol. As the control could not grow exponentially, it gradually died out (***Figure 5d***, open circles).

Sequence replication

The above-mentioned reactions did amplify, but not replicate actual sequence information, as they only contained strands with 0/$\bar{0}$ information domains. To study the replication of arbitrary sequences of binary code, replication reactions with all 16 strands encoding for '0' and '1' were performed. To discriminate sequences encoded in equally sized complexes and deduce error rates, we compared these results to those from different reaction runs with defects, that is lacking one or two of the

hairpin sequences required for the faithful replication of a particular template. Reference reactions contained all 16 strands (0_A, $\bar{0}_A$, 1_A, $\bar{1}_A$, 0_B, …, $\bar{1}_D$) at 100 nM each, and were run for each of three different template sequences ($\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$, $\bar{0}_A\bar{1}_B\bar{0}_C\bar{1}_D$, and $\bar{0}_A\bar{0}_B\bar{1}_C\bar{1}_D$) (***Figure 6***). The product yields were quantified from reaction time traces, extracted by integrating the intensities of all gel bands containing tetramers with the labeled strand 0_A.

Leaving out a single strand (reaction label "+++−", for example omitting 0_D for template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$) reduced the yield of full-size product to about 40 % (***Figure 6a, b***). The non-zero product yield with a missing strand is most likely due to the incorporation of the corresponding strand with an information domain mismatch (here 1_D). This type of mismatch allows the hairpin backbone to form regardless, and the unfaithful product can propagate since both strands needed for an amplification of '1' at position D (1_D and $\bar{1}_D$) are provided.

In particular during the first few cycles, mostly complex $0_A0_B0_C$:$\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ (3:4) was detected in the gel, instead of the desired tetramer product (***Figure 6—figure supplement 1***). This was expected given the lack of strand 0_D and provides an upper limit on the error rate of the full replication. The fact that the full reaction produced almost no complexes 3:4 or 4:3 indicates that the incomplete product was indeed caused by the lack of a particular strand.

Removal of a further strand either directly next to the previous one ('++−−', missing strands 0_C and 0_D) or not ('+−+−', missing strands 0_B and 0_D) reduced the yield of product tetramers even further. Due to the periodic design those two variants represent all defective sets with two missing strands. Replication of the other two templates $\bar{0}_A\bar{1}_B\bar{0}_C\bar{1}_D$ and $\bar{0}_A\bar{0}_B\bar{1}_C\bar{1}_D$ produced very similar results. Product concentrations after six cycles are given in ***Figure 6c*** for each of the three templates as well as an average over the template sequences (horizontal lines). A single defect reduced the yield of tetramer complexes to about 40 %, two defects to 15–20 %, which is close to $0.4 \times 0.4 = 0.16 \simeq 15 - 20$ %, that is the combined probability of two independent mismatches.

Replication fidelity

The observed rate of erroneous product formation can be attributed to the spontaneous background rate (***Figure 4b,c***, ***Figure 5a,b*** and ***Figure 6b***). The reaction '+−+−' (dark green) amplified similarly to the untemplated reference reaction (solid line), as it did not contain any strands that could bind next to each other to the template and form a backbone duplex (***Figure 6b***). For the templated reactions '+++−' and '++−−', templating worked for partial sequences, producing intermediate yields.

The reduction in yield caused by a single defect (i.e. missing strand) to ~40 % (and to ~16 % for two defects) translates into a replication fidelity per information domain of ~60 %. The exact value for the replication fidelity is 62 % and can be calculated from ***Figure 6b*** by extracting the endpoint concentrations (blue vs. yellow line) and calculating $1 - \frac{14\,\mathrm{nM}}{37\,\mathrm{nM}} = 0.62$.

However, this is a worst-case estimation, and the replication fidelity is likely higher due to binding competition. The mutations caused by a single defect ('+++-') in ***Figure 6b*** were imposed by not providing strand 0_D for a template ending with $\bar{0}_D$ and only leaving the option to incorporate 1_D instead. For the full system ('++++'), however, with the presence of the matching strand, there is a binding competition for position D. Since the matching strand preferentially binds, the unfaithful incorporation of the wrong strand would be reduced. A similar effect of competition was observed in a protein-catalyzed ligation reaction (***Toyabe and Braun, 2019***). There, a comparable binding competition lead to a sevenfold decrease of the inferior ligation reaction in the presence of competition (***Figure 2a, b*** therein). Therefore, we expect the real fidelity to be better than above lower bound estimate.

It is interesting to project and compare this per information domain replication fidelity to a per nucleotide replicator (i.e. polymerization). To do so, we define a threshold in the decrease of melting temperature per information domain as the criterion for when the replication mechanism is still functional. Then, we estimate how many point mutations in the information domain can maximally be tolerated to stay within this range of decrease in melting temperature. From this, we can calculate a hypothetical, corresponding per nucleotide fidelity to the measured information domain fidelity.

We compared the properties of the duplex 0:$\bar{0}$ to duplexes 0:$\bar{0}$*, where $\bar{0}$* differs from $\bar{0}$ by K point mutations. We assumed that within the temperature range of this replication mechanism (***Figure 7b***, gray box) a reduction in information domain melting temperature T_m of the mutated

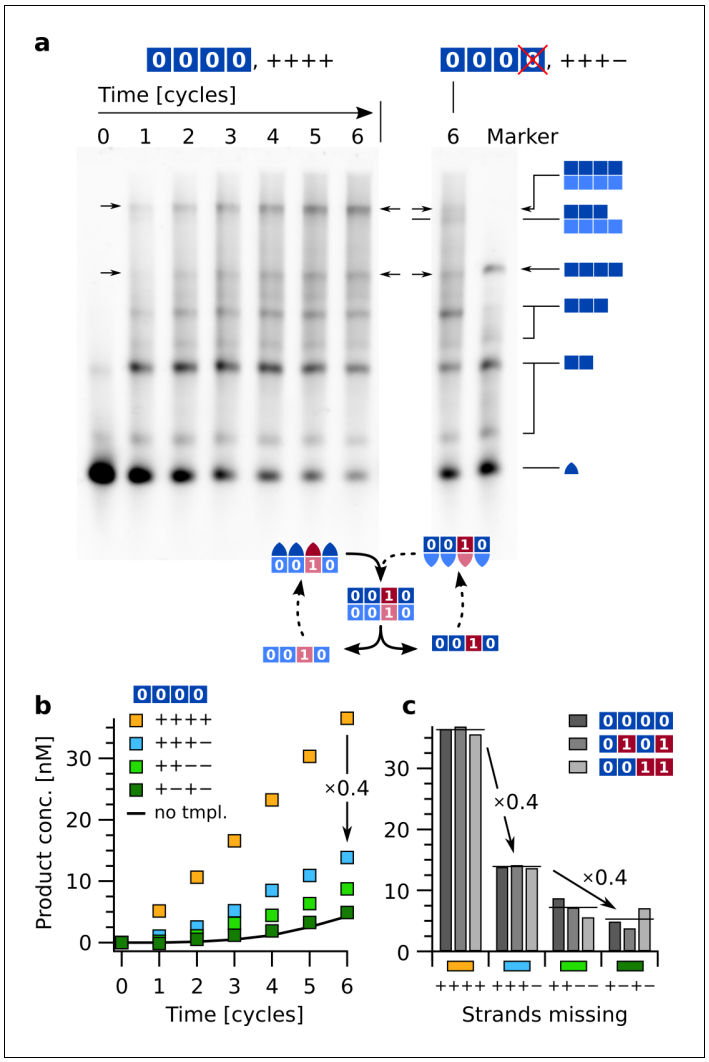

Figure 6. Sequence replication with thermal oscillations and fidelity check by forcing mutations from '0' to '1' at different locations. (**a**) Replication of sequence $0_A0_B0_C0_D$. Reactions were started with 15 nM initial template $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$. All strands ($0_A$, $\bar{0}_A$, 1_A, …, $\bar{1}_D$) were present at 100 nM each. Native-PAGE results comparing the reaction of all 16 strands ('++++') with the reaction lacking strand 0_D ('+++−'). The defective set '+++−' mostly produced 3:4 complexes instead of 4:4 complexes (see schematics on the right). The overall yield of tetramer-containing complexes was greatly reduced. As size reference, the marker lane contained complexes $0_A0_B0_C0_D$, $0_A0_B0_C$, 0_A0_B, and monomers 0_A. The complete gel is presented in ***Figure 6—figure supplement 1***. (**b**) Product concentration over time for the complete sequence network (yellow) and three defective sets with missing strands. Data was integrated by quantitative image analysis from electrophoresis gels using covalent markers on the 0_A-strand counting all product complexes containing tetramers. Mutations of information in the product from '0' to '1' were induced by defective reactions that lacked strands 0_D ('+++−'), 0_C and 0_D ('++−−'), and 0_B and 0_D ('+−+−'). All reactions were initiated with 15 nM of $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$. The solid line shows data from reaction '++++' without template. (**c**) End point comparison of reactions with templates $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ (panels **a**, **b**), $\bar{0}_A\bar{1}_B\bar{0}_C\bar{1}_D$, and $\bar{0}_A\bar{0}_B\bar{1}_C\bar{1}_D$ after six cycles. Horizontal lines indicate averages of the three template sequences. A single missing strand reduced product yield to about 40 %, two missing strands to 15–20 %.

The online version of this article includes the following source data and figure supplement(s) for figure 6:

Source data 1. Source data for sequence replication with thermal oscillations and fidelity check by forcing mutations from '0' to '1' at different locations.

Figure supplement 1. Extended electrophoresis gel image data.

Figure supplement 1—source data 1. Source data for extended electrophoresus gel image data.

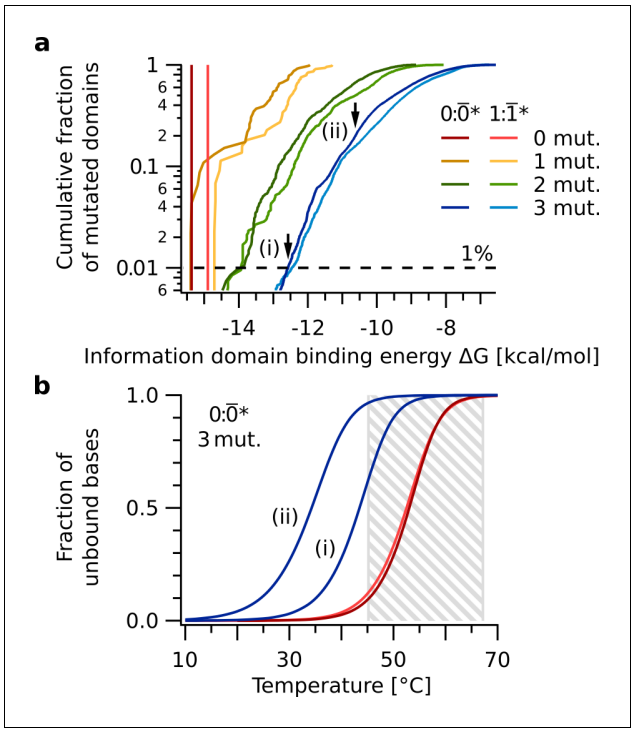

Figure 7. Sequence space analysis of information domain binding. The binding energies quantify the ability of the replication mechanism to discriminate nucleotide mutations. (**a**) Cumulative free energy distributions of information domain duplexes $0{:}\bar{0}$ (red), $1{:}\bar{1}$ (light red), as well as all $0{:}\bar{0}^*$ and $1{:}\bar{1}^*$ with up to three point mutations in $\bar{0}^*$ and $\bar{1}^*$ (yellow, green, blue). 99 % of duplexes $0{:}\bar{0}^*$ with three point mutations have free energies $\Delta G \geq$ -12.5 kcal/mol (dashed line), significantly weaker than that of $0{:}\bar{0}$ (ΔG = -15.4 kcal/mol). (**b**) Melting curves of information domain duplexes $0{:}\bar{0}$ (red), $1{:}\bar{1}$ (light red), and the two duplexes $0{:}\bar{0}^*$ indicated by arrows in panel a. Even the $0{:}\bar{0}^*$ duplex (i) at the low end of the ΔG distribution has a melting temperature of about 10 ˚C below that of $0{:}\bar{0}$. This difference in melting temperature destabilizes binding of the information domain and causes the replication mechanism to reject these sequences in the thermal oscillation regime between T_{base} = 45 ˚C and T_{peak} = 67 ˚C (gray box).

The online version of this article includes the following source data and figure supplement(s) for figure 7:

Source data 1. Source data for information domain binding energy statistics split into information domains containing terminal mutations and those with internal mutations only.

Figure supplement 1. Information domain binding energy statistics split into information domains containing terminal mutations and those with internal mutations only.

Figure supplement 1—source data 1. Source data for sequence space analysis of information domain binding.

duplex $0{:}\bar{0}^*$ by up to 10 ˚C compared to the original duplex $0{:}\bar{0}$ would be tolerated by the replication reaction. This was inferred from the width of the melting transition of duplex $0{:}\bar{0}$ (***Figure 7b***), where a shift of 10 ˚C corresponds to an increase of the unbound fraction from 0.08 at T_{base} = 45 ˚C to 0.66 at 55 ˚C. In terms of free energies of the information domain duplex, this difference corresponds to $\Delta G(0{:}\bar{0}^*) \geq -12.5$ kcal/mol compared to $\Delta G(0{:}\bar{0}) = -15.4$ kcal/mol. 99 % of all duplexes $0{:}\bar{0}^*$, with $\bar{0}^*$ containing three point mutations, met that criterion (***Figure 7a***). Therefore, up to $K = 3$ point mutations can be allowed.

We will assume that the replication did not differentiate between information domain $\bar{0}$ and any information domain $\bar{0}^*$ if $\bar{0}$ and $\bar{0}^*$ differ by less than K point mutations. The fidelity per information domain $p_K(N)$ is given by a cumulative binomial distribution:

$$p_K(N) = \sum_{k=0}^{K-1} \binom{N}{k} p^{N-k} (1-p)^k \tag{2}$$

Here, N is the information domain length, and p the per nucleotide replication fidelity. The reduction in binding energy of the information domain duplex $0{:}\bar{0}$* and subsequent change in melting temperature was used as criterion to define the functionality of the replicator and to translate between a per information domain and a per nucleotide approach. As justified above, we calculate with $K = 3$ mutations within the $N = 15$ bases of the information domain, that is the replication can tolerate up to three mismatches in the information domain. From ***Figure 6*** we extracted a per information domain fidelity of $p_3(15) = 0.62$, and deduce a per nucleotide fidelity of $p = 85$ %. In fact, information domain duplexes $0{:}\bar{0}$* with mutations at two internal bases all show similar properties as information domains with a total of three mutations (***Figure 7—figure supplement 1***). This refinement ($p_2(13) = 0.62$) would increase the per nucleotide fidelity to $p = 90$ %. We therefore estimate that a per nucleotide replication process would need a replication fidelity of 85–90 % to produce sequences with an error rate equivalent to the presented mechanism. Detailed calculations of the per nucleotide fidelities can be found in the supplementary information.

Discussion

A cross-catalytic replicator can be made from short sequences and without covalent bonds under a simple non-equilibrium setting of periodic thermal oscillations. The replication is fast and proceeds within a few thermal oscillations of 20 min each. This velocity is comparable to other replicators (***Kindermann et al., 2005***), cross-ligating ribozymes (***Robertson and Joyce, 2014***), or autocatalytic DNA networks (***Yin et al., 2008***). The required thermal oscillations can be obtained by laminar convection in thermal gradients (***Braun et al., 2003***; ***Salditt et al., 2020***), which also accumulates oligonucleotides (***Mast et al., 2013***). Depending on the envisioned environment, the mechanism could also be driven by thermochemical oscillations (***Ball and Brindley, 2014***) or convection in pH gradients (***Keil et al., 2017***). It should however be noted, that with the current state-of-the-art prebiotic chemistry regarding polymerization and ligation, the creation of >80 nt RNA is not yet understood.

It is likely that a slower prebiotic ligation chemistry could later fix the replication results over long timescales. Such an additional non-enzymatic ligation (***Stadlbauer et al., 2015***) that joins successive strands would relax the constraint that backbone duplexes must not melt during high-temperature steps. Early on, this is difficult to achieve in aqueous solution against the high concentration of water. In order to overcome this competition and to favor the reaction entropically by a leaving group, individual bases are typically activated by triphosphates (***Attwater et al., 2013***; ***Horning and Joyce, 2016***) or imidazoles, which are especially interesting in this context since they can replicate RNA directly (***O'Flaherty et al., 2019***; ***Zhou et al., 2019***). However, the required chemical conditions of enhanced Mg^{2+} concentration hinder strand separation.

The overall replication fidelity is limited by the spontaneous bond formation rate between pairs of hairpin sequences, caused by the interaction of strands in free solution. At lower concentrations, as one would imagine in a prebiotic setting, this rate would decrease at the expense of an overall slower reaction. To some degree and despite ongoing design efforts, such a background rate is inherent to hairpin-fuelled DNA or RNA reactions (***Green et al., 2006***; ***Krammer et al., 2012***; ***Yin et al., 2008***).

The replication mechanism is expected to also work with shorter strands, as long as the order of the melting temperatures of the information domain and the backbone duplexes is preserved. Smaller strands would also be easier to produce by an upstream polymerization process, simply because they contain less nucleotides. In addition, binding of shorter information domain duplexes could discriminate even single base mismatches, resulting in an increased selectivity. It is not straightforward to estimate a minimal sequence length for the demonstrated mechanism. However, it is worth noting that it has been suggested that tRNA arose from two proto-tRNA sequences (***Hopfield, 1978***).

Pre-selection of nucleic acids for the presented hairpin-driven replication mechanism can be provided by highly sequence-specific gelation of DNA. This gel formation has been shown to be most efficient with double hairpin structures very similar to the tRNA-like sequences used in this study (***Morasch et al., 2016***). For our replication system, we have demonstrated this in ***Figure 3*** by showing the spontaneous formation of agglomerates and sedimentation under gravity if all molecules of the assembly are present. This self-selection shows a possible pathway how the system can emerge from random or semi-random sequences, for example in a flow or a convection system where the

molecules are selected as macroscopic agglomerate (**Mast et al., 2013**). Another selection pressure could stem from the biased hydrolysis of double-stranded nucleotide backbones, which favors assembled complexes over the initial hairpins (**Obermayer et al., 2011**).

The replication mechanism could serve as a mutable assembly strategy for larger functional RNAs (**Mutschler et al., 2015**; **Vaidya et al., 2012**). As an evolutionary route toward a more mRNA-like replication product with chemically ligated information domains, the mechanism would be supplemented by self-cleavage next to the information domains that cuts out the non-coding backbone duplexes, followed by ligation of the information domains. Both operations could potentially be performed by very small ribozymatic centers (**Dange et al., 1990**; **Szostak, 2012**; **Vlassov et al., 2005**).

The proposed replication mechanism of assemblies from tRNA-like sequences allows to speculate about a transition from an autonomous replication of successions of information domains to the translation of codon sequences encoded in modern mRNA (**Figure 1a**). Short peptide-RNA hybrids (**Griesser et al., 2017**; **Jauker et al., 2015**), combined with specific interactions between 3'-terminal amino acids and the anticodons, could have given rise to a primitive genetic code. The spatial arrangement of tRNA-like sequences that are replicated by the presented mechanism would translate into a spatial arrangement of the amino acid or short peptide tails that are attached to the strands in a codon-encoded manner (**Schimmel and Henderson, 1994**). The next stage would then be the detachment and linking of the tails to form longer peptides. Eventually, tRNA would transition to its modern role in protein translation. The mechanism thus proposes a hypothesis for the emergence of predecessors of tRNA, independent of protein translation. This is crucial for models of the evolution of translation, because it could justify the existence of tRNA before it was utilized in an early translation process. However, many questions around the evolutionary steps that created translation are still unclear.

Therefore, replication and translation could have, at an early stage, emerged along a common evolutionary trajectory. This supports the notion that predecessors of tRNA could have featured a rudimentary replication mechanism: starting with a double hairpin structure of tRNA-like sequences, the replication of a succession of informational domains would emerge. The interesting aspect is, that the replication is first encoded by hybridization and can later be fixed by a much slower ligation of the hairpins. The demonstrated mechanism could therefore jumpstart a non-enzymatic replication chemistry, which was most likely restricted in fidelity due to working on a nucleotide-by-nucleotide basis (**Robertson and Joyce, 2012**; **Szathmáry, 2006**).

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	0_A	Biomers		P - GCAGCGTTAATTCCCGC GCCTATCGGGAATGTAA CGCAGTGGGTAATAATG ACGATAGCCGTTCGGGA AAAGCGAACGGTATCG
Sequence-based reagent	0_B	Biomers		P - GCAGCGATACCGTTCG CTTTTCCCGAACGGCT ATCGCAGTGGGTAATA ATGAGCGAACTGTCGG TGCTTGCGACAGTGTCGC
Sequence-based reagent	0_C	Biomers		P - GCAGGCGACACTGTCG CAAGCACCGACAGTTC GCCAGTGGGTAATAAT GAGCGGTTCCTTGCGG AGTAGGCAAGGAATCCGC
Sequence-based reagent	0_D	Biomers		P - GCAGGCGGATTCCTTG CCTACTCCGCAAGGAA TCGCCAGTGGGTAATA ATGACGTTACATTCCC GATAGGCGCGGGAATTAACG

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	$\bar{0}_A$	Biomers		P - GCTGCGCATTAACGCG CTTGTCCCGCGTTAAT TGCGCTCATTATTACC CACTCGCTCTCGGCTG TTTTGCCCAGCCGAGCAGCG
Sequence-based reagent	$\bar{0}_B$	Biomers		P – GCTGCGTTGCATTGGC GATCAAAGCCAATGCG AACGCTCATTATTACC CACTCGCAATTAACGC GGGACAAGCGCGTTAATGCG
Sequence-based reagent	$\bar{0}_C$	Biomers		P - GCTGGTTGGAGAAGGC GAACAGCACGCCTTCC CAACCTCATTATTACCC ACTCGTTCGCATTGGC TTTGATC GCCAATGCAACG
Sequence-based reagent	$\bar{0}_D$	Biomers		P - GCTGCGCTGCTCGGCT GGGCAAAACAGCCGAG AGCGCTCATTATTACCC ACTGTTGGGAAGGCGT GCTGTTCGCCTTCTCCAAC
Sequence-based reagent	1_A	Biomers		P - GCAGCGTTAATTCCCG CGCCTATCGGGAATGT AACGCAAAAGAAGAGA AAGACGATAGCCGTTC GGGAAAAGCGAACGGTATCG
Sequence-based reagent	1_B	Biomers		P - GCAGCGATACCGTTCG CTTTTCCCGAACGGCT ATCGCAAAAGAAGAGA AAGAGCGAACTGTCGG TGCTTGCGACAGTGTCGC
Sequence-based reagent	1_C	Biomers		P - GCAGGCGACACTGTCG CAAGCACCGACAGTTC GCCAAAAGAAGAGAAA GAGCGGTTCCTTGCGG AGTAGGCAAGGAATCCGC
Sequence-based reagent	1_D	Biomers		P - GCAGGCGGATTCCTTG CCTACTCCGCAAGGAA TCGCCAAAAGAAGAGA AAGACGTTACATTCCC GATAGGCGCGGGAATTAACG
Sequence-based reagent	$\bar{1}_A$	Biomers		P - GCTGCGCATTAACGCG CTTGTCCCGCGTTAAT TGCGCTCTTTCTCTTC TTTTCGCTCTCGGCTG TTTTGCCCAGCCGAGCAGCG
Sequence-based reagent	$\bar{1}_B$	Biomers		P - GCTGCGTTGCATTGGC GATCAAAGCCAATGCG AACGCTCTTTCTCTTC TTTTCGCAATTAACGC GGGACAAGCGCGTTAATGCG
Sequence-based reagent	$\bar{1}_C$	Biomers		P - GCTGGTTGGAGAAGGC GAACAGCACGCCTTCC CAACCTCTTTCTCTTC TTTTCGTTCGCATTGG CTTTGATCGCCAATGCAACG
Sequence-based reagent	$\bar{1}_D$	Biomers		P- GCTGCGCTGCTCGGCT GGGCAAAACAGCCGAG AGCGCTCTTTCTCTTC TTTTGTTGGGAAGGCG TGCTGTTCGCCTTCTCCAAC

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	0_A – Cy5	Biomers		Cy5 -GCAGCGTTAATTCCCGC GCCTATCGGGAATGTAA CGCAGTGGGTAATAATG ACGATAGCCGTTCGGGA AAAGCGAACGGTATCG
Sequence-based reagent	1_A – Cy5	Biomers		Cy5 - GCAGCGTTAATTCCCG CGCCTATCGGGAATGT AACGCAAAAGAAGAGA AAGACGATAGCCGTTC GGGAAAAGCGAACGGTATCG
Sequence-based reagent	R (random)	Biomers		NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNN
Sequence-based reagent	R (random) – Cy5	Biomers		Cy5 - NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNN
Software, algorithm	NUPACK	nupack.org	https://doi.org/10.1002/jcc.21596	
Software, algorithm	ImageJ	ImageJ http://imagej.nih.gov/ij/	RRID:SCR_002285	
Software, algorithm	ImageJ stabilization plugin	http://www.cs.cmu.edu/~kangli/code/Image_Stabilizer.html		

Strand design

DNA double-hairpin sequences were designed using the NUPACK software package (***Zadeh et al., 2011***). In addition to the secondary structures of the double-hairpins, the design algorithm was constrained by all target dimers. Candidate sequences were selected for optimal homogeneity of binding energies and melting temperatures. Backbone domains connecting consecutive strands (e.g. $0_A0_B0_C$) had to be the most stable bonds in the system, in particular more stable than between a template and a newly formed product complex (e.g. 0_B:$\bar{0}_B$). On the other hand, hairpin melting temperatures had to be low enough to allow for a sufficient degree of thermal fluctuations. To reconcile this with the length of the strands, mismatches were introduced in the hairpin stems. The sequences of all strands are listed in ***Supplementary file 1***.

Thermal cycling assays

All reactions were performed in salt 20 mM Tris-HCl pH 8, 150 mM NaCl with added 20 mM $MgCl_2$. DNA oligonucleotides (Biomers, Germany) were used at 200 nM concentration per strand in reactions containing a fixed-sequence subset of eight strands (e.g. 0/$\bar{0}$ only) and 100 nM per strand in reactions containing all 16 different strands.

Thermal cycling was done in a standard PCR cycler (Bio-Rad C1000). Reaction kinetics were obtained by running each reaction for different run times or numbers of cycles in parallel. The products were analyzed using native PAGE. The time between thermal cycling and PAGE analysis was minimized to exclude artifacts from storage on ice.

Template sequences were prepared using a two-step protocol. Annealing from 95˚C to 70˚C within 1 hr, followed by incubation at 70 ˚C for 30 min. Afterwards, samples were cooled to 2 ˚C and stored on ice. When assembling complexes containing paired information domains (***Figure 2***), samples were slowly cooled down from 70 to 25 ˚C within 90 min before being transferred onto ice. DNA double hairpins were quenched into monomolecular state by heating to 95 ˚C and subsequent fast transfer into ice water.

Product analysis

DNA complexes were analyzed using native polyacrylamide gel electrophoresis (PAGE) in gels at 5 % acrylamide concentration and 29:1 acrylamide / bisacrylamide ratio (Bio-Rad, Germany). Gels were run at electric fields of 14 V/cm at room temperature. Strand $0_A/1_A$ was covalently labeled with Cy5. Cy5 fluorescence intensities were later used to compute strand concentrations. As an additional color channel, strands were stained using SYBR Green I dye (New England Biolabs). Complexes were identified by comparing the products obtained from annealing different strand subsets.

To correctly identify bands in the time-resolved measurements, gels were run with a marker lane. The marker contained strands 0_A (200 nM), 0_B (150 nM), 0_C (50 nM), and 0_D (100 nM), and was prepared using the two-step annealing protocol from 95 to 70 ˚C. The unequal strand concentrations ensured that the sample contained a mixture of mono-, di-, tri-, and tetramers.

Electrophoresis gels were imaged in a multi-channel imager (Bio-Rad ChemiDoc MP), image post processing, and data analysis were performed using a self-developed LabVIEW software. Post-processing corrected for inhomogeneous illumination by the LEDs, image rotation, and distortions of the gel lanes if applicable. Background fluorescence was determined from empty lanes on the gel, albeit generally low in the Cy5 channel.

For the determination of reaction yields, the intensities of all gel bands containing strands of the sequence length of interest were added up. For strings of four strands, these were the single tetramer as well as its complex with di- and tri- and tetramers. Single strands separated from their complements during electrophoresis (***Figure 2*** and ***Figure 6—figure supplement 1***).

Thermal melting curves

Thermal melting curves were measured using either UV absorbance at 260 nm wavelength in a UV/Vis spectrometer (JASCO V-650, 1 cm optical path length), via quenching of the Cy5 label at the 5'-end of strand 0_A (excitation: 620–650 nm, detection: 675–690 nm), or using fluorescence of the intercalating dye SYBR Green I (excitation: 450–490 nm, detection: 510–530 nm). Fluorescence measurements were performed in a PCR cycler (Bio-Rad C1000). Samples measured via fluorescence were at 200 nM of each strand, those measured via UV absorption contained 1 µM total DNA concentration to improve the signal-to-noise ratio. Before analysis of the melting curves (***Mergny and Lacroix, 2003***), data were corrected for baseline signals from reference samples containing buffer and intercalating dye, if applicable.

Self-assembly and sedimentation analysis

The samples were mixed in the replication buffer (150 mM NaCl, 20 mM $MgCl_2$, 20 mM Tris-HCl pH 8) at a total oligomer concentration of 5 µM, that is varying concentration per strand depending on the number of different strands in the configuration (4, 7, or 8). The microfluidic chamber was assembled with a custom cut, 500 µm thick, Teflon foil placed between two plane sapphires (***Figure 3—figure supplement 2***). Three Peltier elements (QuickCool QC-31–1.4-3.7AS, purchased from Conrad Electronics, Germany) were attached to the backside of the chamber to provide full temperature control. The chamber was initially flushed with 3M Novec7500 (3M, Germany) to avoid bubble formation. The samples were pipetted into the microfluidic chamber through the 0.5 mm channels using microloader pipette tips (Eppendorf, Germany). The chamber was then sealed with Parafilm and heated to 95 ˚C for 10 s to fully separate the strands and cooled rapidly (within 30 s) to 25 ˚C. Assembly and sedimentation were monitored for 20 hr on a fluorescence microscope (Axiotech Vario, Zeiss, Germany) with two LEDs (490 nm and 625 nm, Thorlabs, Germany) using a 2.5 x objective (Fluar, Zeiss, Germany). The observed sedimentation was independent of the attached dye and its position (***Figure 3—figure supplement 1c***). Prior to image analysis the image stacks were stabilized using an ImageJ plugin (***Li, 2008***). The ratio of sedimented fluorescence relative to the first frame after heating was used to quantify sedimentation (***Figure 3***). The sedimentation time-traces (***Figure 3b***) were fitted with a Sigmoid function to determine the final concentration increase c/c_0 (***Figure 3c***). The experiment was also performed with random 84 nt DNA strands at 5 µM total concentration to exclude unspecific agglomeration (***Figure 3—figure supplement 1c***).

Acknowledgements

We gratefully acknowledge financial support the Deutsche Forschungsgemeinschaft (DFG) through the TRR 235 Emergence of Life (Project-ID 364653263) and the CRC 1032 NanoAgents (Project-ID 201269156). We thank for funding from the Graduate School 'Quantitative Bioscience Munich' (QBM). We appreciate the fruitful discussions in the Simons Collaboration on the Origins of Life, thank for the measurements by Thomas Rind and acknowledge discussions with Tim Liedl, Christof Mast and Lorenz Keil. We thank Filiz Civril, Adriana Serrão and Thomas Matreux for comments on the manuscript.

Additional information

Funding

Funder	Grant reference number	Author
Deutsche Forschungsgemeinschaft	TRR 235	Alexandra Kühnlein Dieter Braun
Deutsche Forschungsgemeinschaft	Project-ID 364653263	Alexandra Kühnlein Dieter Braun
Deutsche Forschungsgemeinschaft	CRC 1032 (A04) Project-ID 201269156	Simon Alexander Lanzmich Dieter Braun
Deutsche Forschungsgemeinschaft	Student fellowship	Alexandra Kühnlein
Deutsche Forschungsgemeinschaft	Graduate school "Quantitative Bioscience Munich"	Alexandra Kühnlein

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Alexandra Kühnlein, Conceptualization, Data curation, Formal analysis, Visualization, Methodology, Writing - original draft, Writing - review and editing; Simon A Lanzmich, Conceptualization, Data curation, Software, Formal analysis, Visualization, Methodology, Writing - original draft; Dieter Braun, Conceptualization, Software, Funding acquisition, Methodology, Writing - original draft, Writing - review and editing

Author ORCIDs

Alexandra Kühnlein https://orcid.org/0000-0001-9582-6304
Dieter Braun https://orcid.org/0000-0001-7751-1448

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.63431.sa1
Author response https://doi.org/10.7554/eLife.63431.sa2

Additional files

Supplementary files

- Source data 1. Nupack script used for the sequence design.
- Supplementary file 1. Sequences of all DNA strands used. Strand 0_A is 5'-labeled with Cy5, all other strands have a 5'-terminal phosphate. Solid underlines highlight hairpin loops, information domains are indicated by dashed underlines.
- Transparent reporting form

Data availability

No data sets (e.g. sequencing data, clinical trial data etc.) were produced in this study. The source data files (Igor incl. macros) and data analysis (LabVIEW) tools used are provided as supporting fFiles (zip).

References

Adamski P, Eleveld M, Sood A, Kun Ádám, Szilágyi A, Czárán T, Szathmáry E, Otto S. 2020. From self-replication to replicator systems en route to de novo life. *Nature Reviews Chemistry* **4**:386–403. DOI: https://doi.org/10.1038/s41570-020-0196-x

Altay Y, Tezcan M, Otto S. 2017. Emergence of a new Self-Replicator from a dynamic combinatorial library requires a specific Pre-Existing replicator. *Journal of the American Chemical Society* **139**:13612–13615. DOI: https://doi.org/10.1021/jacs.7b07346, PMID: 28910535

Ashkenasy G, Hermans TM, Otto S, Taylor AF. 2017. Systems chemistry. *Chemical Society Reviews* **46**:2543–2554. DOI: https://doi.org/10.1039/C7CS00117G, PMID: 28418049

Attwater J, Wochner A, Holliger P. 2013. In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry* **5**:1011–1018. DOI: https://doi.org/10.1038/nchem.1781, PMID: 24256864

Ball R, Brindley J. 2014. Hydrogen peroxide thermochemical oscillator as driver for primordial RNA replication. *Journal of the Royal Society Interface* **11**:20131052. DOI: https://doi.org/10.1098/rsif.2013.1052

Barany F. 1991. Genetic disease detection and DNA amplification using cloned thermostable ligase. *PNAS* **88**:189–193. DOI: https://doi.org/10.1073/pnas.88.1.189, PMID: 1986365

Bottero I, Huck J, Kosikova T, Philp D. 2016. A synthetic replicator drives a propagating Reaction-Diffusion front. *Journal of the American Chemical Society* **138**:6723–6726. DOI: https://doi.org/10.1021/jacs.6b03372, PMID: 27177046

Bourbo V, Matmor M, Shtelman E, Rubinov B, Ashkenasy N, Ashkenasy G. 2011. Self-assembly and Self-replication of short amphiphilic β-sheet peptides. *Origins of Life and Evolution of Biospheres* **41**:563–567. DOI: https://doi.org/10.1007/s11084-011-9257-y

Braun D, Goddard NL, Libchaber A. 2003. Exponential DNA replication by laminar convection. *Physical Review Letters* **91**:158103. DOI: https://doi.org/10.1103/PhysRevLett.91.158103, PMID: 14611502

Carnall JM, Waudby CA, Belenguer AM, Stuart MC, Peyralans JJ, Otto S. 2010. Mechanosensitive self-replication driven by self-organization. *Science* **327**:1502–1506. DOI: https://doi.org/10.1126/science.1182767, PMID: 20299594

Dange V, Van Atta RB, Hecht SM. 1990. A Mn2(+)-dependent ribozyme. *Science* **248**:585–588. DOI: https://doi.org/10.1126/science.2185542, PMID: 2185542

Dirks RM, Pierce NA. 2004. Triggered amplification by hybridization chain reaction. *PNAS* **101**:15275–15278. DOI: https://doi.org/10.1073/pnas.0407024101, PMID: 15492210

Dolinnaya NG, Sokolova NI, Gryaznova OI, Shabarova ZA. 1988. Site-directed modification of DNA duplexes by chemical ligation. *Nucleic Acids Research* **16**:3721–3738. DOI: https://doi.org/10.1093/nar/16.9.3721, PMID: 3375071

Doudna JA, Couture S, Szostak JW. 1991. A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**:1605–1608. DOI: https://doi.org/10.1126/science.1707185, PMID: 1707185

Engelhart AE, Cafferty BJ, Okafor CD, Chen MC, Williams LD, Lynn DG, Hud NV. 2012. Nonenzymatic ligation of DNA with a reversible step and a final linkage that can be used in PCR. *ChemBioChem* **13**:1121–1124. DOI: https://doi.org/10.1002/cbic.201200167, PMID: 22556064

Green SJ, Lubrich D, Turberfield AJ. 2006. DNA hairpins: fuel for autonomous DNA devices. *Biophysical Journal* **91**:2966–2975. DOI: https://doi.org/10.1529/biophysj.106.084681, PMID: 16861269

Griesser H, Bechthold M, Tremmel P, Kervio E, Richert C. 2017. Amino Acid-Specific, Ribonucleotide-Promoted peptide formation in the absence of enzymes. *Angewandte Chemie* **129**:1244–1248. DOI: https://doi.org/10.1002/ange.201610651

Hopfield JJ. 1978. Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *PNAS* **75**:4334–4338. DOI: https://doi.org/10.1073/pnas.75.9.4334, PMID: 279919

Horning DP, Joyce GF. 2016. Amplification of RNA by an RNA polymerase ribozyme. *PNAS* **113**:9786–9791. DOI: https://doi.org/10.1073/pnas.1610103113, PMID: 27528667

Jauker M, Griesser H, Richert C. 2015. Spontaneous formation of RNA strands, peptidyl RNA, and cofactors. *Angewandte Chemie International Edition* **54**:14564–14569. DOI: https://doi.org/10.1002/anie.201506593, PMID: 26435376

Keil LMR, Möller FM, Kieß M, Kudella PW, Mast CB. 2017. Proton gradients and pH oscillations emerge from heat flow at the microscale. *Nature Communications* **8**:3. DOI: https://doi.org/10.1038/s41467-017-02065-3

Kindermann M, Stahl I, Reimold M, Pankau WM, von Kiedrowski G. 2005. Systems chemistry: kinetic and computational analysis of a nearly exponential organic replicator. *Angewandte Chemie International Edition* **44**:6750–6755. DOI: https://doi.org/10.1002/anie.200501527, PMID: 16187397

Kosikova T, Philp D. 2017. Exploring the emergence of complexity using synthetic replicators. *Chemical Society Reviews* **46**:7274–7305. DOI: https://doi.org/10.1039/C7CS00123A, PMID: 29099123

Krammer H, Möller FM, Braun D. 2012. Thermal, autonomous replicator made from transfer RNA. *Physical Review Letters* **108**:238104. DOI: https://doi.org/10.1103/PhysRevLett.108.238104, PMID: 23003995
Lee DH, Granja JR, Martinez JA, Severin K, Ghadiri MR. 1996. A self-replicating peptide. *Nature* **382**:525–528. DOI: https://doi.org/10.1038/382525a0, PMID: 8700225
Li K. 2008. The image stabilizer plugin for ImageJ. *ImageJ*. 5.0. http://www.cs.cmu.edu/~kangli/code/Image_Stabilizer.html
Li Y, Breaker RR. 1999. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2′-Hydroxyl Group. *Journal of the American Chemical Society* **121**:5364–5372. DOI: https://doi.org/10.1021/ja990592p
Mariani A, Bonfio C, Johnson CM, Sutherland JD. 2018. pH-Driven RNA strand separation under prebiotically plausible conditions. *Biochemistry* **57**:6382–6386. DOI: https://doi.org/10.1021/acs.biochem.8b01080, PMID: 30383375
Mast CB, Schink S, Gerland U, Braun D. 2013. Escalation of polymerization in a thermal gradient. *PNAS* **110**:8030–8035. DOI: https://doi.org/10.1073/pnas.1303222110, PMID: 23630280
Mergny JL, Lacroix L. 2003. Analysis of thermal melting curves. *Oligonucleotides* **13**:515–537. DOI: https://doi.org/10.1089/154545703322860825, PMID: 15025917
Morasch M, Braun D, Mast CB. 2016. Heat-Flow-Driven oligonucleotide gelation separates Single-Base differences. *Angewandte Chemie International Edition* **55**:6676–6679. DOI: https://doi.org/10.1002/anie.201601886, PMID: 27060490
Mutschler H, Wochner A, Holliger P. 2015. Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry* **7**:502–508. DOI: https://doi.org/10.1038/nchem.2251, PMID: 25991529
Nguyen DT, Saleh OA. 2017. Tuning phase and aging of DNA hydrogels through molecular design. *Soft Matter* **13**:5421–5427. DOI: https://doi.org/10.1039/C7SM00557A, PMID: 28702663
O'Flaherty DK, Zhou L, Szostak JW. 2019. Nonenzymatic Template-Directed synthesis of Mixed-Sequence 3′-NP-DNA up to 25 nucleotides long inside model protocells. *Journal of the American Chemical Society* **141**:10481–10488. DOI: https://doi.org/10.1021/jacs.9b04858, PMID: 31180644
Obermayer B, Krammer H, Braun D, Gerland U. 2011. Emergence of information transmission in a prebiotic RNA reactor. *Physical Review Letters* **107**:018101. DOI: https://doi.org/10.1103/PhysRevLett.107.018101, PMID: 21797574
Orgel LE. 2004. Prebiotic chemistry and the origin of the RNA world. *Critical Reviews in Biochemistry and Molecular Biology* **39**:99–123. DOI: https://doi.org/10.1080/10409230490460765, PMID: 15217990
Patzke V, McCaskill JS, von Kiedrowski G. 2014. DNA with 3′-5′-disulfide links–rapid chemical ligation through isosteric replacement. *Angewandte Chemie International Edition* **53**:4222–4226. DOI: https://doi.org/10.1002/anie.201310644, PMID: 24623660
Paul N, Joyce GF. 2002. A self-replicating ligase ribozyme. *PNAS* **99**:12733–12740. DOI: https://doi.org/10.1073/pnas.202471099, PMID: 12239349
Pino S, Costanzo G, Giorgi A, Di Mauro E. 2011. Sequence complementarity-driven nonenzymatic ligation of RNA. *Biochemistry* **50**:2994–3003. DOI: https://doi.org/10.1021/bi101981z, PMID: 21361363
Robertson MP, Hesselberth JR, Ellington AD. 2001. Optimization and optimality of a short ribozyme ligase that joins non-Watson-Crick base pairings. *RNA* **7**:513–523. DOI: https://doi.org/10.1017/S1355838201002199, PMID: 11345430
Robertson MP, Joyce GF. 2012. The origins of the RNA world. *Cold Spring Harbor Perspectives in Biology* **4**:a003608. DOI: https://doi.org/10.1101/cshperspect.a003608, PMID: 20739415
Robertson MP, Joyce GF. 2014. Highly efficient self-replicating RNA enzymes. *Chemistry & Biology* **21**:238–245. DOI: https://doi.org/10.1016/j.chembiol.2013.12.004, PMID: 24388759
Rodbard D, Chrambach A. 1970. Unified theory for gel electrophoresis and gel filtration. *PNAS* **65**:970–977. DOI: https://doi.org/10.1073/pnas.65.4.970, PMID: 4191703
Rohatgi R, Bartel DP, Szostak JW. 1996. Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3′-5′ phosphodiester bonds. *Journal of the American Chemical Society* **118**:3340–3344. DOI: https://doi.org/10.1021/ja9537134, PMID: 11539268
Rubinov B, Wagner N, Matmor M, Regev O, Ashkenasy N, Ashkenasy G. 2012. Transient fibril structures facilitating nonenzymatic self-replication. *ACS Nano* **6**:7893–7901. DOI: https://doi.org/10.1021/nn302223v, PMID: 22856322
Sadownik JW, Philp D. 2008. A simple synthetic replicator amplifies itself from a dynamic reagent pool. *Angewandte Chemie* **120**:10113–10118. DOI: https://doi.org/10.1002/ange.200804223
Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell Anemia. *Science* **230**:1350–1354. DOI: https://doi.org/10.1126/science.2999980, PMID: 2999980
Salditt A, Keil LMR, Horning DP, Mast CB, Joyce GF, Braun D. 2020. Thermal habitat for RNA amplification and accumulation. *Physical Review Letters* **125**:048104. DOI: https://doi.org/10.1103/PhysRevLett.125.048104, PMID: 32794805
Schimmel P, Henderson B. 1994. Possible role of aminoacyl-RNA complexes in noncoded peptide synthesis and origin of coded synthesis. *PNAS* **91**:11283–11286. DOI: https://doi.org/10.1073/pnas.91.24.11283, PMID: 7972050
Schulman R, Yurke B, Winfree E. 2012. Robust self-replication of combinatorial information via crystal growth and scission. *PNAS* **109**:6405–6410. DOI: https://doi.org/10.1073/pnas.1117813109, PMID: 22493232

Sharp SJ, Schaack J, Cooley L, Burke DJ, Söll D. 1985. Structure and transcription of eukaryotic tRNA genes. *Critical Reviews in Biochemistry* **19**:107–144. DOI: https://doi.org/10.3109/10409238509082541, PMID: 3905254

Sievers D, von Kiedrowski G. 1994. Self-replication of complementary nucleotide-based oligomers. *Nature* **369**: 221–224. DOI: https://doi.org/10.1038/369221a0, PMID: 8183342

Stadlbauer P, Šponer J, Costanzo G, Di Mauro E, Pino S, Šponer JE. 2015. Tetraloop-like geometries could form the basis of the catalytic activity of the most ancient ribooligonucleotides. *Chemistry - a European Journal* **21**: 3596–3604. DOI: https://doi.org/10.1002/chem.201406140

Szathmáry E. 2006. The origin of replicators and reproducers philosophical transactions of the royal society B. *Biological Sciences* **361**:1761–1776. DOI: https://doi.org/10.1098/rstb.2006.1912

Szostak JW. 2012. The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry* **3**:2. DOI: https://doi.org/10.1186/1759-2208-3-2

Tjivikua T, Ballester P, Rebek J. 1990. Self-replicating system. *Journal of the American Chemical Society* **112**: 1249–1250. DOI: https://doi.org/10.1021/ja00159a057

Toyabe S, Braun D. 2019. Cooperative ligation breaks sequence symmetry and stabilizes early molecular replication. *Physical Review X* **9**:011056. DOI: https://doi.org/10.1103/PhysRevX.9.011056

Turk RM, Illangasekare M, Yarus M. 2011. Catalyzed and spontaneous reactions on ribozyme ribose. *Journal of the American Chemical Society* **133**:6044–6050. DOI: https://doi.org/10.1021/ja200275h, PMID: 21438575

Ura Y, Beierle JM, Leman LJ, Orgel LE, Ghadiri MR. 2009. Self-assembling sequence-adaptive peptide nucleic acids. *Science* **325**:73–77. DOI: https://doi.org/10.1126/science.1174577, PMID: 19520909

Vaidya N, Manapat ML, Chen IA, Xulvi-Brunet R, Hayden EJ, Lehman N. 2012. Spontaneous network formation among cooperative RNA replicators. *Nature* **491**:72–77. DOI: https://doi.org/10.1038/nature11549, PMID: 23075853

Vlassov AV, Kazakov SA, Johnston BH, Landweber LF. 2005. The RNA world on ice: a new scenario for the emergence of RNA information. *Journal of Molecular Evolution* **61**:264–273. DOI: https://doi.org/10.1007/s00239-004-0362-7, PMID: 16044244

von Kiedrowski G. 1986. A Self-Replicating hexadeoxynucleotide. *Angewandte Chemie International Edition in English* **25**:932–935. DOI: https://doi.org/10.1002/anie.198609322

von Neumann J. 1951. The general and logical theory of automata. In: Jeffress L. A (Ed). *Cerebral Mechanisms in Behavio*. Wiley. p. 1–41.

Walton T, DasGupta S, Duzdevich D, Oh SS, Szostak JW. 2020. In vitro selection of ribozyme ligases that use prebiotically plausible 2-aminoimidazole-activated substrates. *PNAS* **117**:5741–5748. DOI: https://doi.org/10.1073/pnas.1914367117, PMID: 32123094

Yin P, Choi HM, Calvert CR, Pierce NA. 2008. Programming biomolecular self-assembly pathways. *Nature* **451**: 318–322. DOI: https://doi.org/10.1038/nature06451, PMID: 18202654

Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, Dirks RM, Pierce NA. 2011. NUPACK: analysis and design of nucleic acid systems. *Journal of Computational Chemistry* **32**:170–173. DOI: https://doi.org/10.1002/jcc.21596, PMID: 20645303

Zhang DY, Turberfield AJ, Yurke B, Winfree E. 2007. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science* **318**:1121–1125. DOI: https://doi.org/10.1126/science.1148532, PMID: 18006742

Zhou L, Kim SC, Ho KH, O'Flaherty DK, Giurgiu C, Wright TH, Szostak JW. 2019. Non-enzymatic primer extension with strand displacement. *eLife* **8**:e51888. DOI: https://doi.org/10.7554/eLife.51888, PMID: 31702557

Appendix 1

Calculation of fidelity rate

Through the experiments shown in ***Figure 6*** we already know that the replication fidelity per information domain is 62 %. Now, we want to assume that the presented replication mechanism would translate into a base-by-base replication and look at (i) how tolerant would the replication be to point mutations at the information domain and (ii) given that threshold, how good would a base-by-base replication have to do to perform equally well, that is what per nucleotide fidelity would it need to have.

Question (i) is answered in ***Figure 7***, where we see that on the 15 nt information domain we can allow up to three base mismatches to stay within the bounds of the temperature cycling (gray box, ***Figure 7b***). In order to calculate how the measured replication fidelity per information domain translates into a hypothetical replication fidelity per nucleotide we assume a cumulative binomial distribution:

$$p_K(N) = \sum_{k=0}^{K-1} \binom{N}{k} p^{N-k}(1-p)^k$$

We know that the overall likelihood to get a 'correctly' replicated information domain is 62 %. From ***Figure 7*** we know that in a base-by-base replication, 'correctly' means with up to three mismatches. Therefore, we must find the number of combinatorial possibilities of spatially distributing 0, 1 or 2 mismatches on the 15 nt information domain (using $N = 15$ nucleotides and allowing up to $K = 3$ mismatches). Using this, we can determine the probability p for a success, that is the correct replication of a single nucleotide, to meet the $p_K(N) = 0.62$ overall likelihood.

For $K = 3$ and $N = 15$, we measure the replication fidelity per information domain to be $p_{K(N)} = 0.62$. Therefore, we calculate:

$$\begin{aligned}
&\textstyle\sum_{k=0}^{K-1}\binom{N}{k}p^{N-k}(1-p)^k = \sum_{k=0}^{2}\binom{15}{k}p^{15-k}(1-p)^k && = 0.62\\
&\binom{15}{0}p^{15}(1-p)^0 + \binom{15}{1}p^{14}(1-p)^1 + \binom{15}{2}p^{13}(1-p)^2 && = 0.62\\
&1p^{15} + 15p^{14}(1-p)^1 + 105p^{13}(1-p)^2 && =\\
&= p^{15} + 15p^{14} - 15p^{15} + 105p^{13} - 210p^{14} + 105\,p^{15} && =\\
&= 91p^{15} - 195p^{14} + 105p^{13} && = 0.62\\
&p = 0.853 && = 85\,\%
\end{aligned}$$

From the information domain energy statistics shown in ***Figure 7—figure supplement 1***, one can see that strands with two internal mutations behave nearly identical to strands with a total of three mutations (accepting internal and terminal mutations). Therefore, we simplify the calculation and only consider internal mutations.

Accordingly, we calculate for $K = 2$ and $N = \#all\,bases - \#terminal\,bases = 15 - 2 = 13$ and a per information domain fidelity $p_{K(N)} = 0.62$:

$$\begin{aligned}
&\textstyle\sum_{k=0}^{K-1}\binom{N}{k}p^{N-k}(1-p)^k = \sum_{k=0}^{1}\binom{13}{k}p^{13-k}(1-p)^k && = 0.62\\
&\binom{13}{0}p^{13}(1-p)^0 + \binom{13}{1}p^{12}(1-p)^1 && = 0.62\\
&1p^{13} + 13p^{12}(1-p)^1 && =\\
&= p^{13} + 13p^{12} - 13p^{13} && =\\
&= -12p^{13} + 13p^{12} && = 0.62\\
&p = 0.900 && = 90\,\%
\end{aligned}$$

Therefore, a comparable base-by-base replication would need a per nucleotide fidelity of 85–90 % to perform equally well as the presented replication mechanism.

B. Associated Publications

M. Morasch, J. Liu, C. F. Dirscherl, A. Ianeselli, A. Kühnlein, K. LeVay, Ph. Schwintek, S. Islam, M. K. Corpinot, B. Scheu, D. B. Dingwell, P. Schwille, H. Mutschler, M. W. Powner, C. B. Mast, D. Braun. **Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules.** *Nat. Chem.*, 11: 779-788, 2019.

T. Matreux, K. LeVay, A. Schmid, P. Aikkila, L. Belohlavek, Z. Caliskanoglu, E. Salibi, A. Kühnlein, C. Springsklee, B. Scheu, D. B. Dingwell, D. Braun, H. Mutschler, C. B. Mast. **Heat flows in rock cracks naturally optimize salt compositions for ribozymes.** *Nat. Chem., accepted.*

A. Ianeselli, D. Tetiker, J. A. C. Stein, A. Kühnlein, C. B. Mast, D. Braun, T.-Y. D. Tang. **Non-equilibrium conditions inside rock pores drive fission, maintenance and selection of coacervate protocells.** *Submitted, in resubmission at Nat. Chem*.

nature chemistry

ARTICLES

https://doi.org/10.1038/s41557-019-0299-5

Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules

Matthias Morasch[1], Jonathan Liu[1], Christina F. Dirscherl[1], Alan Ianeselli[1], Alexandra Kühnlein[1], Kristian Le Vay[2], Philipp Schwintek[1], Saidul Islam[3], Mérina K. Corpinot[3], Bettina Scheu[4], Donald B. Dingwell[4], Petra Schwille[2], Hannes Mutschler[2], Matthew W. Powner[3], Christof B. Mast[1] and Dieter Braun[1]*

Non-equilibrium conditions must have been crucial for the assembly of the first informational polymers of early life, by supporting their formation and continuous enrichment in a long-lasting environment. Here, we explore how gas bubbles in water subjected to a thermal gradient, a likely scenario within crustal mafic rocks on the early Earth, drive a complex, continuous enrichment of prebiotic molecules. RNA precursors, monomers, active ribozymes, oligonucleotides and lipids are shown to (1) cycle between dry and wet states, enabling the central step of RNA phosphorylation, (2) accumulate at the gas–water interface to drastically increase ribozymatic activity, (3) condense into hydrogels, (4) form pure crystals and (5) encapsulate into protecting vesicle aggregates that subsequently undergo fission. These effects occur within less than 30 min. The findings unite, in one location, the physical conditions that were crucial for the chemical emergence of biopolymers. They suggest that heated microbubbles could have hosted the first cycles of molecular evolution.

Life is a non-equilibrium system. Through evolution, life has created a complex protein machinery to maintain the non-equilibrium of crowded molecules inside dividing vesicles. Based on entropy arguments, equilibrium conditions are unlikely to have triggered the evolutionary processes at the origin of life[1]. External non-equilibria had to have been provided for the accumulation, encapsulation and replication of the first informational molecules. These can locally reduce entropy, give rise to patterns[2] and lean the system towards a continuous, dynamic self-organization[3]. Non-equilibrium dynamics can be found in many fluid systems, including gravity-driven instabilities in the atmosphere[4], the accumulation of particles in nonlinear flow[5,6] and shear-dependent platelet activation in blood[7]. Our experiments discuss whether gas–water interfaces in a thermal gradient could have provided such a non-equilibrium setting for the emergence of life on early Earth.

Non-equilibrium systems in the form of heat flows were a very common and simplistic setting found ubiquitously on early Earth[8]. Hydrothermal activity is considered to have been abundant and intimately linked to volcanic activity[9]. Water was thus circulating through the pore space of volcanic rocks, which was formed by magmatic vesiculation (primary origin) and fractures (secondary origin). These systems have been studied as non-equilibrium driving forces for biological molecules in a variety of processes[10–17].

Gases originating from degassing of deeper magma bodies percolate through these water-filled pore networks. At shallow depths bubbles are formed by gases dissolved in water and the formation of vapour where sufficient heat is supplied by the hydrothermal system. The bubbles create gas–water interfaces, which previously have been discussed in connection with atmospheric bubble–aerosol–droplet cycles[18], the adsorption of lipid monolayers and DNA to the interface[19,20] and the formation of peptide bonds[21].

In the absence of a temperature gradient, evaporation of a drop of water on a surface exhibits the so-called 'coffee-ring effect'[22]. Upon evaporation, molecules in the drop are accumulated at its rim by capillary flow. After complete evaporation, a ring of concentrated material is deposited. In the inverted setting studied here, a gas bubble is immersed in water (Fig. 1) and a temperature gradient drives this process continuously.

Results

Accumulation at the gas–water interface. Experimentally, bubbles were created by filling a 240-µm-thick, corrugated microfluidic chamber with solution (Supplementary Fig. 1). As the solution could not fill all the cavities, pinned gas bubbles were created. At higher temperatures, bubbles were found to also form spontaneously anywhere in the system as a result of outgassing. These bubbles were often not restricted by their surrounding geometry and moved along the heated surface.

Accumulation of molecules at these heated gas–water interfaces is caused by a continuous evaporation–recondensation water cycle. The interfaces are held in a constant non-equilibrium state, leading to a steady-state coffee-ring effect that does not end in a fully dry state. Here, we observe six physico-chemical processes, which have all individually been suggested to be relevant for the emergence of prebiotic evolution and are co-located in a single non-equilibrium system: (1) enhanced catalytic activity of ribozymes, amplified by the accumulation of oligonucleotides and ions, (2) condensation of

[1]Physics Department, Center for Nanoscience, Ludwig-Maximilians-Universität München, Munich, Germany. [2]Max Planck Institute of Biochemistry, Martinsried, Germany. [3]Department of Chemistry, University College London, London, UK. [4]Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, Munich, Germany. *e-mail: dieter.braun@lmu.de

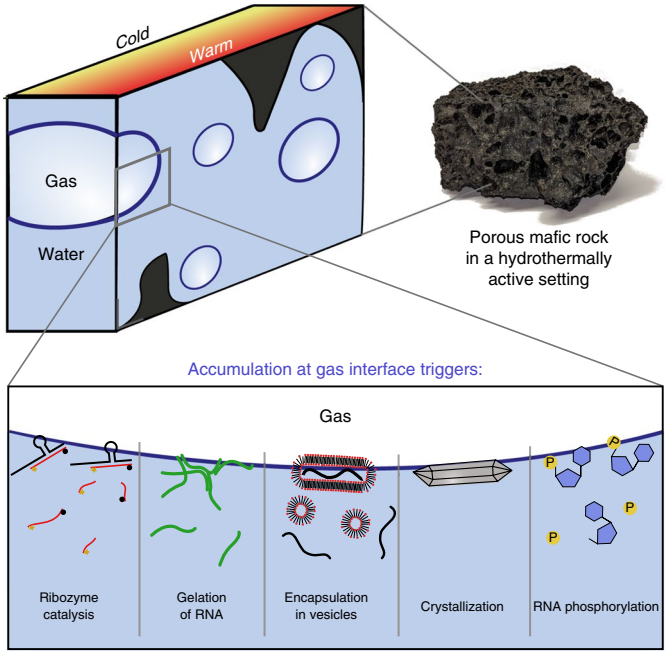

Fig. 1 | DNA accumulation at gas bubbles in a thermal gradient. Volcanic rocks in shallow hydrothermal settings are subjected to water cycles in their pore spaces, which are of both primary (magmatic vesicles) and secondary (fractures) origin. Gases, originating from magma degassing at depth, percolate through the water. Heat supplied by the hydrothermal system causes vaporization. At the gas microbubbles, molecules are accumulated by the continuous capillary flow on the warmer side of the gas–water interface. As shown experimentally, this environment can enhance the catalytic activity of ribozymes, trigger the formation of a hydrogel from self-complementary RNA, encapsulate oligonucleotides such as aptamers in vesicle aggregates, trigger their subsequent fission, drive the crystallization of ribose aminooxazoline (RAO)—a prebiotic RNA precursor—and initiate the phosphorylation of RNA nucleosides.

self-complementary RNA 36mers into millimetre-sized hydrogels, (3) vesicle aggregation at the bubble interface along with encapsulation of oligonucleotides in aqueous phases with up to 18-fold enhanced concentration, (4) fission of the vesicle structures in the adjacent micro-convection, (5) formation of euhedral 300 µm crystals from the RNA precursor ribose aminooxazoline (RAO) around bubbles (which also act as seeds for new bubbles) and (6) dry–wet cycles enhancing, for example, the phosphorylation of nucleosides, created by fluctuating and moving interfaces (Fig. 1 and Supplementary Videos 1–6). All six mechanisms were established within 30 min and, importantly, operated in continuous contact with bulk water.

Here, length- and temperature-dependent accumulations were measured at low-salt conditions (0.1-fold PBS buffer: 13.7 mM NaCl, 0.27 mM KCl, 1 mM phosphate buffer). However, DNA and RNA gelation measurements were performed equally well under physiological conditions (1-fold PBS buffer: 137 mM NaCl, 2.7 mM KCl, 10 mM phosphate buffer).

To observe the dynamics at the interface, bubbles were created as described above, using fluorescently labelled molecules in solution (Fig. 2a(i)). The front and back sides of the chamber were heated and cooled, respectively, to generate a temperature gradient. The system was monitored through the warm side using a fluorescence microscope. The optical axis ran along the temperature gradient (Supplementary Figs. 1 and 2) and the quantitative fluorescence was captured with a charge-coupled device (CCD) camera.

Initially, the chamber was filled with a solution of 200 nM 6-Carboxyfluorescein (FAM)-labelled 132-base single-stranded DNA (ssDNA) oligomer in 0.1-fold PBS buffer. When no temperature gradient was applied to the system ($T_{warm} = T_{cold} = 10\,°C$), we observed no accumulation of DNA near the gas–water interface. The fluorescence signal exhibited a constant small peak at the observed interface, possibly due to a slight adsorption of the DNA to the gas–water interface (Fig. 2a(i), 0 s). Heating one side of the chamber ($T_{warm} = 30\,°C$, $T_{cold} = 10\,°C$) resulted in the rapid accumulation of DNA in a small area on the warm side at the contact line (Fig. 2a(ii), dashed red box, Supplementary Video 1). The chamber-averaged fluorescence at the contact line increased within 6 min by ~12-fold compared to the bulk fluorescence (Fig. 2b). No accumulation was observed on the cold side.

We calculated the local concentration at the contact line from the ratio of the meniscus and bulk fluorescence and the geometry of the curved gas–water interface. Because the fluorescence was averaged over the chamber by the microscope objectives, a 60-fold higher concentration in addition to the higher fluorescence is inferred due to the thinner size of the accumulation region (~4 µm) compared to the 240-µm-wide chamber. Therefore, from the observed 12-fold increase in fluorescence, we estimated a concentration increase by a factor of 700, corresponding to 140 µM DNA concentration in the meniscus when starting from a 200 nM bulk solution (Fig. 2c). Simulations suggested that without thermophoresis this accumulation would be only slightly higher, showing that it does not play a significant role in the accumulation process (Fig. 2c, red dotted line).

Further analysis from experiment and theory showed that the accumulation was caused by the focused evaporation of water at the tip of the meniscus[23] (Fig. 2d, orange). A continuous flow of water into the meniscus dragged the molecules with it; because they could not evaporate, they could only escape by diffusion against the one-way capillary flow. The flow was visualized by filling the chamber with a suspension of 200-nm-diameter FAM-labelled polystyrene beads. By particle tracking, we measured the velocity profile in the meniscus (Fig. 2d,e and Supplementary Video 1). Beads moved towards the accumulation region near the hot side of the chamber at the contact line.

We attribute this to capillary flow, which superseded the comparatively weak bulk buoyant convection (Fig. 2e). Temperature gradients have also been demonstrated to create Marangoni flows[23], in which water is drawn from the warm to the cold side of an interface due to a surface tension gradient. This was difficult to observe because the main temperature gradient was along the viewing axis. However, we observed strong lateral flows at the interface when accumulating vesicles. We attribute these to lateral Marangoni flows and therefore assume also a combination of Marangoni flows and convection along the interface (Fig. 2d, green). Without this flow, our simulation predicts a larger accumulation (Fig. 2c, blue dotted line).

The water that evaporated near the warm chamber wall was found to condense on the cold wall, forming small water droplets. As a result, the gas bubble had less space, expanded, and moved the gas–water interface. Once the condensed 'rain' droplets had grown, merged and re-entered into the bulk solution, driven by surface tension, the interface moved back to its initial position (Supplementary Video 1). These fluctuations of the contact line triggered a drying and recondensation of the molecules at the location where the capillary flow initially accumulated them.

A fluid dynamics model was used to describe the main features of the accumulation. We considered a two-dimensional (2D) section perpendicular to the interface. Because the width of the channel was smaller than the capillary length for the water and gas[24], the interface geometry was approximated by an arc of a circle.

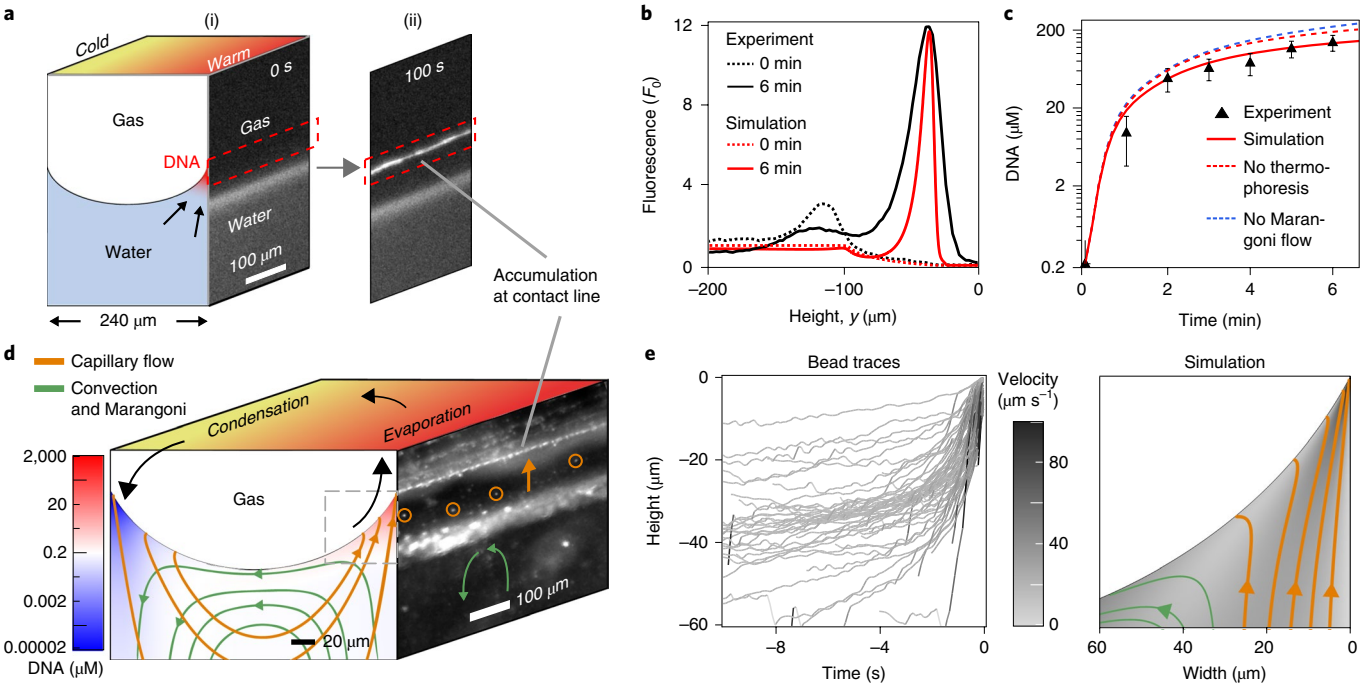

Fig. 2 | DNA accumulation by capillary flow. a, (i) Gas–water interface imaged by fluorescence microscopy with the contact line area outlined with a red dashed line. The initial higher fluorescence intensity at the contact line originated from adsorbed DNA at the gas–water interface. (ii) After 100 s, the 132mer DNA has accumulated in water near the contact line at the warm temperature side of the chamber (T_{warm} = 30 °C, T_{cold} = 10 °C, Supplementary Video 1). **b**, The fluorescence profile reached a 12-fold increase as compared to the bulk fluorescence within 6 min. **c**, Due to the confined meniscus geometry, the accumulated concentration was significantly higher. The 12-fold increase in fluorescence corresponded to a 700-fold increase in DNA concentration, consistent with simulation results (red solid line). Without thermophoresis (red dotted line) or Marangoni flows (blue dotted line), simulations predict a slightly higher accumulation. Error bars were estimated from fluorescence analysis averaged over an ~200 µm interface width at three positions along the contact line from the shown example measurement. **d**, Fluid flow near the contact line measured by single-particle tracking (Supplementary Video 1). Capillary flow (orange) pulled the beads upwards toward the contact line as water mainly evaporated at the tip of the meniscus. Marangoni and convection flows (green) shuttled the beads between the hot and cold sides. Superposed is the logarithmic concentration profile obtained from the simulation for the accumulation in **a**. **e**, Single-particle tracking of the capillary flow. Peak flow velocities reached 50 µm s^{-1} in the last second of the water flow before its evaporation (Supplementary Fig. 3). T_{cold} = 10 °C, T_{warm} = 40 °C.

The contact line of the interface was pinned in the simulation[25,26], motivated by the observation that the accumulation kinetics was generally faster than the fluctuating movement of the interface.

The model superposed four water flows, providing the boundary condition for the accumulation of DNA: (1) capillary flow at the meniscus, (2) diffusion of water vapour between the interface and the gas bubble, (3) convection of water and (4) Marangoni flow along the interface. The relative strength of the Marangoni flow—a free parameter due to the unknown presence of surface-active molecules[27]—was adjusted to fit the velocities measured in the experiment.

The interplay of all four flows led to the accumulation of molecules at the meniscus. As only water evaporated on the warm side, dissolved molecules were continuously dragged towards the contact line, where their concentration depended on back-diffusion and the speed of the capillary flow. Convection and Marangoni flow provided a constant cycling of water and new material towards the accumulation region.

DNA accumulation was measured experimentally for various temperature differences and DNA lengths (Supplementary Fig. 4). The accumulation rose with increasing temperature difference (ΔT), reaching a 4,000-fold increase for ΔT = 40 °C. We also found that smaller (15mer) DNA molecules accumulated three times less effectively than larger (132mer) DNA, which was attributed to their higher diffusion coefficient. The model predicted a multi-fold accumulation of mono- and divalent ions (Supplementary Fig. 4), resulting in a higher salt concentration at the meniscus.

Enhanced ribozyme catalysis at the interface. The accumulation of larger biomolecules as well as ions at the interface makes it a powerful mechanism to enhance the activity of functional nucleic acids. To test this, we monitored the activity of the Hammerhead ribozyme[28,29], which cleaves a 12mer RNA substrate strand. The substrate and magnesium concentration determined its activity[30], and both were accumulated in a 30 °C gradient (T_{warm} = 10 °C, T_{cold} = 40 °C) at low bulk concentrations (0.1 µM Hammerhead, 0.5 µM substrate, 0.4 mM $MgCl_2$, Fig. 3a). As a control, chambers without gas interfaces were studied. A FAM dye and a black hole quencher were attached on opposite sides of the substrate, inhibiting the fluorescence of the FAM dye. On cleavage, the dye was not quenched anymore and could be detected by fluorescence microscopy. Figure 3a,b and Supplementary Video 2 show the average fluorescence of the chamber over time. The substrate strands were cleaved predominantly at the interface, as seen by the rise of fluorescence there. From here, the cleaved strands were frequently ejected into the bulk solution. After 25 min, samples were extracted from the bulk fluid and analysed by polyacrylamide gel electrophoresis (PAGE, Fig. 3c). We detected up to 50% concentration of the cleaved substrate (bottom band, Supplementary Fig. 5) in a chamber with bubbles.

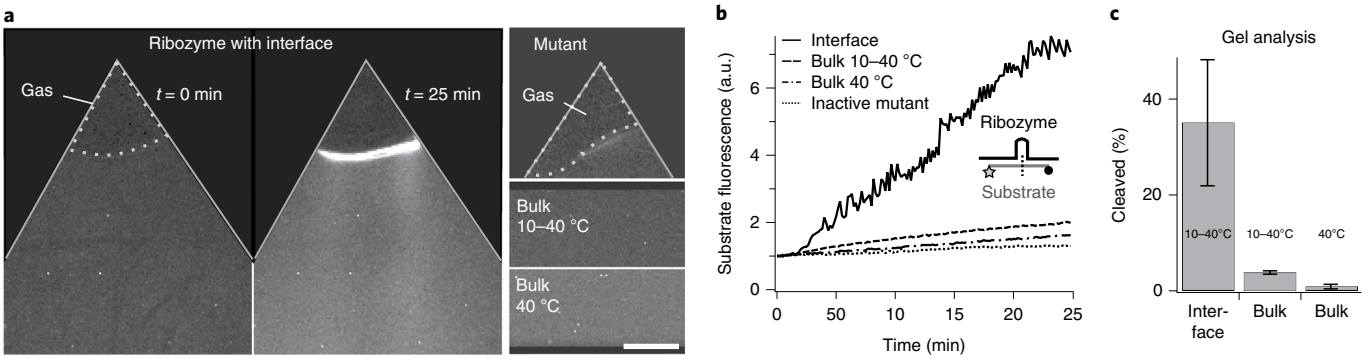

Fig. 3 | Ribozyme catalysis triggered by an interface. Fluorescence microscopy and PAGE analysis of Hammerhead ribozyme activity. **a**, Hammerhead in a chamber with a gas–water interface after 0 and 25 min (left and right, respectively; T_{warm} = 40 °C, T_{cold} = 10 °C, 0.1 μM ribozyme, 0.5 μM substrate, 0.4 mM $MgCl_2$). Fluorescence increased strongly at the interface over time (Supplementary Video 2). Without the interface (bottom right), the Hammerhead shows little activity, both in the same temperature difference and at a homogeneous 40 °C; an inactive mutant (top right) showed no significant increase in fluorescence. Scale bar, 500 μm. **b**, Overall fluorescence in the bulk fluid from **a** over time. The illustration shows the ribozyme (black) and substrate (grey) with the cleavage site (dotted line), dye (star) and quencher (black circle). **c**, PAGE analysis results with and without the interface after 25 min (gels in Supplementary Fig. 5). We found a significant increase in cleaved product with the interface (35 ± 13%, five measurements) compared to bulk samples in the same temperature gradient (3.79 ± 0.35%, three measurements) or at 40 °C (0.86 ± 0.49, three measurements). Error bars show s.d. from normalized gel band intensities.

In the same temperature gradient, but without interfaces, only 3.8% of the substrate was cleaved. This decreased even further when the bulk temperature was set homogeneously to 40 °C, where only very little activity could be observed. When using an inactive mutant version of the ribozyme, the fluorescence did not increase noticeably (Fig. 3a, top right; Fig. 3b, dotted line), demonstrating that it did not originate from enhanced hydrolysis at the interface. This shows that the accumulation mechanism can enhance the catalytic activity of ribozymes, while maintaining low-salt conditions in the bulk solution. It increases the turnover—the number of strands cleaved per ribozyme—by the accumulation of substrates and ions.

RNA/DNA gelation. The length selectivity observed in nucleic acid accumulation could increase the concentration of self-complementary oligonucleotides to the point that hydrogel-forming concentrations could be reached (Fig. 4a). Self-complementary strands were found to form a macroscopic, millimetre-sized hydrogel, a process previously shown for DNA in a thermophoretic accumulation chamber[15]. We observed the formation of hydrogels for self-complementary DNA and RNA and both for GC-only (Fig. 4 and Supplementary Video 3) and AT-only sequences (Supplementary Fig. 6). All oligonucleotides were end-labelled and HPLC purified to minimize the presence of free dye in the experiments.

Starting from uniformly distributed DNA in bulk solution (Fig. 4a(i), initial concentration 10 μM; 1-fold PBS buffer), the DNA was quickly accumulated at the interface once the temperature gradient was established. Within 8 min, a hydrogel had formed, after which it quickly detached from the interface and entrained in the convection flow (Fig. 4a(ii)). The hydrogel nature of the DNA was checked by increasing the temperature from 30 °C, below the melting point of the self-complementary sequences, to 70 °C, where the hydrogel dissolved into the convection flow (Fig. 4a(iii)).

The formation of hydrogels from self-complementary RNA was also observed. We co-accumulated two different sequences in a 20 °C temperature difference (Fig. 4b). The red fluorescence channel monitored a 36mer non-complementary ssRNA strand and the green fluorescence channel a self-complementary GC-only 36mer ssRNA strand with three self-complementary binding sites. Based on simulations (Nupack, www.nupack.org), the hydrogel-forming strands bind to each other and form a network of polymers (Fig. 4b, right). Over the course of the experiment, both the non-complementary and self-complementary strands accumulated at the same interface. However, after 21 min, we only observed a hydrogel for the self-complementary green RNA. No hydrogel could be observed for the non-complementary red RNA, which also accumulated near the surface but was not forming large-scale structures (Supplementary Video 3). Replacing the GC-only strand with a 60mer AU-only RNA with similar self-complementarity gave the same results (Supplementary Fig. 6). This demonstrates that gel formation and separation of the strands is not dominated by G-quadruplex formation.

The red and green strands separated macroscopically based only on their sequence. The self-complementary strands remained at a local high concentration in the hydrogel, which offered reduced hydrolysis rates due to its predominantly double-stranded nature. This sequence-selective gelation was also similarly found for DNA (Supplementary Fig. 6).

DNA encapsulation in vesicle aggregates. A key requirement for the emergence of cellular life is the encapsulation of molecules at increased concentration relative to their more dilute external environment. Fatty acids are potential candidates that could separate nucleic acids in vesicles, possibly incorporating phospholipids into their membranes over time[31,32]. For the encapsulation, three autonomous processes need to occur: (1) accumulation of oligonucleotides to meaningful concentrations, (2) accumulation of vesicles to trigger their aggregation or fusion and (3) the combination of both in one location to encapsulate oligonucleotides into vesicular structures. Here, we show that heated gas–water interfaces could fulfil these requirements. The accumulated vesicles do not necessarily form larger, round vesicles, but aggregate. However, we demonstrate in the following that these aggregates enclose DNA and RNA in an aqueous phase, allowing their binding and folding. Shear flows close to the interface as well as in the convection flow were shown to divide these aggregates to form smaller and more round structures.

To introduce lipids in a homogeneous manner, the chamber was filled with small 100-nm-sized vesicles prepared from a 10 mM oleic acid solution (0.2 M Na-bicine, 1 mM EDTA, pH 8.5) and 2 μM DNA. These initially small vesicles appeared as a continuous

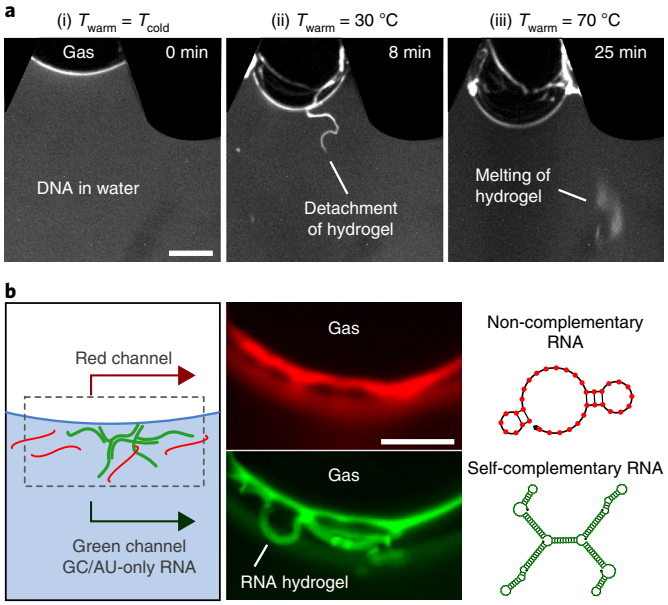

Fig. 4 | Sequence-selective gelation of RNA and DNA. Fluorescence microscopy of DNA and RNA revealed the formation of hydrogels at the interface. **a**, Gelation of DNA. (i) Before applying the temperature gradient, DNA accumulated at the interface only due to slight surface adsorption to the gas–water interface. (ii) After applying the temperature gradient, a DNA hydrogel formed and detached from the interface (Supplementary Video 3). (iii) At T_{warm} = 70 °C, the hydrogel melted and redistributed the DNA back into the bulk fluid. We estimated the DNA concentration in the hydrogel to be 100 µM. Scale bar, 500 µm. **b**, Gelation of RNA. In a single experiment, non-complementary 36mer ssRNA (red) was accumulated with self-complementary GC-only 36mer ssRNA (green). Both strands accumulated at the interface (T_{warm} = 30 °C, T_{cold} = 10 °C), but only the self-complementary RNA formed an elongated, fibrous hydrogel. The same behaviour was found for DNA and AU-only RNA (Supplementary Fig. 6). Scale bar, 125 µm.

background (Fig. 5a,b and Supplementary Video 4). After turning on the temperature gradient, we observed the accumulation of these vesicles together with DNA within 10 min at the interface (T_{warm} = 70 °C, T_{cold} = 10 °C). The vesicles aggregated together and formed larger clusters. It should be noted that this aggregation and cluster formation were strongly increased if ~0.1% 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine was added to the lipids (present here to also label the lipids). Interestingly, the co-accumulated DNA was encapsulated into these vesicle aggregates, in which we found an up to 18-fold increase in oligonucleotide concentration compared to the bulk solution. These aggregates were shuttled into the convection and frequently formed thread-like structures. Close to the interface, we observed strong flows (Supplementary Video 4), which we attribute to Marangoni flows. These could originate from lateral temperature gradients across the chamber due to inhomogeneous heating or differences in the thermal conductivity of the chamber material and water. Aggregates were observed to divide and split into smaller compartments in these flows and convection (Supplementary Fig. 4).

Vesicles formed from the phospholipid 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) (3.6 mM DOPC, 0.2 M Na-bicine, 1 mM EDTA, pH 8.5) showed a similar behaviour, but they aggregated more strongly at the interface, and externally applied flow across the microfluidics was sometimes necessary to remove them from the interface. The accumulated 100-nm-sized vesicles formed larger and more spherical structures compared to the oleic acid aggregates, and encapsulated DNA equally well. They also underwent fission due to shear stress in the convection flow (Fig. 5c and Supplementary Video 4).

The vesicle aggregates (oleic acid or DOPC) concentrated around smaller gas bubbles (150 µm in diameter), inducing a clustering of DNA (Supplementary Fig. 7). This was not observed in the absence of lipids. Lipid vesicles therefore significantly enriched the local DNA concentration.

The co-location of DNA and lipids raises the question of whether the oligonucleotides are in an aqueous phase that allows, for example, the folding of RNA or binding of DNA, and whether these compartments are protected from their surrounding. To test the former, we accumulated the RNA aptamer 'Broccoli'[33] (Fig. 5d), which folds around the fluorophore DFHBI-1T and increases its fluorescence, with DOPC vesicles (1 µM aptamer, 10 µM DFHBI-1T, 3.6 mM DOPC, 50 mM HEPES, pH 7.6, 100 mM KCl, 1 mM $MgCl_2$). Here, we used T_{warm} = 40 °C and T_{cold} = 10 °C to avoid RNA and fluorophore degradation. Again, the RNA, fluorophore and vesicles accumulated at the interface, leading to the formation of aggregates that were visible both in the lipid as well as the fluorophore colour channel. Replacing the Broccoli aptamer with a non-binding RNA strand led to a more than 100-fold reduced fluorescence of the fluorophore. This shows that the aptamer was folded inside the aggregates.

To demonstrate that the accumulated material was protected inside the aggregates, a 72mer double-stranded DNA (dsDNA) was accumulated with DOPC vesicles (Fig. 5e, 7.1 µM DNA, 3.6 mM DOPC, 0.2 M Na-glycineamide pH 8.5, 6 mM $MgCl_2$, 1 mM $CaCl_2$). Each pair of dsDNA strands thereby contained a FAM dye on one strand and a 5-Carboxy-X Rhodamin (ROX) dye opposite to it on the other strand. These dyes form a Förster resonance energy transfer (FRET) pair, in which the excited FAM dye can transfer energy to the ROX dye, which then fluoresces. As this energy transfer works only in close proximity of the dyes, we could use it to measure the amount of dsDNA inside the vesicle aggregates. The higher the signal—between 0 and 1—the more dsDNA was present. After formation of aggregates at the interface, most of the solution was extracted and DNase I was added (0.5 units per ~10 µl). The solution was then put back into the chamber and the FRET signal was observed at 37 °C. The DNase digested the DNA strands outside the aggregates, where the signal quickly dropped (Fig. 5e, Supplementary Video 4 and Supplementary Fig. 7). The FRET signal of the aggregates reduced only slightly, probably as a consequence of the digestion

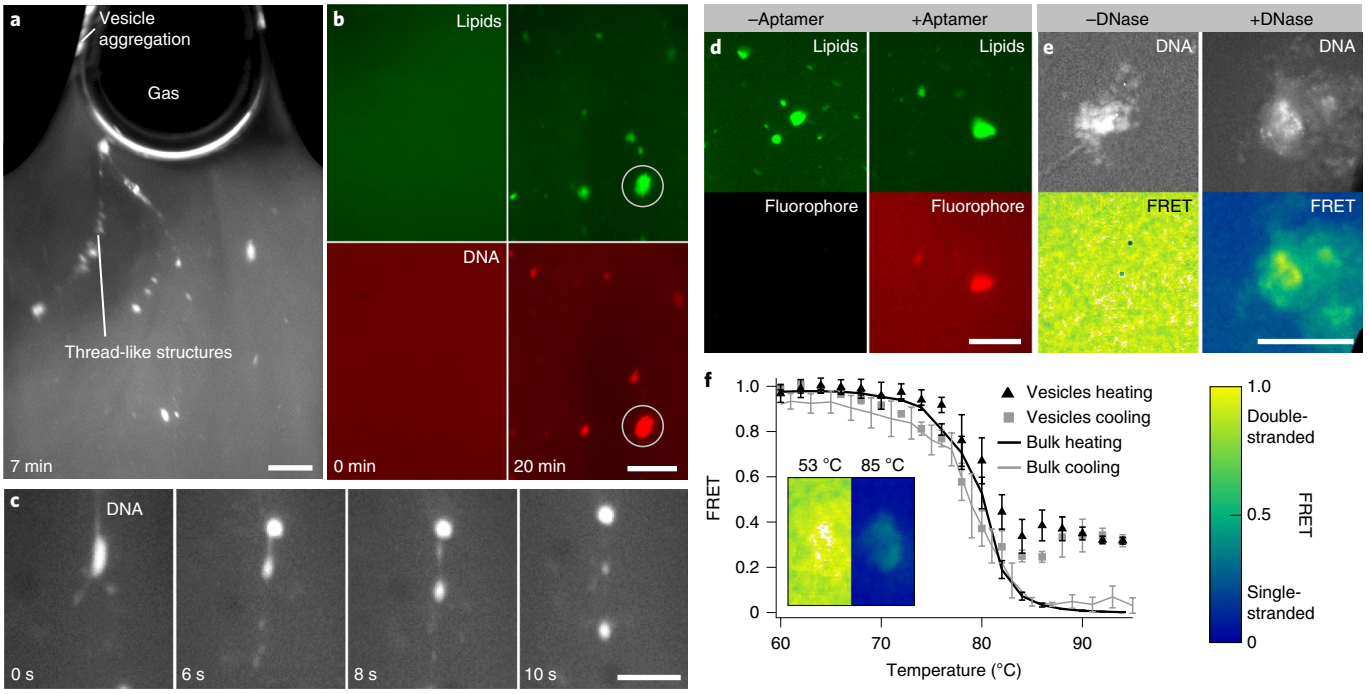

Fig. 5 | DNA and RNA encapsulation and protection in vesicles formed at the interface. Fluorescence microscopy images of DNA/RNA and lipid colour channels. **a**, Lipid channel, with 100 nm sized oleic acid vesicles accumulated at the gas–water interface. These formed aggregates of vesicles that were ejected into the bulk and formed thread-like structures (Supplementary Video 4). **b**, Comparison of lipid and DNA colour channels (green, lipid; red, DNA), showing that lipids and DNA co-localize, meaning that the 100 nm-sized vesicles have formed larger clusters and encapsulated DNA at up to 18-fold enhanced concentration compared to the bulk. **c**, DNA channel, with DOPC vesicles accumulated under the same conditions and also encapsulated DNA at enhanced concentrations. Here, a fission of the DOPC vesicles containing DNA was observed in the shear flow. **d**, Encapsulation of fluorophore DFHBI-1T with a non-binding RNA (left) and the Broccoli aptamer (right), which increases its fluorescence. It shows that RNA has folded in an active conformation inside the vesicles. **e**, FRET analysis of dsDNA in aggregates before (left) and after (right) addition of DNase I. DNA inside the aggregates was protected from the DNase (high FRET), whereas it was digested in the bulk solution (low FRET, Supplementary Video 4). **f**, FRET melting curve of dsDNA inside and outside the aggregates, starting from 1.0 (all DNA double-stranded) at 60 °C to 0 (all DNA single-stranded) at 95 °C outside and 0.3 inside the aggregates. Error bars indicate s.d. measured for a larger area of normalized signal of the shown example measurement. Scale bars, 200 µm (**a**,**b**), 100 µm (**c**), 250 µm (**d**,**e**).

of solution above/below or DNA sticking to their surface. We thus show that the aggregates protect the molecules inside them.

The FRET signal was also used to observe the melting of dsDNA inside the chamber (Fig. 5f). After accumulation (7.1 µM DNA, 3.6 mM DOPC, 0.2 M Na-glycineamide pH 8.5, 11 mM NaCl, 0.22 mM KCl, 0.8 mM phosphate buffer), the chamber was heated to 95 °C, during which the FRET signal reduced to 0 in the bulk and ~0.3 in the aggregates. The remaining signal could stem from aggregated DNA or DNA that is enclosed and stabilized by lipids in a way that did not allow the strands to fully unbind. After cooling, the FRET signal returned to 1 for both aggregates and the bulk solution. The combination of aptamer and FRET analysis demonstrates that oligonucleotides are encapsulated inside the aggregates in an aqueous phase that allows them to melt and fold.

The above experiments assumed the presence of uniformly 100-nm-sized vesicles at the beginning. We also explored the behaviour of the system when it initially contained a range of vesicle sizes, up to ~30 µm in diameter (Supplementary Fig. 8). These exhibited a similar, but often slower accumulation behaviour. Within 20 min, the system again started to form vesicle clusters that contained enhanced DNA concentrations. The formation of oleic acid aggregates was also observed in the absence of DNA, indicating that they were not the result of a DNA/lipid interaction (Supplementary Fig. 8).

Crystallization at gas bubbles. The building blocks for the synthesis of single nucleotides, such as the prebiotic RNA precursor ribose aminooxazoline (RAO), accumulated near gas bubbles to concentrations that triggered its crystallization. For RAO, a crystallization is of fundamental interest because it can be both diastereoisomerically purified by selective sequential crystallization of its precursors and enantiomerically enriched by conglomerate crystallization, where the two enantiomers (D- and L-) of RAO crystallize into discrete independent domains[34,35]. To trigger controlled crystal growth, RAO would need to be accumulated slowly around a growing bubble. In previous experiments, bubbles were artificially created. Here we used elevated temperatures (70 °C warm side, 10 °C cold side) to trigger the spontaneous formation of a bubble. Subsequently, RAO accumulated and crystallized around it (Fig. 6a,b). Therefore, a 40 mM solution of D-RAO, 2.8-fold below the saturation concentration at T_{warm} = 70 °C (~110 mM), was used to fill a chamber without a corrugated geometry and did not create a gas–water interface. For the crystallization to occur at the warm side of the chamber, RAO would have needed to accumulate several fold to overcome the nucleation energy barrier[36]. We monitored the fluorescence of 1 µM Cy5 added to the solution, which co-accumulated at the gas–water interface but was not incorporated into the crystals.

An initially small bubble formed on the warm side and accumulated RAO around it. Within 40 min, the bubble grew while continuously increasing the RAO concentration at its warm side. The crystal shown in Fig. 6a was found at the location where the bubble had formed. No crystals were found on the cold side of the chamber. An X-ray crystal structure determination confirmed that crystals grown on the warm side were indeed D-RAO (Fig. 6c).

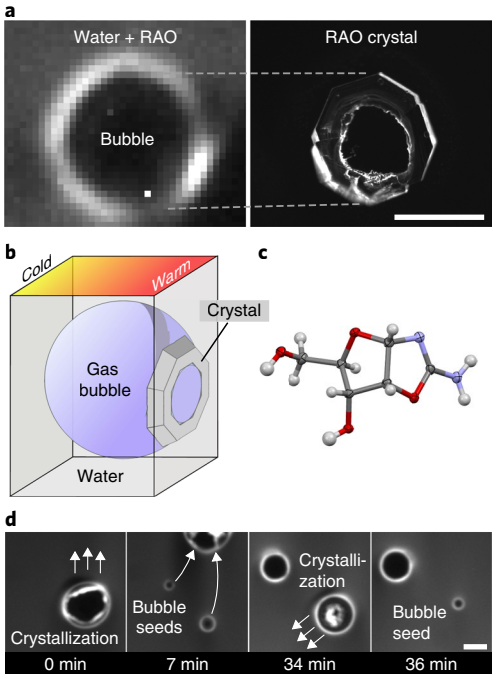

Fig. 6 | Crystallization and bubble movement. a, RAO accumulated from 40 mM bulk concentration around a growing bubble at the warm side of the chamber (70 °C). After 37 min, an RAO crystal formed below the bubble at the water–gas–solid interface, as indicated. **b**, The crystal shape suggests that it formed successively as the bubble grew, consistent with simulation results (Supplementary Fig. 9). **c**, X-ray crystal structure analysis of a D-RAO crystal grown at a gas–water interface. **d**, Moving bubbles leave behind crystals (7 min) that act as seeds for new bubbles (34 min, Supplementary Video 4). Scale bars, 250 µm.

Interestingly, the formation of RAO crystals led to the later re-formation of gas bubbles near small cavities—a known and well described process of heterogeneous nucleation in supersaturated solutions[36]. We observed this crystal-induced bubble formation in experiments at high T_{warm} = 70 °C (Fig. 6d and Supplementary Video 4). When a gas bubble (around which RAO crystals had formed) moved away, it left behind accumulated material, possibly crystal cavities still filled with gas, which facilitated the growth of new daughter bubbles at the same location. When these daughter bubbles increased in size sufficiently, they started to accumulate RAO, formed crystals and eventually moved away. The remaining crystals again hosted the growth of new bubbles and began the bubble-induced crystallization cycle again.

Dry–wet cycles and phosphorylation by moving interfaces. If bubbles were not confined by their geometry, they tended to move upwards in the chamber (Fig. 7a) due to buoyancy forces. This movement led to a continuous cycling of dry–wet conditions (Fig. 7a), as accumulated material close to the interface entered the bubble, dried and was rehydrated when the bubble moved away. At the same time, accumulated material at the trailing edge was dragged along if the bubble moved slowly enough, keeping high molecule concentrations in the vicinity of the interface (Supplementary Video 6).

At a hydrophilic surface material such as silicon dioxide (quartz), used so far on the cold side in the experiments, the contact angles of the condensed droplets were small, and they re-entered quickly into the bulk water, leading to only little movement of the gas–water interface. To trigger many wet–dry cycles, we placed a Teflon foil on the warm and cold sides of the chamber. This led to the formation of many droplets on the cold side (Fig. 7b, small round structures inside the dark gas region, Supplementary Video 6), which, if large enough, could also be in contact with the warm side. In this setting, we tested the phosphorylation of nucleosides, a reaction that usually requires dry conditions at elevated temperatures. The reaction is known to be most effective when the solution is dried at 100 °C, a scenario hard to reconcile with typical RNA-world conditions[37,38]. We observed a 20 times more efficient phosphorylation of cytidine nucleosides (Fig. 7c, T_{warm} = 60 °C, T_{cold} = 37 °C, 240 mM cytidine and ammonium dihydrogen phosphate, 2.4 M urea, similar to the phosphorylation used in ref. [37]) compared to bulk water at average temperatures of 50 °C and 60 °C. The found enhanced in situ phosphorylation reaction would improve the recycling of hydrolysed RNA.

Discussion

We have found a general accumulation mechanism of molecules and small vesicles at gas bubbles subjected to heat flow in water. A temperature gradient across a gas–water interface created a continuous process of evaporation and condensation of water between the warm and cold sides. This moved molecules towards the bubble interface and increased their concentration by several orders of magnitude, depending on their diffusion coefficients. Because the contact line between gas and water was mobile, dry–wet cycles were created, in many conditions as often as twice per minute. The molecules studied here have been discussed as prebiotic candidates before the emergence of life[35,37,39]. Fluorescently labelled analogues were used to probe their concentration. We did not elaborate on the types of gas used in this study, because, due to their low concentrations, they did not modify the surface tension and water evaporation significantly[40].

Our simulation captured the basic characteristics of this accumulation and validated the experimental results. A 4,000-fold DNA accumulation was reached both in experiment and simulation. In comparison with thermophoretic traps, with which the setting could be combined, the accumulation at the interface occurred significantly faster, on the timescale of minutes rather than hours[11].

We used different temperature gradients adapted to the different scenarios. The magnitude of accumulation depended strongly on the applied temperature difference (Supplementary Fig. 4), which was kept at high values in Figs. 3 and 4 (10–40 °C and 10–30 °C, respectively) to accumulate RNA strongly, but still under cold conditions to keep hydrolysis insignificant. Vesicles aggregated under similar conditions (Fig. 5d), but with faster convective flows from a higher temperature difference (10–70 °C), shuttled aggregated vesicles more efficiently away from the interface for downstream analysis. The creation of crystals from RAO (Fig. 6) required the larger temperature gradients required for an efficient accumulation of the comparably small molecules. The phosphorylation chemistry of monomers (Fig. 7) profited from an overall enhanced temperature.

The co-accumulation of small ions offers additional reactivity. Divalent salts such as Mg^{2+} were predicted by our model to accumulate by a factor of 4–5 for temperature differences of 20–30 °C. In high-salt environments, this could enhance the hydrolysis of, for example, accumulated RNA, but the larger temperature differences also increase the movement of the interface by the recondensation of water, decreasing the time accumulated molecules spend dried at high temperatures. On the other side, the enhanced salt concentrations would trigger ribozymatic activity at the interface while the molecules in the bulk are protected by low salt concentrations from hydrolysis.

The system provided, in a single setting, a network of widely different reactions, connected by the fast diffusive transport of these molecules through water between different microbubbles. Ligation[41] chemistry to drive replication could be made possible by

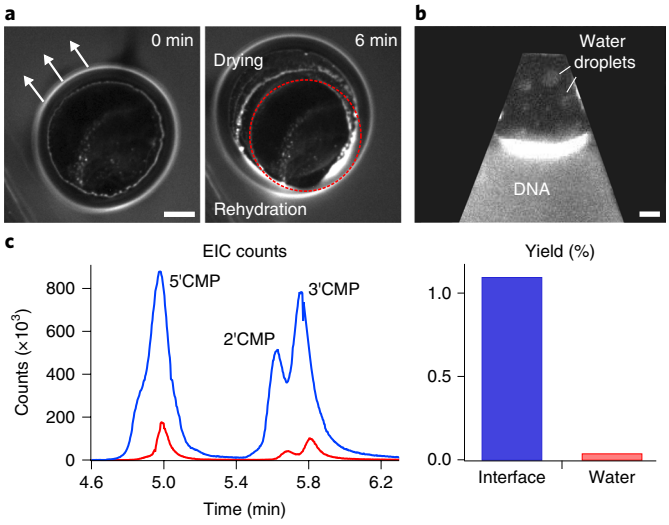

Fig. 7 | Dry–wet cycles and phosphorylation of nucleosides. a, Movement of bubbles in the chamber left behind dried material in the front that was rehydrated at the back side of the bubble, leading to dry–wet cycles (left, 0 min; right, 6 min). Dotted circle, the position of the accumulation region at $t = 0$ min. **b**, Using a hydrophobic cold side in the chamber increased the speed of dry–wet cycling by enhancing the formation of circular water droplets on the cold side. As they 'rained' into the bulk water, the interface moved periodically, triggering wet–dry cycles at the interface (Supplementary Video 6). **c**, Cytidine nucleosides were phosphorylated in the chamber shown in **b**. Over 12 h, the phosphorylation was, on average, 20 times more effective with the interfaces present (blue, $T_{warm} = 60$ °C, $T_{cold} = 37$ °C; Cytidine monophosphate (CMP) yield, 1.1 ± 0.73% from 12 repeats) compared to bulk water at 60 °C (red; CMP yield, 0.042 ± 0.02% from five repeats) or 50 °C (0.056 ± 0.0015% from 15 repeats). Left, extracted-ion chromatogram (EIC) counts (EIC) for CMP peaks after HPLC-MS analysis. Right, yields measured by HPLC and ultraviolet detection. Scale bars, 100 µm (**a**), 250 µm (**b**).

the high local concentrations of substrates, while the drying process in close proximity would trigger the necessary phosphorylation[37,38] and activation chemistry[42] to drive the polymerization[43] of monomers. The dry–wet cycles were implemented continuously at the moving interface and molecules were retained at high concentrations at the interfaces after rehydration. In the adjacent water, molecules thermally cycled between warm and cold via laminar thermal convection.

At the interface, a strongly enhanced ribozymatic activity was demonstrated by the accumulation of the Hammerhead ribozyme. Its activity at the interface clearly dominates the turnover of substrate in the system, with up to 50% of the product being cleaved in a chamber with a gas interface (compared to 3.8% in the bulk). This can be attributed to higher local concentrations of ribozyme, substrate and $MgCl_2$ at the interface. Product strands are shuttled back into the bulk solution, where they are protected against the higher salt conditions. The mechanism thereby provides a way to enhance catalytic activity and increase the efficiency of RNA-catalysed processes.

Complex sequence phenotypes of RNA with several self-complementary sites have shown a sequence-selective formation of hydrogels at the interface. These hydrogels maintained a high local RNA concentration in water, an interesting setting to support efficient RNA catalysis[44,45], also because the high amount of hybridization in the hydrogels could protect oligonucleotides from hydrolysis[46], even in challenging salt concentrations.

In the presence of lipids such as oleic acid or DOPC, accumulation at the heated gas–water interface led to a continuous encapsulation of oligonucleotides into vesicle aggregates. The local DNA concentration inside these structures increased by a factor of up to 18 compared to the bulk solution. Folded RNA aptamers also accumulated inside the aggregates and dsDNA was shown to melt and re-anneal, demonstrating that aqueous phases readily exist inside the aggregates. As shown with DNase, the vesicles protected the encapsulated DNA from the bulk solution. The encapsulation of oligonucleotides into lipid membranes is considered to be one of the key elements for more complex life and it has been suggested that lipids facilitated the assembly and polymerization of monomers[47,48]. In the convection flow, the vesicles were subjected to temperature cycles and shear forces that led to vesicle fission. DOPC also produced vesicular structures, showing that modern phospholipids[39] could also accumulate and encapsulate oligonucleotides in the shown conditions.

For the prebiotic synthesis of RNA, crystallization at microbubbles would enable the purification of sugar–nucleobase precursors[34] and possibly also their chiral amplification by the enhanced growth of conglomerate RAO crystals[35]. Interestingly, the sites of crystal formation later triggered, again, the formation of gas bubbles, showing a self-selection for crystallization conditions. If the seed crystal was homochiral, the subsequent bubble formation could accumulate and promote the assembly of more homochiral molecules at the same location. Finally, we found that the prebiotically important dry chemistry of nucleoside phosphorylation was enhanced by the gas interface: cytidine formed CMP 20 times more effectively compared to aqueous conditions.

To conclude, the experiments showed multiple modes of condensation, enrichment, accumulation and increased catalysis at heated gas microbubbles. This led to the physicochemical assembly and localization of prebiotic molecules—such as RNA precursors, lipids and ribozymes. We argue that this accumulation of molecules at a gas–water interface was a robust feature of natural microfluidic systems in porous volcanic rocks in aqueous environments, a setting likely to be ubiquitous on early Earth[9]. The simultaneous occurrence of six synergistic mechanisms for the accumulation and processing of prebiotic molecules, all operating in close proximity, fulfils the requirements for early life to connect a cascade of core reactions in the same non-equilibrium setting.

The setting presented here could therefore have largely helped in an informational polymer world, in which the first simple replicators were evolving. Following a synthesis of life's first building

blocks, the accumulation at gas–water interfaces offers a mechanism to select polymers and enhance their catalytic activity. The shown continuous encapsulation dynamics of the accumulated molecules at the interface offers a pathway for the emergence of the cellular processes of life. Further experiments will test how this setting can host replication and selection towards early molecular evolution.

Methods

For the experiments in Fig. 2, a thermal chamber was sandwiched from a thin (>240 µm) layer of polyethylene terephthalate (PETG) plastic film deposited on a silicon wafer using a 3D printer (Ultimaker 2) in a funnel shape to facilitate a gas–water interface (Supplementary Fig. 1). A sapphire (Al_2O_3) block sealed the chamber. The chamber was filled through two 240-µm-thick borosilicate capillaries (Vitrocom). The system was annealed at 150 °C and the sapphire pressed down, fixing the thickness to 240 µm. Polydimethylsiloxane was deposited to seal the chamber. For most other experiments, chambers were built with an ultraviolet-curable resin (Photocentric 3D Daylight Resin, flexible, colour amber) through a master obtained by laser printing on a transparency film. Spacers between master and resin defined a chamber thickness of 150 µm or 250 µm. After illumination, sapphire windows were placed on top of the chambers and sealed with resin. Microfluidic access was provided through holes in the sapphire. Similarly, the chamber used for the phosphorylation and Hammerhead experiments was built by replacing the resin with a 254-µm-thick Teflon foil from which the structure was cut out using a cutting plotter. For the phosphorylation, additional Teflon foils were placed on the warm and cold sides of the chamber to mimic a hydrophobic surface. The temperature gradient was produced by heating the sapphire block through copper fixtures using heater cartridges for 3D printers and cooling the silicon side with a water bath. Temperature sensors and proportional–integral–derivative (PID) software maintained the temperatures. The temperature inside the chamber was calculated from the chamber geometry and known material constants with finite element methods (Comsol).

Fluorescence was measured with a fluorescence microscope (Zeiss Axio) through the transparent sapphire heating block using Mitutoyo infinity corrected long working distance objectives (×2 and ×10) and a Zeiss Fluar ×5 objective. The accumulation of DNA was detected with 200 nM FAM-labelled 132-base ssDNA in 0.1-fold PBS (see Supplementary Information for sequence). The silicon substrate was maintained at 10 °C, with the copper heaters initially turned off. The experiment began with the copper heaters set to maintain a desired temperature. Background intensity levels were obtained from the non-fluorescing gas region. The bulk fluorescence signal was obtained from an area far from the free interface. The accumulated DNA fluorescence ratio was averaged perpendicular to the interface and divided by the bulk fluorescence. The flow was visualized with 200 nm FAM-labelled polystyrene beads in 0.1× PBS. The positions and velocities of the beads were tracked using ImageJ.

Simulation protocol. Simulations were performed using the finite element software COMSOL v4.4. The 2D model solved the convective heat equation, molecule diffusion equation and Navier–Stokes equations perpendicular to the contact line. Marangoni flows were established by implementing a stress boundary condition of the fluid velocity at the interface and by introducing a temperature-dependent surface tension. The water vapour concentration was simulated in the gas region above the interface by a diffusion–convection equation. The gas velocity was calculated from the temperature profile. Its velocity, combined with diffusion, caused an efficient net mass transport of vapour away from the interface into the gas bubble. By coupling a temperature-dependent vapour concentration boundary condition to the interface, a velocity boundary condition for water at the interface was imposed by the state equation of water. This resulted in capillary flow and evaporative mass transport of the water vapour, allowing vapour to enter and escape the water through evaporation or condensation. Thermophoresis was introduced via a thermophoretic drift term in the convection–diffusion equation describing the DNA concentration, and the Soret coefficients of the DNA were taken from experimental data[49]. To incorporate the time lag of the heating process, the sapphire temperature was measured over time and incorporated as a polynomial function for T_{warm}.

The above simulation was solved over time, resulting in the time evolution of accumulation at the contact line. To fit the simulation to the observed data, the surface tension dependency on temperature was set within observed values[50] and fine-tuned as a free parameter using the bead tracking data. The geometry of the interface itself was adjusted to moderately tune the DNA accumulation dynamics. Concentrations were determined by averaging the top ~10 µm of the meniscus at the hot side and comparing this to the simulated and experimental fluorescence, establishing a relation between fluorescence and average tip concentration.

To simulate the bubble shown in Supplementary Fig. 9, the simulation was transferred to an axial-symmetric geometry with spherical coordinates. The simulation around a small bubble was closed and gravity was pointed downwards. This removed the slow convection of water in the simulation, which, as expected from our modelling, did not change the accumulation characteristics. Ion and salt diffusion coefficients were taken from refs. [51–53].

Data availability

The data supporting the findings of this study are available within the paper and its Supplementary Information. Additional information and files are available from the corresponding author upon reasonable request. X-ray crystallographic data were also deposited at the Cambridge Crystallographic Data Centre (CCDC) under CCDC deposition no. 1847429.

Code availability

The complete details of both simulations are documented in the html report and mph simulation files in the Supplementary Information.

Received: 7 September 2018; Accepted: 21 June 2019;
Published online: 29 July 2019

References

1. Schrödinger, E. *What is Life? The Physical Aspect of the Living Cell* (Cambridge Univ. Press, 1944).
2. Cross, M. C. & Hohenburg, P. Pattern-formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851–1112 (1993).
3. Bodenschatz, E., Pesch, W. & Ahlers, G. Recent developments in Rayleigh–Benard convection. *Annu. Rev. Fluid Mech.* **32**, 709–778 (2000).
4. Fritts, D. C. & Alexander, M. J. Gravity wave dynamics and effects in the middle atmosphere. *Rev. Geophys.* **41**, 1003 (2003).
5. Eaton, J. K. & Fessler, J. R. Preferential concentration of particles by turbulence. *Int. J. Multiph. Flow* **20**, 169–209 (1994).
6. Götzendorfer, A., Kruelle, C. A., Rehberg, I. & Svenšek, D. Localized subharmonic waves in a circularly vibrated granular bed. *Phys. Rev. Lett.* **97**, 198001 (2006).
7. Chen, J. & Lopez, J. A. Interactions of platelets with subendothelium and endothelium. *Microcirculation* **12**, 235–246 (2005).
8. Moore, W. B. & Webb, A. A. G. Heat-pipe Earth. *Nature* **501**, 501–505 (2013).
9. Arndt, N. T. & Nisbet, E. G. Processes on the young Earth and the habitats of early life. *Annu. Rev. Earth Planet. Sci.* **40**, 521–549 (2012).
10. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc. Natl Acad. Sci. USA* **103**, 19678–19682 (2006).
11. Baaske, P. et al. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl Acad. Sci. USA* **104**, 9346–9351 (2007).
12. Niether, D., Afanasenkau, D., Dhont, J. K. G. & Wiegand, S. Accumulation of formamide in hydrothermal pores to form prebiotic nucleobases. *Proc. Natl Acad. Sci. USA* **113**, 4272–4277 (2016).
13. 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).
14. 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).
15. Morasch, M., Braun, D. & Mast, C. B. Heat-flow-driven oligonucleotide gelation separates single-base differences. *Angew. Chem. Int. Ed.* **55**, 6676–6679 (2016).
16. Keil, L. M. R., Möller, F. M., Kieß, M., Kudella, P. W. & Mast, C. B. Proton gradients and pH oscillations emerge from heat flow at the microscale. *Nat. Commun.* **8**, 1897 (2017).
17. Budin, I., Bruckner, R. J. & Szostak, J. W. Formation of protocell-like vesicles in a thermal diffusion column. *J. Am. Chem. Soc.* **131**, 9628–9629 (2009).
18. Lerman, L. Potential role of bubbles and droplets in primordial and planetary chemistry. *Orig. Life Evol. Biosph.* **16**, 201–202 (1986).
19. Ariga, K. & Hill, J. P. Monolayers at air–water interfaces: from origins-of-life to nanotechnology. *Chem. Rec.* **11**, 199–211 (2011).
20. Eickbush, T. H. & Moudrianakis, E. N. A mechanism for the entrapment of DNA at an air–water interface. *Biophys. J.* **18**, 275–288 (1977).
21. Griffith, E. C. & Vaida, V. In situ observation of peptide bond formation at the water–air interface. *Proc. Natl Acad. Sci. USA* **109**, 15697–15701 (2012).
22. Deegan, R. D. et al. Capillary flow as the cause of ring stains from dried liquid drops. *Nature* **389**, 827–829 (1997).
23. Smith, K. A. On convective instability induced by surface-tension gradients. *J. Fluid Mech.* **24**, 401–414 (1966).
24. Batchelor, G. K. *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, 1973).
25. Deegan, R. D. Pattern formation in drying drops. *Phys. Rev. E* **61**, 475–485 (2000).
26. Larson, R. G. Transport and deposition patterns in drying sessile droplets. *AIChE J.* **60**, 1538–1571 (2014).
27. Savino, R., Paterna, D. & Favaloro, N. Buoyancy and marangoni effects in an evaporating drop. *J. Thermophys. Heat Transf.* **16**, 562–574 (2002).
28. Drobot, B. et al. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nat. Commun.* **9**, 3643 (2018).
29. Weinberg, M. S. & Rossi, J. J. Comparative single-turnover kinetic analyses of trans-cleaving hammerhead ribozymes with naturally derived non-conserved sequence motifs. *FEBS Lett.* **579**, 1619–1624 (2005).

30. Dahm, S. C. & Uhlenbeck, O. C. Role of divalent metal ions in the hammerhead RNA cleavage reaction. *Biochemistry* **30**, 9464–9469 (1991).
31. Zhu, T. F. & Szostak, J. W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.* **131**, 5705–5713 (2009).
32. Budin, I. & Szostak, J. W. Physical effects underlying the transition from primitive to modern cell membranes. *Proc. Natl Acad. Sci. USA* **108**, 5249–5254 (2011).
33. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).
34. Islam, S., Bučar, D.-K. & Powner, M. W. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.* **9**, 584–589 (2017).
35. Anastasi, C., Crowe, M., Powner, M. W. & Sutherland, J. D. Direct assembly of nucleoside precursors from two- and three-carbon units. *Angew. Chem. Int. Ed.* **45**, 6176–6179 (2006).
36. Jones, S. F., Evans, G. M. & Galvin, K. P. Bubble nucleation from gas cavities—a review. *Adv. Colloid Interface Sci.* **80**, 27–50 (1999).
37. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
38. Lohrmann, R. & Orgel, L. E. Urea–inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* **171**, 490–494 (1971).
39. Gibard, C., Bhowmik, S., Karki, M., Kim, E.-K. & Krishnamurthy, R. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nat. Chem.* **10**, 2012–2017 (2017).
40. Chow, Y. T. F., Maitland, G. C. & Trusler, J. P. M. Interfacial tensions of the (CO_2+N_2+H_2O) system at temperatures of (298 to 448) K and pressures up to 40 MPa. *J. Chem. Thermodyn.* **93**, 392–403 (2016).
41. Sosson, M. & Richter, C. Enzyme-free genetic copying of DNA and RNA sequences. *Beilstein J. Org. Chem.* **14**, 603–617 (2018).
42. Forsythe, J. G. 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).
43. 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).
44. Vaidya, N. et al. Spontaneous network formation among cooperative RNA replicators. *Nature* **491**, 72–77 (2012).
45. Mutschler, H., Wochner, A. & Holliger, P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nat. Chem.* **7**, 502–508 (2015).
46. Soukup, G. A. & Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308–1325 (1999).
47. Toppozini, L., Dies, H., Deamer, D. W. & Rheinstädter, M. C. Adenosine monophosphate forms ordered arrays in multilamellar lipid matrices: insights into assembly of nucleic acid for primitive life. *PLoS One* **8**, e62810 (2013).
48. Rajmani, S. et al. Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig. Life Evol. Biosph.* **38**, 57–74 (2008).
49. Reineck, P., Wienken, C. J. & Braun, D. Thermophoresis of single stranded DNA. *Electrophoresis* **31**, 279–286 (2010).
50. Vargaftik, N. B., Volkov, B. N. & Voljak, L. D. International tables of the surface tension of water. *J. Phys. Chem. Ref. Data* **12**, 817–820 (1983).
51. Lide, D. R. *CRC Handbook of Chemistry and Physics* (CRC Press, 2001).
52. Li, Y. & Gregory, S. Diffusion of ions in sea water and in deep-sea sediments. *Geochim. Cosmochim. Acta* **33**, 703–714 (1974).
53. Fell, C. J. D. & Hutchison, H. P. Diffusion coefficients for sodium and potassium chlorides in water at elevated temperatures. *J. Chem. Eng. Data* **16**, 427–429 (1971).

Acknowledgements

The authors thank L. Keil for help with data analysis. Financial support from the Simons Foundation (318881 to M.W.P. and 327125 to D.B.), the German Research Foundation (DFG) through CRC/SFB 235 Project P07 and SFB 1032 Project A04, DFG Grant BR2152/3-1 and the US–German Fulbright Program is acknowledged. H.M. is supported by the MaxSynBio consortium, which is jointly funded by the Federal Ministry of Education and Research of Germany and the Max Planck Society. H.M. and K.L.V. are supported by the Volkswagen Initiative 'Life?—A Fresh Scientific Approach to the Basic Principles of Life'. A.K. is supported by a DFG fellowship through the Graduate School of Quantitative Biosciences Munich.

Author contributions

M.M., J.L., C.F.D., A.K., A.I. and Ph.S. performed the experiments. M.M., J.L., K.L.V., S.I., B.S., D.B.D., H.M., Pe.S., M.W.P., C.B.M. and D.B. conceived and designed the experiments. M.M., J.L., K.L.V., S.I., M.K.C., H.M., M.W.P. and D.B. analysed the data. M.M., J.L. and D.B. wrote the paper. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-019-0299-5.

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

Correspondence and requests for materials should be addressed to D.B.

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