# Oligonucleotide Selection and Replication in Early Earth Environments

**Philipp Josef Schwintek**


München 2024

# Oligonucleotide Selection and Replication in Early Earth Environments

**Philipp Josef Schwintek**

**Dissertation**

zur Erlangung des Grades
Doktor der Naturwissenschaften

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

vorgelegt von
Philipp Josef Schwintek
aus München, Deutschland

München, den 10.10.2024

Erstgutachter: Prof. Dr. Dieter Braun
Zweitgutachter: Prof. Dr. Christoph Weber
Tag der mündlichen Prüfung: 25.11.2024

# Zusammenfassung

Die molekulare Evolution von Nukleinsäuren mit ihrer dualen Fähigkeit, genetische Informationen zu speichern und ihre eigene Replikation zu katalysieren, spielt eine zentrale Rolle in der Erforschung des Ursprungs des Lebens. Die Auswahl funktionsfähiger Oligonukleotide aus Pools sowie deren Replikation waren jedoch mit erheblichen Herausforderungen verbunden. In dieser Arbeit werden zwei Schlüsselmechanismen untersucht, die diese Prozesse in der frühen Erdumgebung begünstigt haben könnten. Kapitel 1 befasst sich mit der Entstehung von funktionellen Oligonukleotiden durch einen molekularen Selektionsmechanismus. Zyklische Prozesse wie tägliche Temperaturschwankungen werden als Auslöser für sequenzspezifische Phasentrennung vorgeschlagen. Anhand von Experimenten mit DNA wird gezeigt, dass in der sedimentierten dichten Phase eine sequenzspezifische Anreicherung stattfindet, die insbesondere kurze 22-mer DNA-Sequenzen bevorzugt, die durch Basenpaarung eine bienenwabenartige Sekundärstruktur bilden. Dieser Selektionsmechanismus, der gegenüber Änderungen der Pufferbedingungen robust ist und mit Finite-Elemente-Simulationen übereinstimmt, verändert die Sequenzverteilung in Oligonukleotidpools im Zeitverlauf. Bei Anwendung auf ein realistisches Durchflussszenario in einem porösen Gesteinsimitat wird beobachtet, dass sich bestimmte Sequenzen gegen einen verdünnenden Abfluss schützen und konzentrieren, was ein präbiotisches Modell für die Entstehung autokatalytischer, selbstreplizierender Oligonukleotide darstellt. Kapitel 2 befasst sich mit der Replikation von Nukleinsäuren, die durch Verdünnung und Schwierigkeiten bei der Strangtrennung in Gegenwart divalenter Ionen wie $Mg^{2+}$ beeinträchtigt wird. Herkömmliche Lösungen beinhalten Temperaturgradienten, die zwar effektiv sind, aber zum Zerfall der Moleküle führen können und nur begrenzt zur Verfügung stehen. Hier wird eine isotherme geologische Umgebung untersucht, in der ein Gasfluss über eine offene Gesteinspore in Verbindung mit einem Wasserzufluss die Ansammlung von Biomolekülen und die zyklische Strangtrennung durch lokale Salzkonzentrationsschwankungen antreibt. Experimentelle Daten und Simulationen zeigen, dass dieses System Nukleinsäuren bis zu 30-fach anreichert und zirkuläre Strömungsmuster hervorruft, die eine Strangtrennung ohne die Notwendigkeit einer Erhitzung ermöglichen. Dieser Mechanismus, der durch Fluoreszenzbildgebung und Monte-Carlo-Simulationen unterstützt wird, ermöglicht die Replikation von Nukleinsäuren unter isothermen Bedingungen, was durch die Zugabe von Taq-Polymerase als Modellreplikator nachgewiesen wurde. Diese Entdeckung erweitert das Spektrum der plausiblen Umgebungen für frühes Leben auf unserem Planeten erheblich.

# Abstract

The molecular evolution of nucleic acids, with their dual capacity to store genetic information and to catalyze their own replication, plays a central role in origin of life research. However, the selection of functional oligonucleotides from pools, as well as their replication were required to overcome significant challenges. This thesis explores two key mechanisms that could have facilitated these processes in early Earth environments. Chapter 1 focuses on the emergence of functional oligonucleotides through a molecular selection mechanism. Cyclic processes like daily temperature oscillations are proposed to trigger sequence-specific phase separation. Through experiments with DNA, it is shown that sequence-specific enrichment occurs in the sedimented dense phase, particularly favoring short 22-mer DNA sequences that form a honeycomb-like secondary structure via base-pairing. This selection mechanism, supported by finite element simulations and robust to buffer condition changes, shows an increase of pool sequence bias over time. When applied to a realistic flow-through scenario in a porous rock mimic, specific sequences shelter and concentrate against a diluting outflux, providing a prebiotic model for the emergence of auto-catalytic, self-replicating oligonucleotides. Chapter 2 addresses the challenge of nucleic acid replication, which is hampered by dilution and the difficulty of strand separation in the presence of divalent ions like $Mg^{2+}$. Conventional solutions involve temperature gradients that, while effective, can lead to molecular degradation and are environmentally limited. Here, an isothermal geological environment is investigated, where a gas flow across an open rock pore with an inflow of water, drives the accumulation of biomolecules and cyclic strand separation through local salt concentration oscillations. Experimental data and simulations demonstrate that this system accumulates nucleic acids up to 30-fold and induces circular flow patterns, which facilitate strand separation without the need for heating. This mechanism, supported by evidence from fluorescence imaging and Monte Carlo simulations, enables the replication of nucleic acids under isothermal conditions, demonstrated by using Taq Polymerase as a model replicator. This finding significantly expands the range of plausible environments for early life on our planet.

# List of Acronyms and Abbreviations

**DNA** Deoxyribonucleic acid

**RNA** Ribonucleic acid

**ssDNA** Single-stranded DNA

**dsDNA** Double-stranded DNA

**nt** Nucleotides

$T_m$ Melting temperature

**HPLC** High-performance liquid chromatography

**SP** Sequence Pair

**LLPS** Liquid-liquid phase separation

**PAGE** Polyacrylamide Gel Electrophoresis

**PNAS** Proceedings of the National Academy of Sciences

**PCR** Polymerase chain reaction

**FRET** Förster Resonance Energy Transfer

**LHB** Late Heavy Bombardment

**NASA** National Aeronautics and Space Administration

**PTFE** Polytetrafluoroethylene

**mer** The length of the oligonucleotide

**TRIS** Tris(hydroxymethyl)aminomethan

**HEPES** 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

**CHES** N-cyclohexyl-2-aminoethanesulfonic acid

# Contents

# Introduction

The origin of life remains an unsolved puzzle for research. In order to specify when exactly life could have originated on our planet, it is important to understand its early geological history. Around 4.54 billion years ago Earth accreted from a proto-planetary disk [1]. There is evidence however, that only 200-800 million years later, around 4.52-4.42 billion years ago, a Mars sized planitesimal called "Theia" struck early Earth [2, 3]. In this highly energetic process, the surface was molten and any geological evidence from the time before erased. Debris from the collision later accreted to form the moon and accordingly, this event has been referred to as the "Moon-forming impact".

Still, at this point, Earth would remain uninhabitable for a long time. Studies show that the magma ocean could have cooled down and solidified around 4.4 billion years ago [4]. This time frame coincides with the oldest evidence of water, believed to be engraved in ancient zircons. At relatively low temperatures and when in contact with liquid water, the heavier isotope of oxygen ($^{18}O$) is partitioned preferentially into zircons compared to the lighter variant $^{16}O$. As zircons formed from magma preserve the $^{18}O/^{16}O$ ratio, water is estimated to already have been present on earth's surface as early as 4.4 billion years ago [5].

500 million years later, 3.9 billion years ago, a short yet intense bombardment of asteroids is hypothesized to have hit the Hadean Earth, possibly prolonging the period of time in which Earth was uninhabitable. The Apollo program brought back samples from moon craters, which were dated using $^{40}Ar/^{39}Ar$ isotopes, suggesting multiple 100 km sized asteroid impacts. As Earth's gravitational cross-section is around 20 times larger, a much higher impact rate is to be expected. This late heavy bombardment (LHB) is however disputed, as the Argon ratios from the Apollo samples may have been interpreted incorrectly [6]. Nevertheless, such an event would have made life on Earth almost impossible, setting the boundary of earliest life on earth to around 3.9 billion years ago.

Unfortunately for the research surrounding the emergence of life, the earliest evidence of life itself dates back only 3.7 billion years: Fossilized stromatolite-like structures, which are known to be formed by photosynthetic organisms making up microbial mats, were found at the Isua supracrustal belt in Greenland [7]. Additionally, graphite globules formed from sedimentary rock that had incorporated remains of organisms with depleted $^{13}C/^{12}C$ signatures suggest that cellular life had already developed at this point. This sets the time frame in which life originated on the Archean Earth to 3.7 - 3.9 billion years ago [8].

At this point, the atmosphere is suggested to be rich in $CO_2$, providing a strong greenhouse effect counterbalancing the around 30% weaker young sun to keep temperatures on earth high enough for liquid water to exist [9]. This hypothesis is primarily supported by the zircon record as mentioned above, showing that these rocks must have been in contact with liquid water at that time. Another possibility is the so called "Snowball Earth". As $CO_2$ is highly reactive with silicates [10], it would have been removed from the atmosphere and subducted into the crust after the solidification of Earth's mantle. If the $CO_2$ was removed too quickly from the atmosphere during this process, then the equilibrium temperature on Earth would be around -23°C, far below today's 14°C. However, since there are records of

the formation of zircons at this time, at least some part of the Earth must have been above freezing temperatures as these rocks form around 644-801 °C [11].

During the Archean (4.0-2.5 billion years ago) the planet is believed to be mostly covered in water, with volcanic islands of high activity. The atmosphere on this Archean Earth is currently agreed to have been devoid of $O_2$ and comprised of high levels of $CO_2$ and $N_2$ as well as some reducing gases such as $CH_4$ and $H_2$, providing the planet with a strong greenhouse effect [12].

In 1952, a milestone experiment conducted by Stanley Miller showed that through heat and electric discharges, similar atmospheric compounds can spontaneously form amino acids and other simple organic building blocks [13]. Future studies have shown a plethora of possible non-enzymatic pathways that can then build ribose sugars, nucleotides and small peptides from similar simple starting compounds [14–17]. Nucleic acids, such as RNA and DNA, are at the heart of modern biology and are used by every organism on earth. They are used to store the genetic information needed to make proteins, which are responsible for translating and replicating them.

This gave rise to the "chicken and egg paradox" of the origin of life, as both proteins and nucleic acids are codependent, making it unclear which of them might have developed first on early Earth. NASA defines life as a "self-sustaining chemical system capable of Darwinian evolution" [18]. With the discovery of functional RNA sequences, known as ribozymes, and their ability to store genetic information and replicate, RNA therefore meets the essential criteria for early chemical life. It is believed that replication cycles solely carried out by RNA were possible without any protein involvement. The resulting "RNA-World" hypothesis eliminates the "chicken and egg" problem between proteins and RNA. While at this point amino acids may have already existed as suggested by co-evolution theories [19] and supported by experiments such as Urey-Miller's, they might have acted only as modern co-factors and not played a central role in early replication cycles. Only later, it is believed that RNA adopted amino acids and peptides as their more efficient catalysts. Recent studies show for example that peptide synthesis can be performed directly on non-canonical RNA bases [20], laying the basis for the early tRNA analogues, modern biology's amino acid carrying adapter molecule [21].

In this hypothetical "RNA-World" framework, individual nucleotides, the building blocks of nucleic acids, polymerize into short oligomers through various activation steps [22]. Once formed, these sequences can pair with complementary counterparts via Watson-Crick base pairing [23]. This formation of double strand helices not only provides resistance to hydrolysis [24] but also enables replication reactions on a template strand through ligation or base-by-base polymerisation. Previous studies have demonstrated the feasibility of this process using prebiotic activation chemistries like 2',3'-phosphate [25] or imidazole [26]. Through these non-enzymatic condensation events small oligonucleotides then grew longer, thus carrying more information.

But the complexity of nucleic acids does not only depend upon their sequence length (number of bases, considered the primary structure) but also on the folds they form through base pairing (considered the secondary structure) [27]. These can be simple, such as hairpins loops or more complex such as pseudo-knots [28, 29]. Such conformations are a key aspect for the catalytic function of ribozymes, which are believed to have acted as RNA-replicators before proteins were deployed for this task. Through directed evolution, previous research already discovered a variety of different ribozymes, such as Sun Y [30], class I ligase [31] and even a ribozyme polymerase with a fidelity of over 90% [32].

The non-enzymatic replication of nucleic acids, however, faces many problems on early Earth. Firstly, reagents must be at sufficiently high concentrations for such reactions to work, requiring the environment to provide shelter from large diluting reservoirs such as the ocean. Nucleic acids also easily degrade in aqueous solutions, especially at high temperatures and salt concentrations [33], as well as through UV-irradiation [34]. Furthermore, after a strand has been replicated, the product strand has to be separated from the template again to initiate another round of copying. At high salt concentrations, however, this is a difficult task, as melting temperatures of nucleic acids can rise above the boiling point of water [35]. To circumvent the aforementioned challenges, molecular life was therefore dependent on a driving environment at constant thermodynamic non-equilibrium to sustain chemical reaction networks [36].

In this work, two different approaches to tackle these challenges are presented.

Firstly, nucleic acids possess the intrinsic ability to phase-separate into various dense phases such as liquid crystals, hydrogels, or coacervates. In Chapter 1 the phase separation of nucleic acids through the formation of double stranded networks will be demonstrated to act as a Darwinistic selection pressure, by sheltering them from threats like dilution. In a mimic of a prebiotic rock pore environment, short oligomers (<30nt) are shown to form a dense phase in a sequence selective manner, accumulating at the bottom of the rock pore through sedimentation.

In chapter 2, an abundant and prebiotically plausible environment is shown to support nucleic acid replication. Previous studies have demonstrated how temperature gradients on the microscale offer numerous advantages for nucleic acid replication and accumulation. Their abundance on early Earth is however limited. Thus finding environments that do not depend on steep temperature gradients is of great importance for widening the range of possible early life habitats. The scenario at hand uses a mere gas flux across a water-filled rocky pore as the energy input, resulting in a micro-environment that enables dissolved nucleic acids to accumulate and periodically strand separate through salt oscillations.

# 1 Sequence self-selection by cyclic phase separation

## 1.1 Summary

The emergence of functional oligonucleotides on early Earth required a molecular selection mechanism to screen for specific sequences with prebiotic functions. Cyclic processes such as daily temperature oscillations were ubiquitous in this environment and could trigger oligonucleotide phase separation. Here, sequence selection based on phase separation cycles is proposed, realized through sedimentation in a system subjected to the feeding of oligonucleotides. Using theory and experiments with DNA, sequence-specific enrichment in the sedimented dense phase is shown, in particular of short 22-mer DNA sequences. The underlying mechanism selects for complementarity, as it enriches sequences that tightly interact in the dense phase through base-pairing. This process is demonstrated to be highly sequence specific, robust to buffer condition changes and follows predictions from finite elements simulations. This mechanism also enables initially weakly biased pools to enhance their sequence bias or to replace the previously most abundant sequences as the cycles progress. When implemented in a realistic flow-through scenario in a porous rock mimick, specific sequences sheltering and up-concentrating against a diluting outflux through sedimentation were observed. Our findings provide an example of a selection mechanism that may have eased screening for the first auto-catalytic self-replicating oligonucleotides.

This chapter was published by Barlotucci, Serrão and Schwintek et. al [37] in PNAS and is here adapted and reprinted in part with permission from PNAS. It was based on a collaborative project with Giacomo Bartolucci and Christoph A. Weber. Full article attached in the List of Publications.

## 1.2 Motivation

Selection pressures drive evolution on early chemical systems. Describing the influence of selection pressures on prebiotic mixtures is a central area of research in the origins of life field. Here, liquid liquid phase separation (LLPS) of nucleic acids is shown acting as a Darwinistic selection mechanism to shelter oligonucleotides from various threats of the environment such as dilution.

LLPS is a process in which a homogenous liquid transitions from one isotropic liquid-like state into at least two others. A liquid may even have more than two such states and these can coexist simultaneously [38]. Interestingly, nucleic acid mixtures are capable of LLPS, separating into a dense and dilute phase through different mechanisms, such as hydrogen bonding or electrostatic interactions. It has been shown that they can form a variety of liquid-like states.

A prominent example are coacervates. These condensed liquid-like droplets, usually formed by the partitioning of oppositely charged polymeric molecules, which interact with each other through electrostatic interactions stronger than the solvation interactions with the solvent. These dynamics lead to the formation of a dense phase, the coacervates, mostly constituted of the charged polymer and of a dilute phase, mostly constituted of the solvent. Coacervates are an important example of membraneless organelles, and are a form of dense phase also found inside living cells in modern biology [39]. The conditions and composition of the dense phase inside the coacervate can vary substantially from the surrounding, preventing specific types of molecules from entering, while others are able to permeate the phase boundary [40]. This enables chemical living systems to produce micro-environments in which for example enzymatic activity is favored [41]. Since nucleic acids, such as DNA and RNA, are capable of forming coacervates [42–44], this mechanism presents a great model for intracellular, as well as prebiotic compartmentalization.

Furthermore, nucleic acids can self-assemble into liquid crystals [45]. Liquid crystals present anisotropy of certain properties without a three-dimensional crystal lattice, i.e. have ordered structure like a crystal in two dimensions and flow like a liquid in the third dimension. At concentrations surpassing a certain threshold (typically above 100 mg/ml or in the mM regime), nucleic acids form a variety of different liquid-crystal dense phases, triggered by stacking and base pairing interactions [46]. At these high concentrations, the formation of liquid crystals could have played an important role in protecting information carrying polymers over long periods of time from environmental stress, such as radiation or starvation of building blocks [47]. Even in modern biology, *E. coli* cells, when starved of any food source, show spontaneous DNA packing through reversible liquid crystal phase separation in order to protect its genome from degradation [48]. It also was found that from a pool of short random oligomers (<16nts) liquid crystal formation was promoting complementarity among those sequences during crystallization, generating conditions for prebiotic ligation reactions [49]. At higher lengths however, random strands form liquid crystals without any sequence sepcificity. Concentrations of DNA/RNA in the mM-regime are difficult to achieve in a prebiotic setting on early earth. Therefore phase transitions that do not require such high concentrations are of great importance for the origin of life research.

Another liquid-like phase that nucleic acids can transitions to are so called hydrogels. Hydrogels are characterized by a porous three-dimensional structure which still retains a high percentage of the solvent (10%) as interstitial fluid [50]. This non-equilibrium, kinetically trapped state, is the consequence of nucleic acids being able to hybridize in a sequence

specific manner [51]. Through hydrogen bonding of the individual bases along a DNA/RNA strand or crosslinker molecules, such as ethylene glycol, these polymers form networks swollen in aqueous solution [52]. As these networks grow sufficiently large, they partition into a dense phase. This phase separation is thermo-reversible and highly dependent on local salt concentration and pH [53], because base-pairing of two nucleic acid strands requires positively charged ions to counterbalance the repulsion of the negatively charged phosphate backbones. Hydrogels even allow for enzymes to penetrate the phase boundary and perform catalytic function within and at the surface of the gel, depending on the respective melting temperature of the gel [54]. Recent studies show that they are efficient for cell-free protein expression when the respective genes are part of the gel scaffold and therefore up-concentrated and readily available [55]. Futhermore, hydrogels can act as a stabilizing cytoskeleton for lipid membranes, significantly increasing their interfacial tension, elastic modulus and shear resistance [56].

Since hydrogels are large networks of nucleic acid secondary structure, the selection of sequence motifs that form such networks would represent an important step in nucleic acid evolution. Notably, secondary structures, such as hairpins, are at the catalytic core of many ribozymes [57]. Gelation of DNA/RNA is also highly sequence specific, even separating single base differences [58]. Furthermore, complementary nucleic acid strands are bound by strong interactions of about 5 $k_bT$ per base, resulting in a low saturation concentration for phase separation to occur [59]. This provides a way to form a concentrated dense phase of highly specific oligonucleotide sequences.

The above mentioned LLPS mechanisms of nucleic acids provide stabilizing and catalytic properties for cells, protective advantages for chemical systems through compartmentalization, as well as up-concentration of nucleic acids. In the context of origin of life research, this becomes especially important once the mechanism is sequence selective through base-pairing dependence. If harsh environmental conditions threaten the survival of chemical systems, then the ability to shelter from these dangers through phase separation would present a Darwinistic advantage. Sequences that are able to undergo LLPS then would possess the required fitness to outcompete others.

In a prebiotic environment, such as an under-water rocky pore [60–62], phase separation can act as a selection mechanism for oligonucleotides. Figure 1.1(a) shows a schematic of such an environment. An under-water porous rock, for example near hydrothermal vent systems, can be subjected to a flux of bio-molecules, such as nucleic acids [63, 64]. Once these molecules reach the surrounding ocean, their concentration would drastically decrease to a point where no chemical reactions are possible any longer. Therefore, a mechanism that can retain chemical systems inside the rock cracks is of major importance for their survival.

As nucleic acids phase separate, the respective dense phase precipitates out of the surrounding solution and sediments under gravity. Figure 1.1(a) shows how this effect can select sequences capable of LLPS from an influx of many oligonucleotide sequences. Without phase separation, the concentration of any sequence inside the displayed pore would be equal to the pool composition of the influx. However, when a specific sequence is prone to forming large networks that sediment towards the bottom of the pore, then its concentration inside the pore increases the more sequences are recruited into the dense phase. Figure 1.1(b) shows how the initial state composition inside the pore of many different sequences (here shown in different colors) changes upon the formation of a dense and a dilute phase within equilibration time "$t_{eq}$". The dense phase then sediments towards the bottom, while the dilute phase rests above. Through a continuous or discrete flux across the pore, the dilute phase is then replaced with

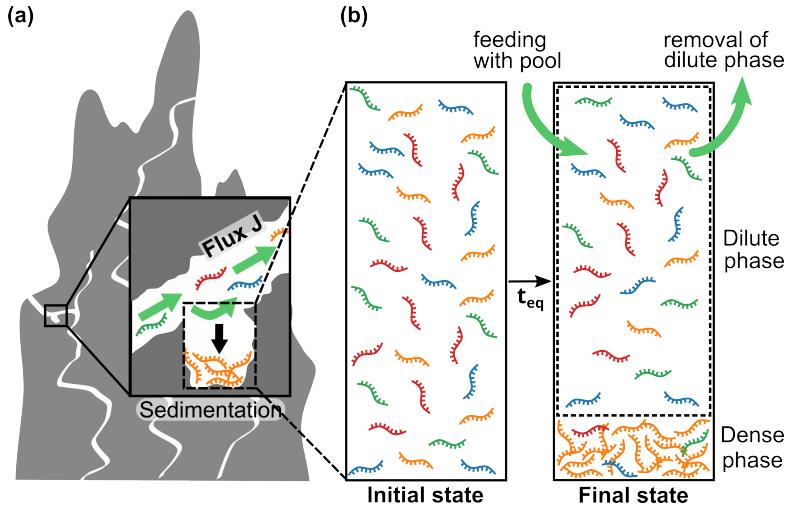


**Figure 1.1:** *Schematic of the proposed principle of sequence selection.* **(a)** Sketch of a prebiotic rock through which a flux of oligonucleotides can protrude with flux J. Sequences capable of phase separation sediment to the bottom of pores along the rock cracks, while others are flushed out. **(b)** Simplified view of the sequences inside a rock pore. Initially, the sequences are homogeneously distributed. During equilibration time $t_{eq}$, sequences are able to sediment to the bottom of the pore, forming a dense phase. Other sequences stay in the dilute phase, which is continuously replaced by the composition of the initial state, represented here by the feeding with pool.

the composition of the influx (pool). Over time, sequences that cannot phase separate would rest at pool concentration, while sequences capable of LLPS up-concentrate at the bottom of the pore. In the following work, this process is shown to efficiently select for specific sequences.

## 1.3 Results and Discussion

### 1.3.1 Sequence Design and Folding

To experimentally confirm and demonstrate phase separation of oligonucleotides as a selection mechanism requires the design of a sequence pool to start with. The sequence space for random oligonucleotides of length $L$ with $m$ different nucleotides is proportional to $m^L$, so in the case of DNA/RNA: $4^L$. To our knowledge, the shortest nucleic acid sequences for which phase separation into condensates has been observed so far, are 20 - 25 nt in length [65]. While this length regime is a likely scenario for a prebiotic soup sequence pool, the amount of possible sequences is $4^{20} \approx 10^{11}$. Studying a pool this large is not feasible in an experimental setting. For this reason a pool of a few selected sequences to showcase the selection mechanism in a model system was designed. Rather than using strands which are not expected to form large network structures through base-pairing by the nature of their sequence composition, sequences were chosen, which should by design be able to phase separate. Here, the aim was to use a honeycomb like structure as the desired network.

Several experimental systems were designed with DNA as a model oligomer. In contrast to RNA, DNA is less prone to hydrolysis and therefore more stable. Also, the selection mechanism of phase separation through hydrogel formation relies solely on base-pairing, which is very

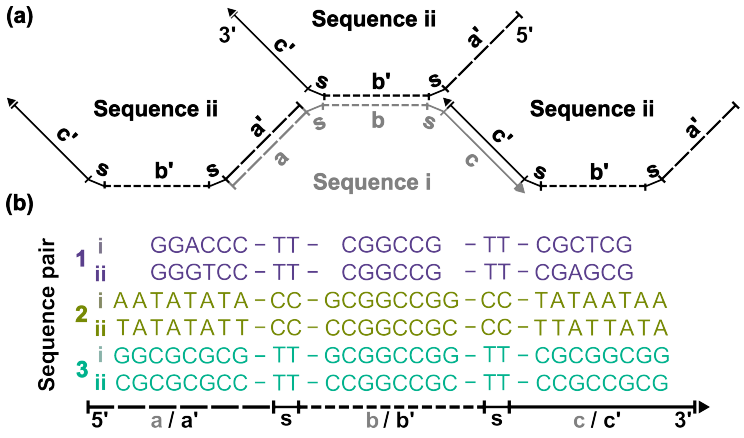


**Figure 1.2:** *Sequence design.* **(a)** Sequences were designed in pairs. Sequence i is composed of three segments a, b and c, separated by spacers s. Its corresponding partner sequence ii, contains the reverse complements of each segment a', b' and c' in reversed order. This prevents the formation of a complete double strand and enforces binding in a honeycomb structure offering multiple binding sites for more sequences to bind. **(b)** Composition of all 3 sequence pairs designed.

similar for both [66]. Each system was comprised of a DNA sequence pair (SP). Figure 1.2 shows the three sequence pairs. Each SP consists of strand i comprised of 3 binding segments a, b and c and strand ii with segments a', b' and c' respectively. These regions are individually reverse-complementary, namely a' is the reverse complement of a. Importantly, sequence i is not the reverse complement of sequence ii. By keeping the same order of segments (5'- a, b and c -3') in the partner sequence ii (5'- a', b' and c' -3'), the two strands cannot form a complete double strand structure. This allows each strand i to bind to 3 strands ii and vice versa (See Figure 1.2(a)). The honeycomb-like binding scheme enables cooperative binding. For every sequence ii bound to the first i, one more binding site will be available. In particular, the binding sites available are $b = 3 + n$, with $n$ representing the amount of sequences bound to the first one. Thus, the more sequences are bound, the faster additional sequences can find available binding sites. To ensure the strands to be flexible enough through minimizing angular constraints, the segments are separated by dimeric spacer segments s, either composed of TT or CC, chosen such that they do not interact with the other segments [51]. While the palindromic center segments b and b' were kept GC-only, the outer "arm"-segments a, a', c, and c' contained different base-compositions from SP to SP. Figure 1.2(b) shows the exact sequence of each SP. SPs 2 and 3 are 28 nt in length and are made up solely by AT or GC, respectively. In contrast, SP 1 contains only 22 nt. However in its arms, all four nucleobases ATGC were used.

In order to find matching sequence pairs, the NUPACK design tool 3.2.2 [67] (For a detailed code see Appendix Section 1.A, Figure 1.19) was utilized. After mutating the output sequences multiple times, the online NUPACK analysis tool was used to check how well the sequence pairs form the desired honeycomb network. Figure 1.3 shows the folding predictions with the segments indicated respectively. All three SPs in theory show the ability to form the desired structure, in which the binding follows the schematic from Figure 1.2(a).

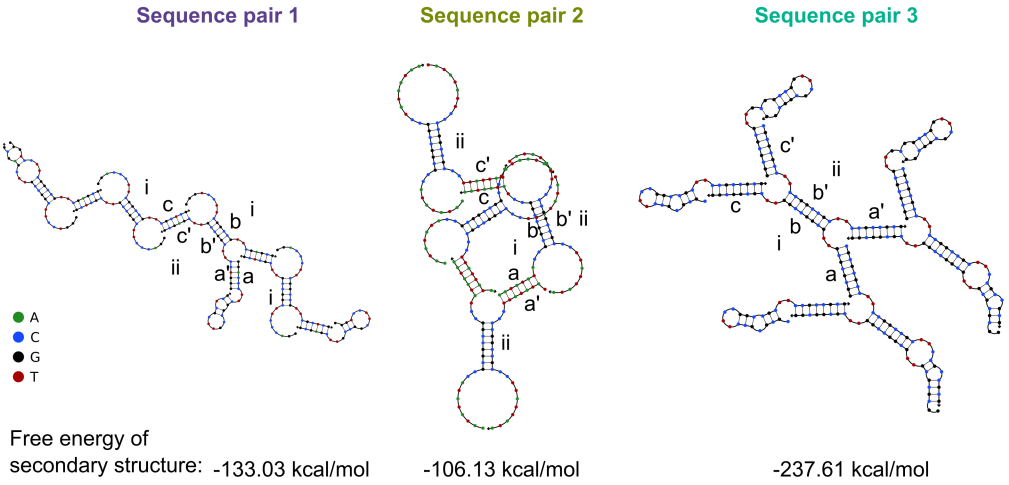


**Figure 1.3:** *NUPACK folding prediction.* Using the NUPACK analysis tool, based on a binding energy data set, each sequence pair was analysed for folding. Settings were 15 °C, 125 mM NaCl, 10 mM $MgCl_2$ and 2 µM per strand. The secondary structure allowed up to 10 strands. Labels a, b, c, etc. correspond to the sequence segments shown in Figure1.2. Each sequence pair shows the desired secondary structure, allowing for 1 additional sequence to bind for every sequence already attached.

### 1.3.2 Thermal Melting Curves

As hydrogel formation is thermo-reversible, determining the melting temperature for each SP is key to design experimental procedures. A melting curve describes the hybridization behaviour of nucleic acids relative to temperature. The characteristic melting temperature $T_m$ of a strand denotes the temperature at which half of the strands are in duplex (dsDNA or dsRNA) form. Below that temperature, more than half of the strands are bound to their respective counterpart, while above it, most are in their single stranded form (ssDNA or ssRNA). The melting temperature of a strand is dependent on the length of the oligomer, as well as its GC content. In general , the longer the sequence and the higher the GC content, the higher the $T_m$ tends to be. Usually, UV-absorbance spectroscopy at 260 nm, fluorescence measurements with intercalating dyes or attached fluorophores are used to determine $T_m$ [68]. For the melting curves performed in this work, the intercalating dye SYBR Green I was used [69], which exhibits a >1000-fold brighter fluorescence response when bound to a duplex nucleic acid structure compared to a single stranded one. Figure 1.4, shows the thermal melting curves performed in a thermocycler. The sample composition was 25 µM of each DNA strand, 5x Sybr Green I, 10 mM TRIS Buffer pH 7, 125 mM NaCl and 10 mM MgCl2. For more details on the melting curve measurements, see Materials and Methods Section 1.5.4. By fitting the data of independent triplicates, the melting temperature was found to be 57°C for SP 1, 71°C for SP 2, and 65°C for SP 3. As expected, the melting temperature of SP 1 is the lowest of the three, as it contains the shortest sequences of only 22 nt, compared to the 28 nt of SP 2 and 3.

### 1.3.3 Sedimentation Behaviour

While the SPs were designed to form networks and therefore phase separate, it is important to verify experimentally whether a dense phase is observable and for which experimental

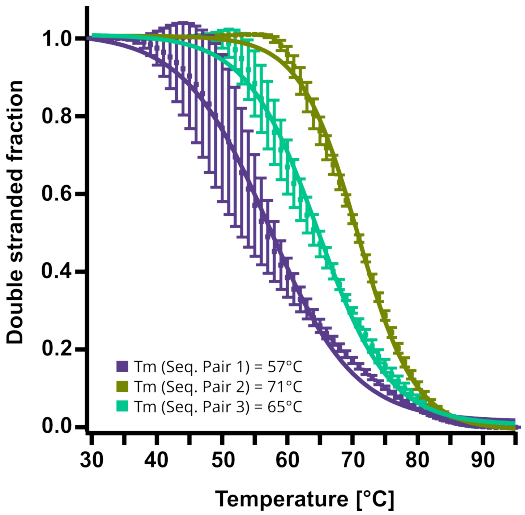


**Figure 1.4:** *Melting curves of each sequence pair.* Conditions used were 25 µM of each DNA strand, 10 mM TRIS Buffer pH 7, SYBR Green I 5X, 125 mM NaCl and 10 mM $MgCl_2$. Data points are the average of the each triplicate respectively. Error bars correspond the standard deviation of each data point calculated from independent triplicates. Points were fitted using a sigmoidal function, with the melting temperature as a fitting parameter.

conditions. For this purpose, time lapse fluorescence microscopy was used in combination with the intercalating dye SYBR Green I to observe the phase separation process in situ. A self-built fluorescence microscopy setup was used in order to be able to simultaneously observe multiple samples allowing efficient screening of parameters (See Materials and Methods Figure 1.15). The samples were observed inside a 500 µm thin microfluidic chamber (See Materials and methods Figure 1.16). The temperature was controlled using a Peltier element. As the DNA networks grow larger in size, and eventually large enough to become visible by fluorescence microscopy, they precipitate out of solution and subsequently sediment towards the bottom of the microfluidic chamber. Figure 1.5(a) shows micrographs of the sedimentation experiments. Each sequence pair was observed in a separate chamber, with the conditions fixed at 25 µM of each DNA strand, 5x Sybr Green I, 10 mM TRIS Buffer pH 7, 125 mM NaCl and 10 mM $MgCl_2$ (A screening of other buffer conditions can be found in Section 1.3.5). To ensure homogeneous starting conditions for each experiment, the samples were first heated to 65 °C. Usually, heating to 95 °C is preferred, as it ensures complete melting of all strands. However, in this chamber, this lead to strong evaporation, altering the initial conditions of the experiment. Also, the thermal melting curves in Figure 1.4 show that at around 65 °C most strands are in their ssDNA form. The samples were then slowly cooled down to 15 °C at a rate of 6 K/min, where the temperature was then held constant for at least 3 h of incubation. This temperature is well below $T_m$ for all SPs, resulting in most sequences to be in dsDNA form. Choosing a higher incubation temperature leads to less amounts of dsDNA, resulting in smaller hydrogel aggregates (See Section 1.8).

In Figure 1.5(a), micrographs of 3 time points for SP 1 are shown. "0 h" corresponds to the moment at which the final temperature of 15 °C has been reached, all subsequent time points are relative to this time. At first, the strands are homogeneously distributed in the bulk of the chamber, without any condensates visible. 10 minutes after the incubation temperature

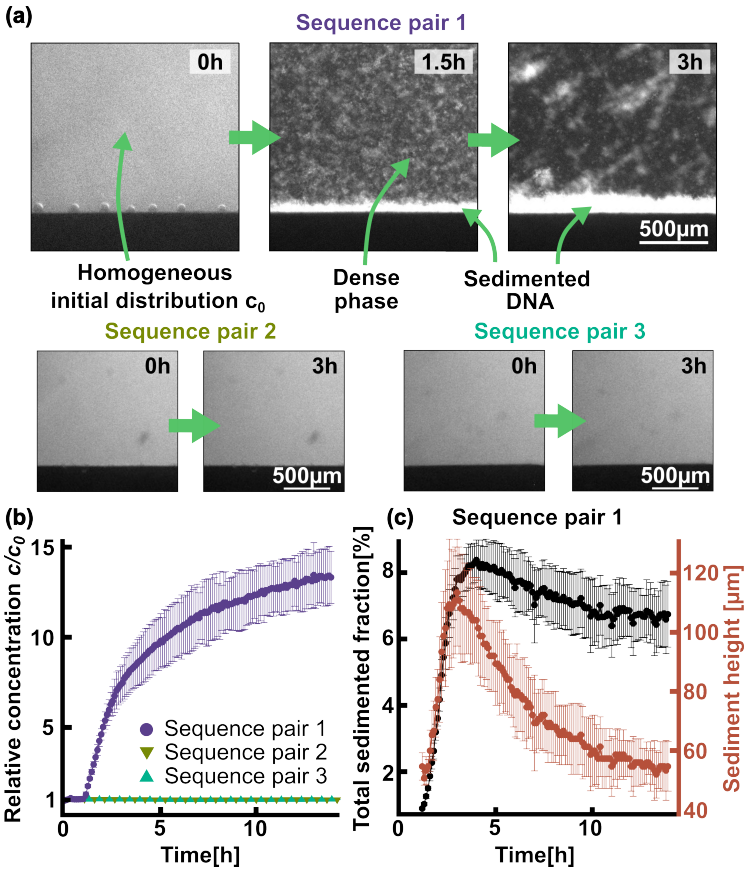


**Figure 1.5:** *Sedimentation behaviour of each sequence pair.* **(a)** Fluorescence micrographs of the individual sequence pairs at different time points. Images were taken in a 500 µm thick microfluidic chamber with 25 µM of each strand, 10 mM TRIS-HCL pH 7, 10 mM $MgCl_2$, 125 mM NaCl and 5 X SYBR Green I. Samples were cooled from 65 °C to 15 °C and monitored through fluorescence timelapse imaging. After cooling, sequence pair 1 forms a dense phase in the form of aggregates, which sediment towards the bottom of the chamber. Sequence pair 2 and 3 do not show this behaviour. **(b)** Concentration over time recorded for all sequences at the bottom of the chamber. Sequence pair 1 shows up 13-fold up-concentration compared to the initial state, while 2 and 3 remain at the pool concentration $c_0$. **(c)** SYBR Green I fluorescence is recorded to analyse the sediment formation. The total sedimented fraction of DNA plateaus after 5 h at 6 % to 8 %, while the sediment height starts to contract after 4h from 110 µm down to 50 µm. Data points are the average of independent triplicates and error bars their standard deviation.

was reached however, SP 1 starts to form DNA condensates, growing in size and sedimenting towards the bottom of the chamber. This behavior could not be observed for SP 2 and SP 3. As both of these sequence pairs are made up of a binary alphabet (AT or GC), non-specific binding becomes likely. Subsequently, alternative, non desired, secondary structures are formed which can lead to locked state. Hairpins, internal loops or G-quadruplexes then hinders interactions with other strands, preventing the hydrogel formation [70]. After about 1.5 h, SP 1 formed aggregates of a few µm in size. As the condensates sedimented to the bottom the chamber, the data from the fluorescence micrographs was used to scrutinize the accumulated concentration and gel properties using a self-written Labview program (See

Materials and Methods Figure 1.14). Figure 1.5(b) shows the concentration at the bottom of the pore, relative to the initial concentration $c_0$. Averaging independent triplicates over time reveals an increase of relative concentration for SP 1 up to 13-fold. In contrast, SP 2 and SP 3 do rest at the initial concentration of 25 µM.

After analysing the fluorescence data for SP 1 in detail, the growth of the sediment as well as the total fraction of DNA that sedimented could be observed. Figure 1.5(c) illustrates how both of these parameters vary over the course of the experiment. After 3 h of experiment, the total sedimented fraction saturates at around 8%, while the sediment height reaches its maximum of 100 µm. Afterwards, the gel contracts slowly down to around 50 µm. This behaviour has been observed in literature and results from the dynamic rearrangement of the arms of the sequence pairs [51]. While the kinetically trapped hydrogel first forms, not all arms of the strands are hybridized to their corresponding partner. This constrains the gel and results in unfavorable conformations, which can be resolved through rearrangement. This results in the contraction of the hydrogel over time, until the optimal conformation has been reached. This is well visible in Figure 1.5(c) (red data points): after all DNA has sedimented, the gel slowly approaches optimal conformation, decreasing to about half the initial size.

### 1.3.4 Sequence Dependence

To further investigate the sequence dependence of the phase separation through hydrogel formation, the composition of SP 1 was varied to build 5 similar SPs of equal length. In four of them, a binary GC-only alphabet was used (except for the spacers), while in the fifth, the full 4-letter alphabet was again used in the arms (a and c). Figure 1.6(a) and (b) show the sequences used. SP 4 utilizes the full alphabet (A/T highlighted in red). Figure 1.6(c) displays the micrographs of the sedimentation experiment for each individual SP after 7h. Sample conditions were kept equal to the experiments in Figure 1.5, however with increased NaCl content of 500 mM to ensure that the lack of salt could not be the reason for no hydrogel formation.

No phase separation was observed for any of the cases, except for SP 4. In agreement with our previous measurements, the SP with all four bases used in the arms was capable of phase separation. Even though SP 5 is the exact replica of SP 4, with the red highlighted A and T replaced by C and G, no phase separation could be observed. Even after multiple changes of the binary sequences (SP 6-8) or removal of the spacers (SP 7), no sedimentation took place. These results underline the sequence specificity and selectivity of the phase separation process.

### 1.3.5 Influence of pH and Salts

After finding out that the system is highly sequence specific, understanding how other parameters, such as ion concentrations and pH, influence the sedimentation process, became of high interest. To obtain this information, a screening of pH and salt concentrations was performed using SP 1 only. By fixing the buffer conditions to 10 mM TRIS pH 7 first, the NaCl and $MgCl_2$ concentrations were varied between 0-250 mM and 0-10 mM, respectively. Figure 1.7(a) shows micrographs of each experiment. After cooling from 65 °C to 15 °C with subsequent incubation at 15 °C, sedimentation could not be observed in the cases with 0 mM MgCl and NaCl lower than 250 mM. This shows the importance of $Mg^{2+}$ for LLPS in this system. At high levels of $Mg^{2+}$, however, the stability of non-canonical base pairs, such as G-G or A-A

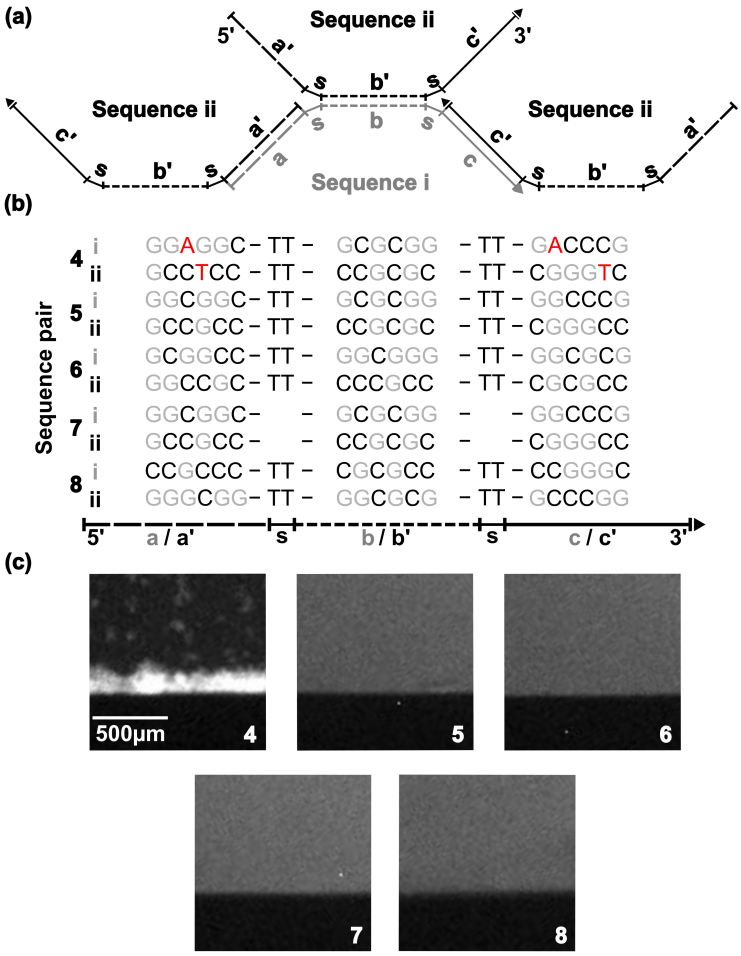


**Figure 1.6:** *Sequence dependence of phase separation.* **(a)** Scheme of sequence design as displayed in Figure 1.2. SP i is comprised of three segments a, b and c, while SP ii is made up of their reverse complements a', b' and c'. Their order is however reversed, such that the formation of a complete double strand is prohibited and the honeycomb binding structure is forced. **(b)** List of investigated SPs. Note that all three Sections are made up GC only, while the spacers s are TT. Only sequence pair 4 contains A/T in sections a/a' and c/c' (highlighted in red). **(c)** Fluorescence micrographs of each individual sequence pair at the bottom of the sedimentation chamber after 7h. After cooling from 60°C to 10°C the samples were given 7h to sediment. Sample composition: 10 mM TRIS pH7, 10 mM $MgCl_2$, 500 mM NaCl, 50 µM of each DNA strand and 5X SYBR Green I. Only SP 4 shows phase separation and subsequent sedimentation.

for example, is strongly influenced, which could lead to different secondary structures than predicted [71]. In the scope of this work, higher concentrations of $Mg^{2+}$ were not investigated, could however show interesting consequences on the sequence specificity of the system. To examine the pH dependence of the system, the salt concentrations were fixed at the buffer concentration used in all previous experiments: 10 mM $MgCl_2$ and 125 mM NaCl. The pH was then changed to 5.5 using MES, 7 using HEPES and 9 using CHES (Figure 1.7(b)). These buffering agents were chosen to ensure the respective buffering capacity at the desired pH. For pH 7 and 9, phase separation triggered sedimentation as before. However at pH 5.5, no

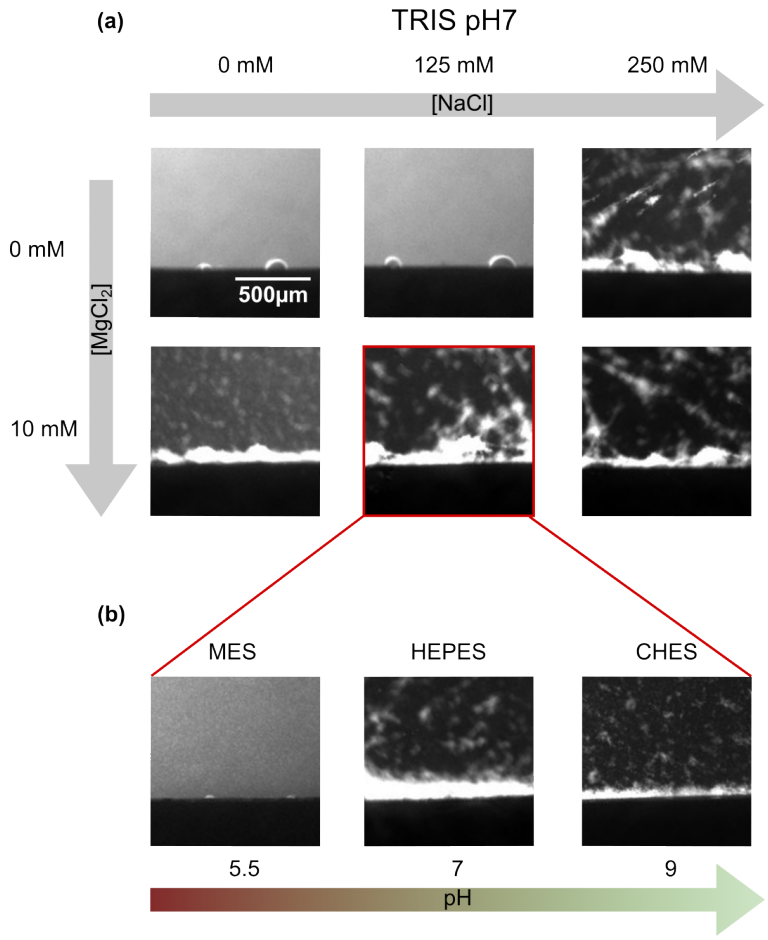


**Figure 1.7:** *pH and salt screening of the sedimentation behaviour of sequence pair 1.* **(a)** Salt screening at pH 7 (10 mM TRIS-HCL), 25 µM of each DNA strand, 5 X SYBR Green I and varying $MgCl_2$ and NaCl concentrations. Micrographs were taken at the bottom of the chamber 3h after cooling from 65°C to 15°C. Sedimentation could not be observed at NaCl concentrations below 250 mM combined with 0 mM $MgCl_2$. **(b)** Keeping the salt concentration at 125 mM NaCl and 10 mM $MgCl_2$, the pH was varied from 5.5 to 9. Sedimentation could only be observed for pH 7 and 9.

DNA aggregates were formed. This was to be expected, as base-pairing of DNA is hindered in acidic environments [66].

### 1.3.6 Sedimentation Dynamics

After pinpointing the chemical restrictions on hydrogel formation and sedimentation via phase separation, the question remains, whether the environment is posing geometric restrictions on this phenomenon. In particular, can sedimentation be observed in rock pores of any size, or is there a minimum size required to observe this effect? To answer this question, the sinking DNA aggregates were approximated as sinking spheres to show that there is indeed a minimum chamber height for this effect to be non-negligible, by applying Boltzmann statistics. Assuming the aggregates to behave like spherical particles with small Reynolds numbers

in a viscous fluid, subject to a drag force. Stokes' law for the frictional force is given by:

$$F_d = 6\pi\mu R v\,, \tag{1.1}$$

It describes the friction exerted on a spherical particle of radius $R$ sinking at speed $v$ with a surrounding liquid of dynamic viscosity $\mu$. The gravitational force dragging the sphere downwards is proportional to the density difference to its surrounding liquid:

$$F_g = \Delta\rho \frac{4}{3} g\pi R^3\,, \tag{1.2}$$

$$\text{with} \quad \Delta\rho = \rho_p - \rho_f\,. \tag{1.3}$$

Where $\rho_p$ and $\rho_f$ describe the density of the particle and the surrounding fluid, respectively. With this, the terminal velocity of the sinking sphere can be calculated by setting $F_d = F_g$:

$$v = \frac{2\Delta\rho g R^2}{9\mu}\,, \tag{1.4}$$

$$\text{and rearranged for } \Delta\rho\text{:} \quad \Delta\rho = \frac{9v\mu}{2gR^2}\,. \tag{1.5}$$

Since the velocity is quadratically dependent on the particle's radius, it is the most important parameter to measure experimentally. However, the maximum resolution of our fluorescence microscope was 4 µm/pixel, setting the lower end for the smallest observable aggregates. Aggregates below 4 µm in radius could not be measured.

In order to obtain statistics about the velocity distribution of sinking DNA aggregates, a LabView script that tracks the sinking particles across the time lapse fluorescence microscopy images was written. Figure 1.8 shows the measured particle speed distribution. This experiment required us to change the final incubation temperature from 15 °C to the room temperature of 25 °C. The temperature difference of 10 °C from chamber to surrounding air would have otherwise led to convective flows, significantly changing the flow speed distribution. Other parameters were kept equal to the previous experiments (125 mM NaCl, 10 mM $MgCl_2$, 5X Sybr Green I, 25 µM of SP 1, 10 mM TRIS pH 7). The increased temperature however, lead to particles not growing past 4-16 µm in diameter.

Tracing particles between 50 and 100 min after the temperature spike of 65 °C revealed an average sinking speed of 0.73 µm/s, while the highest observed speeds are around 2.5 µm/s. As reliable measurements of precise particle sizes are limited due to the resolution of our camera, a constant particle radius of 3 µm was assumed for further calculations.

As the Boltzmann energies at room temperature of the particles scale with the density difference to surrounding water, equation (1.5) was used to deduce the relative density of the sinking spheres, using measured speeds. The dynamic viscosity of 0.9 mPas for water at 25 °C was taken from Ref. [72]. The turquoise axis in Figure 1.8 shows the effective density ($\rho_p/\rho_f$) for all measured flow speeds. On average, particles are around 1.03 X denser than the surrounding water.

With the calculated parameters in place, it is now possible to use the Boltzmann distribution to calculate the geometric restrictions on the system. By comparing the gravitational potential energy of sinking particles to thermal fluctuations, the minimum chamber height, or characteristic sedimentation length $l$, required for sedimentation to be observable, can

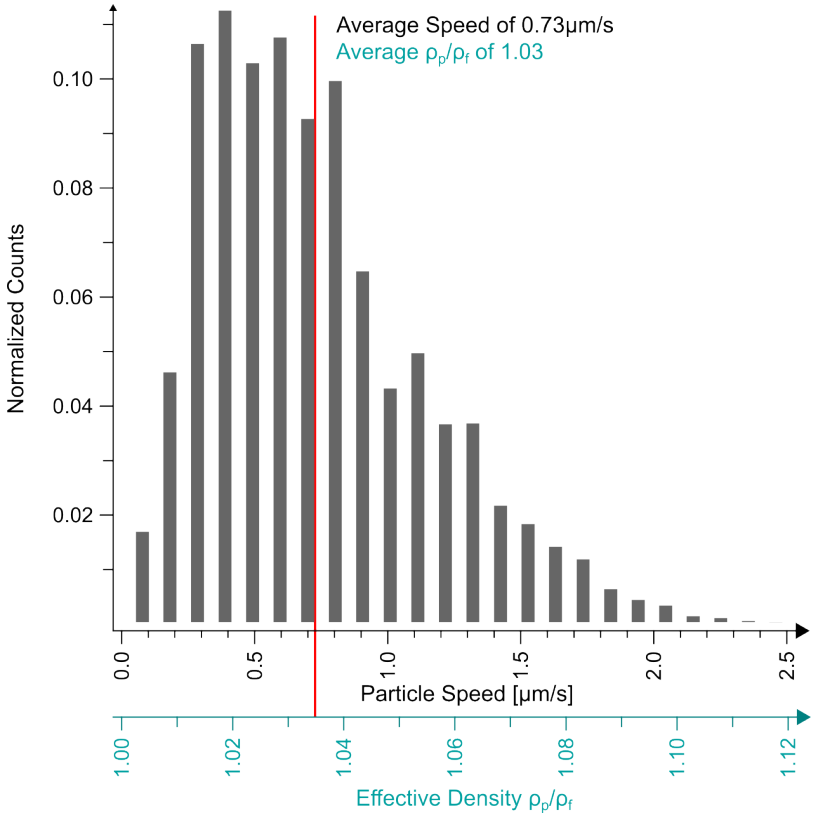


**Figure 1.8:** *Statistics of the sedimentation dynamics.* A self written particle tracker LabView program was used to measure the speed of sinking DNA aggregates. Histogram displays the particle speeds as well as their effective density ratio compared to water (turqoise axis). The vertical red line shows the average speed and density. This time the temperature was cooled from 65 °C to 25 °C to avoid convection fluxes. Sample conditions: 125 mM NaCl, 10 mM MgCl, 5 X SYBR Green I, 10 mM TRIS pH 7 and 25 µM of each DNA strand.

be calculated. While the thermal fluctuations are solely dependent on the temperature, the potential energy of sinking particles scales with volume $V$ and relative density:

$$p \propto \exp\left\{\frac{\Delta\rho \cdot V \cdot g \cdot h}{k_B T}\right\} \tag{1.6}$$

$$\propto \exp\left\{h \cdot \frac{1}{l}\right\}, \tag{1.7}$$

$$\tag{1.8}$$

with $l$ denoting the the sedimentation length:

$$l = \frac{k_B T}{\Delta\rho \cdot V \cdot g}\,. \tag{1.9}$$

Inserting the average parameters obtained from the particle tracking ($\Delta\rho = 1.03\rho_f = \mathbf{1.03} \cdot 997\frac{kg}{m^3}$ and $V=\frac{4}{3}\pi * R^3$ with $R=3\mu m$), one obtains a characteristic sedimentation length of

$$l = \frac{k_B T}{\Delta\rho \cdot V \cdot g} \tag{1.10}$$

$$= \frac{k_B \cdot 298.15K}{1.03\rho_f \cdot \frac{4}{3}\pi \cdot (3\mu m)^3 \cdot g} \tag{1.11}$$

$$= 12.5\mu m\,. \tag{1.12}$$

This means that sedimentation is only observable in a chamber higher than $l = 12.5\ \mu m$. If the chamber is smaller, then thermal fluctuations outweigh the gravitational effects of the density difference and sedimentation would be a negligible effect. For the plausibility of the selection of sequences via sedimentation, prebiotic rock cracks are assumed to be in the order of millimeters, at minimum. The chamber height used in our experiments was about 10 mm, exceeding $l$ by multiple orders of magnitude. Here, sedimentation plays a key role. Even if the relative density of the aggregates was 10 to 1000 times smaller, then our chamber height would not exceed the characteristic sedimentation length $l$.

The same logic allows us to estimate the minimal aggregate radius for sedimentation to occur in this geometry. As $l$ depends cubically on the radius $R$, a 10-fold decrease in aggregate radius results in a 1000-fold larger sedimenation length of around 10mm. This means in the experimental setup, sedimentation is only observable once DNA hydrogel aggregates have grown larger than 0.3 µm in radius.

### 1.3.7 Discrete Cycles of Phase Separation

Having understood the underlying properties of LLPS in the case of hydrogel formation and subsequent sedimentation, one can can go back to the initial hypothesis of sequence selection via cyclic phase separation. Based on the high sequence specificity of phase separation via hydrogel formation presented in Section 1.3.4, an experimental model to test its capabilities as a selection mechanism was developed. Orienting closely to the scheme portrayed in Figure 1.1, the continuous replacement of the dilute phase with the pool in a rocky pore environment was approximated with discrete re-feeding steps. As fluxes vary due to pressure changes, day-night cycles, etc., this seemed appropriate and facilitates the experimental procedure, while focusing on the key aspects of the system. Figure 1.9 shows depiction of the experimental scheme. Starting from the initial pool containing two competing sequence pairs and buffer (**0.**), the sample is slowly cooled down to 15 °C to allow for strand annealing and network formation. This leads to phase separation and subsequent sedimentation. The sequence pair that can perform LLPS accumulates at the bottom, while the others stay at initial concentration. A sediment is formed at the bottom of the tube (**1.**). Because removing the complete dilute phase above the sediment is not feasible in an experimental environment, the tube was divided into a top and bottom fraction, accounting for 50 % of the total volume each. One can then remove the top fraction (**2.**) via pipetting and replace it with the initial pool (**3.**). The composition of the extracted top fraction can be directly analysed via HPLC (See Section 1.5.3). The sediment in the bottom fraction, however, can only be analysed after all cycles have already completed. Through indirect mass-balance calculation, however, the concentration of DNA at the bottom fraction of the vial can be inferred (See Figure 1.10 By heating to 95 °C, a homogeneous mixture (**4.**) similar to the initial pool starting point is

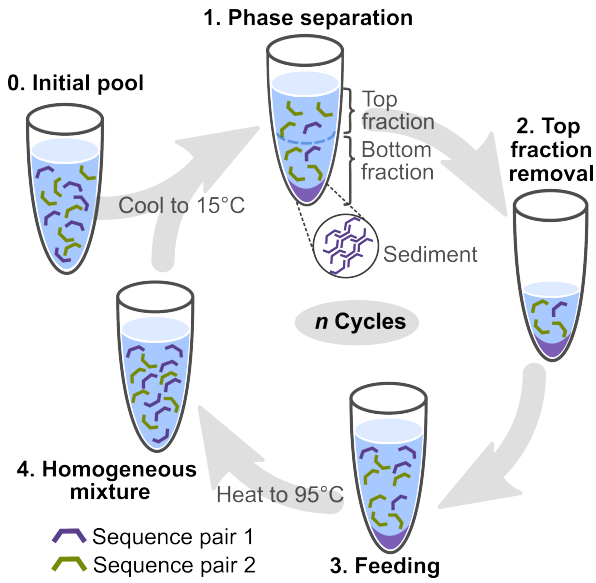


**Figure 1.9:** *Scheme of the experimental protocol for discrete cycles of phase separation.* After cooling to 15 °C and sedimentation to the bottom fraction, the top fraction (half the total volume) is removed and then replaced by the same volume of the initial pool. The sample is then heated to 95 °C to ensure a homogeneous mixture, completing one cycle of phase separation.

reached, avoiding any kinetically trapped states to not participate in the next cycle, while also ensuring fast relaxation to thermodynamic equilibrium. This completes cycle $n$ of phase separation with re-feeding.

To experimentally determine the concentration in the bottom fraction after each cycle $n$ described above, mass balance calculations were used. First, the concentration of the top fraction in each cycle is measured using HPLC with UV-Absorbance (See Section 1.5.3). This value together with concentration of the initial concentration $c_0$ can now be used to calculate the concentration of DNA in the bottom fraction for each cycle as shown in Figure 1.10.

Starting at step **1.** from Figure 1.9, the sequence pairs (SP) phase separate into hydrogels and sediment towards the bottom of the vial. This process conserves the total mass, as no new molecules are being added yet:

$$\bar{c}_n \, 2V_{\frac{1}{2}} = c_{\text{top},n} \, V_{\frac{1}{2}} + c_{\text{bottom},n} \, V_{\frac{1}{2}} \,, \qquad \text{and thus} \quad c_{\text{bottom},n} = 2\bar{c}_n - c_{\text{top},n} \,, \tag{1.13}$$

Here $V_{\frac{1}{2}}$ denotes half of the total volume of the vial. In every extraction step, exactly half of the initial volume is replaced, so it is convenient to use this as the main parameter. The respective concentration of DNA in the top and bottom half at each cycle $n$ are $c_{\text{top},n}$ and $c_{\text{bottom},n}$, respectively. $c_{\text{bottom},n}$ denotes the total concentration of the bottom fraction including the sediment together with some fraction of dilute phase. The exact concentration of sediment can therefore not be determined in each cycle $n$ exactly. Only in the final step, when all feeding cycles have been performed already, the sediment can be extracted exclusively to be measured. Furthermore, $\bar{c}_n$ describes the overall concentration of DNA in the whole volume of the vial $V = 2V_{\frac{1}{2}}$.

In step **2.** the top fraction is removed and replaced with the pool concentration $c_{pool}$ in step **3.**. The mass balance for this step can be written as:

$$c_{\text{pool}}\, V_{\frac{1}{2}} + c_{\text{bottom},n}\, V_{\frac{1}{2}} = \bar{c}_{n+1}\, 2V\,. \tag{1.14}$$

Using equation (1.13), $\bar{c}_{n+1}$ can be written as a function of known measurable values $c_{\text{top},n}$, $c_n$ and $c_{\text{pool}}$:

$$\bar{c}_{n+1} = \bar{c}_n + \frac{c_{\text{pool}} - c_{\text{top},n}}{2}\,, \tag{1.15}$$

$$c_{\text{bottom},n+1} = 2\bar{c}_n + c_{\text{pool}} - c_{\text{top},n} - c_{\text{top},n+1}\,. \tag{1.16}$$

At $n = 0$, $\bar{c}_n$ is equal to the pool concentration $c_{pool}$ and the top fraction concentrations $c_{\text{top},n+1}$ are measured in each cycle. This allows for recursive calculation of all bottom fraction concentrations $c_{\text{bottom},n+1}$ using equation (1.16).

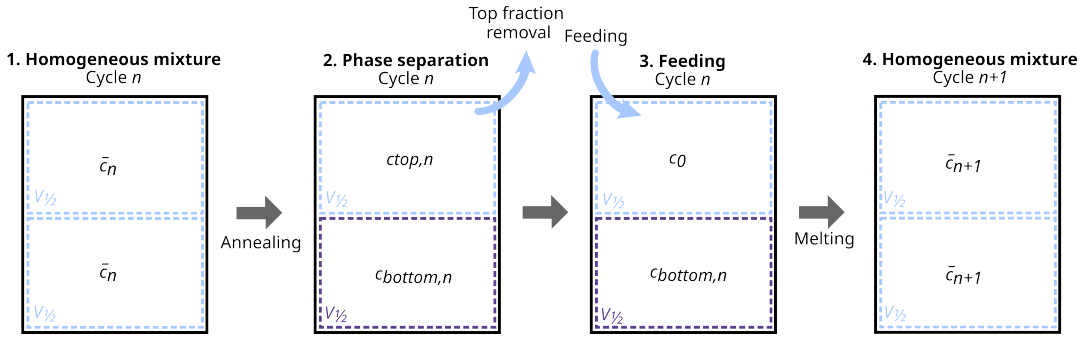


**Figure 1.10:** *Schematic of the feeding cycles illustrating the evolution of the concentrations in the bottom and top fraction.* The average concentration in the chamber at cycle n $c_n$ is split into $c_{top,n}$ and $c_{bottom,n}$ after annealing. Sedimentation increases the concentration in the bottom and $c_{top,n}$ is then replaced with pool concentration $c_0$. After melting the sediment, the solution homogenizes again yielding the chamber to be at the new concentration $c_{n+1}$. By analyzing $c_{top,n}$ in HPLC, $c_{bottom,n}$ can be inferred indirectly.

Having established the framework to analyse the individual concentrations of sequence pairs over multiple cycles of phase separation, the development of two sequence pairs mixed together is experimentally followed over the course of 6 feeding cycles. Figure 1.11 shows the corresponding data from two different initial pools containing SP 1 + 2 and SP 2 + 3 (inlet), respectively. Since phase separation could be observed for SP 1 but not for SP 2, SP 1 was expected to be selected for over the course of the cycles. As a control experiment, SP 2 and 3 were mixed together, as phase separation has not been observed for either of them. The initial pools were initialized at 125 mM NaCl, 10 mM MgCl, 5 X SYBR Green I, 10 mM TRIS pH7 and 25 µM of each DNA strand from the two sequence pairs.

Starting from this mixture (denoted as black data point in Figure 1.11(a)), the concentration in the bottom and top fraction over 6 feeding cycles was monitored using HPLC UV-Absorbance. In the control experiment with SP 2+3 (inlet) the DNA concentration in the bottom and top fraction stayed around pool concentration of 25 µM over the course of the entire experiment, with slopes around 2 µM/cycle. For the system containing SP 1 and 2, however, the bottom fraction showed a significant linear increase in SP 1 with a slope of (10.2 ± 0.4) µM/cycle (purple). At the same time, the measured top fraction of SP 1 stayed below the initial concentration of

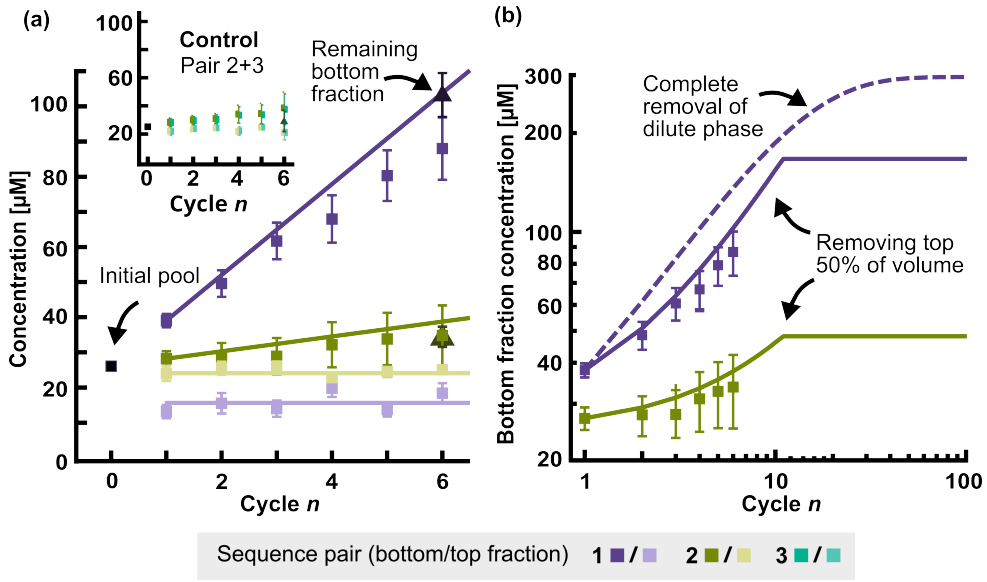


**Figure 1.11:** *Quantification of discrete cycles of phase separation with refeeding.* **(a)** The initial pool contains 125 mM NaCl, 10 mM MgCl, 5X SYBR Green I, 10 mM TRIS pH7 and 25 µM of each DNA strand of two sequence pairs. After each removal of the top fraction (See Figure 1.9), the DNA concentration of each strand in the extracted supernatant is quantified using HPLC. The concentration in the bottom fraction is calculated using conservation of mass (See Figure 1.10). While the concentration of sequence pair 1 increases linearly every cycle with $10.2 \pm 0.4$ µM/cycle, sequence pair 2 remains constant around pool concentration. The inlet graph displays the control result with sequence pair 2 and 3, in which no enrichment could be observed. **(b)** Solid lines are theoretical predictions, this time for higher cycle number up to 100. The data points are the same as in (a). Saturation occurs once the the bottom fraction of the chamber is filled with sediment. The dotted lines shows the prediction if the complete dilute phase would be removed, resulting in the steepest increase in concentration.

$c_{pool}$ at 14.7 µM. This is to be expected, since the material incorporated into the dense phase is now missing in the dilute phase of the top fraction. Meanwhile, SP 2 did not follow this trend and the bottom fraction only weakly enriched during the cycling. This low increase can be attributed to adhesion of DNA strands to the vial surfaces, which cannot be experimentally prevented completely. Error bars correspond to independent triplicates. Most of the error was expected to be caused by pipetting, when for example the pipette tip reaches down too deep into the bottom fraction, disrupting the sediment and removing DNA from there. After the 6th cycle, the remaining bottom fraction was measured and is shown as a black triangle in the graph of Figure 1.11(a).

Using the theoretical predictions described in appendix 1.B, the findings could be extrapolated to larger numbers of cycles up to n=100. Figure 1.11(b) shows the same experimental results as in (b) but plotted together with theoretical predictions. After around 20 cycles, selection has approximately doubled. Since the bottom fraction of the vial is limited in volume, the selection with partial removal of the dilute phase saturates, once the entire bottom fraction is filled with sediment. In an ideal case, with the entire dilute phase removed (Figure 1.11(b), dotted line) in each step, the phase separating sequence pair can enrich 2-fold better against partial top fraction removal to values reaching 10-fold the initial pool concentration $c_{pool}$.

### 1.3.8 Influence of SYBR Green I

Because SYBR Green I is an intercalating dye, possibly interfering with the network formation of the sequences analysed, its influence on the phase separation process was investigated. For this purpose, the experiment of cycles of phase separation (Figure 1.11) was repeated without adding SYBR Green I to the sample composition and only 5 instead of 6 feeding cycles. Figure 1.12 shows the corresponding HPLC data points at 260 nm UV detection of the bottom and top fraction for SP 1 + 2 and SP 2 + 3 (inlet). The initial pool composition (denotes as black data point) for both systems was: 125 mM NaCl, 10 mM MgCl, 10 mM TRIS pH 7 and 25 µM of each DNA strand of two sequence pairs.

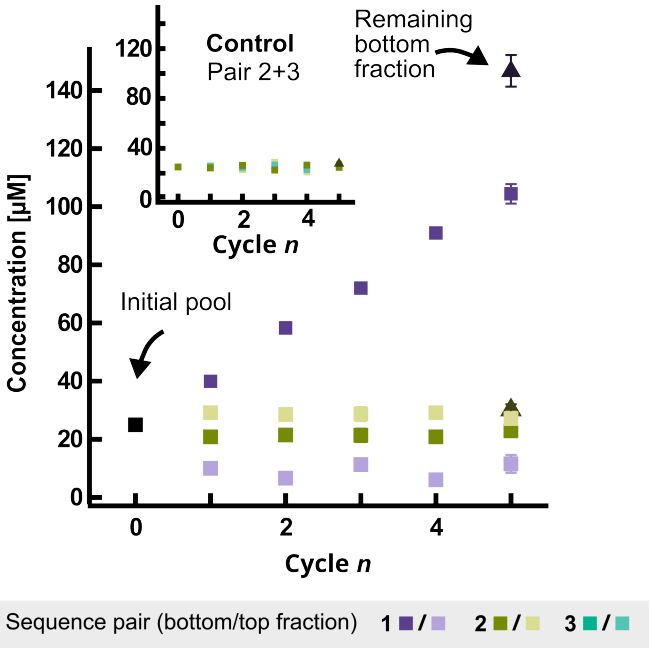


**Figure 1.12:** *Influence of SYBR Green I on the phase separation with feeding cycles.* Without adding SYBR Green I, like in Figure 1.11, sequence pair 1 is the only one that increases in concentration. However, the increase per cycle is 1.6 times higher without SYBR Green I, with a slope of $16.1 \pm 0.3$ µM/cycle. All other parameters were kept equal to before with the initial pool being 125 mM NaCl, 10 mM MgCl, 10 mM TRIS pH 7 and 25 µM of each DNA strand of two sequence pairs.

The previously observed behaviour of the different SPs remained the same. SYBR Green I does not seem to have prohibited SP 2 or 3 from forming the required networks for phase separation. Again, SP 1 increased its concentration in the bottom fraction linearly over the course of the 5 feeding cycles. Its slope, however, is about 1.6 fold higher at (16.1 ± 0.3) µM/cycle compared to the case with SYBR Green I added (10.2 ± 0.4 µM/cycle). As described above, SYBR Green I is an intercalating dye, binding to the minor groove site of a DNA double helix [73]. This can interfere with the secondary structure, as well as change its respective melting temperature [74]. The same interaction has been observed for other intercalating dyes as well [75]. This influence on the phase separation behaviour, however, is necessary in order to be able to use fluorescence microscopy for observation.

### 1.3.9 Flow-through Implementation

The scenario of discrete feeding cycles with sufficient waiting time in between steps, offering the system time to relax to phase equilibrium, is hard to imagine on an early Earth. A rather plausible scenario would involve flows of nucleotides, salts, and other prebiotically plausible molecules through rock pores, as depicted in Figure 1.1 (a). Previous studies have shown how

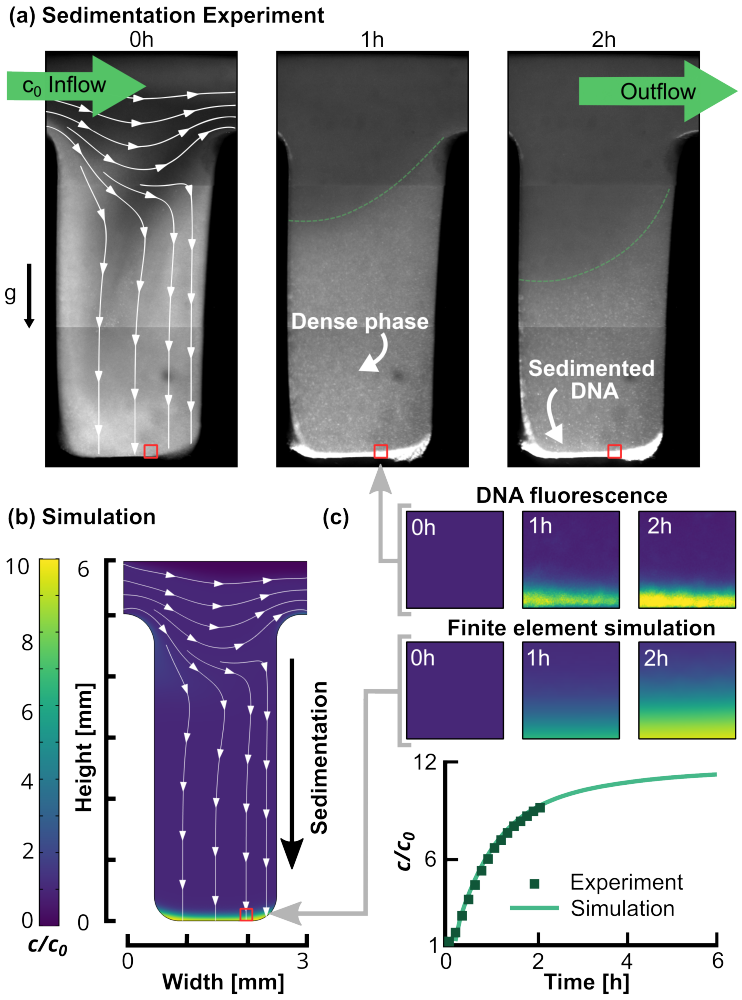


**Figure 1.13:** *Continuous cycles of phase separation with feeding flow* **(a)** Fluorescence micrographs of different time points from the sedimentation experiment in a microfluidic flow-through chamber. A feeding flow of 2 nL/s at concentration $c_{pool}$ containing sequence pair 1 streams across a 5mm deep and 500 µm thick pore. Through phase separation, DNA sediments towards the bottom of the pore. **(b)** 2D finite elements simulation of the chamber matches well with experiments using a diffusion constant of 5 µm$^2$/s for the aggregates and 0.1 µm/s sedimentation speed. **(c)** Relative maximum concentration $c/c_0$ plotted over time agrees well with results from the simulation. Data points are taken from the 200 µm squared cutouts from red rectangles which show that experiment and simulation are in agreement.

gases and liquids can easily percolate through cracks in hydrothermal vent structures like glassy basalt [62, 76]. In such an environment, dilution is an important selection pressure, as

any information stored in oligonucleotides is lost upon reaching a large reservoir such as the ocean. Thus sequence must quickly phase separate and sediment to escape the deadly outflux, leading to the degradation through hydrolysis or loss of information by dilution. Here, a continuous implementation of the cycles of phase separation is introduced. A microfluidic chamber connected to an in- and outflux is used to mimick the aforementioned prebiotic rock pores, subject to feeding flows (See Materials and Methods Section 1.5.6, Figure 1.17). The experimental chamber was cut out from a teflon sheet of the desired thickness of 500 µm and sandwiched between two sapphire plates. Teflon (PTFE or Polytetrafluoroethylene) was chosen, because it does not react with the reagents used in this study and is hydrophobic, reducing undesired surface effects [77]. The results were then compared to fluid flow theory using COMSOL Multiphysics.

This microfluidic chip was subjected to a continuous feeding flow of a pool containing SP 1 and the same buffer conditions as before (125 mM NaCl, 10 mM MgCl, 10 mM TRIS pH7 and 25 µM of each DNA strand). Two separate syringe pumps were used, with the two liquids only meeting inside the microfluidic chip directly. One contained DNA and SYBR Green I, while the second contained the salts and buffer. This ensured that sedimentation could only happen in the chip and not outside of the system, for example inside the syringes. With an inflow speed of about 2 nL/s, the pool then flows across a pore (3mm wide, 5mm deep and 500 µm thick). Figure 1.13(a) shows fluorescence micrographs of this pore with the indications for the flow direction at 0h, 1h and 2h of experiment.

As the pool flows across the pore, sequences capable of phase separation, can sediment to the bottom of the pore and are therefore protected from the outflow towards the ocean. The flowlines in Figure 1.13(a) are taken from simulations and represent the final flow comprised of an interplay between influx- and sedimentation speed. As shown in the micrographs, up-concentration of SP 1 at the bottom of the pore was observed, despite the continuous outflux across the pore. Analyzing the fluorescence intensity over time at the bottom of the pore, an 8-fold concentration increase of SP 1 relative to the pool inflow $c_0$ was observed. Using finite elements simulations, the Navier-Stokes equations for a fully meshed 2D projection of the pore (Figure 1.13(b)) were solved. With a sedimentation speed of 0.1 µm/s and a fixed diffusion constant of 5 $\mu m^2/s$ , the simulation yielded good agreement with experimental data (Figure 1.13(c)). The sedimentation speed was simulated to be maximal, when the surrounding concentration is low, decreasing with a stepfunction (See Materials and Methods Figure 1.18) towards 0, when approaching higher concentration regimes. This ensures that the particle stops when reaching the bottom of the pore.

## 1.4 Conclusion

In this work, it was shown that oligonucleotides that possess the ability to form condensed phases, have an evolutionary advantage to other sequences, not having this selection mechanism. Experimentally subjecting a pool of multiple short DNA sequence pairs to cyclic feeding steps, the supernatant phase was repeatedly replaced with the initial pool composition. This lead to the growth of a condensed phase for specific sequences, while other sequences were depleted relative to the pool. Over the course of multiple feeding cycles, the enriched sequences were therefore selected over others.

The designed sequence pairs were only 22-28 nt in length, showcasing the prebiotic relevance of this selection mechanism. While 22 nt long sequences with 6 nt long binding regions

could phase separate at room temperature, 28 nt long sequences with base pairing regions of 8n t did not. Counterintuitively, this selection mechanism prioritizes sequence over length, in stark contrast to for example charged mineral surfaces [78]. A single base mutation was demonstrated to be enough to inhibit phase separation, underlining the sequence specificity of the process. Analogous to previous literature, characteristic hydrogel features of the condensed phase were observed, in particular the contraction of the gel towards a thermodynamically more stable structure [51]. Different buffer conditions of the selection mechanism were screened and it was shown that the rock pore must be larger than a characteristic height in order for sedimentation to be observable.

In addition, a more realistic implementation of the feeding cycles was realized, where a continuous flow of the pool crosses a through a microfluidic mimick of a rock pore. Despite the outflux, phase separation and subsequent sedimentation enabled sequences to remain in the pore. The findings were matched with data from finite elements simulations with a 2D projection of the pore.

The selection mechanism works robustly for short oligonucleotides, suggesting its relevance at the origin of life, where the recruitment of short peptides, such as RNA and DNA, from prebiotic pools through Darwinian evolution led to the formation of a longer, information carrying, sequence pool. Sequences that would normally face numeruous threats, such as dilution, are accumulated. Inside the condensed phase, they could have then been sheltered from chemical degradation reactions, such as catalytic cleavage [54], or from hydrolysis due to promoted duplex formation [24]. Notably, a correlation between catalytic sequences the phase separation in functional ribozyme polymerases has already been studied earlier [79].

## 1.5 Materials and Methods

### 1.5.1 DNA Strands

DNA stocks were purchased in dry state from biomers.net in HPLC-grade purity and kept at -20 °C for storage after dilution with nuclease free water (Ambion nuclease-free water from Invitrogen) to 200 µM. Before every experiment, the strands were denatured at 95 °C for 2 minutes.

### 1.5.2 Sample Preparation

For the sedimentation experiments, as well as for the cycles of phase separation, the initial pool samples were prepared in Eppendorf tubes using 15 µL of 25 µM of each respective DNA strand of the SP, 10 mM TRIS-HCL Buffer at pH 7, 5 X SYBR Green I (Excitation at 450-490 nm, Emission at 510-530 nm), 125 mM NaCl and 10 mM $MgCl_2$. To ensure homogeneous starting conditions and to avoid kinetically trapped hybridization states prior to the start of the experiment, the solutions were always heated to 95 °C for 2 minutes. The samples were then cooled down to 65 °C for 10 s and then slowly cooled down to 15 °C (ramp rate: 6K per minute) to allow for hybridization and consequent phase separation. At this final temperature, the samples was then incubated for at least 3 h in order to give the sediment enough time to form. The temperature protocols were performed in a standard thermocycler (Bio-Rad CFX96 Real-Time System). For the feeding cycle experiments (Figure 1.11), 7.5 µL of the dilute phase – half of the intitial volume – was removed after the respective sedimentation step. This volume was removed carefully by pipetting only at the center of the meniscus to avoid removing material from the sediment. Afterwards, 7.5 µL of the initial pool stock was added to the remaining bottom fraction. Repeating the aforementioned temperature steps, completes one cycle $n$.

### 1.5.3 Sample Analysis via HPLC

Ion-pairing reverse-phase High-Performance Liquid Chromotography (HPLC) experiments were carried out on a column liquid chromatography system equipped with an auto-sampler and a bio-inert quaternary pump (Agilent 1260 Infinity II Bio-Inert Pump G5654A, Agilent Technologies). A C18 capillary column (AdvanceBio Oligonucleotide 4.6x150 mm with particle size 2.7 µm, Agilent) was used to perform reverse-phase liquid chromatography. The temperature of the autosampler was set to 4°C. The mobile phases consisted of two eluents. Eluent A was HPLC water (Sigma-Aldrich), 200mM 1,1,1,3,3,3, -Hexafluoro-2-propanol (HFIP) (Carl Roth GmbH), 8mM Triethylamine (TEA) (Carl Roth GmbH). Eluent B was a 50:50 (v/v) mixture of water and methanol (HPLC grade, Sigma Aldrich, Germany), 200 mM HFIP, 8 mM TEA. The injection volume for each measurement was 100 µL. The samples were eluted with a gradient of 1 % B to 58.6 % B over the course of 45 minutes with a flow rate of 1 mL/min. Prior to the gradient, the column was flushed with 1 % B for 5 minutes. Retention times were analyzed via a UV Diode Array Detector (Agilent 1260 Infinity II Diode Array Detector WR G7115A) at 260 nm with a bandwidth of 4nm. Samples were diluted for HPLC loading in the following manner: 7.5µL of sample, 105µL nuclease-free water and 75µL of a 5M Urea solution. They were heated to 95°C for 2 minutes afterwards to ensure de-hybridization of the strands and dissolution of any sediment. Then, 105 µL of the diluted samples were transferred into N9

glass vials (Macherey-Nagel GmbH) and stored at 4 °C in the auto-sampler of the HPLC-MS system (1260 Infinity II, Agilent Technologies) until injection.

### 1.5.4 Thermal Melting Curves

To measure the melting temperature of DNA strands, thermal melting curve measurements were performed. The SYBR Green I fluorescence was used to measure the double stranded fraction as a function of temperature inside standard thermocyclers with fluorescence readout (Bio-Rad CFX96 Real-Time System). The sample compostion was kept equal to the composition used for all previous experiments: 15 µL of 25 µM of each respective DNA strand of the SP, 10 mM TRIS-HCL Buffer at pH 7, 5 X SYBR Green I, 125 mM NaCl and 10 mM $MgCl_2$. The temperature of the sample was then varied from 30 to 95 °C with a ramprate of 0.1 °C/s, yielding sigmoidal fluorescence profiles, with two plateaus. Each curve was repeated thrice by pipetting three independent samples, yielding triplicates. Additionally, the background signal was corrected for by using a mixture without any DNA present, only with buffers and SYBR Green I. The melting curves were then analyse using a self-written LabView script based on the background correction from [68]. First the signal from the background fluorescence is subtracted from the fluorescence of the sample. Afterwards, the lower and higher baseline (linear) functions are determined and used for the baseline adjustment. These correspond to fully bound and fully unbound duplex states, respectively. The corrected data are then exported to "Igor Pro 6.37" and fitted with a sigmoidal function, where the midpoint fitting parameter corresponds to the melting temperature $T_m$ (see Figure 1.4). The $T_m$ of all the sequence pairs was below the denaturing temperature used in the thermal protocol (95 °C, see Figure 1.9), ensuring the mixture is homogeneous before triggering phase separation through cooling.

### 1.5.5 Sedimentation Analysis

To analyse the properties of the sediment inside the microfluidic chamber using fluorescence readout, a LabView program was designed (See Figure 1.14 for a screenshot). With it, the fluorescence intensity of the timelapse images of the bottom of the chamber are quantified. Using a background measurement, as well as knowing the initial concentration of the dissolved DNA, one can measure the maximum concentration relative to the initial pool. Furthermore the sediment height over time can be read out, giving crucial insights about the dynamics of the hydrogel. For fluorescence detection, SYBR Green I was used, which scales linearly with the amount of dsDNA in solution [73], thus is appropriate for quantifying the dense phase of DNA at hand. The first image is taken before the temperature protocol described in Section 1.5.2, yielding a concentration reference for all further calculations. ll the remaining micrographs are divided by this one to obtain relative concentration $c/c_0$, where $c_0$ is the sequence concentration in the chamber immediately after flushing, i.e. the pool concentration. Once the temperature protocol has finished and the system rests at the incubation temperature of 15 °C, the sediment at the bottom of the chamber starts to grow in size. When a sediment is present, the relative concentration through the sediment is obtained by measuring the concentration along a defined line perpendicular to the wall of the well, which was parametrized by $x$. The maximum of relative concentration occurs at the center of the sediment. Sediment height is determined along the $x$-direction from the center concentration $c_{\max}$ until the value has reached a $0.5\, c_{\max}$. The sediment height is then

the distance where $c > 0.5\,c_{\max}$. The relative average sediment concentration is calculated by averaging over all points for which $c > 0.5\,c_{\max}$. Using the sediment height $h_{\text{sed}}$ and average relative concentration $\bar{c}/c_0$, the total amount of sedimented material was calculated $N_{\text{sed}}$ through Eq. (1.17), where $c_0$ is the initial concentration, $L$ is the length of the chamber and $d$ is the depth of the chamber:

$$N_{\text{sed}} = \left(\frac{\bar{c}}{c_0}\right) c_0\, h_{\text{sed}}\, L\, d\,. \tag{1.17}$$

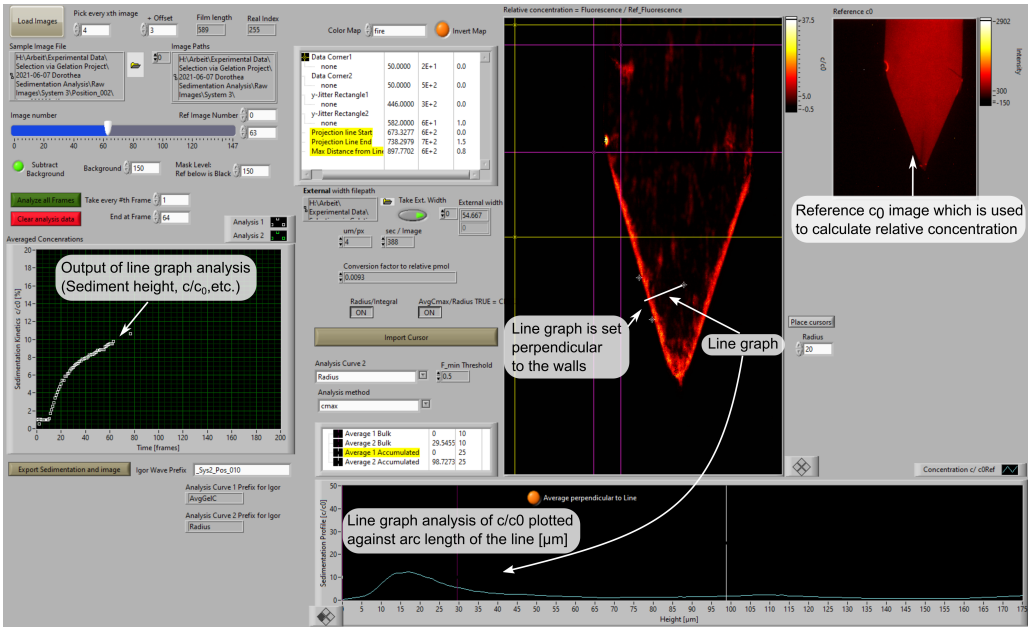


**Figure 1.14:** *Screenshot of self-written LabView program for sedimentation analysis.* For each frame taken by the microscope, the maximum concentration along a line graph perpendicular to the chamber walls is extracted. Background and initial concentration are used to calculate $c/c_0$ values.

### 1.5.6 Experimental Setups

For the fluorescence measurements throughout this work, two different microfluidic chambers were built and mounted on a self-built fluorescence microscopy setup. A sketch of the self-built fluorescence microscope is displayed in Figure 1.15. In this setup, a stepper motor is used to move the microfluidic chamber relative to the camera, such that multiple locations on the microfluidic chip could be images consecutively. The setup was controlled by a self-written LabView software and operated inside black curtains to reduce noise from stray light.

Figure 1.16 shows the microfluidic chamber used to capture timelapse fluorescence images of the sedimentation process. As a precise temperature control was required to execute the temperature protocol, this chamber was connected to a Peltier heat element and controlled remotely. The chamber was comprised of a 500 µm thick teflon cutout, forming multiple wells able to contain 15 µL of sample each. When sandwiched between two sapphire plates and attached to the peltier element in the back, this configuration enabled us to have full

visual control of the experiment. This structure was held together by a steel frame and was attached to an aluminium back, which was held at 15 °C by a water bath (JULABO). During

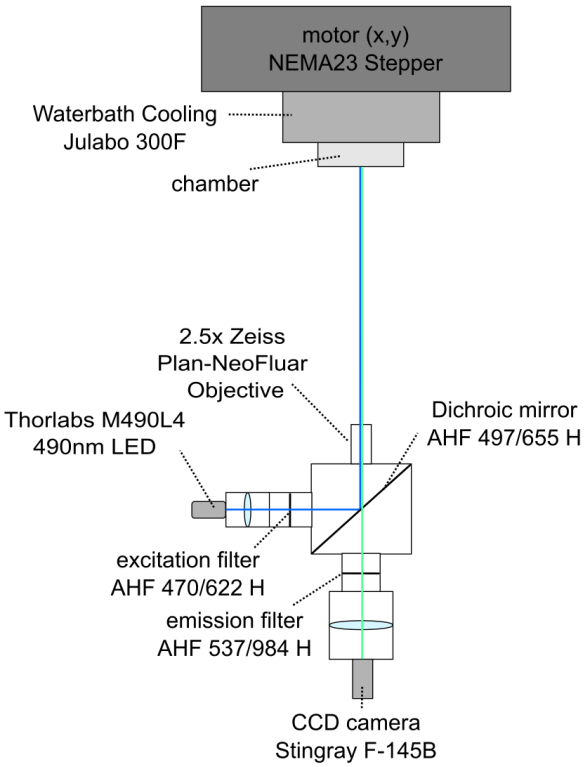


**Figure 1.15:** *Sketch of the self-built fluorescence microscope.* Stepper motor, temperature, LEDs and camera are operated and triggered by a self-written LabView program. The waterbath is set to a constant temperature, while the actual chamber temperature is controlled by a Peltier-element (See Figure 1.16).

the temperature protocol, the individual wells were sealed using Parafilm in order to prevent evaporation at higher temperatures. At temperature above 65 °C however, the evaporation could not be prevented, prompting the reduction of the maximum temperature spike to 65 °C. In the case of the flow-through experiment, the chamber was required to be accessible by syringe pumps during fluorescence measurements. For this purpose, another microfluidic chamber was designed (See Figure 1.17). The out- and inlet of the microfluidic chamber were handled by holes in the back sapphire as well as in the aluminium backings. Using Nemesys syringe pumps, a constant flow of intial pool could be achieved in this manner. Two syringes were used, one containing DNA and SYBR Green I, the other containing TRIS Buffer, NaCl and $MgCl_2$. To maintain constant temperature throughout the experiments, the front temperature of the chamber was controlled by resistor heaters, while the back by the water bath.

### 1.5.7 Finite Elements Simulations

To gain further understanding of the system a 2D-Simulation projection of the flow chamber solved using Finite elements simulations. The experimental chamber was designed using an Autodesk CAD software (Inventor 2019) and then cut out of a teflon sheet. The same 2D drawing was then exported to COMSOL Multiphysics 5.4 using the Inventor live-link plugin.

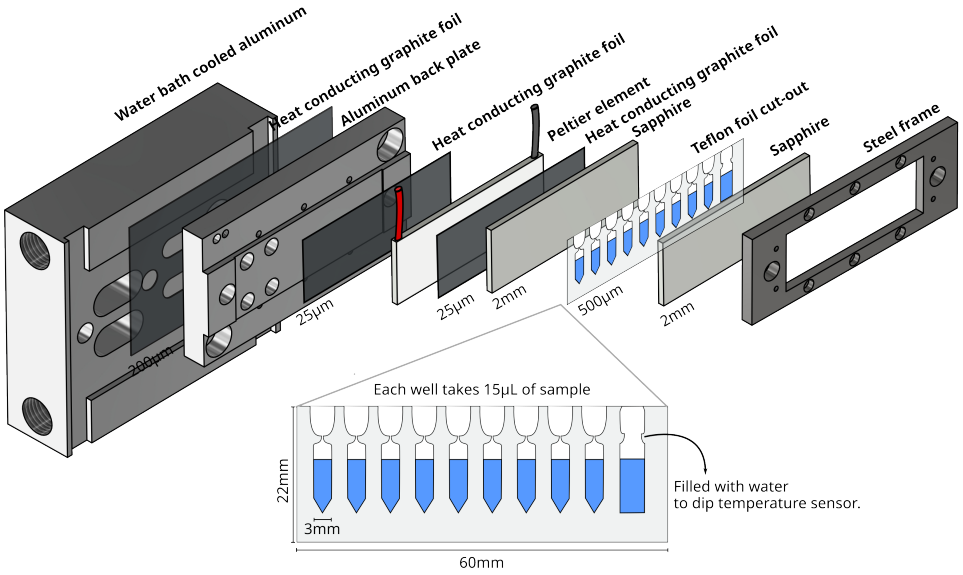


**Figure 1.16:** *Exploded view of the chamber used for sedimentation experiments.* From left to right: the water cooled aluminium backing is connected with a 200 µm thin graphite foil to the aluminium backplate of the chamber to ensure maximum heat conduction. On this backing, a peltier element is used for temperature control and again connected with 25 µmgraphite foils. Then, 2 2 mm thick sapphires sandwich the teflon cutout of the chamber, held together by a steel frame.

The geometry is therefore matching the experimental chamber in the two dimensions $x$ and $y$. Since the flows and the sedimentation of material occur only in $x$- and $y$-direction, omitting the $z$-axis allowed us to effectively screen parameters focusing on the key dynamics of the system.

The inlet and outlet of the well were emulated defining a constant normal inflow speed of 2 µm/s as well as an outlet with a pressure boundary condition. Assuming a constant 20 °C across the whole geometry, stationary laminar flow was solved with Navier-Stokes equations for conservation of momentum and continuity equation for conservation of mass. Since the flow speed of 2 µm/s is very low (Mach number M < 0.3), the flow can be considered as incompressible. Thus, the density is assumed to be constant and the continuity equation reduces to the condition:

$$\nabla \cdot \boldsymbol{u} = \mathbf{0}\,. \tag{1.18}$$

with $\rho$ denoting the mass density and $\boldsymbol{u}$ the velocity field. The Navier-Stokes equation then reduces to

$$\rho(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \nabla \cdot [-p\boldsymbol{I} + \eta(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T] + \boldsymbol{F} = \mathbf{0}\,, \tag{1.19}$$

with $p$ being pressure, $\boldsymbol{I}$ the unity tensor, $\eta$ the fluid dynamic viscosity and $\boldsymbol{F}$ the external forces applied to the liquid. The reference pressure was set to 1[atm], the reference temperature was 20°C and all surfaces are described as non-slip boundary conditions.

Assuming the laminar flow to be stationary the transport of a diluted species $c_i$ was described by combining convection and diffusion effects:

$$\frac{\partial c_i}{\partial t} + \nabla(-D_i \nabla c_i + \boldsymbol{u} c_i) = 0\,, \tag{1.20}$$

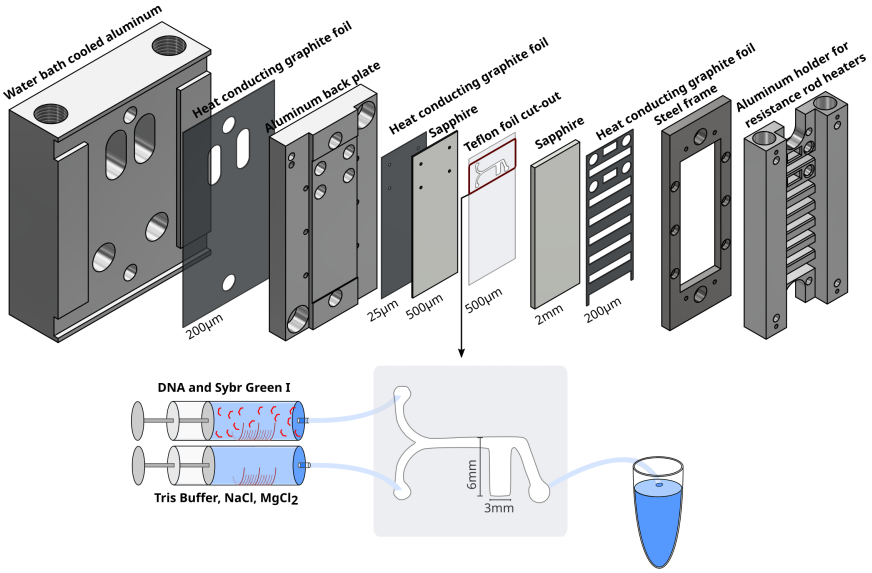


**Figure 1.17:** Exploded view of the chamber used for flow-through experiments. Similar setup to Figure 1.16. Here, the temperature is controlled using waterbath in the back and resistance heaters in the front. This enables access to the chamber using syringe pumps from the back through in- and outlets. DNA and SYBR Green I are loaded in a seperate syringe as buffer and salts such that they only meet inside the chamber to prevent sedimentation inside the syringes. Flowspeed is controlled using NEMESYS syringe pumps.

where $D_i$ denotes the diffusion coefficient of species $i$. Sedimentation of oligomers was simulated assuming a downwards oriented flow speed $v_{sed}$ affecting diluted species only. With increasing local relative concentration $\frac{c}{c_0}$, the flow speed decreases from 0.1µm/s to 0, using an inbuilt smoothened heavyside step function, which is shown in Figure 1.18. The $y$-component of the velocity field $\boldsymbol{u}$ used to describe the convective movement of diluted species $c_i$ then reads:

$$u_y = v - v_{sed} \quad \text{with} \quad v_{sed} = v_0 \cdot step\left(\frac{c_i}{c_0}\right). \tag{1.21}$$

Here, $v_{sed}$ ensures that molecules cannot sediment into an area, where the local concentration has already reached $c_i/c_0 = 10$. This maximum density was estimated using the fluorescence data from the experiments. Choosing the center of the step function to be at $c/c_0 = 7$ gave best results matching the experiments. The sedimentation speed of maximum 0.1 µm/s was chosen through observation of the distance traveled in between images for the very first visible phase separation DNA aggregates. The aggregates sediment quicker, the larger they become. These kinetics were not simulated. Rather, the system was treated as if it was filled with those small aggregates from the beginning, sedimenting with $v_{\text{sed}}$. Realistically, these aggregates do not diffuse, since they are not soluble any longer. However, the simulation breaks down, once the diffusion coefficient becomes too close to 0. Since the condensates were treated like a diluted species, the lowest diffusion constant possible of 5 $\mu m^2/s$ was chosen. Meshing was done using an automatic finer physics-controlled mesh (see Figure 1.18. After setting the parameters as above, a complete numerical solution could be found. A complete set of all input parameters can be found in Table 2.3.

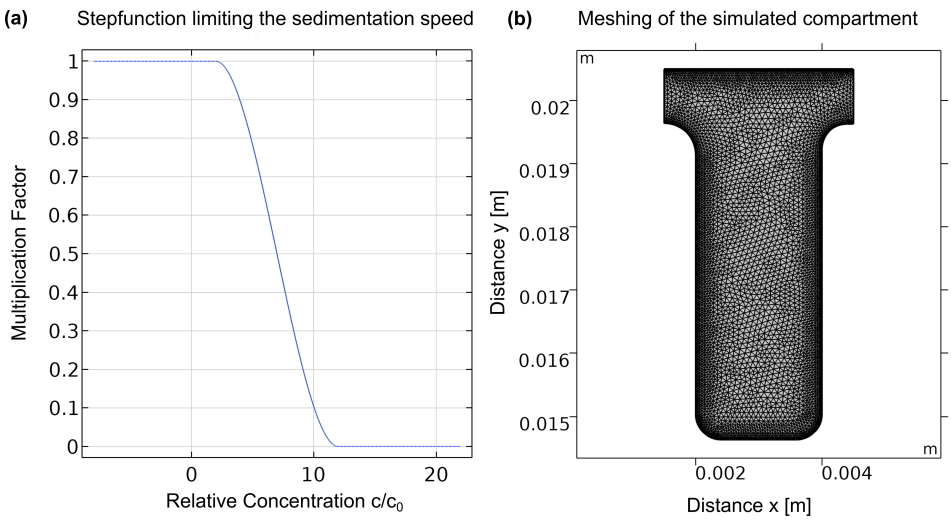


**Figure 1.18:** *Sedimentation speed stepfunction and meshing of the geometry.* **(a)** Stepfunction relating local $c/c_0$ values to a factor between 0 and 1. This value is then multiplied to the sedimentation speed, thus slowing the particle down once it reaches the bottom of the chamber where DNA accumulated already. This ensures the sediment to grow normally. **(b)** Meshing of the geometry. In each triangular, the partial differential equations are solved with their respective boundary conditions to adjacent triangulars.

| Parameter | Value | Description |
|---|---|---|
| $v_{\text{in}}$ | 2 $\mu m/s$ | Inflow velocity |
| $D$ | 5 $\mu m^2/s$ | Diffusion constant of species c |
| $v_0$ | 0.1$\mu m/s$ | Sedimentation velocity |
| $c_0$ | 25 $\mu M$ | Initial concentration of species c |
| Well width | 2 $mm$ | Width of the simulated pore |
| Well height | 5 $mm$ | Height of the simulated pore |
| Cutoff | 7 | Position of the center of the stepfunction |
| Smoothing | 10 | width of the stepfunction |

**Table 1.1: Parameters used for the numerical calculation for the system with continuous feeding flow.** Final set of parameters used to simulate the flow-through experiment from Figure 1.13. Water-specific parameters such as dynamic viscosity or density were taken from inbuilt features of COMSOL Multiphysics 5.4..

# Appendix

## 1.A Nupack Design Code

Figure 1.19 shows an exemplary version of the code used to design the sequence pairs. Multiple rounds of designing were done and afterwards, the secondary structure was checked using the Nupack analysis tool online until a sequence pair was found that fulfils the required secondary structure.

```
temperature[C] = 37.0
material = dna

# domains
domain a1 = S2
domain a2 = W2
domain a3 = S2
domain b1 = S2
domain b2 = W2
domain b3 = S2
domain c1 = S2
domain c2 = W2
domain c3 = S2
domain s = T2

# strands
strand s1 = a1 a2 a3 s b1 b2 b3 s c1 c2 c3
strand s2 = a3* a2* a1* s b3* b2* b1* s c3* c2* c1*

# complexes
complex ca = s1 s2
complex cb = s1 s2
complex cc = s1 s2

# target structures
ca.structure = D6(U16 +) U16
cb.structure = U8 D6( U8 + U8) U8
cc.structure = U16 D6(+ U16)

# tubes
tube tub = ca cb cc
tub.ca.conc[M] = 1e-6
tub.cb.conc[M] = 1e-6
tub.cc.conc[M] = 1e-6

prevent = AAAA,CCCC,GGGG,UUUU,AA,GGG
stop[%] = 10
```

**Figure 1.19: Exemplary Nupack design code.** For DNA at a temperature of 37 °C, two strands were designed by segmentation into smaller domains ($a$, $b$ and $c$). In the # domains section, the $a$, $b$ and $c$ are split into even smaller domains of similar nucleotide content (S being G or C, and W being A or T). In this example code, TT spacers were defined via "s = T2". In "# strands", the sequences of the system are defined, reading from 5′ to 3′. The * denotes complementarity. Hybridization constraints are shown in "# target structures", where the 3 possible binding interactions from "# complexes" between strand 1 and strand 2 are defined further. "D6U16" for example denotes a 6 base pair region followed by a 16 unpaired region. Preventing specific sequence-patterns, such as AAAA or GGGG, above a certain relative amount, can be useful to avoid for example G-quadruplex structures or unwanted stacking.

## 1.B Multi-component phase separation subject to cyclic material exchanges

Here, the concentration changes in a mixture of volume $V$ that is composed of $M$ different oligonucleotide sequences is described when subjected to periodic exchange of material with a pool with fixed composition $\boldsymbol{c}_{\text{pool}}$. We introduce an $M$-dimensional vector $\bar{\boldsymbol{c}}(t)$, where the components of this vector are concentrations of sequences. Starting from the initial state $\bar{\boldsymbol{c}}(t_0)$, we perform $N$ exchange cycles, where each cycle is labeled with $n = 1, \ldots, N$ and composed of the two following steps:

*Phase separation step*: The homogeneous mixture of concentration $\bar{\boldsymbol{c}}(t_n)$ phase-separates into two coexisting phases. We denote the concentrations of the dense and dilute phase as $\boldsymbol{c}^{\text{I}}(t_n)$ and $\boldsymbol{c}^{\text{II}}(t_n)$, respectively. The volume of the dense phase is $V^{\text{I}}(t_n)$ and thus the dense phase occupies the volume $(V - V^{\text{I}}(t_n))$. Mass and particle numbers are conserved during the phase separation step:

$$\bar{\boldsymbol{c}}(t_n) = \left[ \frac{V^{\text{I}}(t_n)}{V} \boldsymbol{c}^{\text{I}}(t_n) + \frac{V - V^{\text{I}}(t_n)}{V} \boldsymbol{c}^{\text{II}}(t_n) \right] . \tag{1.22}$$

*Partial dense phase removal step:* A constant fraction of the volume of the dense phase, $\alpha(V - V(t_{n-1}))$, with the relative fraction $0 < \alpha < 1$, is replaced by the same volume taken from the pool, $\boldsymbol{c}_{\text{pool}}$. The average composition thus changes according to

$$\bar{\boldsymbol{c}}(t_{n+1}) = \left[ \frac{V^{\text{I}}(t_n)}{V} \boldsymbol{c}^{\text{I}}(t_n) + \frac{V - V^{\text{I}}(t_n)}{V} \Big( \alpha \boldsymbol{c}_{\text{pool}} + (1 - \alpha) \boldsymbol{c}^{\text{II}}(t_n) \Big) \right] . \tag{1.23}$$

For the general case where the initial average concentration is not equal to the average concentration of the pool, $\bar{\boldsymbol{c}}(t_0) \neq \boldsymbol{c}_{\text{pool}}$, we determine the phase compositions $\boldsymbol{c}^{\text{I}}(t_n)$ and $\boldsymbol{c}^{\text{II}}(t_n)$, and the phase volumes $V^{\text{I}}(t_n)$ at each cycle time $t_n$. Note that during the selection kinetics, the average concentration $\bar{\boldsymbol{c}}(t_n)$ approaches the tie line passing through the pool concentration vector $\boldsymbol{c}_{\text{pool}}$.

For the special case where the initial average concentration is equal to the average concentration of the pool, $\bar{\boldsymbol{c}}(t_0) = \boldsymbol{c}_{\text{pool}}$, we can obtain an analytic solution, even for an arbitrary number of different components $M$. Only in this case, $\boldsymbol{c}^{\text{I}}$ and $\boldsymbol{c}^{\text{II}}$, remain constant in time since the average concentration moves along the tie line defined by the pool. The iteration rule Eq. (1.23) simplifies to:

$$\bar{\boldsymbol{c}}(t_{n+1}) = \left[ \lambda(t_n) \boldsymbol{c}^{\text{I}} + (1 - \lambda(t_n)) \left( \alpha c_{\text{pool}} + (1 - \alpha) \boldsymbol{c}^{\text{II}} \right) \right] , \tag{1.24}$$

where $\lambda(t_n) = V^{\text{I}}(t_n)/V$ is the relative volume of the dense phase. Using particle conservation in Eq. (1.22) for the pool concentration $\boldsymbol{c}_{\text{pool}}$, remaining within the special case $\boldsymbol{c}_{\text{pool}} = \bar{\boldsymbol{c}}(t_0)$, the interaction rule becomes:

$$\bar{\boldsymbol{c}}(t_{n+1}) = \Big[ \alpha\lambda_0 + (1 - \alpha\lambda_0)\lambda(t_n) \Big] \boldsymbol{c}^{\text{I}} + \Big[ 1 - \alpha\lambda_0 - (1 - \alpha\lambda_0)\lambda(t_n) \Big] \boldsymbol{c}^{\text{II}} , \tag{1.25}$$

where $\lambda_0 = \lambda(t_0)$. Using particle conservation at time step $t_{n+1}$,

$$\bar{\boldsymbol{c}}(t_{n+1}) = \lambda(t_{n+1})\boldsymbol{c}^{\mathrm{I}} + (1 - \lambda(t_{n+1}))\, \boldsymbol{c}^{\mathrm{II}}\,, \tag{1.26}$$

we can identify the term in the first bracket of Eq. (1.25), as $\lambda(t_{n+1})$, the size of the dense phase at $t_{n+1}$, and obtain a recursion relation:

$$\lambda(t_{n+1}) = a\,\lambda(t_n) + b\,, \tag{1.27}$$

where $a = 1 - \alpha\lambda_0$ and $b = \alpha\lambda_0$. For large times, the system reaches a stationary state

$$\lambda(t_\infty) = \frac{b}{1-a} = 1\,. \tag{1.28}$$

This can be used to rewrite the recursion in terms of $\delta\lambda(t_n)$ with $\delta\lambda(t_n) = \lambda(t_\infty) - \lambda(t_n) = 1 - \lambda(t_n)$. As a result, $\delta\lambda(t_{n+1}) = a\,\delta\lambda(t_n)$. Its solution reads $\delta\lambda(t_n) = \delta\lambda_0\, a^n$ that can be written as

$$\lambda(t_n) = 1 - (1 - \lambda_0)\,(1 - \alpha\lambda_0)^n\ . \tag{1.29}$$

This solution completely determines the evolution of the mean volume fraction:

$$\bar{\boldsymbol{c}}(t_n) = \Big[1 - (1 - \lambda_0)\,(1 - \alpha\lambda_0)^n\Big]\boldsymbol{c}^{\mathrm{I}} + (1 - \lambda_0)\,(1 - \alpha\lambda_0)^n\, \boldsymbol{c}^{\mathrm{II}}\,. \tag{1.30}$$

The characteristic number of iterations required to converge to the stationary state $\lambda(t_\infty) = 1$ (and, consequently, $\bar{\boldsymbol{c}}_\infty = \boldsymbol{c}^{\mathrm{I}}$) is

$$n_{\mathrm{c}} = -\frac{1}{\log a} = -\frac{1}{\log[\alpha(1-\lambda_0)]}\,. \tag{1.31}$$

# 2 Prebiotic gas flow environment enables isothermal nucleic acid replication

## 2.1 Summary

Nucleic acid replication is a central mechanism at the origin of life. However, this process encounters significant challenges on early Earth. Dilution, for instance, poses a threat to achieving the necessary concentrations for prebiotic reactions. In addition, bound duplex structures of nucleic acids require a mechanism for strand separation to achieve exponential replication, especially in the presence of divalent ions such as $Mg^{2+}$. Without a non-enzymatic means for this purpose, further copying is stalled. Recent studies have shown that physical non-equilibria, such as temperature gradients, accumulate and strand-separate nucleic acids. This solution however, requires heating, which leads to molecular degradation, especially for RNA. Additionally, the requirement for strong temperature gradients further limits the availability of the environment. In this work, an isothermal geological environment was studied that solves the problem of dilution through accumulation of biomolecules and also provides the means for cyclic strand separation by oscillations in salt concentration. This prebiotically plausible and abundant system consists of a gas flow at an open rock pore connected to an influx of water. The water is continuously being dried up at the water-gas interface, while the momentum transfer of the gas flow across stirs the solution, inducing circular flow patterns. Experimental data shows that evaporation accumulated nucleic acids up to 30-fold, in agreement with the fluid dynamic simulations. Fluorescent micrometre-sized beads were observed in this system, revealing circular flows of up to 1 mm/s caused by the gas flow. Imaging a labelled FRET pair, reveals how this vortex causes a clear separation of double- and single-stranded regions of DNA. Then, a Monte Carlo simulation was performed which followed a dissolved strand of DNA in this environment. Along its path, salt oscillations were observed, reaching approximately 3 times the initial concentration. The protein *Taq* polymerase is then used as a model system to show that these salt oscillations can periodically separate nucleic acid strands, allowing them to replicate under isothermal conditions. As expected, replication could only be found when an initial template sequence was provided and when the gas flow was switched on. Fluorescence clearly showed that the replicated molecules were accumulated at the interface. The found replication at a likely abundant isothermal environment significantly broadens the range of potential environments for early chemical life on Earth.

This chapter was published by Schwintek et. al in eLife (https://doi.org/10.7554/eLife.100152.1) as a reviewed preprint and is here adapted and reprinted in part with permission from eLife. Full article, as submitted to bioRxiv, attached in the List of Publications.

## 2.2 Motivation

One of the pillars of life on earth is its ability to replicate, mutate and thus adapt to the surrounding environment. As early chemical life likely started outside of compartments, such as the cells today, geological boundary conditions dictated its evolution. An environment with harsh chemical conditions, such as extreme pH or temperature, threatens molecules with degradation and therefore forces any chemical system to adapt through Darwinian evolution. The common theory on how this concept translates to very early chemical life, is the so called "RNA World" [80]. Nucleic acids, such as DNA and RNA, are the most important information carrying polymer to life as we know it. Every organism on Earth stores and multiplies its genetic information using nucleic acids. Functionality, however, is mostly carried out by proteins, long chains of amino acids. Even the replication and synthesis of nucleic acids is performed by proteins.

However, through the discovery of ribozymes – functional RNA strands capable of catalytic function – the concept of life primarily based on RNA became imaginable. Because of their ability to both store genetic information, as well as catalyzing their own replication, it is commonly believed that the "RNA World" preceded cellular life. These abilities allow them to mutate and evolve, enabling them to adapt to diverse environments and eventually encode, build and utilize proteins as the catalysts used in modern life.

Such chemical life is however difficult to imagine in a prebiotic soup scenario in a large reservoir such as the ocean. Dilution poses a significant obstacle, since such prebiotic reactions require sufficiently high concentrations of their reagents to work [81]. The ocean lacks local sources of energy to compensate for diffusion and drive reaction pathways out of equilibrium, resulting in dilution of building blocks for chemical reactions due to homogeneity [82, 83]. Large basins are therefore not suited to harbor early molecular life.

Research has thus focused on identifying plausible early Earth environments to meet the criteria. Local physical non-equilibria, for example, have shown the ability to up-concentrate molecules, such as nucleic acids, in a variety of different geological settings [84]. Thermal gradients in rock pores on the microscale can accumulate molecules by many orders of magnitude through an interplay of thermophoresis and convection [85]. As molecules differ in their thermo-diffusive behaviour, this accumulation is drastically charge and size dependent, making it possible to even purify certain solutes through networks of rock cracks subjected to a temperature gradient [86]. When coupled with air-water interfaces, these temperature differences also enable phosphorylation, crystallization or gelation of RNA [60]. An interplay of capillary flows, convection and Marangoni flows enables not only accumulation directly at the interface, but can even lead to the encapsulation in vesicles or trigger ribozyme catalysis.

Many prebiotic synthesis reactions require both wet and dry conditions to work. Central to the RNA world emergence is the polymerization of nucleotides to form longer RNA strands. It has been suggested that warm little ponds on young forming continents emerging from the ocean, or hydrothermal vent settings can promote this process through wet-dry cycles [87–89]. Furthermore, the drying process at the air-water interface has shown to trigger spontaneous polymerisation of 2′,3′-cyclic ribonucleotides at elevated pH [90]. In this context, foams can maximize the surface to volume ratio further amplifying the efficacy of these processes [91].

Additionally, mineral surfaces themselves are charged allowing charged polymers like RNA to adsorb. This adsorption process not only enables synthesis reactions [92], but is also length selective, preferentially binding longer polymers compared to shorter ones due to entropy [93]. As a final example, the eutectic phase in freezing seawater can also up-concentrate reagents

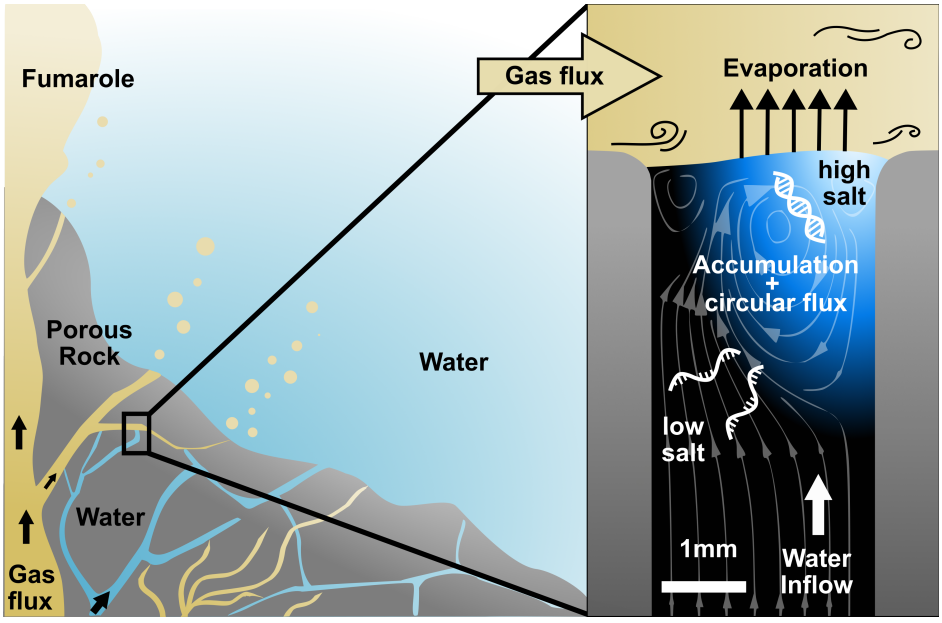


**Figure 2.1:** *Sketch of the studied geological environment.* Scenario in which water, containing biomolecules, is flushed towards a gas flux streaming across. This induces circular currents (See flow lines) and accumulation due to the evaporation of pure water, while dissolved molecules stay in bulk. The induced salt gradient promotes strand separation of dissolved nucleic acids periodically, enabling exponential replication. On early earth, in the vicinity of volcanoes or underwater degassing events, many of such environments can be imagined. The experiments in this work aim to replicate this environment on the microcscale by subjecting a defined volume of sample to a continuous, diluting, influx of pure water, with an airflux brushing across.

and favor synthetic reactions [94, 95]. As the seawater cools down, tiny pure water ice crystals form, leaving behind liquid brine in between, where dissolved molecules and salts are highly concentrated.

However, the accumulation of salts and molecules comes at a cost. The replication process of nucleic acids is reliant on base pairing and leaves the replicated strand attached to its complementary copy. Without separation, further copying is stalled and the reaction cannot proceed. Strand separation becomes increasingly difficult after accumulation because the melting temperature of oligonucleotides is strongly dependent on the local salt concentration [96]. Despite high $Mg^{2+}$ concentrations being required for replication and catalytic activity [97], they elevate the melting temperature of nucleic acid duplex structures to levels surpassing the boiling point of water [35]. Oligonucleotides readily hydrolyze into nucleotide fragments under these conditions, rendering periodic temperature spikes — often relied upon as a primary strand separation mechanism — more detrimental than beneficial [33]. Therefore, other mechanisms are required at the origin of life to separate RNA or DNA strands with minimal thermal stress, and at best combined with an environment where supplied bio-molecules are accumulated from the environment and trapped for long periods of time. Examples have used pH oscillations to drive nucleic acid strand separation, which can be caused either by differential thermophoresis of ionic species or by periodic freeze-thaw cycles [98–100]. Also, dew droplet cycles in a rock pore subjected to a temperature gradient can periodically melt

strands by transiently lowering the salt concentration [101, 102]. Heated gas-water interfaces were also shown to promote many prebiotic synthesis reactions [60, 90, 103].

The above scenarios require temperature gradients or thermal cycling. This creates degradation stress for nucleic acids and limits the scenarios to geological settings with a thermal gradient. Here, a simple and ubiquitous scenario is investigated in which a water flux through a rock pore is dried by a gas flux at constant temperature (Figure 2.1). This can be found in the vicinity of underwater degassing events, where gases percolate through rocks to reach the surface, or in porous rocks at the surface exposed to atmospheric winds [104, 105]. Such a setting would be very common on volcanic islands on early Earth which also offered the necessary dry conditions for RNA synthesis [17].

## 2.3 Results and Discussion

### 2.3.1 Flow Dynamics

To experimentally investigate the scenario depicted in Figure 2.1, a microfluidic chamber made up of a cutout from a teflon sheet was designed matching the geometry. Teflon was chosen for the reasons already described in 1. The geometry was comprised of an upward channel that guides a flow of water towards a perpendicular channel above. At their point of intersection, water evaporates as the air flux brushes across. The channels were 1.5 mm in width and 250 µm in depth, sandwiched between two sapphire plates to allow for imaging using fluorescence microscopy with a self built setup (See Figure 2.11, SETUP A). A complete sketch of all components used to make up the chamber, see Materials and Methods Section 2.12.

Using an Elveflow AF1 dual pump system, a precise and constant flux of air was guided through the perpendicular channel, while Nemesys syringe pumps ensure a constant influx of water from below. At 45 °C, the water inflow rate was adjusted to match the evaporation rate at the interface, such that a stable interface was achieved. In order to experimentally scrutinize how the air flux across the water surface influences molecules in the bulk below, fluorescent bead measurements were deployed first. These 0.5 µm sized spheres allow for individual tracking and speed measurements with the microscope. The left panel in Figure 2.2(a) shows an overlayed stack of 5 images in 50 ms time difference from each other. The resulting image highlights the flowpatterns that can be observed. Far below the interface, the inflow speed of around 4 nl/s ≈ 10 µm/s dictates how fast beads are moving. Closer to the walls, the speed significantly decreases because of the parabolic flow profile (See Figure 2.17). Closer to the interface, however, beads feel the momentum transfer of the air flux across. In the air channel, the flux of 250 ml/min reaches up to 10 m/s. Accounting for the parabolic flow as well as the three orders of magnitude smaller viscosity of air compared to water, the resulting speeds in water only reach up to 1 mm/s right at the interface. To measure this, a self-written particle tracker LabView program was deployed to trace the beads and measure the flow speed distribution. The middle panel of Figure 2.2(a) shows the traces measured in this way, with the color scale displaying local flow speeds. Directly at the interface, the speeds were too fast for the camera to trace, such that there is a black spot visible.

Using COMSOL Multiphysics, a 2D projection of the chamber was designed (See Appendix 2.D). Using the finite elements method, Navier Stokes equations were solved both for air and water. Additionally the evaporation of water at the interface was simulated as well. The

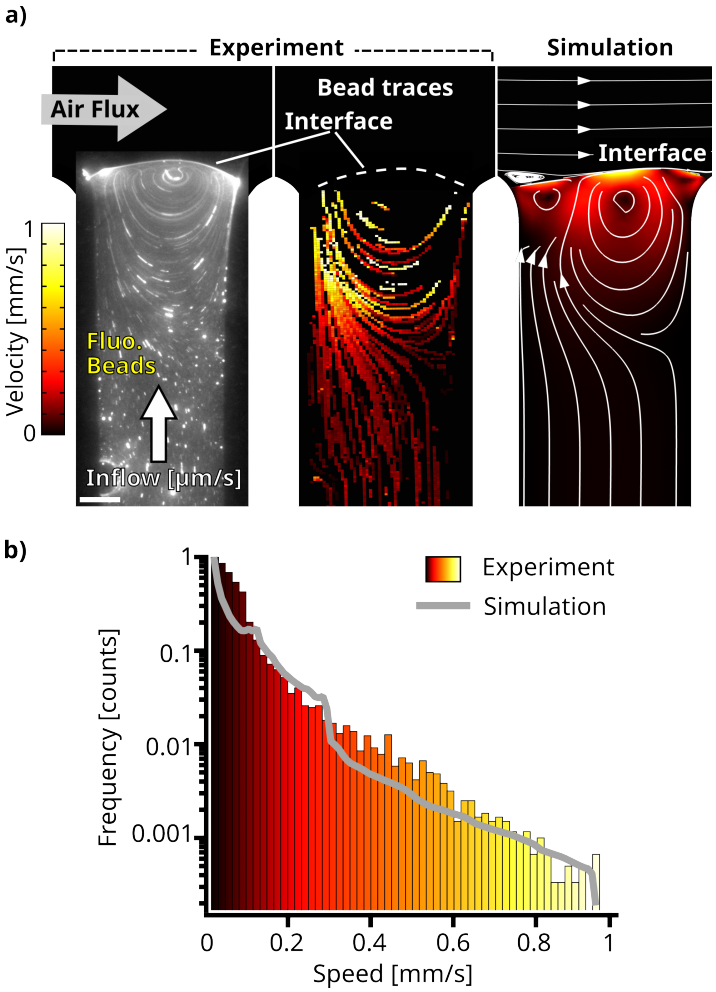


**Figure 2.2:** *Flow dynamics at the air-water interface.* **(a)** (left) Micrograph of fluorescent bead measurement. 0.5 µm beads are flushed towards the inerface at 10 µm/s and reveal a vortex flow pattern generated by a perpendicular air-flux of ≈ 10 m/s at 45 °C. (middle) Tracing of beads reveals its flowspeed distribution reaching up to 1 mm/s close to the interface. (left) Finite elements simulation of the same geometry in 2D. **(b)** Histogram comparing flowspeeds from experiment and simulation.

right panel in Figure 2.2(a) shows the simulated flow speeds, with the same color scale as for the experimentally traced beads. Figure 2.2(b) shows qualitative agreement between the measured flows from experiment and simulation. Important to note is that most beads are very slow and move with speeds around 15 µm/s, dictated by the inflow. In both the simulation as well as experiment, a circular flow pattern (vortex) right below the interface was observed. The flow lines in Figure 2.2(a) show how the upwards water flux reaches the interface on one side of the vortex, whereas on the opposite side, the beads are pushed back down into the bulk.

### 2.3.2 Molecule Accumulation

With the gained understanding of the induced flows in the system, the question of how biomolecules such as DNA and RNA are affected, became of interest. As only water evaporates

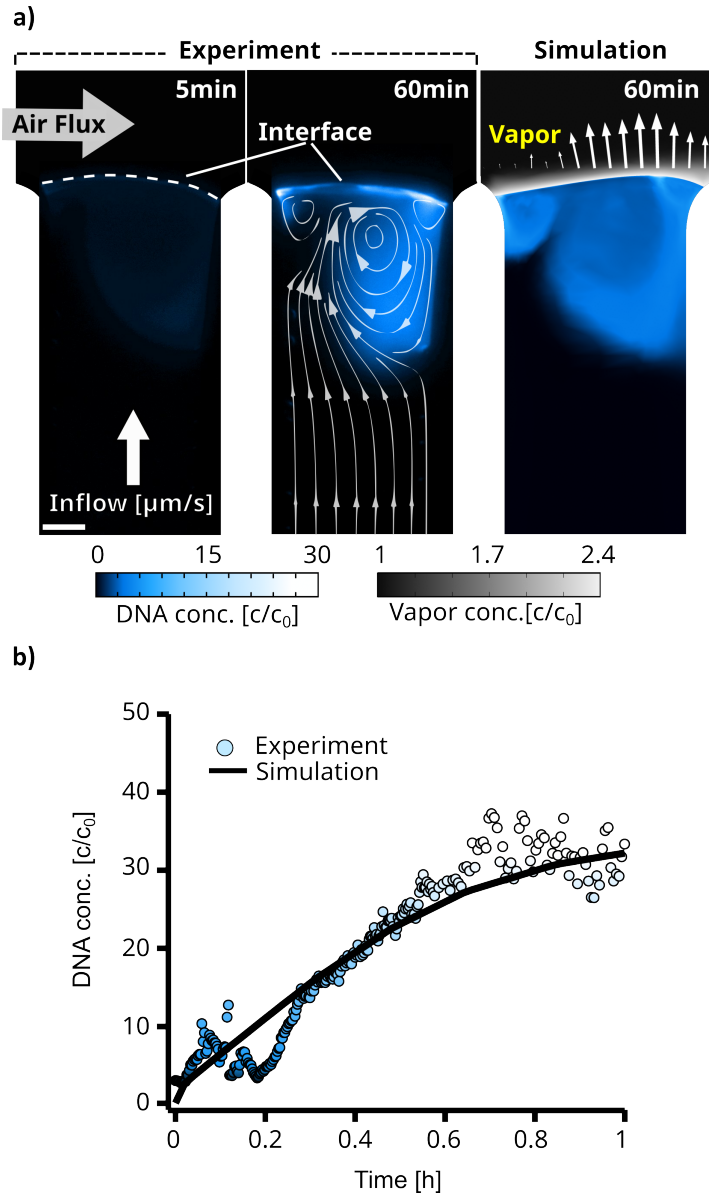


**Figure 2.3:** *Accumulation of DNA at the air-water interface.* **(a)** Fixed volume of fluorescently labelled 63nt long DNA was imaged over time, while being flushed to the air-water interface by a pure water inflow. After 1h, the concentration reaches up to 30-fold compared to the initial $c_0$. Overlaying the simulated flow lines with the micrograph of the experiment (middle panel) shows good agreement between numerical and experimental data. Blue color scale shows accumulation fold for experiment and simulation at the same time, while the gray color scale shows the simulated vapor concentration right above the interface. Arrows in right panel proportionally show evaporation speed along the interface. **(b)** Maximum relative concentration of experiment and simulation measured over time. Reaching up to 30 X the initial concentration, the values follow the simulation well.

at the surface, but molecules are forced to remain in bulk, an accumulation of molecules close to the interface is to be expected. To investigate this experimentally, a 63 nt long DNA strand labelled with FAM (abs. 490 nm, em. 520 nm) was dissolved in water at 5 µM. 10 µl of

this solution were then slowly pushed towards the interface at 45 °C, analogous to the bead experiments before. Behind the 10 µl sample, pure water was injected. Despite this continuous dilution scenario, DNA accumulated at the interface, as only water could evaporate. The left panel in Figure 2.3 (a) shows the observed accumulation using the labelled DNA after 5 and 60 minutes. While in the beginning almost no accumulation can be observed, after one hour, DNA up-concentrated to 30 fold. This accumulation is not only localized at the interface, but also below in the bulk. In the region where the flows create a vortex, DNA is trapped and accumulates there. Meanwhile, the pure water flux from below makes its way upwards to the interface on the left hand side of the vortex. This arrangement was observed to be stable for most parts of the experiment. Through pressure and flow fluctuations, however, the conformation changes and is more turbulent, until a new steady state is reached.

These results were also confirmed with finite elements simulations. For this, a diffusion constant of 95 $\mu m^2/s$ was taken from [85]. The right panel in Figure 2.3(a) shows the simulated state after one hour. Note how similar the accumulation dynamics are comparable to the experimental data. The overlayed flowlines in the middle panel are taken from the simulation and show how high concentration of DNA coincides with vortex flows. The evaporation speed along the interface boundary was also simulated and varies between 10.5 to 5 µm/s (See Figure 2.18). The arrows in the right panel in Figure 2.3(a) are proportional to the evaporation speed and the grey color scale represents the vapor concentration relative to the initial $c_0$.

The maximum accumulation fold of the 63mer DNA was followed over time for both simulation and experiment (See Figure 2.3(b)). Starting to saturate after one hour, DNA accumulated up to 30 fold with simulation and experiment showing good agreement. This is striking, since the accumulation takes place against a constant diluting pure water inflow from below. This creates a micro environment in which DNA is concentrated far above the surrounding average concentration, possibly enabling chemical reactions that would otherwise not work.

### 2.3.3 Hybridization Dynamics using FRET

The most important feature of nucleic acids is their ability to hybridize into double helices via base-pairing and hence form secondary structures. Especially for replication, this concept is crucial, as any copying mechanism requires base-pairing to read out the information stored in the strand that is to be replicated. Usually, an elevation of temperature can separate strands but is accompanied with a higher risk for hydrolysis. The chosen isothermal setting requires changes in salt concentration for this process. More specifically, the circular fluid flow at the interface provided by the gas flux, together with Brownian motion, was expected to drive cyclic strand separation by forcing nucleic acid strands through areas of varying salt concentrations.

To investigate how the system at hand influences the hybridization dynamics of dissolved DNA or RNA strands, a DNA FRET (Förster Resonance Energy Transfer) pair was deployed. DNA was chosen as a model system for the more prebiotically plausible RNA, since DNA is less fragile and is very similar in base-pairing to RNA, which is the main feature relevant to this work [66]. For this purpose, a second fluorescence microscopy setup using an OPTOSPLIT and a dual illumination protocol was built (See Figure 2.11 SETUP B). With this, both dyes could be illuminated individually at the same time, while the camera simultaneously receives two separate images, one for each emission wavelength. Through this, FRET can be calculated per pixel (See Appendix Section 2.B), accounting for background and cross-talk between the two

channels (Figure 2.15). A high FRET signal indicates that two DNA strands are bound, while a low FRET signal indicates that the strands are separated. In this way, FRET becomes an indirect measure of the salt concentration, since a low salt concentration will induce strand separation due to the reduced ionic shielding of the charged DNA or RNA backbones. The two 24 nt long complementary DNA strands had the FRET pair dyes, ROX and FAM, positioned centrally on the opposing strands. First, melting curves using FRET were performed for

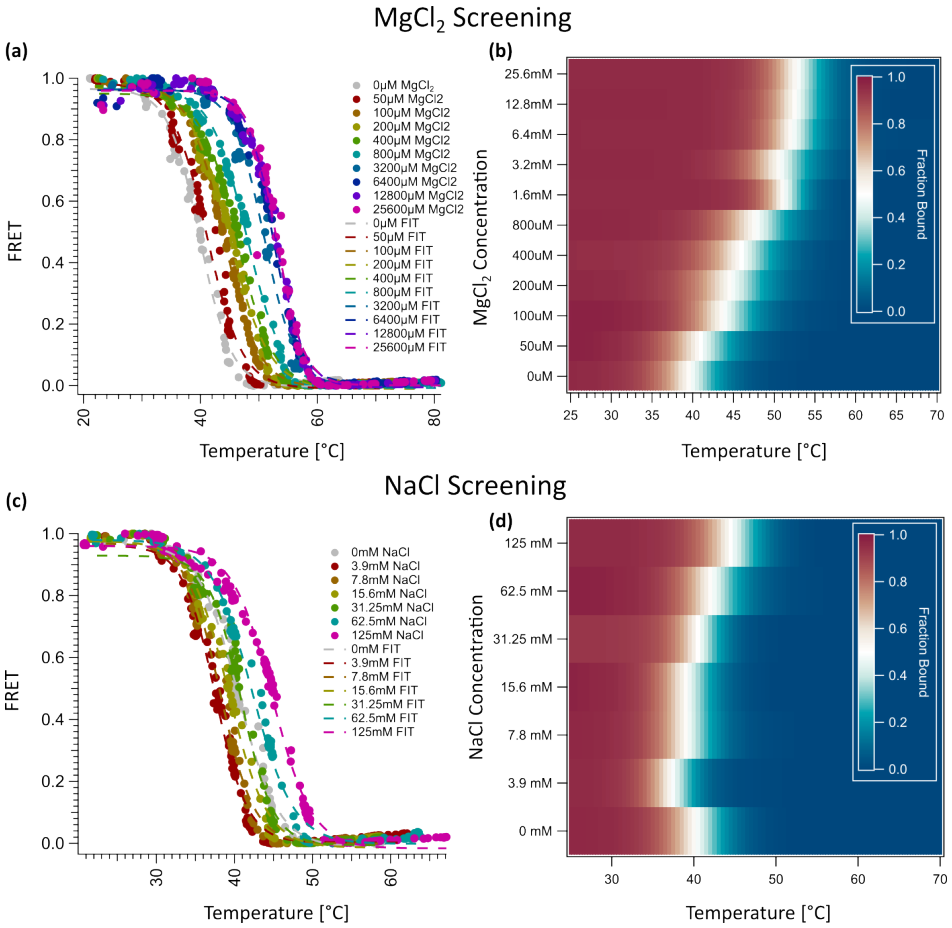


**Figure 2.4:** Melting curves performed in the FRET setup using the 24mer strands labelled with ROX and FAM respectively. **(a)** Raw data of the melting curves of different $MgCl_2$ concentrations with $\alpha$ and $\beta$ already applied for normalization. **(b)** Data from a displayed as a heatmap. The white areas display a fraction bound of 0.5 corresponding to the melting temperature $T_m$. Note that small oscillations in $Mg^{2+}$ strongly influence the melting temperature, which can enable strand separation at isothermal settings. **(c)** Raw data of the melting curves of different NaCl concentrations with $\alpha$ and $\beta$ already applied for normalization. **(d)** Data from c displayed as a heatmap. The white areas display a fraction bound of 0.5 corresponding to the melting temperature $T_m$.

varying $MgCl_2$ and NaCl concentrations. Figure 2.4(a) and (c) show the corresponding melting curves as a function of temperature, while (b) and (d) display the data as a heatmap as a function of temperature and salt concentration. The temperature used in the following FRET experiments was 45 °C. For both NaCl and $MgCl_2$, this temperature results in a low fraction bound below 0.5. When the salt concentration is increased, more and more strands can hybridize. As visible in the heatmaps, varying NaCl concentrations have a far lower impact on

this behaviour than $MgCl_2$, because $MgCl_2$ carries twice the positive charge. For the following experiment in the chamber, FRET will be used as an indirect measure of salt, in particular $MgCl_2$, concentration.

1 µl Sample (10 mM TRIS at pH 7, 50 µM $MgCl_2$, 3.9 mM NaCl, and 5 µM of each DNA strand) was injected into the chamber at 45 °C and flushed towards the interface by pure water with all other conditions same as before. The accumulation of salts and DNA close to the interface was then expected to enable hybridization, while the vortex flows locally vary the concentration of salts and DNA, inducing strand separation.

Figure 2.5 (a) shows micrographs of different time points from the experiment. Initially, after 3 min, the sample containing DNA and salts is still homogeneously distributed in the chamber, leading to a weak FRET signal, indicating that most DNA strands are unbound (blue signal). This was expected, since the $MgCl_2$ concentration of 50 µM at 45 °C is not sufficient for double strand formation (See Figure 2.4(b). Closer to the interface, DNA begins to weakly accumulate, resulting in slightly higher FRET values in the area where the flows are circular. Again, the flow lines were taken from the simulation and overlayed, highlighting the coincidence of accumulation and vortex flows. Only 4 minutes later, a clear separation of high FRET (green) and low FRET (blue) areas can be observed. Note that white pixels at the interface are produced by dried labelled DNA, were FRET can not be calibrated and can therefore assume values above 1. After 16 minutes, the interface has already moved, causing DNA to be left in dry state, when the interface recedes downwards. However, the separation of low and high FRET areas remains stable, as also shown in Figure 2.5(b), where the FRET values in the highlighted ROIs is plotted over time.

These results give rise to the question on whether this system could periodically melt and re-hybridize nucleic acid strands, enabling exponential replication. The right panel in Figure 2.5(a) shows the idea conceptually. In **1.** nucleic acids can hybridize in areas of high salt concentration. If primers are present and a replication mechanism is in place, then the double strand can be completed through base by base polymerisation. This leads to **2.** a fully bound double helix of parent and daughter strand. Without strand separation, the reaction would stall here, as no new primers could bind to the dsDNA. However, if the vortex flows are change local salt and DNA concentration sufficiently, then in **3.** strand separation could occur, finishing the replication cycle.

### 2.3.4 Monte Carlo Random Walk

To test this concept, the accumulation of $MgCl_2$ was simulated analogously to the previous finite elements simulations using a diffusion constant of 705 $\mu m^2/s$ at 45 °C (See Appendix 2.D). A single DNA strand was then followed theoretically, using a Monte Carlo random walk simulation (See Appendix Section 2.C). An exemplary simulated path is plotted on top of the $MgCl_2$ simulation from COMSOL Multiphysics shown in Figure 2.6. Starting from a region of low magnesium concentration, the strand is quickly flushed upwards into the circular flows. As its path is mostly dominated by the vortex flows, Brownian motion allows it to diverge from these paths, drastically changing the local salt concentration periodically. Figure 2.6(b) shows the local salt and DNA concentration plotted along its path. Notably, the concentration of product or template strand as well as the local salt concentration drop multiple times within 30 minutes by 3-4 X the initial value $c_0$. As shown in Figure 2.4, this change in local $MgCl_2$ concentration has a strong impact on the melting temperature of dissolved nucleic

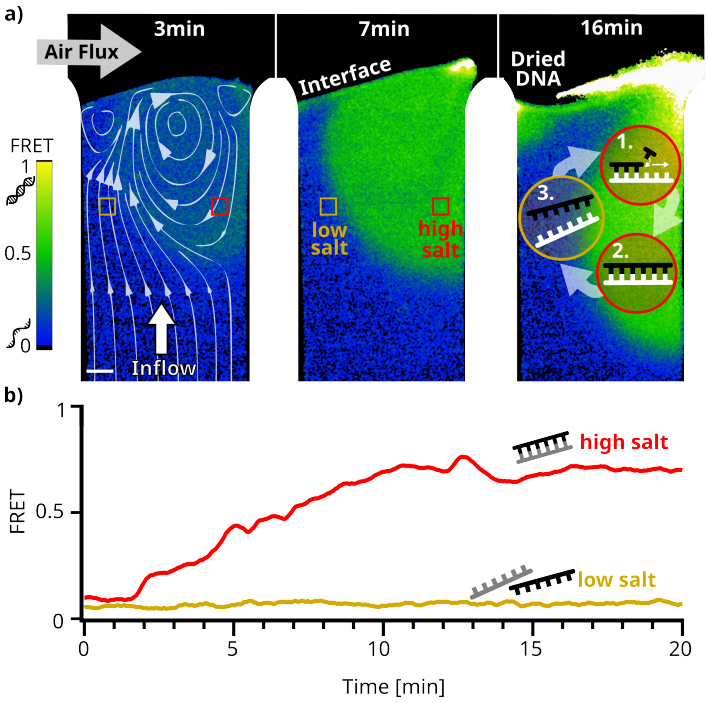


**Figure 2.5:** *Förster Resonance Energy Transfer (FRET) measurements.* **(a)** Micrographs of 24nt long DNA FRET pair in the microfluidic chamber at 45 °C. 1 µl sample (5 µM DNA, 10 mM TRIS pH7, 50 µM MgCl2, 3.9 mM NaCl) was subjected to a 3 nl/s diluting upflow of pure water and a gas flow of 230 ml/min across. The vortex (here shown by simulated flow lines) overlays with regions of high FRET signal, indicating dsDNA. This setting provides the possibility for replication reactions by (**1+2**) strand replication in the high salt region and (**3**) strand separation of template and replicate in the low salt region. **(b)** Measurement of FRET signal over time in the low salt and high salt ROI (indicated in left and middle panel). After initial accumulation of salt at the interface, high and low salt regions remain stably separated.

acid strands. This could enable exponential replication in this isothermal and prebiotically plausible environment.

### 2.3.5 Isothermal Replication with PCR

Nucleic acids and salts accumulated near the interface, but far from the interface, in the bulk below, the concentrations remained vanishingly low due to the diluting inflow of pure water. The air flux induced an accumulation pattern of vortices in which molecules were trapped. The salt and DNA concentration changed cyclically, resulting in possible periodic strand separation of nucleic acids. To put to a test, whether isothermal replication is possible in this gas flux driven rock pore environment, a model system using *Taq* Polymerase was designed. Taq DNA Polymerase was chosen, because it does not have a protein-based strand separating mechanism. This way, the exponential replication can only proceed, if an external mechanism can separate the strands after each round of copying. Starting with a 51mer template and two 30mer primer strands, each with a 5′-AAAAA overhang for detection, the reaction is expected to form a 61mer replicate (Detailed scheme in Figure 2.7), the same length as the DNA used in the random walk model in Figure 2.5. In this way, the original 51mer

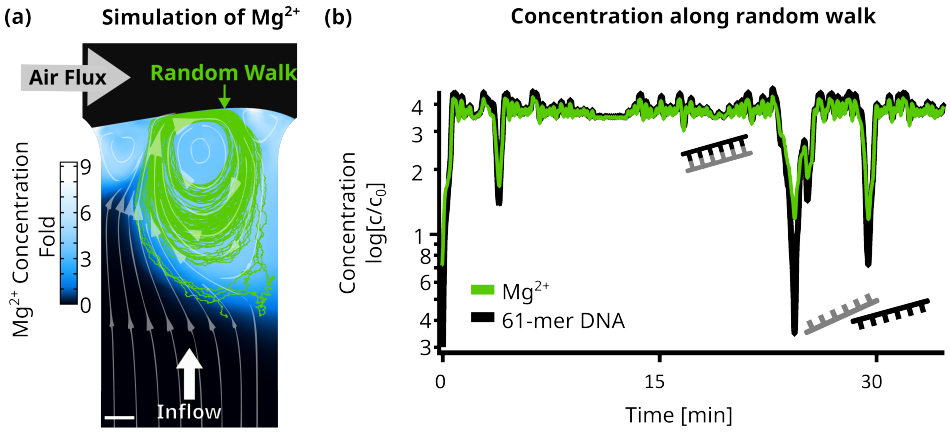


**Figure 2.6: (a)** Comsol simulation of $Mg^{2+}$ ions (D = 705 $\mu m^2/s$ in the chamber agreed with the FRET signal in Figure 2.5 and showed up to 9-fold salt accumulation at the interface. The path of a 61mer DNA molecule from a random walk Monte Carlo model is shown by the green lines, while the white flow lines are taken from the simulation. **(b)** Concentrations along the DNA molecule path in (a) show oscillations relative to the initial concentration of up to 3-fold for $Mg^{2+}$ and 4-fold for 61mer DNA. This could enable replication cycles, as the vortex provides high salt concentrations for replication near the interface, while drops in salt and template concentrations regularly trigger strand separation.

template is not reproduced, which facilitates detection with gel electrophoresis, as bands do not overlap. As *Taq* polymerase has an efficiency peak around 72 °C, a temperature (68 °C) close to that value was chosen. Above 68 °C, bubble formation due to degassing became unavoidable in the microfluidic setting. Sample conditions were 10 µl of the reaction mix (0.25 µM primers, 5 nM template, 200 µM dNTPs, 0.5 X PCR buffer, 2.5 U *Taq* polymerase, 2 X SYBR Green I). This reaction mixture was then exposed to a constant pure water influx of 5 nl/s towards the gas-water interface, matching the rate of evaporation at the interface.

In order to monitor the replication in real time, the intercalating dye SYBR Green I was used. It binds to the minor grove site of the DNA double helix and fluoresces three orders of magnitude stronger, when bound to a DNA in duplex form [69]. As more template and product are formed, SYBR Green I fluorescence was expected to increase, as more duplex DNA strucures will be present. Figure 2.8(a) shows a reduced version of the scheme displayed in Figure 2.7. After the 51mer templates are all elongated by the primers with overhang, the replication becomes exponential, as all produced strands serve as templates for the next replication cycle.

Through the oscillations in salts and DNA observed along the random walk, the 61mer product strand was expected to be able to separate from its respective template strand, enabling exponential replication. Figure 2.8(b) shows fluorescence micrographs of the reaction in the chamber at different time points. Initially, minimal fluorescence is seen. This indicates that the replicated templates are below the detection limit of SYBR Green. Then, the air flux as well as the water inflow were switched on and calibrated such that the interface remained stable. After two hours, SYBR Green fluorescence began to increase, indicating that replication was taking place and the products are forming duplex structures. When the exact same experiment was performed without the air and water fluxes, no fluorescence increase

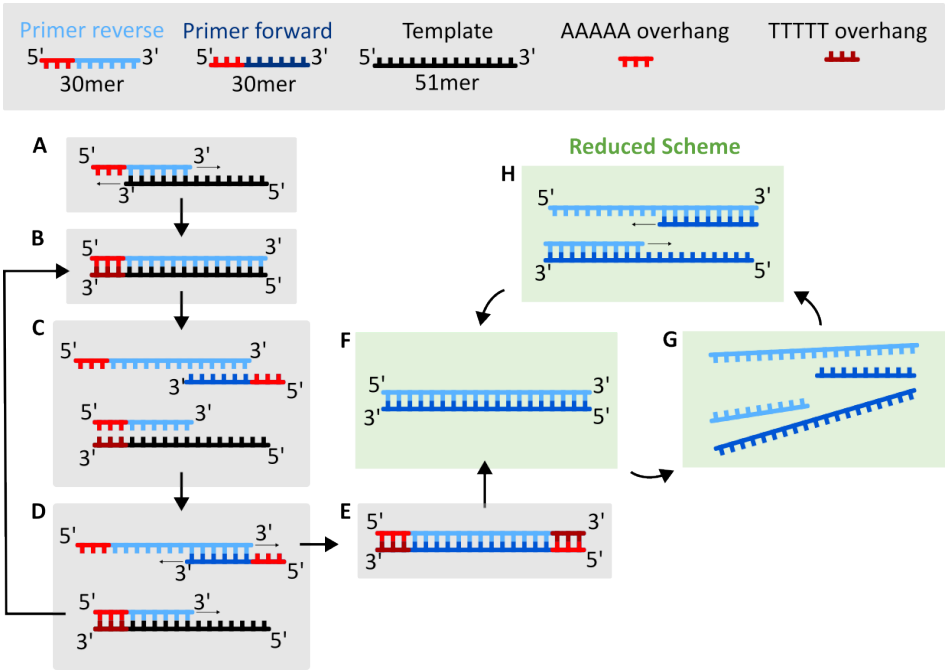


**Figure 2.7:** Sketch of the replication pathway with *Taq* polymerase. **A** The 30mer reverse primer binds to the 51mer template. The 5-A overhang remains unbound. **B** *Taq* polymerase adds nucleotides from 3′ to 5′ and a complete double strand is formed, now containing 56 base pairs. **C** After strand separation, the newly formed 56mer intermediate product with a 5-A overhang at the 5′ end and the intermediate 56mer template with a 5-T overhang at the 3′ end are bound by the primers. **D** In one case, *Taq* elongates the forward primer bound to the intermediate product and proceeds to step E. In the other case, elongation of the reverse primer bound to the intermediate template leads back to step B. **E** The result of the extension is a new product and a new template of 61 bases each. From here, the cycle enters the "reduced scheme" referred to in Figure 2.8. The newly formed products, together with the original primers, now replicate exponentially: **F** The double strand formed in E can now be considered as product and template for the reduced scheme. **G** After de-hybridization, both primers can anneal to template and product. **H** *Taq* extends again from 3′ to 5′, forming two new double strands of template and product, doubling their amount and completing one cycle of exponential amplification.

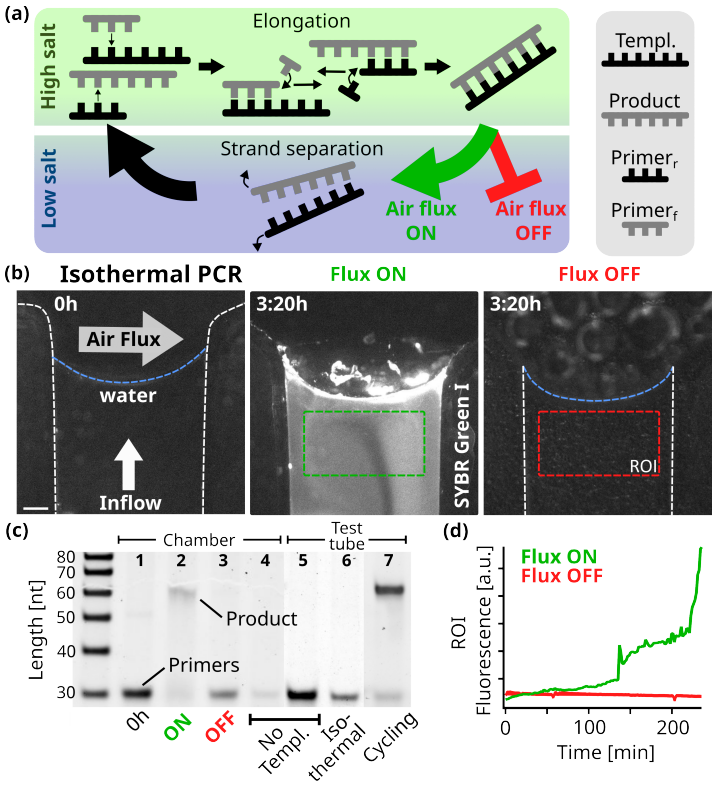


**Figure 2.8:** *Isothermal Replication.* **(a)** Sketch of the replication reaction pathway. In high salt regime (green), primers anneal and elongation by *Taq* Polymerase takes place. Only with air-flux ON, strand separation can take place in low salt regime (blue) and the reaction can proceed. **(b)** Fluorescence micrographs of the PCR reaction in the chamber at isothermal 68 °C. With an influx of 5 nl/s and 250 ml/min gas flux across, the reaction shows strong increase in SYBR Green I fluorescence over time indicating replication. This is not observed, when the fluxes are turned off. Sample conditions: 0.25 µM primers, 5 nM template, 200 µM dNTPs, 0.5 X PCR buffer, 2.5 U *Taq* polymerase, 2 X SYBR Green I. Scale bar is 250 µm. **(c)** Polyacrylamide Gel Electrophoresis (15 %) of the reactions and neg. controls. After 4 hours in the reaction chamber with air- and water-flux ON, the 61mer product was formed under primer consumption **(2)**, unlike in the equivalent experiment with the fluxes turned OFF **(3)**. At the beginning of the experiment **(1)** or in the absence of template **(4)**, no replicated DNA was detected. The reaction mixture was tested by thermal cycling in a test tube **(5-7)**. As expected, replicated DNA was detected only with the addition of template: **(7)** shows the sample after 11 replication cycles. The sample was also incubated for 4 hours at the chamber temperature (68 °C) yielding no product **(6)**. Primer band intensity variations are caused by material loss during extraction from the microfluidic chamber. **(d)** SYBR Green I fluorescence increased when gas and water flow were turned on, but remained at background levels without flow. Fluorescence was averaged over time from the green and red regions of interest shown in (b). SYBR Green I fluorescence indicates replication, as formed products are able to hybridize.

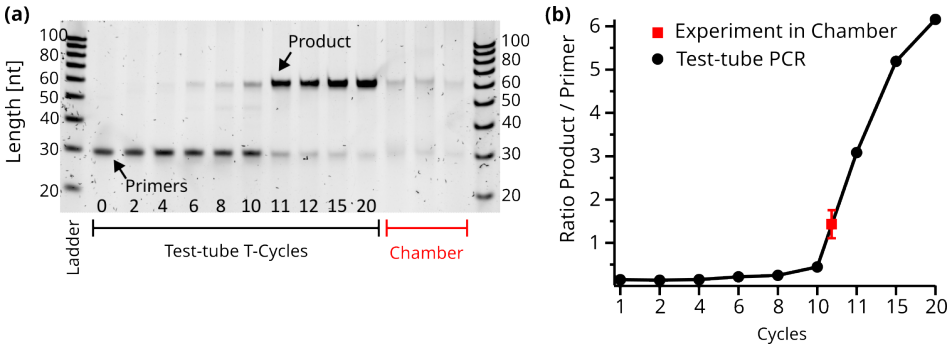


**Figure 2.9:** Comparison of test-tube temperature cycling vs. chamber experiments, to reveal how many full cycles of replication were performed in the chamber. **a)** Full Sample (10 µl of 2.5 U of AllTaq polymerase, 2 X SYBR Green I, 5 nM template, 0.25 µM of each primer, 200 µM of each dNTP and 0.5 X PCR buffer) were subjected to 0 to 20 temperature cycles. The more cycles, the more primers are consumed. The three experiments performed in the chamber show a generally lower gel intensity, probably due to losses during sample extraction from the microfluidic chamber. **b)** Comparison of Product/Primer intensity-ratio of the test-tube sample to the extracted chamber samples. This reveals that in the 4h chamber experiment about 10-11 cycles were performed. Error bar of the experimental data point is the standard deviation of the 3 independent chamber samples in (b).

could be observed. Figure 2.8(d) shows the fluorescence measured over time for both cases, clearly indicating that in the case of air- and water-flux switched on, replication took place.

After 4 hours of experiment, the fluxes were switched off and the sample was extracted from the chamber using the syringe pumps and pipetting once the chamber was opened. Material loss due to adhesion to surfaces or pipetting errors etc., could not be avoided in this case. The extracted samples were then loaded onto a 15 % Polyacrylamide Gel and measured using gel electrophoresis (more details in Materials and Methods Section 2.5.7). The results are shown in Figure 2.8(c). Clearly visible here is that the initial sample (**1**) can only form the 61mer product in the chamber with the fluxes turned on (**2**). Without the fluxes on, no product is formed (**3**). When no template strand is added to the sample, no product is formed in both chamber and test tube experiments (**4+5**). Control experiments were performed in a test tube with cycling temperatures for annealing, elongation and strand separation steps. The exact protocol and all control experiments can be found in Appendix Section 2.A. To confirm that *Taq* requires strand separation for exponential replication, the test tube was held at 68 °C for the same amount of time the chamber sample was run (**6**). No product formation could be observed in this case. Only when the temperature was periodically varied, primers were used up to form the 61mer product strand in the test tube. Notably, the ratio of primer vs. product fluorescence in **2** vs. **7** is different. While in the test tube sample the primer band is faint and the product band is bright, both bands are faint in the chamber sample case. The reason for this is the aforementioned material losses during sample extraction from the chamber. As the same amount of product, template and primer would be lost during extraction errors, the ratio between product and primer brightness on the gel should remain unchanged. To find out how many temperature cycles of annealing, elongation and strand separation the sample underwent in the chamber, multiple control experiments with different amounts of temperature cycles were performed.

Figure 2.9(a) shows the corresponding gel of these experiments, as well as the triplicate performed in the microfluidic chamber. The more cycles are performed, the more primers are used up and converted to product. The product/primer fluorescence ratio from the gel plotted (Figure 2.9(b) against the amount of cycles performed reveals that the sample in the microfluidic chamber underwent an equivalent of around 10-11 temperature cycles within 4 hours.

## 2.4 Conclusion

In this study, a prebiotically plausible and abundant geological environment was investigated to trigger replication of early life. An isothermal setting was considered, where gas flows over an open rock pore filled with water. Previously, thermal gradients have been utilized to separate nucleic acid strands in prebiotic settings, which carries the risk of degradation due to the high temperatures involved. In contrast to this, the combined gas and water flow at an open pore was found to trigger salt oscillations, supporting oligonucleotide replication at constant temperature.

The system was initially probed using fluorescent bead and DNA measurements, with results found to align with fluid dynamics theory through finite element simulations. Tracking of the beads reveals a vortex flow close to the interface, where DNA accumulates. Using a 24 nt long DNA FRET pair, fluorescence microscopy reveals a clear separation of single and double stranded regions. Deploying a Monte Carlo random walk model revealed oscillations in nucleic acid and salt concentrations produced by the interplay between molecular accumulation and interfacial flow. This process has shown to periodically separate nucleic acid strands under chemically gentle conditions.

Given the limitations of RNA-based replication, the environment was further tested with protein-driven DNA replication, leading to the discovery of isothermal replication within this common geological micro-environment. Using a 51 nt long template, 30 nt primers and *Taq* polymerase, about 10 full replication cycles were achieved in 4 hours in the flow chamber. The progress of replication was monitored using SYBR Green I and Gel electrophoresis, revealing product formation in isothermal conditions.

This finding demonstrates that such an environment provides a viable setting for early nucleic acid replication chemistry. Physical non-equilibria, such as steep temperature gradients, often impose restrictive boundary conditions, making the discovery of such settings less likely. However, this isothermal environment significantly expands the range of prebiotic settings that could facilitate replication on early planets. Especially further away from heat sources, such as hydrothermal vents, temperature gradients are rare and shallow. The induced non-equilibrium environment by a mere gas flux could have played a significant role to foster early life on earth.

## 2.5 Materials and Methods

### 2.5.1 Bead Measurements

Fluorescent beads (0.5 µm diameter) were purchased from Invitrogen, Eugene, Oregon, USA, Lot: 31373W. Beads were dissolved 1 to 2000 in nuclease free water before 10 µl of this solution was then loaded into the chamber, with a 3 nl/s upward pure water flux flushing them towards the interface with the 230 ml/min gas flux streaming across. Images were captured using a 50 ms exposure time, resulting in approximately 20 fps. The setup did not allow very fast beads to be traced, as the maximum frame rate of 20 fps did not capture particles faster than about 1 mm/s. The 2D map of traced velocities in Figure 2.2 therefore has some dark spots near the interface where the fastest beads are located.

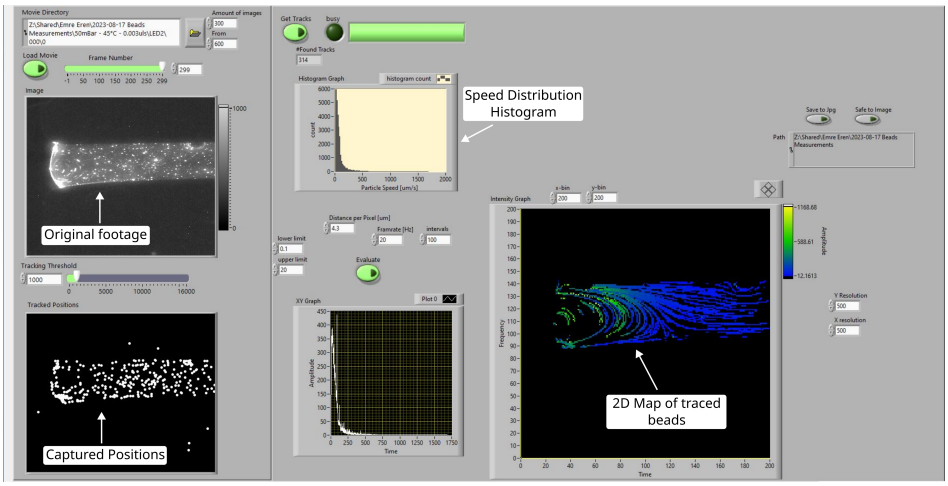


**Figure 2.10:** *Screenshot of self-written LabView particle tracker.* The original fluorescence images taken with 0.5 µm sized beads is loaded and then, after a tracking threshold is set, captured positions are traced over the course of a set amount of images. The corresponding distribution of speeds traced is then displayed in a histogram, as well as in a 2D map.

### 2.5.2 Experimental Setups

All setups used in this work were designed and assembled from components purchased from several companies. Figure 2.11 shows schematics of both setups used. All specifications below are given in nanometers.

Setup A was built to capture fluorescence signals from beads and labeled DNA and was constructed using SM1 lens tubes (Thorlabs), excitation LEDs M490L4 and M625L3-C4 (Thorlabs), excitation filters (470/622H, AHF) and emission filters (497/655H, AHF), and a dual-band dichroic beamsplitter (497/655H, AHF). The camera (Stingray F-145 B/C) was purchased from Allied Vision. An achromatic doublet (AC254-100-A-ML, Thorlabs) was used as the objective.

Setup B was built to capture FRET. A Zeiss Axiotech Vario microscope body was used and equipped with excitation LEDs M590L4 (yellow) and M470L2 (blue) (Thorlabs) with excitation filters BP588/20 and BP482/29 (Thorlabs) coupled into the same beamline via a DC R 475/40 beamsplitter. A dual-band dichroic mirror (505/606 T) and an Optosplit II with a ratiometric

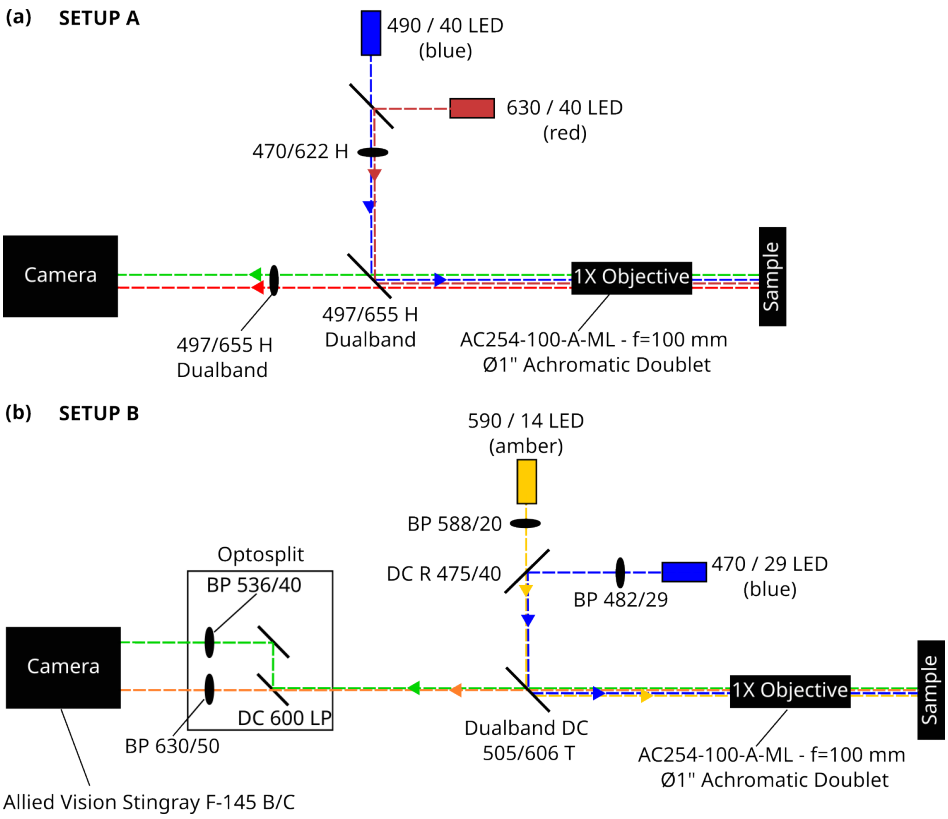


**Figure 2.11:** *Schematic of the fluorescence miroscopy setups used in this work.* **SETUP A** was used for DNA accumulation and bead measurements, where the dyes SYBR Green I or FAM were used. For both excitation is around 490 nm and emission around 520 nm. **SETUP B** was used for FRET measurements, using two alternating LEDs to illuminate donor and acceptor. An OPTOSPLIT is then used to split the resulting image in two parts, depending on the emission wavelength.

filter set (DC600 LP, BP630/50 and BP 536/40) were coupled to the system to capture the FRET signal of the dyes ROX and FAM. Camera and objective are identical to setup A.

For both setups, fluorescence measurements were performed under curtains to ensure low background noise. Temperature was controlled by a JULABO Corio CD water bath connected to the microfluidic chamber (See Materials and Methods Section 2.5.4) by thermally insulated tubing. The temperature was measured directly at the back sapphire surface of the chamber. Gas flow was generated using an AF1 Dual Pump (Elveflow) system with ambient air as the gas source, and gas flow rate was measured with a flow sensor (FS2000, from IDT). Syringes were driven by a Cetoni syringe pump (Nemesys).

### 2.5.3 DNA Strands

DNA stocks were purchased in dry state from biomers.net in HPLC-grade purity and kept at -20 °C for storage after dilution with nuclease free water (Ambion nuclease-free water from Invitrogen) to 100 µM. Before every experiment, the strands were denatured at 95 °C for 2 minutes. For FRET experiments, the strands were then slowly annealed to room temperature before loading them into the chamber for measurement.

| Strand | 5′- Sequence -3′ | Label |
|---|---|---|
| 63-mer | CCAGCCTCCAGTGCCTCGTATCATT GTGCCAAAAGGCACAATGATACGAGGCACTGGAGGCTG | 5′ FAM |
| 24-mer FRET strand 1 | CGTAGTAAATA8CTAGCTAAAGTG | 8 = FAM |
| 24-mer FRET strand 2 | CACTTTAGCTAGA8ATTTACTACG | 8 = ROX |
| 51-mer Template | TTAGCAGAGCGAGGTATGTAGGCGG GACGCTCAGTGGAACGAAAACTCACG | - |
| 30-mer forward primer | AAAAA TTA GCA GAG CGA GGT ATG TAG GCG G | - |
| 30-mer reverse primer | AAAAA CGT GAG TTT TCG TTC CAC TGA GCG T | - |

**Table 2.1:** DNA sequences as ordered from biomers.net.

### 2.5.4 Microfluidic Chamber

The microfluidic chamber used for the experiments in this work was constructed as follows: Two 60x22 mm sapphire plates (front 1 mm thick, back 0.5 mm thick) were used to sandwich a 250 µm thick Teflon foil from which the channels were cut using a Graphtec CE6000-40 Plus plotter. Sapphire plates were used because of their higher thermal conductivity compared to quartz glass (30.3 W/(mK) compared to 1.36 at 300 K) [106, 107]. The back sapphire had four holes to allow gas and water flow in and out of the chamber. The sandwich was held together by the steel frame and aluminum back, which were screwed together with a torque of 0.2 Nm. The aluminum back was held in place on the water bath fixture by magnets. A thin (25 $\mu m$) graphite foil was placed between the sapphire back and the aluminum back to increase the thermal conductivity. Figure 2.12 shows an exploded view of the chamber.

### 2.5.5 Finite Elements Simulations

2D finite element simulations were performed using COMSOL Multiphysics 5.4. Fluid dynamics were simulated by solving the Navier-Stokes equation in two dimensions. Parameters used are available in table 2.3 in the Appendix Section 2.D which also contains the full description of the model used.

### 2.5.6 Förster Resonance Energy Transfer (FRET)

FRET imaging was performed using a second custom-built fluorescence microscopy setup consisting of light-emitting diodes (M470L2, M590L2; Thorlabs) combined by a dichroic mirror on the excitation side, while an Optoplit II with a ratiometric filter set (DC 600LP, BP536/40, BP 630/50) and a Stingray-F145B ASG camera (Allied Vision Technologies) through a 1X objective (AC254 100-A-ML Achromatic Doublet; Thorlabs) detected and

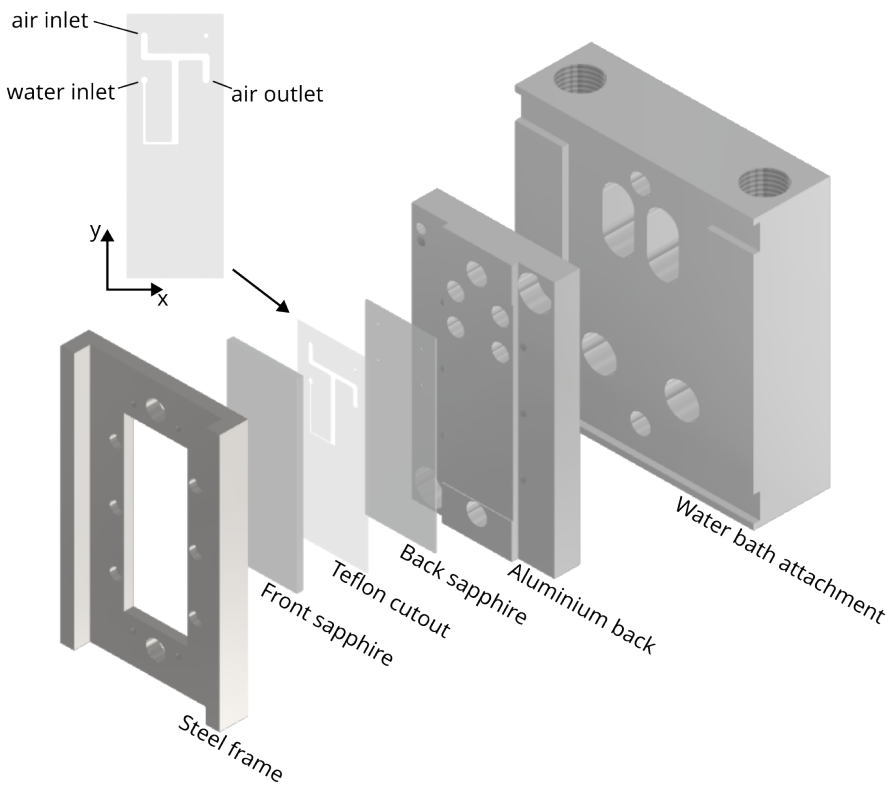


**Figure 2.12:** *Chamber assembly in an exploded view.* The microfluidic chamber cutout of a the teflon foil (250 µm) is sandwiched between a front (1 mm) and back sapphire (1 mm), fixed in place by a steelframe screwed onto an aluminium back with 0.2 Nm. Using magnets, the aluminium back is then attached to the water bath attachment, providing with the temperature control. The air in- and outlet as well as water inlet are accessed by syringe pump fittings.

superimposed both fluorescence emission channels (Appendix 2.B). The DNA sequences used for FRET experiments were: strand 1 5′-CGTAGTAAATAT*FAM*CTAGCTAAAGTG-3′, strand 2 5′-CACTTTAGCTAGAT*ROX*ATTTACTACG-3′. The two labeled complementary strands were diluted from stock solution (100 µM in nuclease-free water) and mixed together to a final concentration of 5 µM in buffer (10 mM TRIS, 50 µM $MgCl_2$, 3.9 mM NaCl, pH7). To promote annealing of the two complementary strands, the solution was heated and slowly cooled from 80 °C to 4 °C (ramp rate of -1 °C per 5 s) in a standard thermocycler (Bio-Rad CFX96 Real-Time System) prior to each experiment. Salt screening in Figure 2.4 was performed in capillaries (0.20 x 2 mm I.D., VITROTUBES, #3520-100, Lot.: 1322J1).

### 2.5.7 Gel Electophoresis (PAGE)

For PAGE and gel imaging, a 15% denaturing (50% urea) polyacrylamide gel with an acrylamide:bis ratio of 29:1 was solidified with TEMED (tetramethylethylenediamine) and ammonium persulfate. 2 µl of sample was mixed with 7 µl of 2 X loading buffer (Orange G, formamide, EDTA), of which 5 µl were loaded onto the gel. Staining was performed with 2X SYBR Gold in 1 X TBE buffer for 5 minutes and the gel was imaged using the ChemiDOC MP imaging station (Bio-Rad).

# Appendix

## 2.A Polyacrylamide Gel Electrophoresis of PCR Reactions

Replication reactions were performed using the AllTaq PCR Core Kit (QIAGEN). Each reaction contained 2.5 U of AllTaq polymerase, 2 X SYBR Green I, 5 nM template, 0.25 µM of each primer, 200 µM of each dNTP and 0.5 X PCR buffer (contains TRIS HCl, KCl, $NH_4SO_4$ and $MgCl_2$). To confirm that the reaction in the chamber followed the predicted scheme, a series of control experiments was performed. The experiment was performed either in the microfluidic chamber or in a test tube inside a thermocycler (Figure 2.13(a)). After a heat activation step of 95 °C for *Taq* polymerase, the temperature was kept constant at 68 °C in the microfluidic chamber, while in the thermocycler, the PCR protocol for *Taq* was followed, in which the sample underwent multiple cycles of replication (Figure 2.13(b)). After 95 °C, the primers are given time to anneal by cooling the sample to 52 °C, followed by a replication step at 68 °C, where *Taq* is at its peak performance. This temperature protocol is then repeated an additional 39 times to complete the 40 cycles, and the sample is then cooled to 4 °C, extracted, and then stored at -20 °C until PAGE analysis. Figure 2.14(a) shows the PAGE results for the test tube

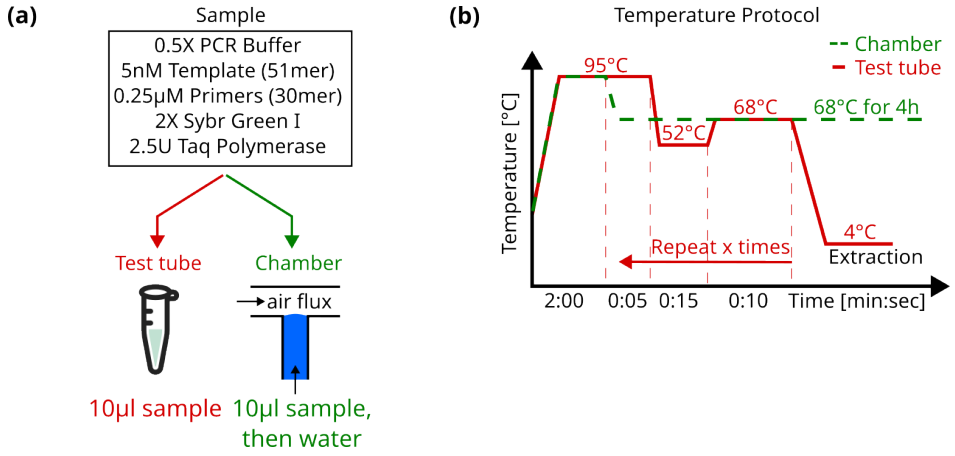


**Figure 2.13:** *Protocols used for PCR experiments in test tube and chamber.* **(a)** 10 µl reaction volume of 0.5 X PCR Buffer, 5 nM Template, 0.25 µM Primers, 2 X SYBR Green I and 2.5 U *Taq* Polymerase are either measured in a reaction tube, or inside a chamber flushed towards the air-water interface with a pure water influx. **(b)** Temperature protocol for the samples in the test tube. A standard Bio-Rad CFX thermocycler was used to perform the procotol. In contrast to the test tube, inside the chamber (green line) the temperature was constant at 68 °C.

samples in the thermocycler. On the left, the triplicate of the full sample (conditions shown in Figure 2.13(a)) shows primer consumption and the formation of a product band in all cases. To be sure that *Taq* is not forming an unwanted side product, such as primer dimers, this experiment was repeated without adding the template strand, and indeed no product strand can be detected. Furthermore, the reaction cannot proceed to generate product without

having both the reverse and forward primers. In the PAGE gel on the right, experiment triplicates are shown without the forward primer, without the reverse primer, or without primers at all. As expected, no product was formed in any of these cases. Without the addition of DNA (no primers and no template), *Taq* polymerase does not form a new strand. As a further negative control, a complete sample in the chamber was kept at isothermal 68 °C, analogous to the experiment, and no product formation could be observed. This is particularly interesting because it shows that without the microfluidic chamber environment, the replication cycle cannot be completed and the reaction is halted, further emphasizing the need for salt cycling in the chamber experiment.

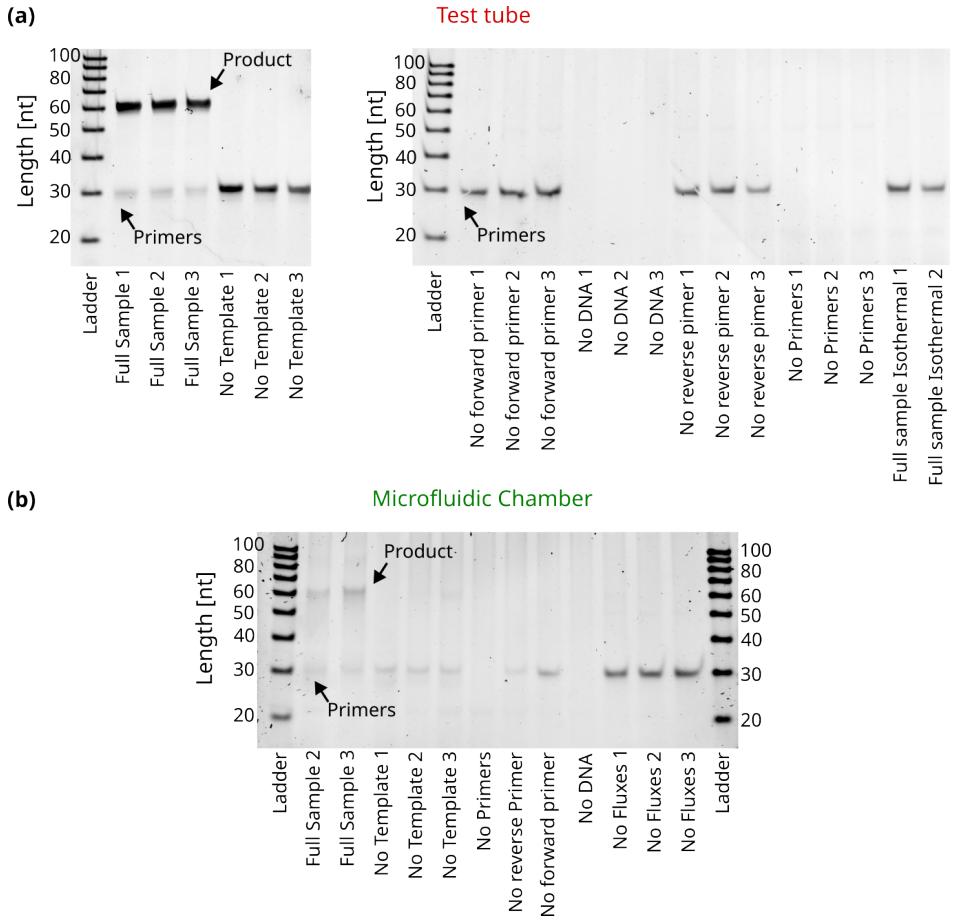


**Figure 2.14:** *PAGE analysis of Taq PCR reactions.* **(a)** PAGE images for the test tube samples. Standard sample composition is the same as in Figure 2.13: 10 µl reaction volume of 0.5 X PCR Buffer, 5 nM Template, 0.25 µM Primers, 2 X SYBR Green I and 2.5 U *Taq* Polymerase. All other samples do not contain individual components, such as template strands or primers. "Full sample isothermal 1&2" have the same temperature protocol as the chamber samples, all other samples were incubated in the temperature protocol of Figure 2.13, red lines. A slight band may be visible around 51 nt caused by the 5 nM of the initial 51mer template. The numbers indicate the n-th replicate of the experiment. **(b))** PAGE image of the chamber control experiments. All experiments were performed in the microfluidic chamber at 68 °C, following the same protocol as before. Primer fluorescence varies between samples, which is caused by variability during sample extraction or primer consumption by *Taq* (in case of "Full Sample").

Figure 2.14(b) shows the PAGE gels of all experiments performed in the microfluidic chamber. While the full sample replicates 2 and 3 show product formation, no product is observed without the addition of the template strand to the reaction mix. When the reaction is run without the reverse primer, forward primer, primers in general, or no DNA at all, no product formation is observed. To show that the replication reaction is only possible when both the gas flow and the water flow are turned on, the experiment was repeated without any fluxes turned on. Here, the full sample was placed in exactly the same microfluidic chamber and kept at isothermal 68 °C as in the other chamber experiments. However, without up-concentration at the interface and without continuous stirring by the gas flow, no product formation was be observed.

## 2.B Förster Resonance Energy Transfer

To measure the FRET signal in the microscope, an alternating illumination protocol was implemented (Supplementary Movie 3). The FRET pair FAM-ROX was excited by two LEDs in rapid succession. The blue LED excited the donor dye (FAM), while the acceptor (ROX) can only be excited indirectly while both dyes are in the FRET region. The yellow LED excited only the acceptor dye (ROX). Individual images of each illumination were captured using an Optosplit II to separate the individual emission wavelengths of FAM and ROX before they reached the camera. This allowed the emission of FAM and ROX to be captured simultaneously for each of the two illuminations, providing four images for each time point: DD, DA, AA, and AD (see table 2.2 for details). The spatially averaged, temperature-dependent, crosstalk- and artifact-corrected FRET signal was calculated using the equation 2.1 [85]. Crosstalk between

| Channel | Excitation | Emission | Label |
|---|---|---|---|
| DD | FAM - 470nm | FAM - 536nm | FAM/ROX |
| DA | FAM - 470nm | ROX - 630nm | FAM/ROX |
| AA | ROX - 590nm | ROX - 630nm | FAM/ROX |
| AD | ROX - 590nm | FAM - 536nm | FAM/ROX |
| $AA_{\boldsymbol{A}}$ | ROX - 590nm | ROX - 630nm | ROX |
| $DA_{\boldsymbol{A}}$ | FAM - 470nm | ROX - 630nm | ROX |
| $DD_{\boldsymbol{D}}$ | FAM - 470nm | FAM - 536nm | FAM |
| $DA_{\boldsymbol{D}}$ | FAM - 470nm | ROX - 630nm | FAM |

**Table 2.2:** Channel definitions for FRET calculation. First capital letter denotes the excitation wavelength (D = Donor, A = Acceptor), second the measured emission wavelength and the subscript stands for the label used in a separate experiment to determine crosstalk related artifacts.

the two channels (aa(T) and dd(T)) was calculated in separate experiments using the same setup parameters with the equation 2.2 and 2.3, respectively. The data used for the crosstalk calculations are shown in Figure 2.15. To test how different salt concentrations affect the FRET signal, melting curves of different salt concentrations were performed in the FRET setup (See Figure 2.4). This revealed that the sodium concentration has little effect on the melting temperature, while $Mg^{2+}$ strongly influences the hybridization state. In the FRET experiment in Figure 2.5, the initial $Mg^{2+}$ concentration was 50 µM at 45 °C. In this state the double stranded fraction is about 0.3. When the salts accumulated at the interface, salt

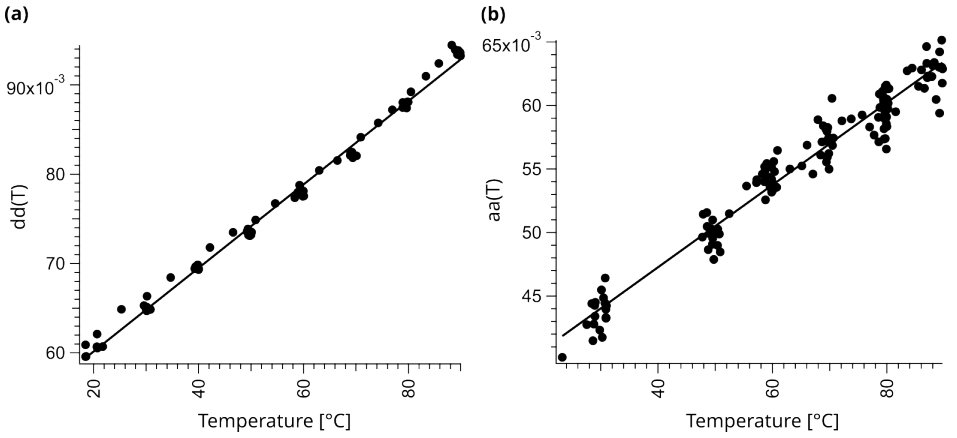


**Figure 2.15:** *Crosstalk between donor and acceptor channel* **(a)**: dd(T) and **b)**: aa(T), plotted as a function of temperature and fitted linearly. For this purpose, separate solutions containing only the donor or acceptor dye were prepared. The respective luminosity in the opposite optosplit channel is then a direct indicator of the crosstalk.

concentrations increased up to 9 fold, strongly changing the double stranded fraction to around 0.8.

$$FRET(T) = \frac{DA(T) - dd(T) \cdot DD(T) - aa(T) \cdot AA(T)}{AA(T)} \tag{2.1}$$

where dd(T) and aa(T) are the non-FRET artifacts (crosstalk):

$$dd(T) = \frac{DA_D(T)}{DD_D(T)} \tag{2.2}$$

and

$$aa(T) = \frac{DA_A(T)}{AA_A(T)} \tag{2.3}$$

Before each experiment, a melting curve of the FRET strands was performed inside the experimental setup chamber. The melting curve was used to normalize the FRET signal to 0 and 1 using the following equation:

$$FRET_{norm}(T) = \frac{FRET(T) - \alpha}{\beta} \tag{2.4}$$

where $\alpha = min(FRET(T))$ and $\beta = max(FRET(T))$.

## 2.C Random Walk Model

For the random walk model, the existing Comsol simulation was first used and the simulation was run with $Mg^{2+}$ ions and a 61mer DNA as the dilute species (the diffusion constant of $705\frac{\mu m^2}{s}$ for $Mg^{2+}$ at 25 °C was taken from [108] and the diffusion constant of 97.04 $\frac{\mu m^2}{s}$ for a 61mer DNA strand from [85]). The simulation was performed in the same chamber, with the same characteristics and settings as in Appendix Section 2.D. The resulting stationary salt and DNA concentration fields after 2 hours were exported as a 2D table with 200 values in x-direction and 200 values in y-direction, representing the whole simulated geometry. The

same was done for the stationary laminar flow field in x-direction and y-direction induced by the air flow across the gas-water interface. Values were linearly interpolated for points between values from the grid.

Then, a self-written LabView script was used to simulate the Brownian motion of a particle with a chosen diffusion constant starting at a random position in the chamber and propagating along the flow vector field in 10ms time steps. The random square displacement of a particle with diffusion constant D reads:

$$x^2 = Dt \tag{2.5}$$

As this particle can move in two directions, left and right this becomes

$$x^2 = 2Dt \tag{2.6}$$

Expanding this into two dimensions, one gets:

$$x_{2D}^2 = x_x^2 + x_y^2 \text{ and therefore: } x_{x,y} = \sqrt{x_x^2 + x_y^2} = \sqrt{2Dt + 2Dt} = \sqrt{4Dt} \tag{2.7}$$

The diffusion constant of a 61mer DNA [85] of $97.04 \cdot 10^{(-12)} \frac{m^2}{s}$ and a random unit vector phi with values between [-1,1] were then inserted. $\vec{u}$ and $\vec{v}$ represent the exported laminar flow field data from Comsol:

$$\text{x displacements: } \sqrt{4 \cdot 97.04 \cdot 10^{(-12)} \cdot dt} \cdot phi + dt \cdot \vec{u} \tag{2.8}$$

$$\text{y displacements: } \sqrt{4 \cdot 97.04 \cdot 10^{(-12)} \cdot dt} \cdot phi + dt \cdot \vec{v} \tag{2.9}$$

The particle was then displaced according to equations 2.8 and 2.9 with a timestep of 10ms for a total of 35 minutes. Along its path, the respective local $Mg^{2+}$ and 61mer DNA concentration were plotted, yielding the graph displayed in main text Figure 2.5(d). The path was overlayed with the original Comsol simulation graphic of the salt concentration distribution in Figure 2.6(a).

## 2.D Finite Elements Simulations

Finite element simulations were performed using COMSOL Multiphysics 5.4, similar to Chapter 1. A 2D geometry was designed using the same parameters as the Teflon cutout for the experimental chamber (Figure 2.16)(a)). Since the dynamics of the experiment take place mainly in the x-y plane, the system was simulated without the z dimension, focusing on the key properties. The geometry is first meshed accordingly (Figure 2.16)(b)) and then coupled to a gas inlet and outlet as well as a water inlet from the bottom, analogous to the experiment. The normal gas inflow has been set to experimentally measured values (about 236 ml/min), resulting in velocities up to about 12 m/s. The gas outlet uses pressure as its boundary condition, releasing as much gas as necessary to maintain a constant pressure. The system is assumed to be laminar, since the velocities don't exceed Mach < 0.3. The transport of water vapor in the gas is coupled at the top, using the stationary velocity field previously established. Simultaneously, the stationary velocity field of the laminar upward flow of water was calculated and coupled with the time-dependent transport of dilute species, in this case dissolved DNA or salts.

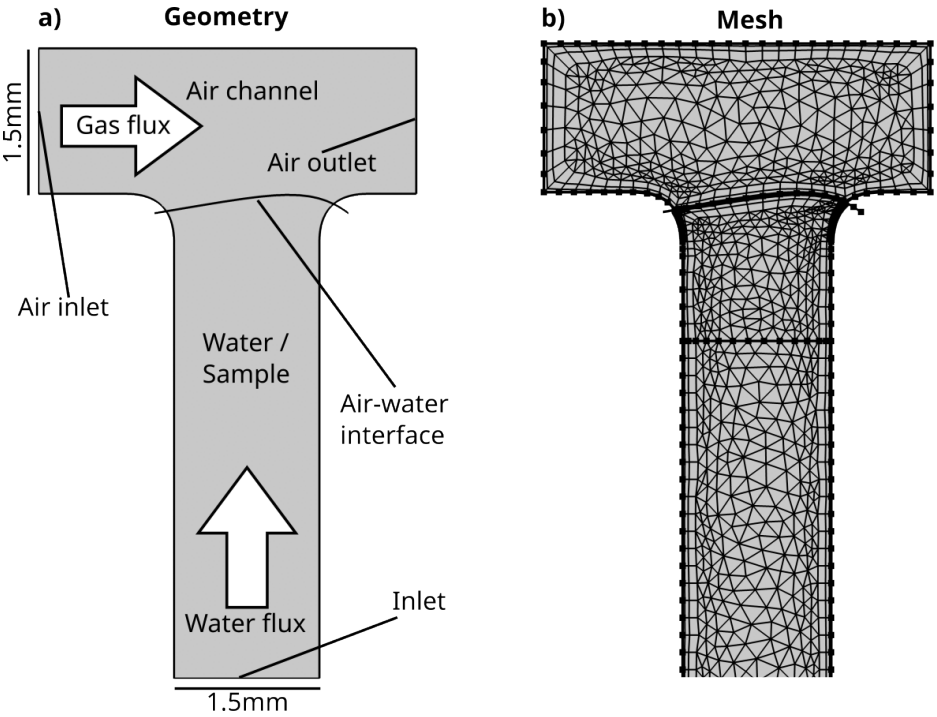


**Figure 2.16:** *Geometry and mesh used for Comsol simulations.* **(a)** The geometry as used for the 2D-simulation. **(b)** Mesh used for the simulation. Multiple boundary layers were introduced, especially at the interface, where the most important dynamics take place.

For both the gas and the water, Navier-Stokes-Equations were solved under the assumption that the flows are laminar due to their relatively low speed to viscosity ratio (Reynolds number). The flows can therefore be considered incompressible, the density constant and the continuity equation reduces equation 1.18, analogously to Chapter 1 Materials & Methods Section 1.5.7. The Navier-Stokes equation then again reduces to equation 1.19. The reference pressure was set to 1 [atm], the reference temperature was 45 °C and all surfaces, except the gas-water

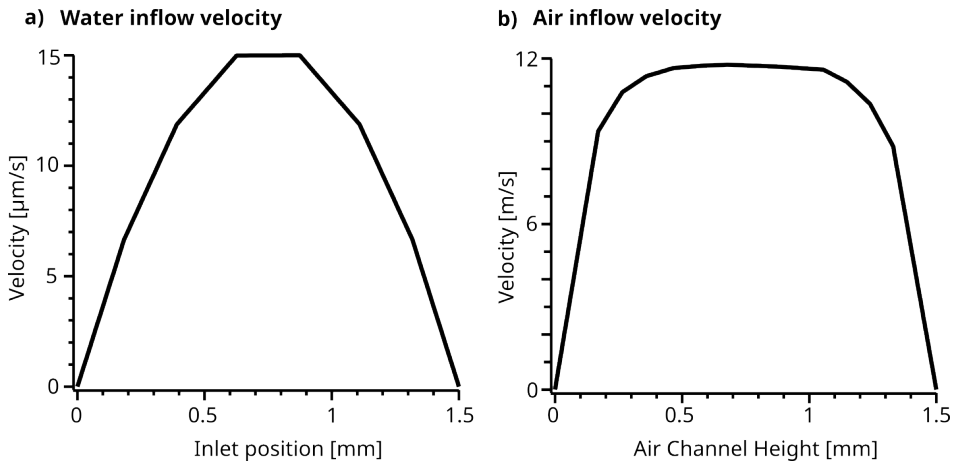


**Figure 2.17:** *Simulated inflow velocities.* **(a)** Inflow velocity of the water channel from below, plotted along the orifice. Due to the parabolic flow profile, the influx velocity is highest in the center around 15 µm/s and drops to 0 µm/s directly at the walls. **(b)** Air inflow velocity of the air channel plotted along the orifice axis. In contrast to the water inflow, the parabolic flow profile is not as pronounced, because of the far lower viscosity.

interface, are described as non-slip boundary conditions. The diffusion dependent transport of diluted species was simulated using Fick's law and convection due to the laminar flow fields like in equation 1.20. Equivalently, equation 1.19 and 1.20 are used for the dynamics of the gas channel. Figure 2.17 shows the inflow velocity of both water and air, plotted perpendicular to the inlet plane. Note the parabolic flow profile caused by the no-slip boundary condition at the channel walls. The boundary condition to combine the gas-flow with the water-flow is embedded in the gas-water interface: The velocity field components in x- and y-direction of the gas- as well as water-flow are required to be equal at the gas-water interface:

$$\boldsymbol{u} = \boldsymbol{u2} \tag{2.10}$$

where $\boldsymbol{u}$ describes the vector field of water, while $\boldsymbol{u2}$ denotes the vector field of the passing gas. The interface acts as a sliding wall, moving in the x-direction of the gas-flux to emulate the momentum transfer of the wind to the water surface. Figure 2.18(a) shows the tangential velocity plotted along the interface. To simulate the evaporation of water into the gas phase, the August equation was used, describing the relation between saturation vapor pressure and temperature:

$$P_{sat} = \exp\left\{20.386 - \frac{5132K}{T}\right\}[mmHg] \tag{2.11}$$

The saturation concentration of water vapor therefore is

$$c_{sat} = \frac{P_{sat}}{R \cdot T} \tag{2.12}$$

,where R denotes the ideal gas constant and T the temperature. At the interface, the boundary condition reads:

$$c_{vapor} = c_{sat}, \tag{2.13}$$

while at the ceiling of the gas-channel, far away from the interface, the concentration is set to:

$$c_{vapor} = h \cdot c_{sat}, \tag{2.14}$$

where h denotes the relative humidity in percent. Furthermore, the speed of evaporation at the interface is proportional to the vapor concentration gradient:

$$\boldsymbol{v}_{\text{evap}} = -D_{vap} \cdot \frac{M}{\rho} \cdot \nabla c_{vap} \tag{2.15}$$

where M represents the molar mass of water, $\rho$ the density of water and $D_{vap}$ the diffusion coefficient of vapor. Figure 2.18 shows a plot of the evaporation speed plotted along the interface.

| **Parameter** | **Value** | **Description** |
|---|---|---|
| $D_{vapor}$ [$m^2$/s] | (21.2E-6)*(1/1[K])*(1 + (0.0071*(T - 273))) | Diffusion Vapor |
| $D_{63mer}$ [µm$^2$/s] | 643*n$^{-0.46}$ = 95.6 | Diffusion 63mer [85] |
| $D_{Mg^{2+}}$ [µm$^2$/s] | 705 | Diffusion $Mg^{2+}$ [108] |
| $H_r$ [%] | 40 | Relative Humidity |
| $p_{sat}$ [mmHg] | exp(20.386 - (5132 K / T)) | Vapor Saturation Pressure |
| $c_{vapor,0}$ [mol/m$^3$] | $H_r$* (0.01 / (R * T)) * $p_{sat}$ | Initial Vapor Concentration |
| $c_{sat}$ [mol/m$^3$] | $p_{sat}$/(R*T) | Vapor Saturation concentration |
| $M_{vapor}$ [kg/mol] | 0.0180 | Molar Mass of Vapor |
| $M_{water}$ [g/mol] | 18.01528 | Molar Mass of Water |
| T [°C] | 45 | Temperature |

**Table 2.3: Parameters used for the finite elements simulation.** Final set of parameters used to simulate the system. Water-specific parameters such as dynamic viscosity or density were taken from inbuilt features of COMSOL Multiphysics 5.4..

At the inlet, the vapor concentration in the gas is set to a constant humidity dependent value (See table 2.3 for a detailed parameter list).

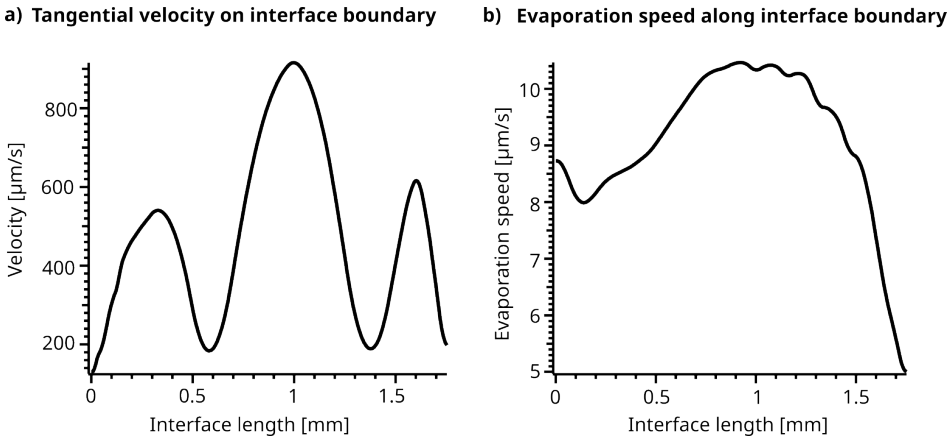


**Figure 2.18:** *Simulated evaporation and flow velocities at the air-water interface* **(a)** Tangential flow velocity at the air-water interface boundary. Since there are multiple circular flows beneath the interface, the direction of the flux changes along the interface, resulting in the varying graph of averaged speed in both directions. The dips in flowspeed indicates a position between two vortices. **(b)** Evaporation speed of water plotted along the air-water interface.

# Conclusion and Outlook

The origin of life research is centered around the chemical evolution of nucleic acids. Not only do they store genetic information, they also are able to perform catalysis. Their most important property, however, is their ability to form Watson Crick base pairs. Within the RNA-World framework, this key functionality enables reading out already stored information and allows for subsequent replication through templated ligation or polymerisation. Moreover, it lets nucleic acids form a variety of secondary structures, which allow them to catalyse reactions such as their own replication.

To bridge the gap from non-functional and short oligomers to long and functional ones, researchers require an understanding of the environments which could have hosted chemical life and investigate their implicit Darwinian selection pressures to drive molecular evolution.

Chapter 1 has shown how the ability of oligonucleotides to phase separate into a dense phase through base pairing can act as a Darwinian selection pressure. While periodic changes to the environment reset the pool composition sequences repeatedly, LLPS proved to select for specific sequence pairs able to form secondary structure. This increased their concentration relative to the pool as cycles progressed, proved to be robust against salt and pH changes of the environment and aligns with Finite elements simulations.

This could have been a key mechanism on early Earth to evolve a pool of random and short oligomers towards higher lengths and complexity, eventually selecting for catalytically active oligonucleotides. Notably, the selection also favored a 4-letter alphabet over a 2-letter one even if the strand is shorter, showcasing how sequence can be more important than length. A sequence made up of only two different nucleobases can bind in many different ways, as its sequence space is smaller and the energy landscape for hybridization is comparably flat. This allows for non-specific secondary structures, such as internal loops, rendering it unable to base pair with further strands. Interestingly, the selection mechanism showed that changing a single base inside a 22 nt long sequence, can can completely change its capability to phase separate.

As life is a process far from equilibrium, however, such an intrinsic selection mechanism like LLPS, is not sufficient to explain the pathway from short oligonucleotide fragments towards longer, complex strands. This process required external energy from the environment. From a large pool of many sequences (and therefore at low individual concentration), phase separation through hybridization seems unlikely to occur without previous enrichment and selection, as the critical concentration for LLPS cannot be reached. Previous studies have shown that temperature gradients across water filled rock pores, or at air water interfaces can enrich oligonucleotides by multiple orders of magnitude in a length selective manner, providing with a solution to the issue. As temperature gradients of such magnitude are however rare, chapter 2 investigated an isothermal environment on early Earth. A water filled pore is kept at non-equilibrium through a mere gas flux streaming across, accumulating biomolecules at the gas-water interface up to 30-fold.

Additionally, FRET measurements revealed a clear separation of double and single stranded regions of DNA. A Monte Carlo random walk model shows that sequences are flushed through

regions of high and low salt and nucleic acid concentrations. Strand separation is a significant hurdle for early nucleic acid replication cycles, as even environments, which accumulate oligonucleotides, suffer from also accumulating salts and especially divalent ions. The isothermal environment at hand separates the strands repeatedly through changes is ion concentration, enabling exponential replication, demonstrated here by Taq DNA polymerase as a model system. A further advantage of the system is the constant possibility for a feeding influx. In this work, a worst case scenario was adopted, in which the system was exposed to a constant dilution scenario. However, it is likely that over time, more and more biomolecules are flushed towards the interface, constantly feeding chemical reaction networks.

In the future, the experiments in chapter 1 and chapter 2 would provide new insights as experiments move towards pure RNA systems, without DNA as a model molecule. Particularly interesting would be observing phase separation with freshly polymerised RNA sequences and studying its impact on sequence and length distribution. The gas flux environment could be used as a means for up-concentrating those sequences in a first step and to allow them to replicate and elongate. Phase separation could then become observable in the same environment, further narrowing down the sequence pool to sequences capable of forming certain secondary structures. The increased temperature instability of RNA will however be an experimental hurdle, especially when exposed to high $Mg^{2+}$ concentrations. Furthermore, testing whether already discovered ribozymes work in this environment would be a big step towards creating a evolving chemical system that is sustained by a non-equilibrium setting and can be constantly fed with the necessary fuel molecules.

# Bibliography


[1] Qingzhu Yin, SB Jacobsen, K Yamashita, J Blichert-Toft, Ph Télouk, and F Albarede. A short timescale for terrestrial planet formation from hf–w chronometry of meteorites. *Nature*, 418(6901):949–952, 2002.

[2] Melanie Barboni, Patrick Boehnke, Brenhin Keller, Issaku E Kohl, Blair Schoene, Edward D Young, and Kevin D McKeegan. Early formation of the moon 4.51 billion years ago. *Science advances*, 3(1):e1602365, 2017.

[3] Mathieu Touboul, Thorsten Kleine, Bernard Bourdon, Herbert Palme, and Rainer Wieler. Late formation and prolonged differentiation of the moon inferred from w isotopes in lunar metals. *Nature*, 450(7173):1206–1209, 2007.

[4] TM Harrison, Janne Blichert-Toft, W Muller, Francis Albarede, Peter Holden, and SJ Mojzsis. Heterogeneous hadean hafnium: evidence of continental crust at 4.4 to 4.5 ga. *Science*, 310(5756):1947–1950, 2005.

[5] Ilya Bindeman. Oxygen isotopes in mantle and crustal magmas as revealed by single crystal analysis. *Reviews in Mineralogy and Geochemistry*, 69(1):445–478, 2008.

[6] Paul D Spudis, Don E Wilhelms, and Mark S Robinson. The sculptured hills of the taurus highlands: Implications for the relative age of serenitatis, basin chronologies and the cratering history of the moon. *Journal of Geophysical Research: Planets*, 116(E12), 2011.

[7] 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.

[8] Ben KD Pearce, Andrew S Tupper, Ralph E Pudritz, and Paul G Higgs. Constraining the time interval for the origin of life on earth. *Astrobiology*, 18(3):343–364, 2018.

[9] Carl Sagan and George Mullen. Earth and mars: Evolution of atmospheres and surface temperatures. *Science*, 177(4043):52–56, 1972.

[10] Norman H Sleep and Kevin Zahnle. Carbon dioxide cycling and implications for climate on ancient earth. *Journal of Geophysical Research: Planets*, 106(E1):1373–1399, 2001.

[11] E Bruce Watson and TM Harrison. Zircon thermometer reveals minimum melting conditions on earliest earth. *Science*, 308(5723):841–844, 2005.

[12] David C Catling and Kevin J Zahnle. The archean atmosphere. *Science advances*, 6(9):eaax1420, 2020.

[13] Stanley L Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, 1953.

[14] Kamila B Muchowska and Joseph Moran. Peptide synthesis at the origin of life. *Science*, 370(6518):767–768, 2020.

[15] Mahipal Yadav, Ravi Kumar, and Ramanarayanan Krishnamurthy. Chemistry of abiotic nucleotide synthesis. *Chemical reviews*, 120(11):4766–4805, 2020.

[16] Seohyun Chris Kim, Derek K O'Flaherty, Constantin Giurgiu, Lijun Zhou, and Jack W Szostak. The emergence of RNA from the heterogeneous products of prebiotic nucleotide synthesis. *Journal of the American Chemical Society*, 143(9):3267–3279, 2021.

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

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

[19] J Tze-Fei Wong. A co-evolution theory of the genetic code. *Proceedings of the National Academy of Sciences*, 72(5):1909–1912, 1975.

[20] Felix Müller, Luis Escobar, Felix Xu, Ewa Węgrzyn, Milda Nainytė, Tynchtyk Amatov, Chun-Yin Chan, Alexander Pichler, and Thomas Carell. A prebiotically plausible scenario of an RNA–peptide world. *Nature*, 605(7909):279–284, 2022.

[21] Venki Ramakrishnan. Ribosome structure and the mechanism of translation. *Cell*, 108(4):557–572, 2002.

[22] Clémentine Gibard, Subhendu Bhowmik, Megha Karki, Eun-Kyong Kim, and Ramanarayanan Krishnamurthy. Phosphorylation, oligomerization and self-assembly in water under potential prebiotic conditions. *Nature Chemistry*, 10(2):212–217, 2018.

[23] James D Watson and Francis HC Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[24] Ke Zhang, Joseph Hodge, Anamika Chatterjee, Tae Seok Moon, and Kimberly M. Parker. Duplex structure of double-stranded RNA provides stability against hydrolysis relative to single-stranded RNA. *Environmental Science and Technology*, 55(12):8045–8053, 2021.

[25] Adriana Calaca Serrao, Sreekar Wunnava, Avinash V Dass, Lennard Ufer, Philipp Schwintek, Christof B Mast, and Dieter Braun. High-fidelity RNA copying via 2', 3'-cyclic phosphate ligation. *Journal of the American Chemical Society*, 146(13):8887–8894, 2024.

[26] Albert C Fahrenbach, Constantin Giurgiu, Chun Pong Tam, Li Li, Yayoi Hongo, Masashi Aono, and Jack W Szostak. Common and potentially prebiotic origin for precursors of nucleotide synthesis and activation. *Journal of the American Chemical Society*, 139(26):8780–8783, 2017.

[27] Steve Minchin and Julia Lodge. Understanding biochemistry: structure and function of nucleic acids. *Essays in biochemistry*, 63(4):433–456, 2019.

[28] Michael Levitt. Folding of nucleic acids. In *Ciba Foundation Symposium 7-Polymerization in Biological Systems*, volume 7, pages 147–171. Wiley Online Library, 1972.

[29] David W Staple and Samuel E Butcher. Pseudoknots: RNA structures with diverse functions. *PLoS biology*, 3(6):e213, 2005.

[30] Rachel Green and Jack W Szostak. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science*, 258(5090):1910–1915, 1992.

[31] David P Bartel and Jack W Szostak. Isolation of new ribozymes from a large pool of random sequences. *Science*, 261(5127):1411–1418, 1993.

[32] Nikolaos Papastavrou, David P Horning, and Gerald F Joyce. RNA-catalyzed evolution of catalytic RNA. *Proceedings of the National Academy of Sciences*, 121(11):e2321592121, 2024.

[33] Yingfu Li and Ronald R Breaker. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2 '-hydroxyl group. *Journal of the American Chemical society*, 121(23):5364–5372, 1999.

[34] Zeljko Jericevic, Ira Kucan, and Robert W Chambers. Photochemical cleavage of phosphodiester bonds in oligoribonucleotides. *Biochemistry*, 21(25):6563–6567, 1982.

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

[36] Erwin Schrödinger. *What is life?: The physical aspect of the living cell*, volume 1. Cambridge university press Cambridge, 1944.

[37] Giacomo Bartolucci, Adriana Calaça Serrão, Philipp Schwintek, Alexandra Kühnlein, Yash Rana, Philipp Janto, Dorothea Hofer, Christof B Mast, Dieter Braun, and Christoph A Weber. Sequence self-selection by cyclic phase separation. *Proceedings of the National Academy of Sciences*, 120(43):e2218876120, 2023.

[38] Hajime Tanaka. General view of a liquid-liquid phase transition. *Physical Review E*, 62(5):6968, 2000.

[39] Salman F Banani, Hyun O Lee, Anthony A Hyman, and Michael K Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nature reviews Molecular cell biology*, 18(5):285–298, 2017.

[40] N Amy Yewdall, Alain AM André, Tiemei Lu, and Evan Spruijt. Coacervates as models of membraneless organelles. *Current Opinion in Colloid & Interface Science*, 52:101416, 2021.

[41] Eugénie Kluczka, Valentin Rinaldo, Angélique Coutable-Pennarun, Claire Stines-Chaumeil, JL Ross Anderson, and Nicolas Martin. Enhanced catalytic activity of a de novo enzyme in a coacervate phase. *ChemCatChem*, page e202400558.

[42] Byoung-jin Jeon, Dan T Nguyen, Gabrielle R Abraham, Nathaniel Conrad, Deborah K Fygenson, and Omar A Saleh. Salt-dependent properties of a coacervate-like, self-assembled dna liquid. *Soft Matter*, 14(34):7009–7015, 2018.

[43] 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.

[44] William M Aumiller Jr and Christine D Keating. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nature chemistry*, 8(2):129–137, 2016.

[45] Michi Nakata, Giuliano Zanchetta, Brandon D Chapman, Christopher D Jones, Julie O Cross, Ronald Pindak, Tommaso Bellini, and Noel A Clark. End-to-end stacking and liquid crystal condensation of 6–to 20–base pair dna duplexes. *Science*, 318(5854):1276–1279, 2007.

[46] Teresa E Strzelecka, Michael W Davidson, and Randolph L Rill. Multiple liquid crystal phases of dna at high concentrations. *Nature*, 331(6155):457–460, 1988.

[47] G Zanchetta, M Nakata, M Buscaglia, NA Clark, and T Bellini. Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *Journal of Physics: Condensed Matter*, 20(49):494214, 2008.

[48] Daphna Frenkiel-Krispin, Smadar Levin-Zaidman, Eyal Shimoni, Sharon G Wolf, Ellen J Wachtel, Talmon Arad, Steven E Finkel, Roberto Kolter, and Abraham Minsky. Regulated phase transitions of bacterial chromatin: a non-enzymatic pathway for generic dna protection. *The EMBO Journal*, 2001.

[49] Tommaso P Fraccia, Gregory P Smith, Giuliano Zanchetta, Elvezia Paraboschi, Youngwooo Yi, David M Walba, Giorgio Dieci, Noel A Clark, and Tommaso Bellini. Abiotic ligation of dna oligomers templated by their liquid crystal ordering. *Nature communications*, 6(1):6424, 2015.

[50] Otto Wichterle and Drahoslav Lim. Hydrophilic gels for biological use. *Nature*, 185(4706):117–118, 1960.

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

[52] Oguz Okay. Dna hydrogels: New functional soft materials, 2011.

[53] Zhongyang Xing, Alessio Caciagli, Tianyang Cao, Iliya Stoev, Mykolas Zupkauskas, Thomas O'Neill, Tobias Wenzel, Robin Lamboll, Dongsheng Liu, and Erika Eiser. Microrheology of dna hydrogels. *Proceedings of the National Academy of Sciences*, 115(32):8137–8142, 2018.

[54] Omar A Saleh, Byoung-jin Jeon, and Tim Liedl. Enzymatic degradation of liquid droplets of dna is modulated near the phase boundary. *Proceedings of the National Academy of Sciences*, 117(28):16160–16166, 2020.

[55] Nokyoung Park, Soong Ho Um, Hisakage Funabashi, Jianfeng Xu, and Dan Luo. A cell-free protein-producing gel. *Nature materials*, 8(5):432–437, 2009.

[56] Chikako Kurokawa, Kei Fujiwara, Masamune Morita, Ibuki Kawamata, Yui Kawagishi, Atsushi Sakai, Yoshihiro Murayama, Shin-ichiro M Nomura, Satoshi Murata, Masahiro Takinoue, et al. Dna cytoskeleton for stabilizing artificial cells. *Proceedings of the National Academy of Sciences*, 114(28):7228–7233, 2017.

[57] Elizabeth A Doherty and Jennifer A Doudna. Ribozyme structures and mechanisms. *Annual review of biochemistry*, 69(1):597–615, 2000.

[58] Matthias Morasch, Dieter Braun, and Christof B Mast. Heat-flow-driven oligonucleotide gelation separates single-base differences. *Angewandte Chemie*, 128(23):6788–6791, 2016.

[59] Diana M Mitrea and Richard W Kriwacki. Phase separation in biology; functional organization of a higher order. *Cell Communication and Signaling*, 14:1–20, 2016.

[60] 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.

[61] LM Barge, E Branscomb, JR Brucato, SSS Cardoso, JHE Cartwright, SO Danielache, D Galante, TP Kee, Y Miguel, S Mojzsis, et al. Thermodynamics, disequilibrium, evolution: Far-from-equilibrium geological and chemical considerations for origin-of-life research. *Origins of Life and Evolution of Biospheres*, 47:39–56, 2017.

[62] Frances Westall, Keyron Hickman-Lewis, Nancy Hinman, Pascale Gautret, Kathleen A Campbell, Jean-Gabriel Bréhéret, Frédéric Foucher, Axelle Hubert, Stephanie Sorieul, Avinash Vicholous Dass, et al. A hydrothermal-sedimentary context for the origin of life. *Astrobiology*, 18(3):259–293, 2018.

[63] 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.

[64] 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.

[65] Siddharth Agarwal, Dino Osmanovic, Melissa A Klocke, and Elisa Franco. The growth rate of dna condensate droplets increases with the size of participating subunits. *ACS nano*, 16(8):11842–11851, 2022.

[66] John SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*, 95(4):1460–1465, 1998.

[67] 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.

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

[69] AI Dragan, R Pavlovic, JB McGivney, JR Casas-Finet, ES Bishop, RJ Strouse, MA Schenerman, and CD Geddes. Sybr green i: fluorescence properties and interaction with dna. *Journal of fluorescence*, 22:1189–1199, 2012.

[70] Daekyu Sun and Laurence H Hurley. The importance of negative superhelicity in inducing the formation of g-quadruplex and i-motif structures in the c-myc promoter: implications for drug targeting and control of gene expression. *Journal of medicinal chemistry*, 52(9):2863–2874, 2009.

[71] Antarip Halder, Rohit Roy, Dhananjay Bhattacharyya, and Abhijit Mitra. Consequences of mg 2+ binding on the geometry and stability of RNA base pairs. *Physical Chemistry Chemical Physics*, 20(34):21934–21948, 2018.

[72] Naef Qasem, Muhammad Mustafa Generous, Bilal Qureshi, and Syed Zubair. A comprehensive review of saline water correlations and data: Part ii—thermophysical properties. *ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING*, 01 2021.

[73] Hubert Zipper, Herwig Brunner, Jürgen Bernhagen, and Frank Vitzthum. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic acids research*, 32(12):e103–e103, 2004.

[74] Haukur Gudnason, Martin Dufva, Dang Duong Bang, and Anders Wolff. Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature. *Nucleic acids research*, 35(19):e127, 2007.

[75] Pauline J Kolbeck, Willem Vanderlinden, Gerd Gemmecker, Christian Gebhardt, Martin Lehmann, Aidin Lak, Thomas Nicolaus, Thorben Cordes, and Jan Lipfert. Molecular structure, DNA binding mode, photophysical properties and recommendations for use of SYBR Gold. *Nucleic acids research*, 49(9):5143–5158, 2021.

[76] Mathieu Colombier, Jeremie Vasseur, Bruce F Houghton, Francisco Cáceres, Bettina Scheu, Ulrich Kueppers, Simon Thivet, Lucia Gurioli, Cristian Montanaro, Arianna Soldati, et al. Degassing and gas percolation in basaltic magmas. *Earth and Planetary Science Letters*, 573:117134, 2021.

[77] WE Hanford and RM Joyce. Polytetrafluoroethylene. *Journal of the American Chemical Society*, 68(10):2082–2085, 1946.

[78] Wei Hua Yu, Na Li, Dong Shen Tong, Chun Hui Zhou, Chun Xiang Cynthia Lin, and Chuan Yun Xu. Adsorption of proteins and nucleic acids on clay minerals and their interactions: A review. *Applied Clay Science*, 80:443–452, 2013.

[79] Annalena Salditt, Lorenz M. R. Keil, David P. Horning, Christof B. Mast, Gerald F. Joyce, and Dieter Braun. Thermal Habitat for RNA Amplification and Accumulation. *Phys. Rev. Lett.*, 125:048104, Jul 2020.

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

[81] Pier Luigi Luisi. Chemistry constraints on the origin of life. *Israel Journal of Chemistry*, 55(8):906–918, 2015.

[82] Nigel Goldenfeld and Carl Woese. Life is physics: evolution as a collective phenomenon far from equilibrium. *Annu. Rev. Condens. Matter Phys.*, 2(1):375–399, 2011.

[83] Nick Lane, John F Allen, and William Martin. How did luca make a living? chemiosmosis in the origin of life. *BioEssays*, 32(4):271–280, 2010.

[84] Alan Ianeselli, Annalena Salditt, Christof Mast, Barbara Ercolano, Corinna L Kufner, Bettina Scheu, and Dieter Braun. Physical non-equilibria for prebiotic nucleic acid chemistry. *Nature Reviews Physics*, 5(3):185–195, 2023.

[85] 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.

[86] Thomas Matreux, Paula Aikkila, Bettina Scheu, Dieter Braun, and Christof B Mast. Heat flows enrich prebiotic building blocks and enhance their reactivity. *Nature*, 628(8006):110–116, 2024.

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

[88] 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.

[89] Bruce Damer and David Deamer. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life. *Life*, 5(1):872–887, 2015.

[90] Avinash Vicholous Dass, Sreekar Wunnava, Juliette Langlais, Beatriz von der Esch, Maik Krusche, Lennard Ufer, Nico Chrisam, Romeo CA Dubini, Florian Gartner, Severin Angerpointner, et al. Rna oligomerisation without added catalyst from 2', 3'-cyclic nucleotides by drying at air-water interfaces. *ChemSystemsChem*, 5(1):e202200026, 2023.

[91] Emre Tekin, Annalena Salditt, Philipp Schwintek, Sreekar Wunnava, Juliette Langlais, James Saenz, Dora Tang, Petra Schwille, Christof Mast, and Dieter Braun. Prebiotic foam environments to oligomerize and accumulate RNA. *ChemBioChem*, 23(24):e202200423, 2022.

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

[93] Ryo Mizuuchi, Alex Blokhuis, Lena Vincent, Philippe Nghe, Niles Lehman, and David Baum. Mineral surfaces select for longer RNA molecules. *Chemical communications*, 55(14):2090–2093, 2019.

[94] Hauke Trinks, Wolfgang Schröder, and Christof K Biebricher. Ice and the origin of life. *Origins of life and evolution of biospheres*, 35:429–445, 2005.

[95] Roscoe Stribling and Stanley L Miller. Template-directed synthesis of oligonucleotides under eutectic conditions. *Journal of molecular evolution*, 32:289–295, 1991.

[96] Carl Schildkraut and Shneior Lifson. Dependence of the melting temperature of dna on salt concentration. *Biopolymers: Original Research on Biomolecules*, 3(2):195–208, 1965.

[97] Arnold Hampel and J_A Cowan. A unique mechanism for RNA catalysis: the role of metal cofactors in hairpin ribozyme cleavage. *Chemistry & biology*, 4(7):513–517, 1997.

[98] Angelica Mariani, Claudia Bonfio, Christopher M Johnson, and John D Sutherland. ph-driven RNA strand separation under prebiotically plausible conditions. *Biochemistry*, 57(45):6382–6386, 2018.

[99] 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):1897, 2017.

[100] Norimichi Takenaka, Masayuki Tanaka, Kenji Okitsu, and Hiroshi Bandow. Rise in the ph of an unfrozen solution in ice due to the presence of nacl and promotion of decomposition of gallic acids owing to a change in the ph. *The Journal of Physical Chemistry A*, 110(36):10628–10632, 2006.

[101] 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.

[102] Annalena Salditt, Leonie Karr, Elia Salibi, Kristian Le Vay, Dieter Braun, and Hannes Mutschler. Ribozyme-mediated RNA synthesis and replication in a model hadean microenvironment. *Nature Communications*, 14(1):1495, 2023.

[103] Alexandra M Deal, Rebecca J Rapf, and Veronica Vaida. Water–air interfaces as environments to address the water paradox in prebiotic chemistry: a physical chemistry perspective. *The Journal of Physical Chemistry A*, 125(23):4929–4942, 2021.

[104] Youxue Zhang. Degassing history of earth. pages 37–69, 2014.

[105] Hananto Kurnio, Subaktian Lubis, and Hersenanto Catur Widi. Submarine volcano characteristics in sabang waters. *Bulletin of the Marine Geology*, 30(2):85–96, 2016.

[106] Elena R Dobrovinskaya, Leonid A Lytvynov, and Valerian Pishchik. Properties of sapphire. *Sapphire*, pages 55–176, 2009.

[107] OA Sergeev, AG Shashkov, and AS Umanskii. Thermophysical properties of quartz glass. *Journal of engineering physics*, 43(6):1375–1383, 1982.

[108] Li Yuan-Hui and Sandra Gregory. Diffusion of ions in sea water and in deep-sea sediments. *Geochimica et cosmochimica acta*, 38(5):703–714, 1974.

# List of Publications

## First and shared first author publications during the doctoral studies (printed in full below)

Bartolucci, G.[1], Calaca Serrao, A.[1], **Schwintek, P.[1]**, Kühnlein, A., Rana, Y., Janto, P., … & Weber, C. A. (2023). Sequence self-selection by cyclic phase separation. Proceedings of the National Academy of Sciences, 120(43), e2218876120

**Schwintek, P.**, Eren, E., Mast, C. B., & Braun, D. (2024). Prebiotic gas flow environment enables isothermal nucleic acid replication.[2] *eLife*, https://doi.org/10.7554/eLife.100152.1.

## Contributing author publications during doctoral studies

Tekin, E., Salditt, A., Schwintek, P., Wunnava, S., Langlais, J., Saenz, J., … & Braun, D. (2022). Prebiotic foam environments to oligomerize and accumulate RNA. *ChemBioChem*, 23(24), e202200423.

Calaca Serrao, A., Wunnava, S., Dass, A. V., Ufer, L., Schwintek, P., Mast, C. B., & Braun, D. (2024). High-Fidelity RNA Copying via 2', 3'-Cyclic Phosphate Ligation. *Journal of the American Chemical Society*, 146(13), 8887-8894.

## Other publications

Morasch, M., Liu, J., Dirscherl, C. F., Ianeselli, A., Kühnlein, A., Le Vay, K., … & Braun, D. (2019). Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nature Chemistry*, 11(9), 779-788.

Edeleva, E., Salditt, A., Stamp, J., Schwintek, P., Boekhoven, J., & Braun, D. (2019). Continuous nonenzymatic cross-replication of DNA strands with in situ activated DNA oligonucleotides. *Chemical Science*, 10(22), 5807-5814.

1 2

[1]contributed equally to this work

[2]Reviewed preprint

PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

# Sequence self-selection by cyclic phase separation

Giacomo Bartolucci[a,b,1], Adriana Calaça Serrão[c,1], Philipp Schwintek[c,1], Alexandra Kühnlein[c], Yash Rana[a,b], Philipp Janto[c], Dorothea Hofer[c], Christof B. Mast[c], Dieter Braun[c,2], and Christoph A. Weber[d,2]

Edited by Dimitar D. Sasselov, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA; received November 8, 2022; accepted September 6, 2023 by Editorial Board Member Herbert Levine

**The emergence of functional oligonucleotides on early Earth required a molecular selection mechanism to screen for specific sequences with prebiotic functions. Cyclic processes such as daily temperature oscillations were ubiquitous in this environment and could trigger oligonucleotide phase separation. Here, we propose sequence selection based on phase separation cycles realized through sedimentation in a system subjected to the feeding of oligonucleotides. Using theory and experiments with DNA, we show sequence-specific enrichment in the sedimented dense phase, in particular of short 22-mer DNA sequences. The underlying mechanism selects for complementarity, as it enriches sequences that tightly interact in the dense phase through base-pairing. Our mechanism also enables initially weakly biased pools to enhance their sequence bias or to replace the previously most abundant sequences as the cycles progress. Our findings provide an example of a selection mechanism that may have eased screening for auto-catalytic self-replicating oligonucleotides.**

molecular selection | phase separation | prebiotic oligonucleotides | molecular origin of life | DNA

Oligonucleotides can catalyse biochemical reactions and store genetic information (1–3). The mechanisms through which functional oligonucleotides became sufficiently abundant are crucial to understanding the molecular origin of life (4). In addition to sequence motifs, sufficient strand length is also a requirement for functional folds (5). Therefore, the assembly of long-chained prebiotic oligonucleotides has been the focus of many recent studies (6–9).

However, the probability of randomly assembling a specific sequence of length $L$ with $m$ different nucleotides is proportional to $m^{-L}$. Since sequence space grows exponentially with sequence length, functional sequences are either not present or too dilute to interact and undergo chemical reactions. It is thus one of the central mysteries of the molecular origin of life how long enough sequences that enable self-replication could be selected from a large random pool of short non-functional oligonucleotides.

Due to the lack of complex biological machinery at the molecular origin of life, various physicochemical selective mechanisms have been considered (10). Examples are biased replication (11, 12), accumulation due to gradients of temperature or salt (13), the accumulation at liquid-vapor interfaces (14), as well as the length selective accumulation at mineral surfaces (15, 16). Multiple of the aforementioned mechanisms may also act synergistically. Another possible mechanism is related to the coexistence of two liquid-like phases. In particular, recent studies have shown that oligonucleotides can phase separate, forming coacervates (17, 18), liquid crystals (19, 20), or hydrogels (21–23), which can lead to a local enrichment of specific oligonucleotides.

An especially elegant case emerges when phase separation is caused directly by the base pairing of sequence segments among oligonucleotides. The strong interactions among complementary oligonucleotide strands (approximately $5\,k_bT$ per base pair) lead to low saturation concentration above which phase separation occurs (24) and an oligonucleotide-dense phase that is composed of strongly correlated sequences. Thus, oligonucleotide phase separation via base pairing provides a mechanism to strongly accumulate a specific set of oligonucleotide sequences.

In a realistic prebiotic environment, such as an under-water rocky pore (14, 25, 26), the phase separation of oligonucleotides can be expected to be subject to periodic, typically daily, changes in the environment. In addition, such systems can exchange oligomers with the environment continuously, for example, by fluid flows (Fig. 1*A*). Without phase separation, the oligomer composition is set by the composition of the environment. However, when the oligomers can phase separate, the oligomer-dense phase can grow by continuously recruiting sequences from the pool. Initially, we focus on periodic flows composed of short spikes separated by long waiting times in which the flow vanishes.

## Significance

A central mystery of the molecular origin of life is the emergence of oligonucleotides, such as RNA, that can self-replicate. In our work, we theoretically study and experimentally verify a simple though minimal mechanism capable of screening for such specific oligonucleotide sequences. This mechanism relies on two physical ingredients ubiquitous on early Earth: cycles of phase separation into oligonucleotide-dense and dilute phases and cyclic oligonucleotide exchange with a surrounding pool. We show that specific sequences can enrich oligonucleotide composition, evolving away from an initial pool. This non-equilibrium selection mechanism may provide the missing link in how specific short-chained peptides, RNA, and DNA sequences were recruited from prebiotic pools steering the assembly of self-replicating oligonucleotides at the molecular origin of life.

The authors declare no competing interest.

This article is a PNAS Direct Submission. D.D.S. is a guest editor invited by the Editorial Board.



[1] G.B, A.C.S, and P.S. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: dieter.braun@lmu.de or christoph.weber@physik.uni-augsburg.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218876120/-/DCSupplemental.

Published October 17, 2023.

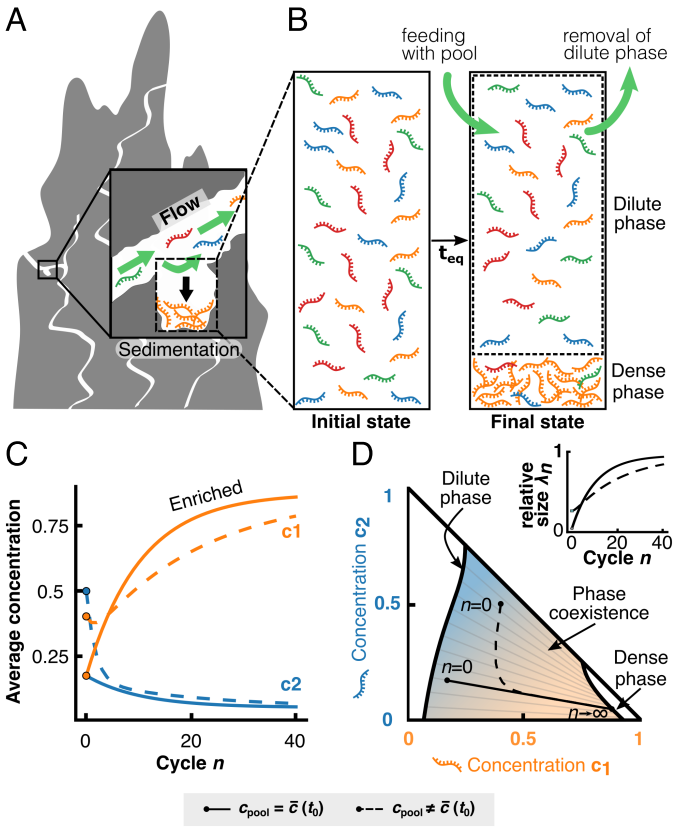


**Fig. 1.** Sequence selection via phase separation under varying feeding conditions. (*A*) Illustration of a time-dependent oligonucleotide flow through porous rocks on early earth. By sedimentation, phase-separated sequences can be enriched in pores, while others are flushed away by the flow. In the case of intermittent flows that are non-zero for periods much shorter than $t_{eq}$ (time to reach phase equilibrium), separated by waiting times much longer than $t_{eq}$, this scenario of sequence exchange can be mimicked via cycles of phase separation. (*B*) At each cycle, the system phase separates into a dense and a dilute phase, then a fraction of the dilute phase is removed and replaced with samples coming from a fixed sequence pool. (*C*) Theoretical result for the selection kinetics of multiple cycles depicted in *B*, discussed in systems composed of solvent and two sequences 1 and 2 with concentrations $c_1$ (orange) and $c_2$ (blue), respectively, see Eq. **1**. Solid and dashed lines correspond to $\alpha$=0.75 and $\alpha$=0.25, where $\alpha$ is the fraction of the dilute phase that gets replaced with the pool at each cycle. The two cases displayed differ also in initial average concentration $\bar{\boldsymbol{c}}(t_0)$ that can be equal to the pool ($\boldsymbol{c}_{pool} = \bar{\boldsymbol{c}}(t_0)$, solid lines) or deviate from it ($\boldsymbol{c}_{pool} \neq \bar{\boldsymbol{c}}(t_0)$, dashed lines). (*D*) The selection kinetics can be represented as a trajectory (solid, dashed black line) in the corresponding phase diagram, where outer white regions correspond to homogeneous mixtures while phase separation occurs in the coloured area. Gray tie lines connect the concentrations of the coexisting phases. During the kinetics, the dense phase grows as indicated by an increase of its relative size $\lambda_n$ (*Inset*).

The duration of the spikes is short compared to the time to reach phase equilibrium $t_{eq}$ while the waiting time between spikes is large compared to $t_{eq}$. In this case, we can mimic the continuous exchange with the pool by a simplified cyclic protocol composed of two different steps: i) A feeding step that corresponds to the replacement of a part of the dilute phase by the pool, followed by ii) a relaxation period toward phase separation equilibrium; see Fig. 1*B*. Later, we examine the opposite regime, i.e., where the flow is constant in time, see *DNA Phase Separation in a Continuous Feeding Flow*.

In this study, we ask whether this recruitment can significantly alter the oligonucleotide composition in the pore and thereby provide a physical mechanism of selection of specific sequences. The key question is how much the oligonucleotide composition of the system can evolve away from the pool, which serves as a reference for the selection process. We investigate this question through theory and experiments using DNA as a model oligomer. We decided to use DNA instead of RNA since DNA is more stable and our study focuses on a selection mechanism that relies on base pairing, which is very similar for both (27). We show that phase separation combined with feeding cycles by a pool indeed provides a strong selection mechanism giving rise to distinct routes for molecular selection of specific oligonucleotide sequences.

## Results and Discussion

**Theory of Cyclic Phase Separation with Feeding.** Here, we first discuss the theory governing the kinetics of an oligonucleotide mixture of volume $V$ which is composed of $M$ different sequences. This system undergoes cycles alternating between i) a period where the material is exchanged with a pool of fixed composition $\boldsymbol{c}_{pool}$, and ii) a period of phase separation (Fig. 1*B*). Specifically, within (ii), the mixture phase separates into a dense and a dilute phase with sufficient time to relax to thermodynamic phase equilibrium, while during the feeding step (i), a fraction $\alpha$ of the dilute phase is replaced by the pool. After $n$ cycles, the average composition of the system is described by the $M$-dimensional vector, $\bar{\boldsymbol{c}}(t_n) = \lambda(t_n)\boldsymbol{c}^{\mathrm{I}}(t_n) + (1-\lambda(t_n))\,\boldsymbol{c}^{\mathrm{II}}(t_n)$, where the vector components are the average concentrations of sequences. Moreover, $\boldsymbol{c}^{\mathrm{I}}(t_n)$ and $\boldsymbol{c}^{\mathrm{II}}(t_n)$ denote the concentrations of the dense and dilute phase, respectively, and $\lambda(t_n) = V^{\mathrm{I}}(t_n)/V$ is the fraction of the system occupied by the dense phase, where $V^{\mathrm{I}}(t_n)$ denotes the volume of the dense phase. The average composition of the system changes with cycle time $t_n$ due to the feeding step (i) and is given by (see *SI Appendix*, section 1 for more information):

$$\bar{\boldsymbol{c}}(t_{n+1}) = \Big[\big(1-\lambda(t_n)\big)\Big(\alpha\boldsymbol{c}_{\mathrm{pool}} + (1-\alpha)\boldsymbol{c}^{\mathrm{II}}(t_n)\Big) + \lambda(t_n)\boldsymbol{c}^{\mathrm{I}}(t_n)\Big], \qquad [1]$$

where $\boldsymbol{c}_{pool}$ is the concentration vector characterizing the composition of the pool that remains constant in time. The fraction of the dense phase $\lambda(t_n)$, and the concentrations of the dense and dilute phase $\boldsymbol{c}^{\mathrm{I}}(t_n)$ and $\boldsymbol{c}^{\mathrm{II}}(t_n)$ at cycle time $t_n$ can be determined by a Maxwell's construction in a $M$-dimensional state space for $\bar{\boldsymbol{c}}(t_n)$ obtained from solving the iteration Eq. **1**. The construction amounts to solving a set of non-linear equations that describe the balance of the chemical potentials and the osmotic pressure between the phases. Their solution requires estimating the sequence-specific interactions among the different oligonucleotides. Details on the numerical method and the determination of interaction matrices are discussed in *SI Appendix*, section 2.

To study the selection of oligonucleotide by cyclic phase separation, we considered the exchange of the oligonucleotide-poor phase by a pool of constant composition $\boldsymbol{c}_{pool}$, where the pool acted as a reference for the selection kinetics. We studied two cases where the pool had the same composition as the initial average concentration at $t = t_0$, $\boldsymbol{c}_{pool} = \bar{\boldsymbol{c}}(t_0)$, or they differed in terms of composition, $\boldsymbol{c}_{pool} \neq \bar{\boldsymbol{c}}(t_0)$. Representative time traces for both cases are shown in Fig. 1*C* for a mixture composed of solvent and two oligonucleotide sequences. We find that for both cases one sequence is enriched while the other sequence decreases in concentration as cycles proceed (orange and blue dashed lines, respectively). These concentration traces can be represented as trajectories of the average concentrations $\bar{\boldsymbol{c}}(t_n)$ in the ternary

phase diagram; Fig 1*D*. For average concentrations in the white region of the phase diagram the system remains homogeneous, while in the coloured region, phase separation occurs. In light gray, we show the tie lines connecting the coexisting dense and dilute phase concentrations. Each average concentration on the trajectory within this coloured region leads to a unique fraction of the dense phase, $\lambda(t_n)$, and a pair of concentrations corresponding to the dense and dilute phase, $\boldsymbol{c}^{\mathrm{I}}(t_n)$ and $\boldsymbol{c}^{\mathrm{II}}(t_n)$, respectively. As cycles proceed, the fraction of the dense phase $\lambda(t_n)$ grows (Fig. 1 *D*, *Inset*). During this growth, sequence 1 is selected over sequence 2. As volume growth saturates at $\lambda(t_\infty) = 1$, the selection process stops and the system settles in a stationary state.

The sequence composition of this stationary state is set by the tie line defined by the pool composition $\boldsymbol{c}_{\mathrm{pool}}$ (straight solid black line in Fig 1*D*). Most importantly, the slope and length of this pool tie line determine if and how well sequences are selected. Only if the pool tie line deviates from the diagonal tie line in the phase diagram, there is sequence selection during the growth of the dense phase. Selection is more pronounced if the pool tie line is longer in the phase diagram since more volume growth can occur. This case can be realized for example by strong interactions among sequences leading to the dilute and dense binodal branches being far apart in the phase diagram. Strikingly, both conditions, non-diagonal pool tie lines and strong sequence interactions, are particularly fulfilled in mixtures of oligonucleotides that can interact via base pairing. From our theoretical study, we conclude that phase separation subject to cyclic feeding can provide a selection mechanism particularly relevant in oligonucleotides mixtures.

**Observation of a Dense Phase of DNA.** To experimentally scrutinize the prerequisite of our proposed selection mechanism, we investigate phase-separation in oligonucleotide mixtures. To this end, we have designed several experimental systems. The constructs were motivated by the theoretical model which suggested that a group of sequences were selected if they interacted strongly among themselves but weakly with other sequences. The three sequence pairs (1 to 3) are composed of sequences with three binding regions each (a, b, and c or a', b', and c') which are separated by dimeric spacer sequences. These segments are individually reverse-complementary, i.e., a' is the reverse complement of a. However, the sequences i and ii are not the reverse complement of each other, because the order of the individual segments is not reversed (since i = 5' a,b,c 3' and ii = 5' a',b',c' 3'). This choice avoids a fully complementary double-stranded structure and allows each sequence to bind to three other sequences, forming a branched structure (Fig. 2 *A* and *B*). See *SI Appendix*, section 4 for a simulation of the secondary structure of each of the sequence pairs with NUPACK. The formation of branched DNA aggregates with short sequences leads to a dense phase. In fact, it was previously shown that mixtures of oligonucleotides are able to form dense phases through mutual base-pairing between long strands (14, 18, 21, 22, 28–30). However, these studies investigated rather long strands and solely ref. (31) studied phase separation of short DNA strands in the length regime of 20 to 25 bp which are more likely in prebiotic soups.

To characterize the phase separation propensity of our designed systems, we used time-lapse fluorescence microscopy (*SI Appendix*, section 13). In particular, we imaged each system over time in thin temperature-controlled microfluidic chambers. For each sequence pair, both strands were at 25 μM. Salt

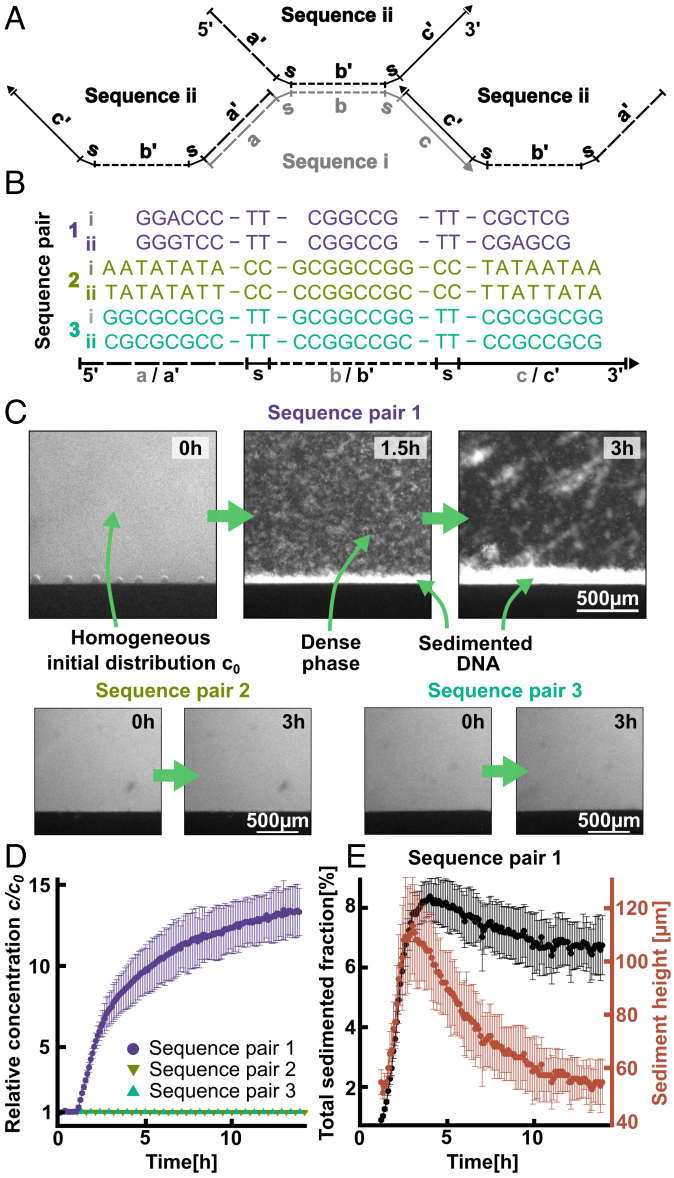


**Fig. 2.** Phase separation and sedimentation behavior of three sequence pairs. (*A* and *B*) Sequence i is composed of three segments a, b, and c with spacer s. Its pair ii consists of reverse complements a', b' and c'. The inverted arrangement of a' and b' creates a network from three binding sites and prevents the formation of a linear double-stranded duplex. (*C*) Fluorescence time laps images in a vertical, 500 μm thin microfluidic chamber to prevent convection flow. Concentrations of strands were 25 μM in a buffer of 10 mM Tris-HCl pH 7, 10 mM $MgCl_2$ and 125 mM NaCl. Fluorescence labelling was provided by 5X SYBR Green I. After cooling from 65 °C to 15 °C, sequence pair 1 phase separated and sedimented to the bottom of the chamber. Sequence pairs 2 and 3 did not form a dense phase, and thus showed a homogeneous fluorescence signal. (*D*) Sequence pair 1 showed an up to 13-fold enhanced relative concentration while sequence pairs 2 and 3 showed no phase separation. (*E*) The sedimentation behaviour of sequence pair 1 is studied by measuring its fluorescence. The total amount of sedimented DNA plateaued at 6 to 8% after 5 h while the sedimented DNA contracted about twofold. The sticking of dense phase DNA to the chamber walls could not be fully prevented. Error bars are SDs of three independent experiments.

concentrations were fixed at 125 mM NaCl and 10 mM $MgCl_2$, pH was controlled using 10 mM TRIS pH 7 buffer, and 5X SYBR Green I was used for fluorescent labelling. Other buffer conditions were also screened (*SI Appendix*, section 6). Prior to each experiment, the solutions were heated inside the microfluidic chamber to 65 °C to ensure homogeneous initial conditions. The samples were then slowly cooled at a rate of 6 K/min to 15 °C and

incubated at that temperature for at least 3 h. This temperature is lower than the melting temperature for all of the sequence pairs, corresponding to a scenario where most of the sequences are bound (*SI Appendix*, section 8). Choosing a higher incubation temperature resulted in smaller aggregates in bulk (*SI Appendix*, section 9 as well as Videos 4 and 5, video description).

Microscope images for all three systems are shown in Fig. 2*C*, where "0 h" corresponds to the moment when the cooling step has reduced the temperature to 15 °C. Within about 10 min, the first dense phase DNA nucleated for sequence pair 1 (Video Description) grew within 1.5 h to a size of a few micrometers and sedimented at speeds between 0.1 to 2.5 μm/s (*SI Appendix*, section 9). As a result, a phase of sedimented DNA accumulated at the bottom of the chamber. Assuming the particles to behave like sinking spheres subjected to Stokes drag, we estimated their densities to be around 3% higher than water. We could thus determine the characteristic sedimentation length to be on the scale of tens of micrometers. For sedimentation to occur, the chamber (pore) height has to exceed the sedimentation length. Inverting this logic, the minimal particle radius for our setup to observe sedimentation would be 0.3 μm. The DNA concentration in the dense phase increased up to 13-fold (Fig. 2*D*).

The total amount of molecules that sedimented saturates at about 8% of the initial material at about 3 h, decreasing then only slightly over time (Fig. 2*E*, black data points). The height of sedimented DNA reached a maximum of about 100 μm at 3 h but then compacted to about half the height (Fig. 2*E*, red data points). Similar behavior has been observed in literature for systems composed of longer DNA strands (21).

No dense phase was observed for sequence pairs 2 or 3, despite their longer segments that suggest stronger binding affinities. This could be caused by non-specific hybridization tendencies of binary (especially G- and C-rich) sequences, which can lead to alternative secondary structures, such as hairpin-rich configurations, internal loops or G-quadruplexes (32). Consequently, the formation of such structures may prevent network formation. We also examined five additional 22-nt sequence pairs similar to pair 1, systematically varying the base composition of the GC-only binding segments. Notably, the inclusion of a single A/T nucleotide in the outer segments a/a' or c/c', thus utilizing the full 4-letter alphabet, proved crucial for phase separation (see *SI Appendix*, section 5 for more details).

**Cycles of Phase Separation and Feeding.** Based on our observation that sequence pair 1 can form a dense DNA-rich phase, we experimentally scrutinize the theoretically proposed selection mechanism shown in Fig. 1 *A–D* that relied on a cyclic material inflow. We subjected the phase-separating DNA to cyclic feeding steps by replacing the *Top* fraction of the dilute phase in the vial with the pool (Fig. 3*A*, steps 1 to 3). The theory suggests that exchanging the complete dilute phase with the pool reaches the final stationary state with minimal amount of cycles (*SI Appendix*, section 1). However, a complete removal of the dilute phase by pipetting turns out to be experimentally difficult since this also risks removing sedimented dense DNA. To avoid kinetically trapped states of phase-separating oligonucleotides, we additionally include annealing and melting steps in the cycle (Fig. 3*A*, steps 3 to 4 and back to step 1). This procedure enabled a fast relaxation to thermodynamic equilibrium after each feeding step.

We investigated a system composed of equal fractions of the sequence pairs 1 and 2 (Fig. 2*B*), where solely the sequence pair 1 showed phase separation before. As control we considered a system composed of sequence pairs 2 and 3 where we could not observe the formation of a dense DNA phase (Fig. 2*B*). We determined the strand concentrations of each system in the *Top* and *Bottom* fractions of the vial using HPLC. We monitored the kinetics over six feeding cycles for the system composed of sequence pairs 1 and 2 (Fig. 3*B*) and compare it to the non-phase-separating control (*Inset* in *B*). Both systems were initialized with equimolar concentrations of the two respective sequence pairs.

We found that the concentrations for the control hardly increased per cycle with slopes about or less than 2 μM/cycle.

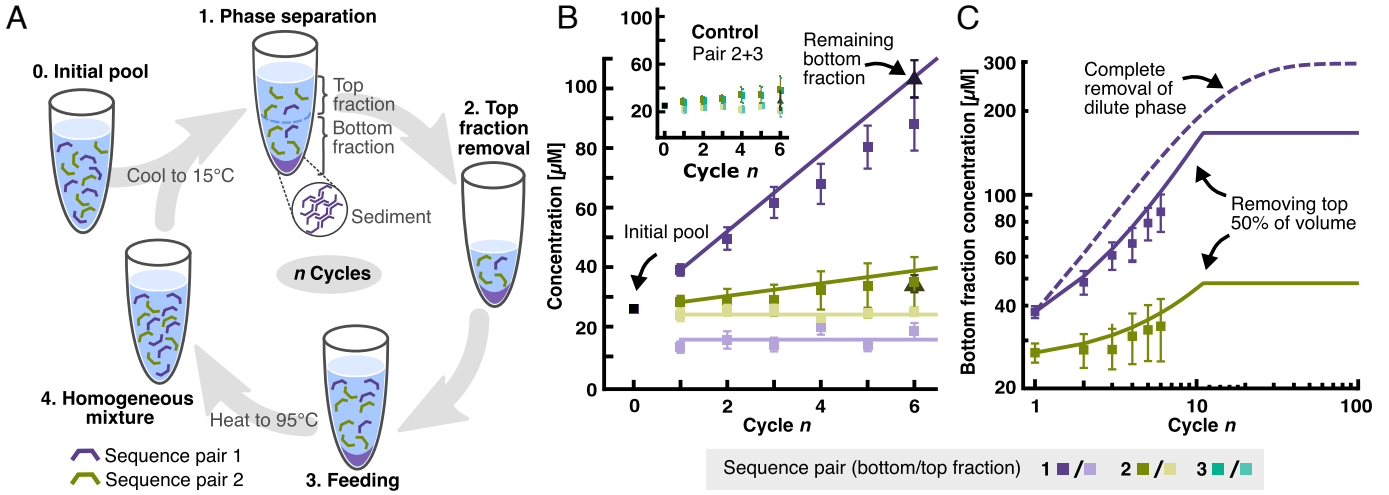


**Fig. 3.** Cycles of phase separation and feeding steps select specific oligonucleotide sequences from the initial pool. (*A*) Cyclic experimental protocol based on Fig. 1*B*. (*B*) The initial pool contained a 25 μM concentration of sequences pairs 1 and 2. After sedimentation, the top half of the volume (Top fraction) was removed ($\alpha = 0.5$) and fed after each cycle with the same volume of the initial pool. Using quantification with HPLC, we found that sequence pair 1 (purple) was enriched while the concentration of sequence pair 2 (green) remained approximately constant. The same flat dynamics was found for a control system using non-phase-separating sequence pairs 2 and 3 (inset, due to significant data overlap markers are halved). In addition, the concentration of all supernatants and the final sediment were measured by absorbance at 260 nm (triangular markers; *SI Appendix*, section 10). (*C*) Solid lines show theoretical predictions. The *Bottom* fraction concentration saturates once the sedimented DNA has filled the *Bottom* fraction of the chamber. If the whole supernatant could have been removed at each step of the cycle, only slightly amplified selection would be predicted (dashed line).

This non-zero increase is probably due to the adhesion of strands to the vial surface. For the phase-separating system with sequence pair 1 we observed that the concentration strongly increased, approximately linear with a slope of about $(10.2 \pm 0.4)$ μM/cycle (purple), while the sequence pair 2 in the mixture got only weakly enriched by the cycling. This observation confirmed that specific sequences could get selected by phase separation from the dilute phase.

Additionally, we also tested whether SYBR Green I changes the general behavior of the sequence pair mixtures over feeding cycles. In its absence, the partitioning of sequence pair 1 to the *Bottom* fraction also occurred, while sequence pair 2 remained constant on both fractions. Interestingly, the linear increase in concentration on the *Bottom* fraction is about 1.6 times higher than in the presence of SYBR Green I but shows very similar results compared to the experiments with SYBR (*SI Appendix*, section 11). However, in order to match theory parameters based on fluorescence measurements, we kept the SYBR concentration constant throughout all experiments.

In the experiments, the selection occurred concomitantly with the growth of the dense phase, which is consistent with our theoretical results. As cycles proceeded, more phase-separating DNA was recruited and led to an increase in the concentration of the *Bottom* fraction (Fig. 3*C*). In contrast, the concentration of the *Top* fraction remained constant at about 14.7 μM (Fig. 3*C*, light purple). A constant *Top* fraction concentration during cycles implies that the system remained on the same tie line while the volume of the sedimented DNA was growing. This corresponds to the simple theoretical scenario where the system is initialized at the pool tie line, as outlined in Fig. 1*D* ($c(t_0) = c_{\text{pool}}$).

We quantitatively compared the experimental results for the *Bottom* fraction concentration with the theoretical model. Since the experimental selection kinetics occurred on a single tie line, the dense and dilute phase concentrations, $\boldsymbol{c}^{\text{I}}$ and $\boldsymbol{c}^{\text{II}}$ remained constant over time, while the sediment size increases. For the dilute phase concentration $\boldsymbol{c}^{\text{II}}$, we use the experimental concentration value of the *Top* fraction. The sediment concentration $\boldsymbol{c}^{\text{I}}$ could be estimated for the theory using the experimental value for the initial average sequence concentration $\bar{c}(t_0)$ and the initial sedimented DNA size $\lambda(t_0)$. Using these values, we find that the theoretical results (solid lines in Fig. 3*B*) agree well with the experimental data points.

Based on the agreement between experiment and theory, we could use the theory to extrapolate the selection kinetics for a larger amount of cycles (Fig. 3*C*, solid lines). For the experimental partial removal of the dilute phase, we found that selection approximately doubled after 20 cycles. The selection kinetics saturate because the sediment has grown to the volume corresponding to the *Bottom* fraction.

Finally, we used the theory to consider the ideal case of the complete removal of the dilute phase. We find that, for this ideal case, the sequence pair can enrich by twofold better than the for *Top* fraction removal and more than 10-fold compared to the initial pool (Fig. 3*C*, dashed lines).

In summary, we have shown experimentally that discrete cycles of feeding, i.e., replacing the dilute phase with a pool, lead to the enrichment of specific sequence pairs through the formation of dense phases, confirming the theoretically proposed selection mechanism.

**DNA Phase Separation in a Continuous Feeding Flow.** The conditions on the early Earth scenario in general deviated from the simplified scenario of periodic feeding cycles separated by waiting times to relax to phase equilibrium. Here, we demonstrate a continuous implementation of the selection mechanism shown in Fig. 3. Using a microfluidic setup, we mimic a continuous exchange of solute and simulate a feeding flow of nucleic acids and salts flowing through a rock pore (Fig. 1*A*). Previous studies have shown that gases and liquids can effectively pass through cracks in hydrothermal structures like shattered glassy basalt, providing one possible implementation of the scenario mentioned above (33, 34). This flow generates a selection pressure, requiring sequence pairs to rapidly phase separate and settle at the pore's bottom to avoid being flushed out, as this would threaten the dilution by a larger reservoir such as the ocean, potentially resulting in hydrolysis-induced death or loss of information.

To test the concept, we designed a microfluidic system with a continuous flow of a pool and compared it to fluid flow theory using COMSOL Multiphysics. In this setup (*SI Appendix*, section 13), a pore (3 mm × 6 mm × 500 μm) at 15 °C is connected to a feeding pool of sequence pair 1 through a channel (See *Left* green arrow in Fig. 4*A*). The results of this experiment are shown in the fluorescence micrographs in Fig. 4*A*. Using an inflow speed of 2 μm/s, we observe upconcentration of sequence

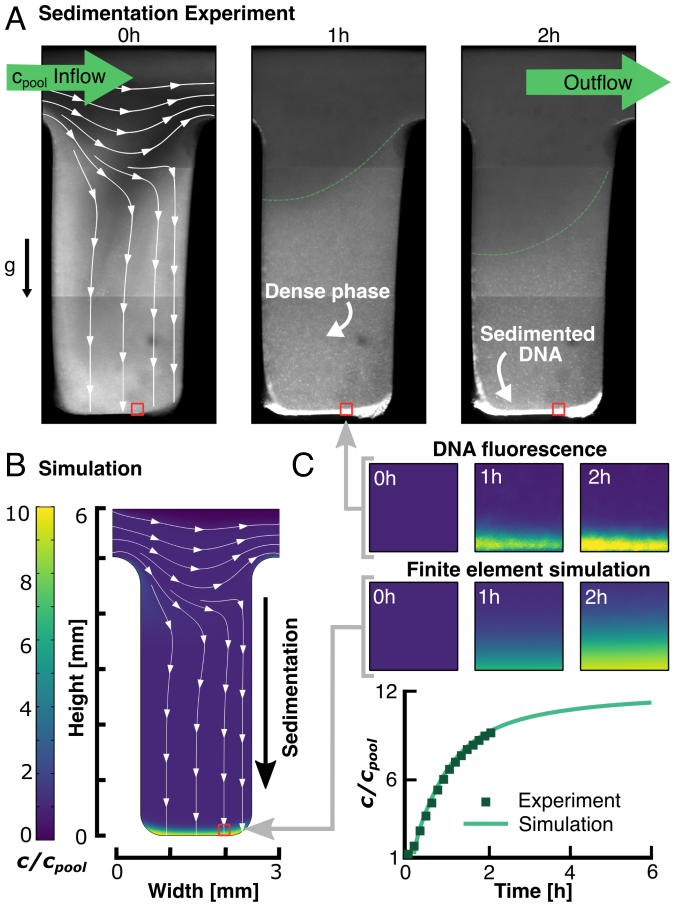


**Fig. 4.** DNA Phase separation with continuous feeding flow. (*A*) Fluorescence time-lapse images of a microfluidic chamber with a 2 μm/s inflow of sequence pair 1 with concentration $c_{\text{pool}}$. DNA that phase separates can sediment, thus is not advected out of the chamber. This implements cycles of phase separation in a system with continuous flow. (*B*) Finite element simulation of fluid flow with sedimentation and diffusion confirms the experimental findings in detail when assuming a downward sedimentation speed of 0.1 μm/s and a diffusion coefficient of 5 μ $m^2$/s. (*C*) The relative concentration $c/c_{\text{pool}}$ at the bottom matches well between experiment and simulation, seen for a 200 μm sized squared cut-out or when plotted over time.

pair 1 by phase separation and sedimentation inside the pore despite the continuous outflow.

In our setup, we found that after 2 h, the concentration of sequence pair 1, $c(t)$, is enriched 8-fold relative to the inflow concentration $c_{\mathrm{pool}}$. The inflow and the choice of pore geometry have to be tuned not to perturb the sedimentation speed of the dense phase DNA (Fig. 4*B*). By numerically solving the hydrodynamic flow equations in addition to sedimentation and diffusion of dense phase DNA, the simulated increase in concentration agrees well with the experiments in both space and time (Fig. 4*C*). In summary, our findings demonstrate that for a sufficiently slow flow (in the 10 μm/s regime), sequences can be effectively selected from the continuously fed pool and accumulate within a pore. This emphasizes the viability of the selection mechanism in a prebiotic context.

**Selection in Pools with Many Sequences.** Up to now, we have investigated specific pools composed of only a few designed sequence pairs for the proof of principle. It remains unclear how robust our proposed selection mechanism is for realistic pools that are formed via polymerization and contain many different sequences. To tackle this question, we use our theory of cyclic oligonucleotide phase separation with discrete feeding cycles and consider pools that could emerge from the polymerization of different units. For simplicity, we focus on sequences of fixed length $L$ composed of two different units, 0 and 1. Both units can be either thought of as two different nucleotides, or as two nucleotide segments of fixed compositions, like the building blocks of the sequences introduced in Fig. 2*A*.

Following ref. 35, sequence ensembles can be characterized by two parameters, the relative composition of the two units $r$ and the blockiness $b_l$. The latter determines the chain correlations of the two units: For $b_l = 1$, the model favors homopolymers (..11.. or ..00..), while for $b_l = -1$, sequences are anti-correlated heteropolymers (..1010..); see *SI Appendix*, section 3*A* for more details on the model. Phase separation of an ensemble of different sequences occurs once the system crosses the cloud point in the phase diagram; for details, see *SI Appendix*, section 3*B*. Subjecting such a phase-separated sequence pool to removal and feeding cycles, we find qualitatively different selection scenarios depending on the parameters $r$ and $b_l$.

For initial pools of low blockiness ($b_l < 0$), we find that the sequence bias of the pool is strongly amplified for a large number of cycles $n$ (Fig. 5*A*). This behavior results from the strong interaction propensity among sequences of the same type. Dominantly, such sequences are recruited from the dilute phase after pool replacement, while other sequences partition into the dilute phase and get subsequently removed by the replacement step. In contrast, for initial pools of high blockiness ($b_l > 0$), the initial sequence bias is completely altered while cycling (Fig. 5*B*).

In particular, homopolymeric sequences are disfavored, while heteropolymeric sequences get more favored as cycles proceed. The reason is that a homopolymeric sequence cannot interact with copies of itself, while more heteropolymeric sequences can. These trends are summarized for largely different sequence ensembles for different values of unit compositions $r$ and blockiness $b_l$ by determining the most abundant sequence before cycling and for a large amount of cycles $n_{\mathrm{f}}$ (Fig. 5*C*).

We identify a domain at low blockiness ($b_l < 0$) where the most abundant sequences are amplified. In this domain, the number of sequences in solution decreases significantly. Thus, a description in terms of a few strongly interacting sequences becomes more and more accurate as cycles proceed.

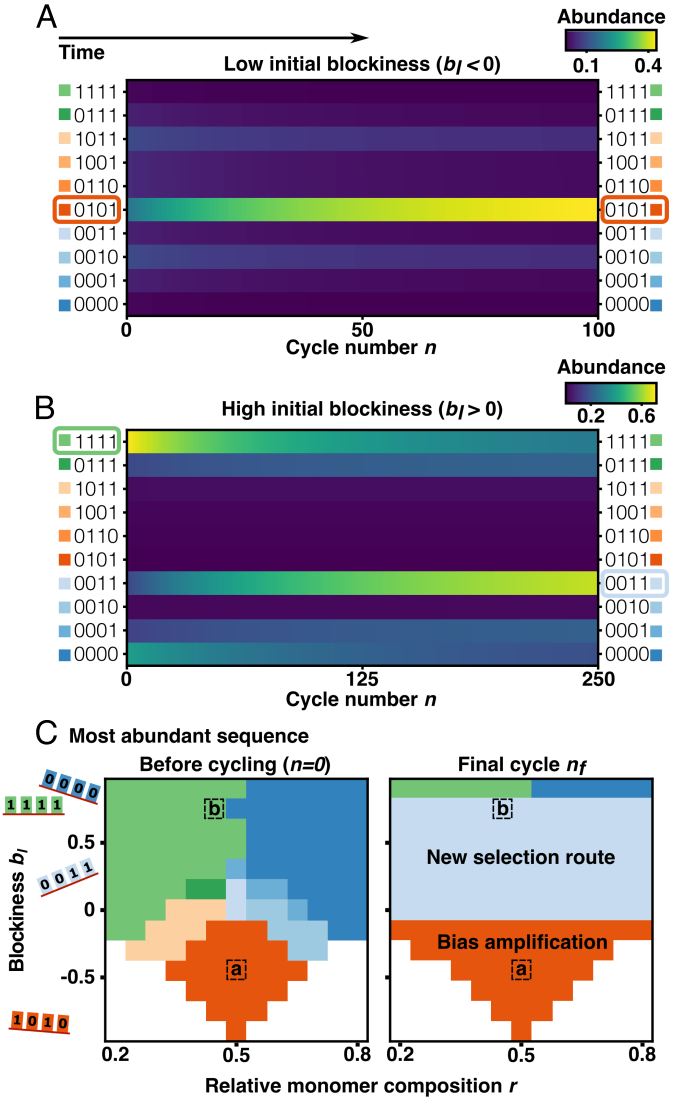


**Fig. 5.** Cycles of phase separation can amplify pool bias or offer an alternative selection route. We consider pools composed of sequence ensembles characterized by two parameters: blockiness $b_l$ and relative composition $r$ of the units 0 and 1. (*A* and *B*) For low initial blockiness ($b_l < 0$), the initially most abundant sequence (here: 0101) gets further enriched. In contrast, for high initial blockiness ($b_l > 0$), homopolymeric sequences (e.g. 1111 and 0000) are depleted with cycling while a more heteropolymeric sequence (0011) is strongly amplified. The color code indicates relative sequence abundances. (*C*) The most abundant sequence is shown for different values of $b_l$ and $r$ before cycling (*Left*) after cycling $n_{\mathrm{f}}$-times (*Right*). Further amplification of the initially most abundant sequence is found for low blockiness ($b_l < 0$), while new selection routes can emerge at high blockiness ($b_l > 0$).

For a quantification of the variation in abundance due to phase separation cycles, see *SI Appendix*, Fig. S4*H*. New selection routes can emerge for sequence ensembles of intermediate or large blockiness ($b_l > 0$). The later regime leads to a switch of the most abundant sequence when subjecting the system to a large number of feeding cycles.

## Conclusion

Here, we showed that the ability of oligonucleotides to phase separate can give rise to an evolutionary selection mechanism if subject to feeding cycles. In particular, replacing the dilute phase with a constant pool composed of different oligonucleotide sequences leads to the growth of a dense phase of specific

sequences while others are depleted relative to the pool. We have quantitatively confirmed our theoretical predictions for discrete cycles by experiments using designed DNA sequences. In addition, we showed that the same mechanism also caused the selection of specific sequences in a prebiotic-relevant scenario where the system is subjected to a continuous flow of the pool.

The key property of the proposed selection mechanism is that it is highly sequence-specific, also in the presence of other interacting sequences. Specifically, sequences that interact strongly with other sequences are enriched in the dense phase while weakly interacting sequences are expelled and thus leave the system through the removal step. A key observation of our work is that the selection mechanism also works for very short oligonucleotides. In our experiments, sequences of 22 nucleotides with base pairing regions of 6 nt form cooperative base-pairing networks at room temperature and phase separate, while others with the length of 28 nucleotides and base pairing regions of 8 nt do not. In contrast to the strong length selectivity of mineral surfaces, for example, this mechanism counter-intuitively prioritizes sequence over length in this case. In the future, it would be interesting to investigate the combined effect of DNA phase separation with other selection mechanisms, particularly how it affects the sequence and length distribution of freshly polymerized oligonucleotide pools (36).

Using theory, we studied realistic, multi-sequence pools that result from polymerization. We found robust and pronounced selection kinetics already for sequences composed of only four segments of nucleotide sequences. We distinguish two qualitative scenarios of sequence selection, where either the initial sequences bias is strongly amplified, or the initial bias is swapped and other sequences are selected.

The robustness of our selection mechanism, particularly for short oligonucleotides, suggests its relevance at the molecular origin of life, where specific short-chained peptides, RNA, and DNA sequences were recruited during their assembly from prebiotic pools. The cyclic removal of weakly interacting sequences can guide the selection of longer sequences which face dilution by the exponentially growing size of sequence space. Moreover, the dense phase could have provided enhanced stability against degrading chemical reactions such as catalytic cleavage (28) or hydrolysis due to the duplex formation (37). In fact, there is a correlation between catalytic sequences and phase separation in functional ribozyme polymerases (38). Ultimately, we expect an enhanced selection propensity when combining self-replication with our selection mechanism relying on base-pairing interactions.

## Materials and Methods

**Strand Design.** DNA oligonucleotide systems were designed using the NUPACK software package 3.2.2 (39). The strands were constrained to contain three binding sites separated by spacers either composed of TT or CC. Systems that formed the intended secondary structure (each strand base-pairing with three other strands) were chosen (*SI Appendix*, section 4). The oligomers were ordered from biomers.net GmbH, in a dry state, with high-performance liquid chromatography purification. The sequences were as follows (5′-3′) - Sequence pair 1, sequence i: GGA CCC TTC GGC CGT TCG CTCG; sequence ii: GGG TCC TTC GGC CGT TCG AGCG; Sequence pair 2, sequence i: AAT ATA TAC CGC GGC CGG CCT ATA ATA A; sequence 2: TAT ATA TTC CCC GGC CGC CCT TAT TATA; Sequence pair 3, sequence i: GGC GCG CGT TGC GGC CGG TTC GCG GCGG; sequence ii: CGC GCG CCT TCC GGC CGC TTC CGC CGCG. All the strands were stored at −20 °C, diluted in nuclease-free water at 200 μM. Before every experiment, the strands were denatured at 95 °C for 2 min.

**Reaction Mixtures.** Initial pools of 15 μL were prepared with 25 μM of each respective DNA strand, 10 mM Tris Buffer-HCl pH 7, 5X SYBR Green I (intercalating dye; excitation 450 to 490 nm, emission 510 to 530 nm), 125 mM NaCl, and 10 mM $MgCl_2$. The mixtures were heated to 95 °C for 2 min to ensure full de-hybridization of the strands. The temperature protocol that allows hybridization and consequent phase separation was *i*. 95 °C for 2 min, *ii*. 65 °C for 10 s, *iii*. cooling to 15 °C (ramp rate: 6 K per minute), iv. 15 °C for at least 3 h. Temperature protocols were performed in a standard thermocycler (Bio-Rad CFX96 Real-Time System). Melting curves were measured in triplicates using the same reaction mixture and temperature profile as for the sedimentation experiments (*SI Appendix*, section 8). Baseline correction using a reference measurement with only SYBR Green I. In the case of feeding cycle experiments, after sedimentation, 7.5 μL of the dilute phase, corresponding to 50% of the initial volume, was removed by carefully pipetting only at the center of the meniscus to avoid removing material from the sediment. Afterward, 7.5 μL of the initial pool stock was added to the remaining *Bottom* fraction. The aforementioned temperature protocol was then repeated, completing one feeding cycle.

**Sedimentation Imaging.** The imaging experiments were performed in a microfluidic chamber containing multiple wells, cut out of 500 μm Teflon foil and sandwiched between two sapphire plates (*SI Appendix*, Fig. S16). The sample volume (about 15 μL per well) was loaded by using microloader pipette tips. The temperature of the chamber was controlled using three Peltier elements. To remove the waste heat from the Peltier elements, a Julabo 300F water bath (JULABO GmbH) was used to cool the back of the chamber. The entire chamber is held in place by screwing a steel frame on top using a homogenous torque of 0.2 Nm. After loading, the wells were sealed with Parafilm to avoid evaporation. Monitoring of the sedimentation was performed using a self-built fluorescence microscope composed of a 490-nm LED (M490L4, Thorlabs), a 2.5x Fluar objective (Zeiss), and the FITC/Cy5 H Dualband Filterset (AHF). Multiple wells could be imaged by moving the chamber perpendicularly to the light axis with two NEMA23 Stepper Motors and a C-Beam Linear Actuator (Ooznest Limited). Images were taken using a Stingray-F145B CCD camera (ALLIED Vision Technologies) connected via FireWire to a computer running a self-written Labview code operating camera, motors, LED's and Peltier elements (*SI Appendix*, Fig. S15). Flow-through experiments were conducted using a similar chamber without Peltier Elements, only using the water bath at homogeneous 15 °C. In this case, the sapphires have holes on the backside, where an outlet and two inlet tubings were attached. Inlet tubing 1 contained 20 uM of each strand of system 3 and 10X SYBR Green I, while inlet tubing 2 contained 20 mM TRIS pH 7, 250 mM NaCl, and 20 mM $MgCl_2$. Flowspeed was adjusted using the Nemesys Controler NEM-B002-02 D (Cetoni GmbH) with two 100 μl syringes. Hardware was controlled using a self-written labview (National Instruments) software (*SI Appendix*, Fig. 4).

**High-Performance Liquid Chromatography (HPLC).** Ion-pairing reverse-phase HPLC experiments were carried out on a column liquid chromatography system equipped with an auto-sampler and a bio-inert quaternary pump (Agilent 1260 Infinity II Bio-Inert Pump G5654A, Agilent Technologies). A C18 capillary column (AdvanceBio Oligonucleotide 4.6x150 mm with particle size 2.7 μm, Agilent) was used to perform reverse-phase liquid chromatography. The temperature of the autosampler was set to 4 °C. The mobile phases consisted of two eluents. Eluent A was HPLC water (Sigma-Aldrich), 200 mM 1,1,1,3,3,3, -Hexafluoro-2-propanol (HFIP) (Carl Roth GmbH), 8 mM Triethylamine (TEA) (Carl Roth GmbH). Eluent B was a 50:50 (v/v) mixture of water and methanol (HPLC grade, Sigma Aldrich, Germany), 200 mM HFIP, and 8 mM TEA. The injection volume for each measurement was 100 μL. The samples were eluted with a gradient of 1% B to 58.6% B over the course of 45 min with a flow rate of 1 mL/min. Prior to the gradient, the column was flushed with 1% B for 5 min. Retention times were analyzed via a UV Diode Array Detector (Agilent 1260 Infinity II Diode Array Detector WR G7115A) at 260 nm with a bandwidth of 4 nm. Samples were diluted for HPLC loading in the following manner: 7.5 μL of sample, 105 μL nuclease-free water, and 75 μL of a 5M urea solution. They were heated to 95 °C for 2 min afterward to ensure de-hybridization of the strands and dissolution of any sediment. Then, 105 μL of the diluted samples were transferred into N9 glass vials (Macherey-Nagel GmbH) and stored at 4 °C in

the auto-sampler of the HPLC-MS system (1260 Infinity II, Agilent Technologies) until injection.

**Finite-Element Simulations.** Simulations were performed in 2D using COMSOL Multiphysics 5.4. The simulation file with all the detailed parameters is given in the supplement in binary format. Additionally, the simulation is given as an auto-generated report in a hierarchical html compressed into a Zip-File. For more detailed information, see *SI Appendix*, section 8.

**Data, Materials, and Software Availability.** The data and the codes that support the findings of this study are available at the following online repository (40).

**ACKNOWLEDGMENTS.** D. Braun acknowledges support from the European Research Council (ERC) Evotrap, Grant No. 787356, the Simons Foundation (Grant No. 327125), the CRC 235 Emergence of Life (Project-ID 364653263), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311, and the Center for NanoScience (CeNS). C. Weber acknowledges the ERC under the European Union's Horizon 2020 research and innovation programme (Fuelled Life, Grant No. 949021) for financial support.

Author affiliations: [a]Division Biological Physics, Max Planck Institute for the Physics of Complex Systems, Dresden 01187, Germany; [b]Center for Systems Biology Dresden, Dresden 01307, Germany; [c]Ludwigs-Maximilian-Universität München and Center for NanoScience, Munich 80799, Germany; and [d]Faculty of Mathematics, Natural Sciences, and Materials Engineering: Institute of Physics, University of Augsburg, Augsburg 86159, Germany

Author contributions: G.B., A.C.S., P.S., C.B.M., D.B., and C.A.W. designed research; G.B., A.C.S., P.S., A.K., Y.R., P.J., D.H., C.B.M., D.B., and C.A.W. performed research; A.K. and C.B.M. contributed new reagents/analytic tools; G.B., A.C.S., P.S., C.B.M., D.B., and C.A.W. analyzed data; and G.B., A.C.S., P.S., D.B., and C.A.W. wrote the paper.

1. M. P. Robertson, G. F. Joyce, The Origins of the RNA World. *Cold Spring Harb. Persp. Biol.* **4**, a003608 (2012).
2. K. Kruger *et al*., Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
3. M. J. Fedor, J. R. Williamson, The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **6**, 399–412 (2005).
4. G. Walter, The RNA world Superlattices point ahead. *Nature* **319**, 618 (1986).
5. K. R. Birikh, P. A. Heaton, F. Eckstein, The structure, function and application of the hammerhead ribozyme. *Euro. J. Biochem.* **245**, 1–16 (1997).
6. C. Deck, M. Jauker, C. Richert, Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).
7. A. Mariani, D. A. Russell, T. Javelle, J. D. Sutherland, A light-releasable potentially prebiotic nucleotide activating agent. *J. Am. Chem. Soc.* **140**, 8657–8661 (2018).
8. T. Walton, W. Zhang, L. Li, C. P. Tam, J. W. Szostak, The mechanism of nonenzymatic template copying with imidazole-activated nucleotides. *Angew. Chem.- Int. Ed.* **58**, 10812–10819 (2019).
9. S. Wunnava *et al*., Acid-catalyzed RNA-oligomerization from 3',5'-cGMP. *Chem. Euro J.* **27**, 17581–17585 (2021).
10. I. Budin, J. W. Szostak, Expanding roles for diverse physical phenomena during the origin of life. *Annu. Rev. Biophys.* **39**, 245–263 (2010).
11. A. V. Tkachenko, S. Maslov, Spontaneous emergence of autocatalytic information-coding polymers. *J. Chem. Phys.* **143** (2015). https://pubs.aip.org/aip/jcp/article/143/4/045102/71438/Spontaneous-emergence-of-autocatalytic-information.
12. A. V. Tkachenko, S. Maslov, Onset of natural selection in populations of autocatalytic heteropolymers. *J. Chem. Phys.* **149** (2018). https://pubs.aip.org/aip/jcp/article/143/4/045102/71438/Spontaneous-emergence-of-autocatalytic-information.
13. C. B. Mast, S. Schink, U. Gerland, D. Braun, Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8030–8035 (2013).
14. M. Morasch *et al*., Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nat. Chem.* **11**, 779–788 (2019).
15. R. Mizuuchi *et al*., Mineral surfaces select for longer RNA molecules. *Chem. Commun.* **55**, 2090–2093 (2019).
16. L. H. de Oliveira *et al*., When RNA meets montmorillonite: Influence of the pH and divalent cations. *Appl. Clay Sci.* **214**, 106234 (2021).
17. W. M. Aumiller, F. Pir Cakmak, B. W. Davis, C. D. Keating, RNA-based coacervates as a model for membraneless organelles: Formation, properties, and interfacial liposome assembly. *Langmuir* **32**, 10042–10053 (2016).
18. Bj. Jeon *et al*., Salt-dependent properties of a coacervate-like, self-assembled DNA liquid. *Soft Matter* **14**, 7009–7015 (2018).
19. M. Nakata *et al*., End-to-end stacking and liquid crystal condensation of 6-to 20-base pair DNA duplexes. *Science* **318**, 1276–1279 (2007).
20. G. Zanchetta, M. Nakata, M. Buscaglia, N. A. Clark, T. Bellini, Liquid crystal ordering of DNA and RNA oligomers with partially overlapping sequences. *J. Phys. Condens. Matter* **20**, 494214 (2008).
21. D. T. Nguyen, O. A. Saleh, Tuning phase and aging of DNA hydrogels through molecular design. *Soft Matter* **13**, 5421–5427 (2017).
22. Z. Xing *et al*., Microrheology of DNA hydrogels. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8137–8142 (2018).
23. Y. Sato, T. Sakamoto, M. Takinoue, Sequence-based engineering of dynamic functions of micrometer-sized DNA droplets. *Sci. Adv.* **6**, eaba3471 (2020).
24. D. M. Mitrea, R. W. Kriwacki, Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Commun. Sig.* **14**, 1–20 (2016).
25. L. M. Barge *et al*., Thermodynamics, disequilibrium, evolution: Far-from-equilibrium geological and chemical considerations for origin-of-life research. *Orig. Life Evol. Biosph.* **47**, 39–56 (2017).
26. F. Westall *et al*., A hydrothermal-sedimentary context for the origin of life. *Astrobiology* **18**, 259–293 (2018).
27. J. SantaLucia Jr, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460–1465 (1998).
28. O. A. Saleh, B. J. Jeon, T. Liedl, Enzymatic degradation of liquid droplets of DNA is modulated near the phase boundary. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16160–16166 (2020).
29. Y. Xing *et al*., Self-assembled DNA hydrogels with designable thermal and enzymatic responsiveness. *Adv. Mater.* **23**, 1117–1121 (2011).
30. S. Biffi *et al*., Phase behavior and critical activated dynamics of limited-valence DNA nanostars. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 15633–15637 (2013).
31. S. Agarwal, D. Osmanovic, M. A. Klocke, E. Franco, The growth rate of DNA condensate droplets increases with the size of participating subunits. *ACS Nano* **16**, 11842–11851 (2022).
32. D. Sun, L. H. Hurley, The importance of negative superhelicity in inducing the formation of G-quadruplex and i-motif structures in the c-Myc promoter: Implications for drug targeting and control of gene expression. *J. Med. Chem.* **52**, 2863–2874 (2009).
33. M. Colombier *et al*., Degassing and gas percolation in basaltic magmas. *Earth Planet. Sci. Lett.* **573**, 117134 (2021).
34. F. Westall *et al*., A hydrothermal-sedimentary context for the origin of life. *Astrobiology* **18**, 259–293 (2018).
35. G. H. Fredrickson, S. T. Milner, L. Leibler, Multicritical phenomena and microphase ordering in random block copolymers melts. *Macromolecules* **25**, 6341–6354 (1992).
36. A. V. Dass *et al*., Rna oligomerisation without added catalyst from 2', 3'-cyclic nucleotides by drying at air-water interfaces. *Chem. Syst. Chem.* **5**, e202200026 (2023).
37. K. Zhang, J. Hodge, A. Chatterjee, T. S. Moon, K. M. Parker, Duplex structure of double-stranded RNA provides stability against hydrolysis relative to single-stranded RNA. *Environ. Sci. Technol.* **55**, 8045–8053 (2021).
38. A. Salditt *et al*., Thermal habitat for RNA amplification and accumulation. *Phys. Rev. Lett.* **125**, 048104 (2020).
39. J. N. Zadeh *et al*., Nupack: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
40. G. Bartolucci, A. Calaça Serrão, P. Schwintek, Supplementary codes. Selection_via_PS. GitHub. https://github.com/GiacoBarto/selection_via_PS. Deposited 4 August 2023.

# Supporting Information for

## Sequence self-selection by cyclic phase separation

**G. Bartolucci, A. Calaça Serrão, P. Schwintek, A. Kühnlein, Y. Rana, P. Janto, D. Hofer, C. Mast, D. Braun, C. Weber**

**Christoph A. Weber**
**christoph.weber@physik.uni-augsburg.de**
**Dieter Braun**
**dieter.braun@lmu.de**

### This PDF file includes:



### Other supporting materials for this manuscript include the following:

## Supporting Information Text

### 1. Multi-component phase separation subject to cyclic material exchanges

Here, we describe the concentration changes in a mixture of volume $V$ that is composed of $M$ different oligonucleotide sequences subjected to periodic exchange of material with a pool with fixed composition $\boldsymbol{c}_{\rm pool}$ . We introduce an $M$-dimensional vector $\bar{\boldsymbol{c}}(t)$, where the components of this vector are concentrations of sequences. Starting from the initial state $\bar{\boldsymbol{c}}(t_0)$, we perform $N$ exchange cycles, where each cycle is labeled with $n = 1, \ldots, N$ and composed of the two following steps:

*Phase separation step*: The homogeneous mixture of concentration $\bar{\boldsymbol{c}}(t_n)$ phase-separates into two coexisting phases. We denote the concentrations of the dense and dilute phase as $\boldsymbol{c}^{\rm I}(t_n)$ and $\boldsymbol{c}^{\rm II}(t_n)$, respectively. The volume of the dense phase is $V^{\rm I}(t_n)$ and thus the dilute phase occupies the volume $(V - V^{\rm I}(t_n))$. Mass and particle numbers are conserved during the phase separation step:

$$\bar{\boldsymbol{c}}(t_n) = \left[ \frac{V^{\rm I}(t_n)}{V} \boldsymbol{c}^{\rm I}(t_n) + \frac{V - V^{\rm I}(t_n)}{V} \boldsymbol{c}^{\rm II}(t_n) \right] . \qquad [1]$$

*Partial dilute phase removal step:* A constant fraction of the volume of the dilute phase, $\alpha(V - V(t_{n-1}))$, with the relative fraction $0 < \alpha < 1$, is replaced by the same volume taken from the pool, $\boldsymbol{c}_{\rm pool}$. The average composition thus changes according to

$$\bar{\boldsymbol{c}}(t_{n+1}) = \left[ \frac{V^{\rm I}(t_n)}{V} \boldsymbol{c}^{\rm I}(t_n) + \frac{V - V^{\rm I}(t_n)}{V} \Big( \alpha \boldsymbol{c}_{\rm pool} + (1-\alpha) \boldsymbol{c}^{\rm II}(t_n) \Big) \right] . \qquad [2]$$

For the general case where the initial average concentration is not equal to the average concentration of the pool, $\bar{\boldsymbol{c}}(t_0) \neq \boldsymbol{c}_{\rm pool}$ (full lines in Fig.1d-e), we determine the phase compositions $\boldsymbol{c}^{\rm I}(t_n)$ and $\boldsymbol{c}^{\rm II}(t_n)$, and the phase volumes $V^{\rm I}(t_n)$ at each cycle time $t_n$; details see Sect. 2. Note that during the selection kinetics, the average concentration $\bar{\boldsymbol{c}}(t_n)$ approaches the tie line passing through the pool concentration vector $\boldsymbol{c}_{\rm pool}$.

For the special case where the initial average concentration is equal to the average concentration of the pool, $\bar{\boldsymbol{c}}(t_0) = \boldsymbol{c}_{\rm pool}$ (dashed lines in Fig.1d-e), we can obtain an analytic solution, even for an arbitrary number of different components $M$. Only in this case, $\boldsymbol{c}^{\rm I}$ and $\boldsymbol{c}^{\rm II}$, remain constant in time since the average concentration moves along the tie line defined by the pool. The iteration rule Eq. (2) simplifies to:

$$\bar{\boldsymbol{c}}(t_{n+1}) = \Big[ \lambda(t_n) \boldsymbol{c}^{\rm I} + (1 - \lambda(t_n)) \Big( \alpha c_{\rm pool} + (1-\alpha) \boldsymbol{c}^{\rm II} \Big) \Big] , \qquad [3]$$

where $\lambda(t_n) = V^{\rm I}(t_n)/V$ is the relative volume of the dense phase. Using particle conservation in Eq. (1) for the pool concentration $\boldsymbol{c}_{\rm pool}$, remaining within the special case $\boldsymbol{c}_{\rm pool} = \bar{\boldsymbol{c}}(t_0)$, the interaction rule becomes:

$$\bar{\boldsymbol{c}}(t_{n+1}) = \Big[ \alpha\lambda_0 + (1 - \alpha\lambda_0)\lambda(t_n) \Big] \boldsymbol{c}^{\rm I} + \Big[ 1 - \alpha\lambda_0 - (1 - \alpha\lambda_0)\lambda(t_n) \Big] \boldsymbol{c}^{\rm II} , \qquad [4]$$

where $\lambda_0 = \lambda(t_0)$. Using particle conservation at time step $t_{n+1}$,

$$\bar{\boldsymbol{c}}(t_{n+1}) = \lambda(t_{n+1}) \boldsymbol{c}^{\rm I} + (1 - \lambda(t_{n+1})) \, \boldsymbol{c}^{\rm II} , \qquad [5]$$

we can identify the term in the first bracket of Eq. (4), as $\lambda(t_{n+1})$, the size of the dense phase at $t_{n+1}$, and obtain a recursion relation:

$$\lambda(t_{n+1}) = a \, \lambda(t_n) + b , \qquad [6]$$

where $a = 1 - \alpha\lambda_0$ and $b = \alpha\lambda_0$. For large times, the system reaches a stationary state

$$\lambda(t_\infty) = \frac{b}{1-a} = 1 . \qquad [7]$$

This can be used to rewrite the recursion in terms of $\delta\lambda(t_n)$ with $\delta\lambda(t_n) = \lambda(t_\infty) - \lambda(t_n) = 1 - \lambda(t_n)$. As a result, $\delta\lambda(t_{n+1}) = a \, \delta\lambda(t_n)$. Its solution reads $\delta\lambda(t_n) = \delta\lambda_0 \, a^n$ that can be written as

$$\lambda(t_n) = 1 - (1 - \lambda_0) \, (1 - \alpha\lambda_0)^n . \qquad [8]$$

This solution completely determines the evolution of the mean volume fraction:

$$\bar{\boldsymbol{c}}(t_n) = \Big[ 1 - (1 - \lambda_0) \, (1 - \alpha\lambda_0)^n \Big] \boldsymbol{c}^{\rm I} + (1 - \lambda_0) \, (1 - \alpha\lambda_0)^n \, \boldsymbol{c}^{\rm II} . \qquad [9]$$

The characteristic number of iterations required to converge to the stationary state $\lambda(t_\infty) = 1$ (and, consequently, $\bar{\boldsymbol{c}}_\infty = \boldsymbol{c}^{\rm I}$) is

$$n_{\rm c} = -\frac{1}{\log a} = -\frac{1}{\log[\alpha(1 - \lambda_0)]} . \qquad [10]$$

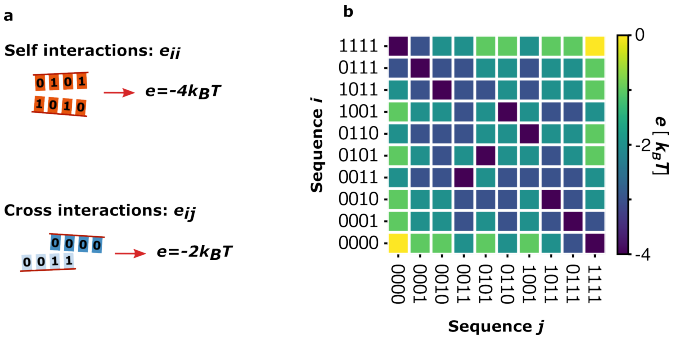


**Fig. S1. Interaction matrix for oligonucleotide sequences**. The interaction among two sequences is estimated identifying the longest complementary part of them and then counting the number of pairs within that region. The energetic scale of each pair is set to be $1k_BT$ and the negative sign indicates that sequences attract each other. **a** Example of energies calculated for a sequence interacting with a copy of itself (top), and two different sequences (bottom). **b** Interaction matrix $e_{ij}$ for all sequences of fixed length $L = 4$.

## 2. Phase separation in multicomponent oligonucleotide mixtures

To describe the phase behaviour of the system at each cycle, we chose the $T$-$V$-$N_i$ ensemble and introduce the free energy density $f(T, c_i) = F(T, V, N_i)/V$, where $F$ is the Helmholtz free energy depending on temperature $T$, volume $V$, and particle number $N_i$ of sequence $i$. The latter are related to concentrations via $c_i = N_i/V$. Specifically, we use the following Flory-Huggins free energy density

$$f = k_B T \left[ \sum_{i=1}^{M+1} c_i \ln(c_i v) + c_\mathrm{w} \ln(c_\mathrm{w} v_\mathrm{w}) + \frac{v^2}{2 v_\mathrm{w} k_B T} \sum_{i,j=1}^{M} e_{ij} c_i c_j \right] , \qquad [11]$$

where $c_\mathrm{w}$ and $v_\mathrm{w}$ denote for the solvent concentration and molecular volume, respectively. We assume that all sequences have the saem molecular volume, $v_i = v$ for $i = 1, \ldots, M$. Furthermore, molecular volumes are are constant making the system incompressible. Thus, the solvent concentration can be expressed as:

$$c_\mathrm{w} = \frac{1 - \sum_{i=1}^{M} c_i v}{v_\mathrm{w}} . \qquad [12]$$

The interaction matrix $e_{ij}$ encode the interactions energies among sequences. We estimated these parameters with the size of the maximum complementary portion of the two sequences, see Fig. S1 a for an example of energy estimation associated to self interactions and cross interactions. For simplicity, we neglect sequence orientation in the interaction matrix, i.e., 0101 and 1010 are considered the same). We also set to zero the interactions between sequences and the solvent $e_{i\mathrm{w}} = 0$. The matrix corresponding to all possible sequences of fixed length $L = 4$ is shown in Fig. S1 b.

From the free energy density Eq. (11), we can calculate the exchange chemical potentials of each sequence and the osmotic pressure:

$$\bar{\mu}_i = \frac{\partial f}{\partial c_i}, \qquad i = 1 \ldots M , \qquad [13a]$$

$$\Pi = -f + \sum_{i=1}^{M} \bar{\mu}_i c_i . \qquad [13b]$$

We searched for the domains in concentration space where the system with average concentrations $\bar{c}_i$ demixes into a solvent poor and a solvent rich phases, the dense and dilute phases, respectively, with concentrations $c_i^\mathrm{I}$ and $c_i^\mathrm{II}$. Number conservation relates both phase concentrations to the average concentration

$$\bar{c}_i = \frac{V^I}{V} c_i^\mathrm{I} + \frac{V - V^I}{V} c_i^\mathrm{II} , \qquad [14]$$

where $V^I$ and $V$ are the volumes of the solvent poor phase and of the system, respectively. The volume of the supernatant phase is $(V - V^I)$. For $\bar{c}_i$ and $V$ fixed, the concentrations of the phases $c_i^\mathrm{I}$ and $c_i^\mathrm{II}$ and the volume $V^\mathrm{I}$ are given by the solution of the following set of $(M + 1)$ equations,

$$\bar{\mu}_i(c_i^\mathrm{I}) = \bar{\mu}_i(c_i^\mathrm{II}) , \qquad i = 1, 2, \ldots M , \qquad [15a]$$

$$\Pi(c_i^\mathrm{I}) = \Pi(c_i^\mathrm{II}) , \qquad [15b]$$

together with the $M$ constraints in Eq. (14). Outside of multi-phase coexistence regions (see Fig. S2 b for a discussion), the solution of the equations above is unique. However, finding the concentrations of the phases is in general a difficult task from a

numerical perspective, in particular for large numbers of sequences $M$ increases. Numerical convergence requires suitable initial guesses. We thus developed a recursive method to find the coexisting phases concentrations $\boldsymbol{c}^{\mathrm{I}}$ and $\boldsymbol{c}^{\mathrm{II}}$, and volume $V^I$ for a given average concentration $\bar{c}_i$. This method is illustrated for a ternary mixture in Fig. S2b. The system is initialized at one edge of the phase diagram, corresponding to e.g. a system composed of sequence 1 and solvent. In this subspace the set of equations in Eq. (15) reduces to two and can be easily solved. We chose as initial average concentration $\bar{\boldsymbol{c}}(0) = (0.5, 0)$ and solve for $\boldsymbol{c}^{\mathrm{I}}(0)$, $\boldsymbol{c}^{\mathrm{II}}(0)$ and $V^I(0)$. We then consider a new average concentration $\bar{\boldsymbol{c}}(1)$ obtained displacing $\bar{\boldsymbol{c}}(0)$ along the direction $\bar{\boldsymbol{c}}(0) - \bar{\boldsymbol{c}}$ (orange arrow in Fig. S2b). The magnitude of the displacement vector is fixed to $\Delta$ and must be chosen sufficiently small. Now we solve for $\boldsymbol{c}^{\mathrm{I}}(1)$, $\boldsymbol{c}^{\mathrm{II}}(1)$ and $V^I(1)$ corresponding to $\bar{\boldsymbol{c}}(1)$, using as initial guess for the solver $\boldsymbol{c}^{\mathrm{I}}(0)$, $\boldsymbol{c}^{\mathrm{II}}(0)$ and $V^I(0)$. We iterate this procedure, letting $\boldsymbol{c}^{\mathrm{I}}(n)$, $\boldsymbol{c}^{\mathrm{II}}(n)$ varying along the dense and dilute binodal branch, respectively (orange line) until the average reaches $\bar{\boldsymbol{c}}$. At this point, the phase concentrations have converged to $\boldsymbol{c}^{\mathrm{I}}$ and $\boldsymbol{c}^{\mathrm{II}}$ (orange dots along the dense and dilute binodal branch, respectively). We then start from the other edges of the phase diagram, corresponding to a system composed only of sequence $i$ and solvent, with $i = 2, \dots M$.

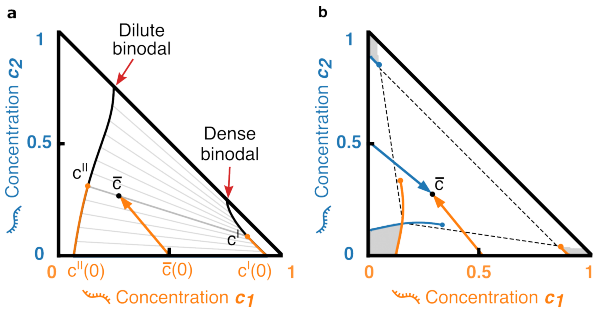


**Fig. S2. a** Determination of the phase concentrations and volume for a target average concentration $\bar{\boldsymbol{c}}$. The system is initialized at e.g. the $c_1$-edge of the phase diagram and moving the average towards $\bar{\boldsymbol{c}}$ iteratively update previous results as guesses for the numerical solver. Here, $e_{11} = -k_BT$, $e_{12} = e_{22} - 0.27k_BT$ and $e_{ij} = 0$ otherwise, as in Fig.1. **b** Identification of three phase coexistence regions as the locus of points $\bar{\boldsymbol{c}}$ for which the phase concentrations are different starting from different edges (orange and blue lines). Here $e_{11} = e_{22} = 5.33k_BT$, and $e_{12} = 1.33k_BT$, $e_{ij} = 0$ otherwise.

Our proposed method is able to determine volume and composition of two coexisting phases associated to a certain average concentration vector $\bar{\boldsymbol{c}}$, provided that there exists a segment joining $\bar{\boldsymbol{c}}$ and the midpoint of any edge that lies entirely in the demixing region (like the orange arrow in Fig. S2b). In other words, any phase-separating domain disconnected from all the corners would not be accessible. Another potential issue with this approach is that it accounts only for coexistence of two phases, neglecting states corresponding to three or higher coexisting phases. Luckily, regions corresponding to multi-phase coexistence can be self consistently detected with our algorithm. In fact, starting from different edges and solving Eq. (15) iteratively, would lead to different phase concentrations, if the average density lies within the multi-phase coexistence region. This is exemplified in Fig. S2c. Here, $\bar{\boldsymbol{c}}$ lies in the three phase coexistence. In this case starting from the $c_1$-only or $c_2$-only edge (orange and blue arrow, respectively) would lead to phase compositions that are different (orange and blue dots, respectively). Thus, in the following studies (see Section 3), we have checked that the average concentration $\bar{\boldsymbol{c}}$ lies in the two-phase coexistence region, making sure that starting from different corners we get the same phase concentrations.

## 3. Polymerization kinetics and cloud point analysis for multi-component sequence pools

We have applied the theory developed in sections 1 and 2 to a system composed of all the possible sequences of length $L$ made by two different units 0 and 1. We focus on the protocol in which the initial sequence pool and the one used to replace the supernatant at each iteration are identical ($\boldsymbol{c}_{\mathrm{pool}} = \bar{\boldsymbol{c}}(t_0)$). In this case, the evolution of the average concentration vector is governed by Eq.1. According to this equation, the kinetics of $\bar{\boldsymbol{c}}$ is determined by the pool concentration $\boldsymbol{c}_{\mathrm{pool}}$, the corresponding phase concentrations, $\boldsymbol{c}^{\mathrm{I}}$ and $\boldsymbol{c}^{\mathrm{II}}$, and the dense phase volume $V^I$ (see Appendix 2).

We now discuss how assembly kinetics of units, which leads to the formation of polymers of length $L$, provides a natural choice of the pool composition, $\boldsymbol{c}_{\mathrm{pool}}$, that is crucial to determine the selection propensity of the mixture.

**A. A minimal model for polymerization kinetics.** To calculate the abundance of each sequence in a mixture, we use a simplified polymerization kinetics model that describes the assembly of units 0 and 1 into sequences of length $L = 4$. This model was first introduced in Ref. (1), and characterizes subsequent addition of units using as input two parameters: $r$, the ratio between the number of 0 units and the total number of units, and the blockiness $b_l$, quantifying the units correlation along the chain. There the authors show that, at steady state, $r$ and $b_l$ are in correspondence with the conditional probabilities $p_{ji}$, with $i, j = 0, 1$, that a unit of type $i$ is followed by a unit of type $j$. In fact, we have

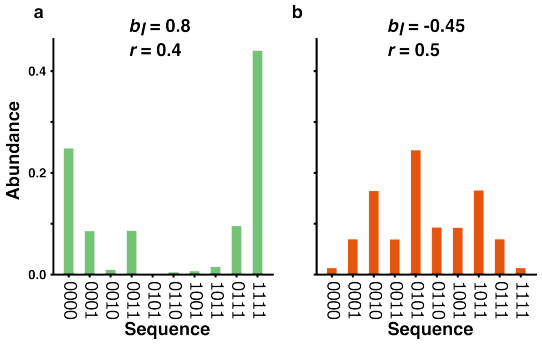


**Fig. S3. Sequence abundances generated via polymerization kinetics**. **a** $b_l = 0.8, r = 0.4$ **b** $b_l = -0.45, r = 0.5$. These are the same parameters as in Fig.5 a and b. Abundances are defined as sequence concentration normalized: $A_i = c_i / \sum_{i=1}^{M} c_i$. In both cases $L = 4$.

$$p_{00} = r(1 - b_l) + b_l \,, \quad [16]$$
$$p_{11} = r(b_l - 1) + 1 \,, \quad [17]$$
$$p_{10} = 1 - p_{00} \,, \quad [18]$$
$$p_{01} = 1 - p_{11} \,. \quad [19]$$

For $b_l = 1$, units of the same type are neighbored leading to only two sequences, i.e., 0000 and 1111. In contrast, for $b_l = 0$ there are no correlations and different units appear randomly. For $b_l = -1$, different units alternate generating a single sequence, i.e., 0101 for $r = 0.5$.

We can now fix $r$ and $b_l$, calculate $p_{ji}$ and use them to construct a long chain of length $L_c$. The sequence pool is then obtained by chopping the long chain into sequences of length $L$. In Fig. S3, we show an example of the abundance of each sequence for different choices of $r$ and $b_l$. Abundances are defined as sequence concentration normalized without taking the solvent into account:

$$A_i = \frac{c_i}{\sum_{i=1}^{M} c_i} \,. \quad [20]$$

For simplicity we neglect sequence orientation in the interaction matrix (see Fig.S2a) and thus grouped sequences which are related via reflection along the mid plane (e.g., 0101 and 1010).

**B. Cloud point determination.** We now discuss how 'accessible' phase separation is for sequence pools of different compositions. To this aim, we determined the minimum total oligomer concentration $c_{\rm tot} = \sum_{i=1}^{M} c_i$ that is needed for the mixture to phase separate for each sequence abundance distribution. This corresponds to locating the cloud point in the phase diagram, defined as the interception between the binodal and the line in which the sequence abundances are constant. The cloud point is a fundamental quantity in particular in experiments since typically only low oligomer concentrations are practically accessible. As discussed in the previous chapter, our model allows us to determine the abundances of all possible sequences given $r$ and $b_l$, thus we aim to locate the cloud point as a function of these two control parameters. The procedure is illustrated in Fig. S4a for a ternary mixture. Here, $e_{11} = -k_BT$, $e_{12} = e_{22} - 0.27k_BT$ and $e_{ij} = 0$ otherwise, as in Fig1 e. The black solid line represents the binodal, while the dashed line is the equi-composition line which corresponds to a constant $c_1/c_2$ ratio. The latter is the two-dimensional analogue of the line defined by fixing the sequence abundances and varying the total oligomer concentration that corresponds to a fixed choice of $r$ and $b_l$. The cloud point is located at the interception between the two lines (red dot indicated with "cp" in Fig. S4a). To locate the cloud point, we repeat the procedure described in Appendix 2 targeting average concentrations on the equi-composition line with decreasing total oligomer concentration $c_{\rm tot}$.

In red, we highlight the cloud point tie line. Its interception with the dense binodal branch gives the dense phase associated with the cloud point, see Fig. S4b. The latter determines the final system composition since we chose a point along the tie line which also corresponds to the average composition of the pool. An example of initial and final sequence abundances, corresponding to $r = 0.4$ and $b_l = 0.8$ (same values as in Fig.5b, in the main text), is displayed in Fig. S4c.

With this method, we compute the point along the fixed-composition line which belongs to the demixing region but is the closest to the cloud point. We compute the corresponding relative dense phase size $\lambda = V^{\rm I}/V$, see Fig. S4d, which is always lower than a few percent. This confirms that the distance between the average density obtained with this method is very close to the dilute phase vector, meaning that we can efficiently locate the could point. In Fig. S4e,f we show the total oligomer concentration $c_{\rm tot}$ at the cloud point and in the corresponding dense phase. As the blockiness decreases, $c_{\rm tot}$ in the dilute phase and the dense phase respectively decrease and increase, in a non-monotonic fashion.

As anticipated, we chose the pool along the cloud point tie line (very close to the cloud point itself, just inside the demixing region). To characterize selection propensity, we introduce the angle between the equi-composition line and the pool tie line

(depicted in Fig. S4b)

$$\theta = \arccos\left(\frac{\boldsymbol{c}^{\mathrm{II}} \cdot \left(\boldsymbol{c}^{\mathrm{I}} - \boldsymbol{c}^{\mathrm{II}}\right)}{\|\boldsymbol{c}^{\mathrm{II}}\| \, \|\boldsymbol{c}^{\mathrm{I}} - \boldsymbol{c}^{\mathrm{II}}\|}\right) . \quad [21]$$

This parameter quantifies how much the final state deviates from the initial pool.

We then introduce the maximum abundance and its variation at the end of the protocols cycles

$$h(t) = \max_{i=1,\dots M} A_i(t) \qquad \text{and} \quad \Delta h = h(t_f) - h(t_0)\,, \quad [22]$$

where $A_i$ are the sequence abundances, see Eq. (20) and graphical representation in Fig. S4c. Positive, large values of $\Delta h$ represent cases in which the final sequence distribution is more peaked around one single sequence, as opposed to the initial distribution, hence strong selection propensity. In Fig. S4g and h we show how these quantifiers vary as a function of $r$ and $b_l$. Pools characterized by slightly negative blockiness values have strong selection propensity, but the sequence distribution is similar before and after phase separation cycles (high $\Delta h$ but low $\theta$). This means that the initial distribution is already dominated by a sequence (0101) and after cycles of phase separation the same sequence remains the most abundant, and becomes even more dominant. This is the scenario that we called “bias amplification" in Fig.5c. Pools characterized by intermediate blockiness values show weaker selection propensity, but as $r$ deviates from $1/2$, the initial and final sequence distributions are different (moderate $\Delta h$ but high $\theta$). In this case the initial distribution is biased towards a particular sequence (0000 or 1111, for example) and after cycles of phase separation the most abundant sequence is another one (0011) which is slightly more dominant than the initial most abundant sequence. This case corresponds to the scenario of a “new selection route", see Fig.5c. Note that in Fig. S4 and in Fig.5 in the main text, $k_B T = 0.8$ in units of a single base pair energy.

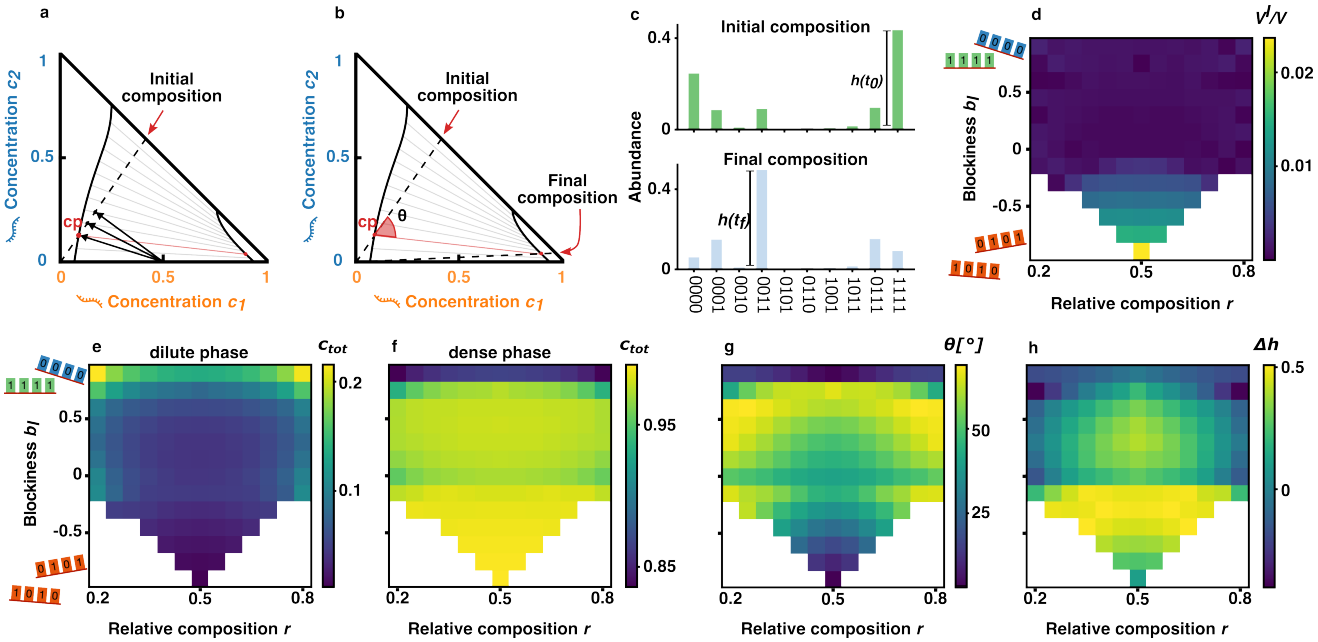


**Fig. S4. Identification and characterisation of the cloud point in a multi-component mixture. a, b** We illustrate the location of the cloud point in the case of a binary mixture with $e_{11} = -k_BT$, $e_{12} = e_{22} - 0.27k_BT$ and $e_{ij} = 0$ otherwise, as in Fig1d. **a** Once the relative sequence concentrations are fixed, the cloud point (red dot with "cp" label) can be located by targeting points with increasing solvent amount and applying the method described in Fig. S2. **b** The angle $\theta$ measures the distance in composition between the dense and dilute phases which, with our choice of protocol, are the final and the initial state of the mixture. In **c**-**h** we focus on the case of $L = 4$ sequences, with the interaction matrix depicted in Fig. S2 and initial composition obtained varying $r$ and $b_l$ as explained in the previous Sec. A. Deviating from the initial composition (high $\theta$), does not always correspond to selection. In fact, to assess whether the final pool has a stronger sequence enrichment than the initial pool, we compare the initial abundances with the final and maximal abundances, indicated in **c** as $h(t_f)$ and $h(t_0)$. Here $r = 0.4$ and $b_l = 0.8$, as in Fig.5). In **d** we check that the cloud point is efficiently located, i.e., the initial pool is characterised by $V^{\mathrm{I}}/V \ll 1$. In **e** and **f** we show the total concentration at the cloud point of the dilute and dense phase, respectively. **g** The angle $\theta$ reveals that the final pool deviates significantly from the initial composition for positive, intermediate blockiness values, i.e., $b_l \simeq 0.5$. **h** $\Delta h$ shows that the final pool has improved in selectivity with respect to the initial pool negative, moderate blockiness values, around $b_l \simeq -0.5$. Here, $k_BT = 0.8$ in units of a single base pair energy.

## 4. NUPACK folding predictions

The sequence pairs were designed to be as short as possible, while still forming a branching network upon hybridization. The goal was to design a sequence pair composed of two sequences containing at least 3 unique binding regions (for the first strand a, b, and c and for the second strand a', b' and c', respectively). See Fig.2 for the binding scheme), such that every strand bound to the first one results in an additional binding site available. This yields $b = 3 + n$ vacant binding sites with $n$ being the number of strands bound to the first one. Therefore, the strand network will grow faster the more strands are already bound. For the design, the Nupack Tool 3.2.2 (2) was used (see Fig. S6 for an exemplary code). While keeping the center part of each sequence restricted to G or C only, the outer sections ("arms") were varied, to either consist of all four bases (Sequence pair 1), A or T only (Sequence pair 2) or GC only (Sequence pair 3).

The resulting output sequences were iteratively mutated afterwards. Upon each mutation of the strands, the sequences were analysed using the Analysis segment of the online NUPACK tool in order to check their ability to form a network (Fig. S5). Specifically, the presence of secondary structures that included strands bound to three other strands were considered indicative of network formation. Spacers of two bases were inserted between the three segments of each sequence. These were chosen to be either TT (Sequence pair 1 and 3) or CC (Sequence pair 2) in order to not have complementarity with the arms and reduce their participation in the overall secondary structure. Inspired by Ref. (3), these spacers allow for more flexibility of the segments by minimizing angular constraints.

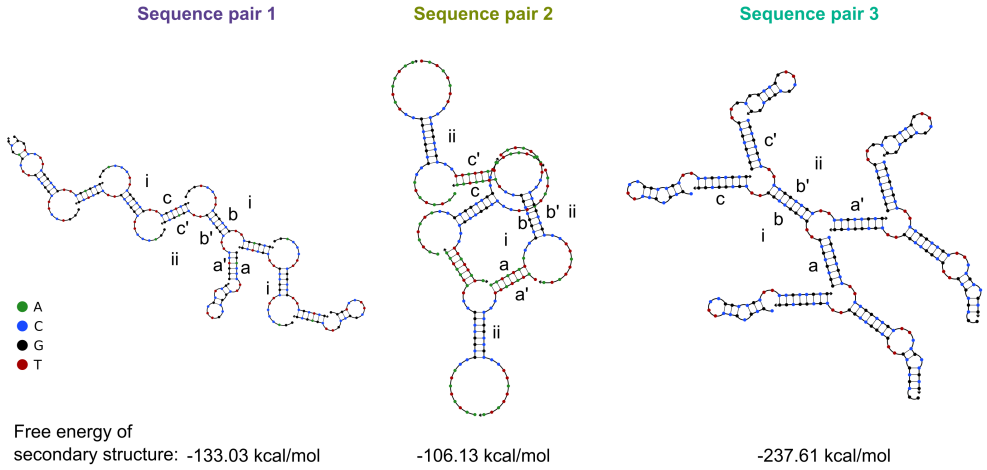


**Fig. S5. Folding predictions of NUPACK.** The sequence pairs 1, 2 and 3 were analysed using the online NUPACK tool, which is based on a binding energy data set (4). Prediction settings were: 15°C, 125mM NaCl, 10mM MgCl2 and 2µM per strand, secondary structure of up to 10 strands. For each system, the strand design leads to the binding pattern necessary to produce a network, in which each strand bound to the first increases the total number of possible binding sites by 1. Labeled parts of the sequences (e.g. i, ii, a, b' etc.) indicate the sequence segments as displayed in Fig.2.

```
temperature[C] = 37.0
material = dna

# domains
domain a1 = S2
domain a2 = W2
domain a3 = S2
domain b1 = S2
domain b2 = W2
domain b3 = S2
domain c1 = S2
domain c2 = W2
domain c3 = S2
domain s = T2

# strands
strand s1 = a1 a2 a3 s b1 b2 b3 s c1 c2 c3
strand s2 = a3* a2* a1* s b3* b2* b1* s c3* c2* c1*

# complexes
complex ca = s1 s2
complex cb = s1 s2
complex cc = s1 s2

# target structures
ca.structure = D6(U16 +) U16
cb.structure = U8 D6( U8 + U8) U8
cc.structure = U16 D6(+ U16)

# tubes
tube tub = ca cb cc
tub.ca.conc[M] = 1e-6
tub.cb.conc[M] = 1e-6
tub.cc.conc[M] = 1e-6

prevent = AAAA,CCCC,GGGG,UUUU,AA,GGG
stop[%] = 10
```

**Fig. S6. Exemplary Nupack design code** For DNA at a temperature of 37°C, two strands were designed by segmentation into smaller domains (a, b and c). In the # domains section, the a, b and c are split into even smaller domains of similar nucleotide content (S being G or C, and W being A or T). In this example code, TT spacers were defined via "s = T2". In "# strands", the sequences of the system are defined, reading from 5' to 3'. The * denotes complementarity. Hybridization constraints are shown in "# target structures", where the 3 possible binding interactions from "# complexes" between strand 1 and strand 2 are defined further. "D6U16" for example denotes a 6 base pair region followed by a 16 unpaired region. The three target structures are also depicted in Fig.2 a. Preventing specific sequence-patterns, such as AAAA or GGGG, above a certain relative amount, can be useful to avoid for example G-quadruplex structures or unwanted stacking.

## 5. Other Sequence pairs

Sequence pairs 2 and 3 did not form a dense phase. Note that these have longer arms (a,a') and (c,c') than sequence pair 1, which should lead to stronger binding interactions through base-pairing, and therefore more stable networks. A possible explanation for this would be that sequence pairs 2 and 3 have arms composed of only two nucleotides (GC or AT, respectively). This leads to an increased number of possible base-pairing interactions, through partial binding of the arms, due to the reduced alphabet. Most of these would be non-specific, i.e. would lead to alternative secondary structures that could inhibit network formation and suppress the growth of dense DNA nuclei.

We also tested 5 additional sequence pairs to have a better understanding of the sequence dependence on the formation of a dense phase. The sequence pairs 4-8 all have 6 nucleotide long segments (like sequence pair 1), which contain only G/C. The exception is sequence pair 4, which is identical to sequence pair 5, but one nucleotide in segments a and c is changed to A/T (highlighted in red in Fig. S7b). Confirming our observations for sequence pair 1, this change triggers sequence pair 4 to phase separate and sediment in contrast to its counterpart sequence pair 5 (See Fig. S7c). This suggests that the selection mechanism is able to separate single base changes and heavily favors sequence including all four nucleotides. Even after changing the GC-only segments multiple times, we could not observe sedimentation (See Fig. S7c - sequence pair 6 and 8). Removing the TT spacers from sequence pair 5, did also not lead to observable phase separation (See sequence pair 7).

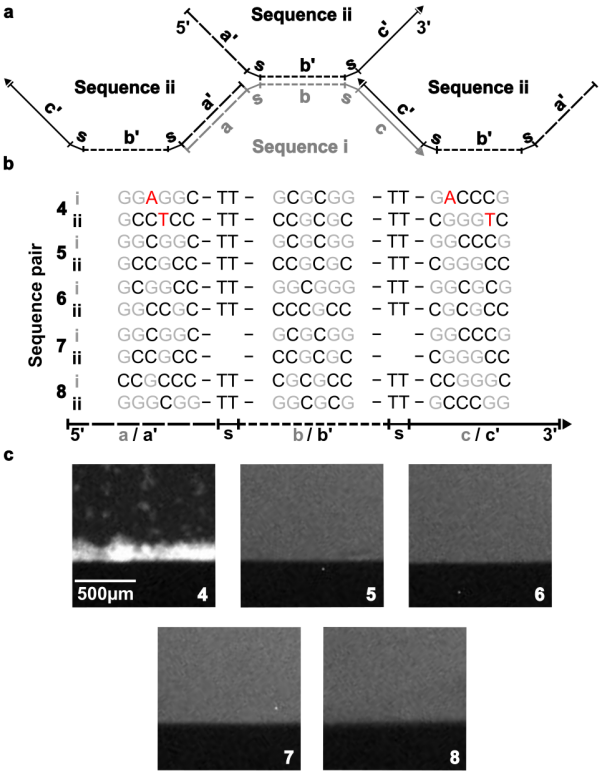


**Fig. S7. Screening of other sequence pairs a**,Sequence i is composed of three segments a, b, and c with spacer s. Its pair ii consists of reverse complements a', b' and c'. The inverted arrangement of a' and b' creates a network from three binding sites and prevents the formation of a linear double-stranded duplex. **b**, Sequence pairs 4 - 8 listed with all G's in grey to guide the eye. A and T are highlighted in red, as sequence pair 4 is the only one, which contains A or T in sections a/a' or c/c', respectively. System 7 does not contain TT spacers between the individual sequence segments. **c** Fluorescence micrographs of sedimentation experiments. Samples were initially heated to 60℃ to ensure dehybridization and then slowly (0.5K/min) cooled to 10℃. Buffer conditions were 10mM TRIS pH7, 10mM MgCl2, 500mM NaCl, 50$\mu$M of each DNA strand and 5X SYBR Green I). Images were taken continuously and the samples were given up to 7 hours to sediment. All micrographs depict the bottom of the well after 7 hours. The only system to phase separate and sediment is sequence pair 4.

## 6. pH and salt concentration screening

In order to understand the dependence of the sequence pair 1 phase separation on the salt and pH conditions a screening was performed. Firstly, we fixed the buffer conditions to 10mM TRIS pH 7 and varied both NaCl and $MgCl_2$ between 0-250mM and 0-10mM respectively. Fluorescence micrographs shown in Fig. S8a). In the absence of $MgCl_2$, phase separation is only observed at the highest concentration of NaCl tested (250mM). In the presence of 10mM $MgCl_2$, phase separation occurs for any concentration of NaCl in the range.

Additionally, the influence of the buffer conditions were tested by fixing the salt concentration to the standard conditions used throughout the manuscript (10mM $MgCl_2$ and 125mM NaCl) and varying the pH between 5.5 and 9 (see Fig. S8b ). Coexisting phases are observed for both pH 7 buffers screened, Tris and HEPES, suggesting that the phase separation is not buffer dependent. Decreasing the pH, by using MES pH 5.5, led to no observable DNA aggregates. Meanwhile, increasing the pH to 9, now buffering with CHES, did not have detrimental effect to the phase separation. Different buffers were used in order to assure buffering capacity for the desired pH range.

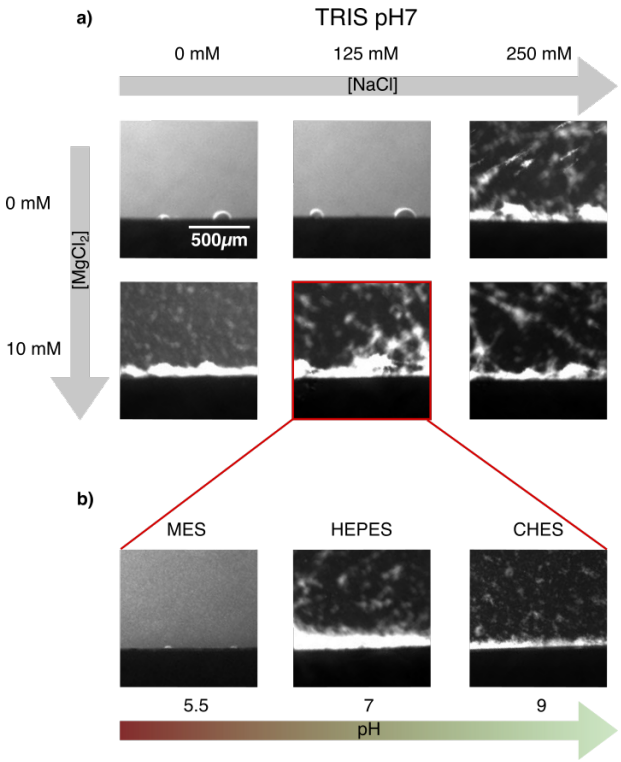


**Fig. S8. pH and salt concentration screening for the phase separation of Sequence pair 1. a**, Mixtures with 25μM of sequence pair 1 in 10mM TRIS buffer pH 7 and varying concentrations of NaCl and $MgCl_2$ were analysed through fluorescence microscopy (5X Sybr Green I was added). The temperature protocol is as described for Fig. 2c, denaturing the samples at 95℃, then fast cooling to 65℃ and lastly cooling to 15℃ at a rate of 6K/min. The micrographs show a cut-out of the bottom of the microfluidic chamber after at least 3 hours have passed since the temperature reached 15℃. **b** Keeping the salt concentration fixed at 125mM NaCl and 10mM $MgCl_2$, the buffer (and consequently the pH) was now varied between 5.5 and 9. Concentration of sequence pair 1 was 25μM, SYBR Green I was 5X and the temperature protocol was kept the same. The micrographs show either homogeneous fluorescence for the conditions no phase separation is observed, or a layer of sedimented DNA at the bottom of the chamber.

## 7. Sedimentation analysis

In order to perform the sedimentation analysis, the time lapse stack of micrographs for each sample is loaded into a self-written Labview script (Fig. S9). Since SyBR Green I fluorescent intensity scales linearly with the amount of dsDNA in solution, it can be used for quantification (5). The first micrograph is acquired before the sedimentation starts. It is used as a reference image after the temperature has reached 15°C. All the remaining micrographs are divided by this one to obtain relative concentration $c/c_0$, where $c_0$ is the sequence concentration in the chamber immediately after flushing, i.e. the pool concentration.

When a sediment is present, the relative concentration through the sediment is obtained by measuring the concentration along a defined line perpendicular to the wall of the well, which we parametrize by $x$. The maximum of relative concentration occurs at the center of the sediment. Sediment height is determined along the $x$-direction from the center concentration $c_{\max}$ until the value has reached a $0.5\,c_{\max}$. The sediment height is then the distance where $c > 0.5\,c_{\max}$. The relative average sediment concentration is calculated by averaging over all points for which $c > 0.5\,c_{\max}$.

Using the sediment height $h_{\text{sed}}$ and average relative concentration $\bar{c}/c_0$, we calculated the total amount of sedimented material $N_{\text{sed}}$ through Eq. (23), where $c_0$ is the initial concentration, $L$ is the length of the chamber and $d$ is the depth of the chamber:

$$N_{\text{sed}} = \left(\frac{\bar{c}}{c_0}\right) c_0\, h_{\text{sed}}\, L\, d\,. \qquad [23]$$

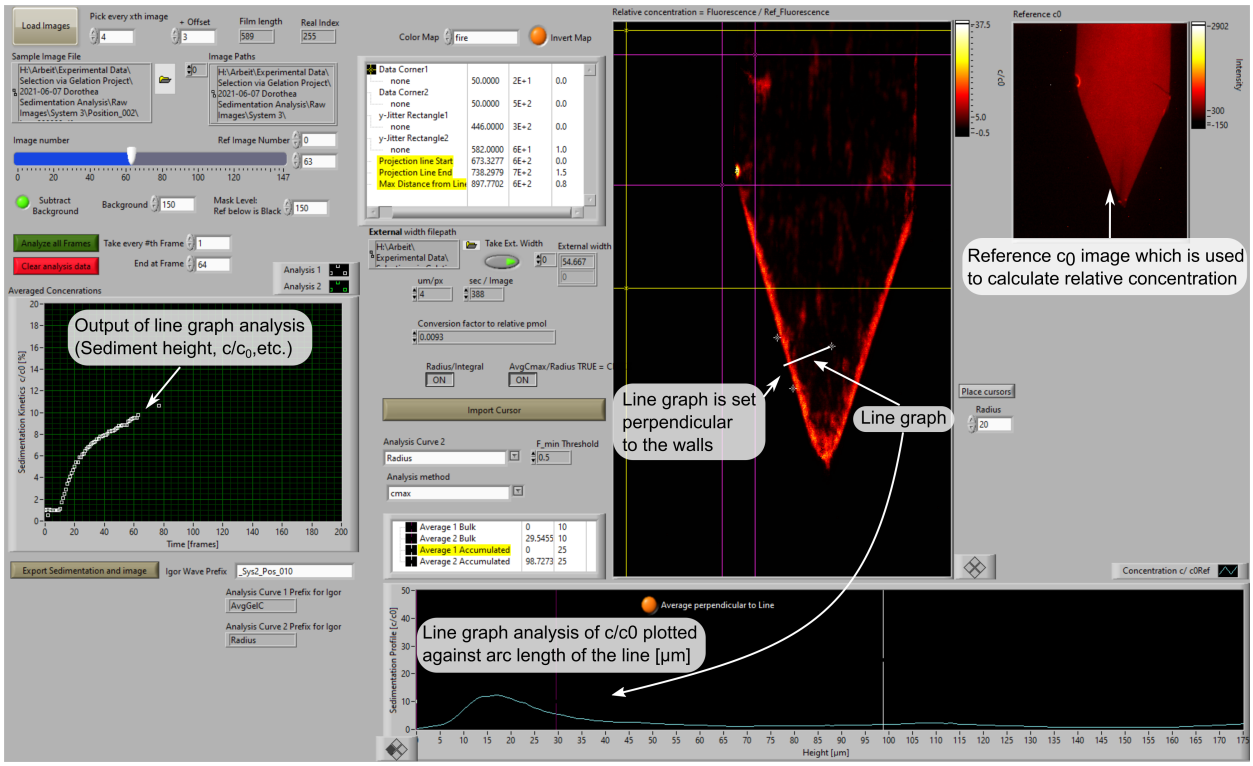


**Fig. S9. Screenshot of the self-written sedimentation analysis software in LabVIEW.** A stack of images is analysed by defining a line perpendicular to the chamber wall ($x$-direction), along which the relative concentration $c/c_0$ is plotted for each image.

## 8. Melting curves

Thermal melting curves were measured using SYBR Green I fluorescence in a thermal cycler with read-out. The samples were mixed the same way as for the sedimentation experiments: 25µM of each DNA strand, 5x Sybr Green I, 10mM Tris Buffer pH 7, 125mM NaCl and 1mM $MgCl_2$. Three independent mixtures were pipetted to provide independent triplicates. Additionally, a reference mixture with Sybr Green, buffer and salts was also measured to correct the data for background signals.

The analysis of the melting curves was done with a self-written Labview script and based on the baseline adjustment described in (6). First the signal from the background fluorescence is substracted from the fluorescence of the sample. Afterwards, the lower and higher baseline (linear) functions are determined and used for the baseline adjustment. These correspond to fully bound and fully unbound duplex states, respectively.

The corrected data are then exported to "Igor Pro 6.37" and fitted with a sigmoidal function, where the midpoint fitting parameter corresponds to the melting temperature $T_m$ (see Fig. S10). The $T_m$ of all the sequence pairs was below the denaturing temperature used in the thermal protocol (95°C, see Fig.3a), so we ensure the mixture is homogeneous before triggering phase separation through cooling.

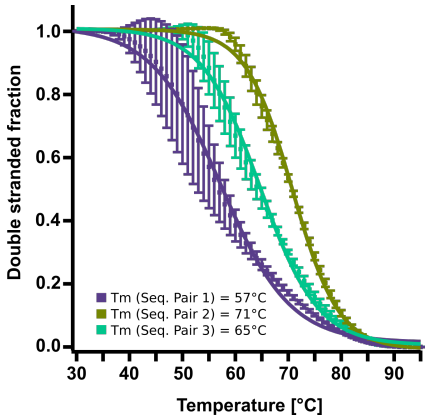


**Fig. S10. Melting curves of sequence pair 1, 2 and 3.** Each point corresponds to the average between independent triplicates. Error bars depict one standard deviation of the mean. The mixtures contained both strands of each system at 25µM each DNA strand, 10mM Tris Buffer pH 7, 125mM NaCl and 10mM MgCl2. SYBR Green I concentration was 5X.

## 9. Sedimentation dynamics

In order to describe the sedimentation process, we assume the particle to behave like sinking spherical objects with very small Reynolds numbers in a viscous fluid upon which a drag force is exerted. Stokes' law for the frictional force is given by

$$F_d = 6\pi\mu R v\,, \qquad [24]$$

where $\mu$ is the dynamic viscosity of the surrounding liquid, $R$ is the radius of the spherical object, and $v$ its flow velocity. The gravitational force on a sinking sphere is per Archimedes proportional to the density difference of the object and the fluid it displaced:

$$F_g = \Delta\rho\frac{4}{3}\pi R^3\,, \qquad [25]$$

$$\text{with} \quad \Delta\rho = \rho_p - \rho_f\,. \qquad [26]$$

Here, $\rho_p$ describes the density of the sphere or particle and $\rho_f$ the density of the displaced liquid. The terminal velocity of a sinking particle is therefore given when $F_d = F_g$:

$$v = \frac{\Delta\rho g R^2}{9\mu}\,, \qquad [27]$$

$$\text{and rearranged for } \Delta\rho\text{:} \quad \Delta\rho = \frac{9v\mu}{2gR^2}\,. \qquad [28]$$

This gives a strong size dependence of $R^2$ for the sedimentation speed of the aggregates. Since the resolution of our fluorescence microscope was $4\mu m/pixel$, the accuracy of measuring particle sizes was limited. Therefore, the smallest observable sedimenting aggregates were about 4 $\mu m$ in size. Whether the sedimentation process starts at that size, we cannot say with certainty. Furthermore, we used a self-written LabView particle tracker program to measure the sedimentation speed distribution. In order to avoid any convective flows caused by cooling the chamber to 15°C, which results in small temperature gradients towards the front of the chamber (at room temperature) we let the chamber cool to room temperature after the temperature

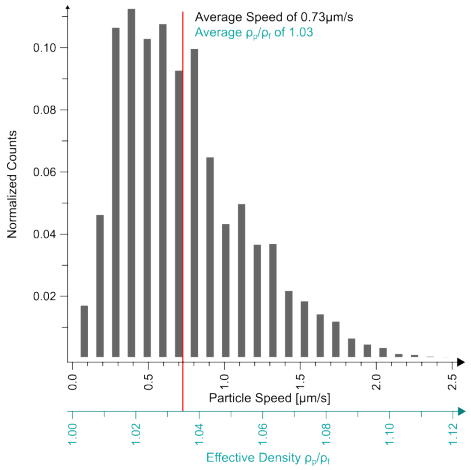


**Fig. S11.** Histogram of particle speeds observed using the self-written particle tracker LabView program. Turquoise axis indicates the corresponding density of the individual particles $\rho_p$ compared to water $\rho_f$. Black axis indicates the measured particle speeds. The red vertical line shows the average for both. The temperature protocol was adjusted to end at roomtemperature (25℃) instead of 15℃ in order to avoid convection. Other parameters remained unchanged: 125mM NaCl, 10mM MgCl2, 5X Sybr Green I, 25uM of sequence pair 1, 10mM TRIS pH7.

spike, for the purpose of sedimentation tracking. The other parameters were the same as in the sedimentation experiments (125mM NaCl, 10mM MgCl2, 5X Sybr Green I, 25uM of sequence pair 1, 10mM TRIS pH7). Interestingly, we observed that with the final temperature of 25°C the aggregates did not grow past a few pixel in size (about 4-16$\mu m$).

We measured particle traces between 50min and 100min after the heat spike and found that the average sedimentation speed of the aggregates is 0.73 $\mu m/s$, while the highest observed speeds are around 2.5$\mu m/s$ (See Fig. S11). Since our camera is unable to pick up more detail in order to reliably measure particle sizes, we assumed them to have a constant radius of 3 $\mu m$. This allows using equation Eq. (28) to convert the individual sedimentation speeds into density. For this assumption, a dynamic viscosity of 0.9 mPas for water was estimated based on Ref. (7) and the the radius $R$ was kept constant. As displayed in Fig. S11, the average density of sedimenting aggregates was found to be 1.03 times higher than the density of water. For the observed particle density, we calculated the minimal chamber height or sedimentation length, $l$, that would be necessary for sedimentation to occur. For this, we compared the gravitational energy of the particles to Boltzmann energies at room temperature. Boltzmann fluctuations depend solely on temperature, while the potential energy of the particles scales with $R^3$ and linearly with $\Delta\rho$. The Boltzmann distribution reads:

$$p \propto \exp\left\{\frac{\Delta\rho \cdot V \cdot g \cdot h}{k_B T}\right\} \quad [29]$$

$$\propto \exp\left\{h \cdot \frac{1}{l}\right\}, \quad [30]$$

$$[31]$$

with $l$ denoting the the sedimentation length:

$$l = \frac{k_B T}{\Delta\rho \cdot V \cdot g}\,. \quad [32]$$

Inserting the average parameters obtained from the particle tracking ($\Delta\rho = 1.03\rho_f = 1.03 \cdot 997\frac{kg}{m^3}$ and $V{=}\frac{4}{3}\pi * R^3$ with $R{=}3\mu m$), we obtain a characteristic sedimentation length of

$$l = \frac{k_B T}{\Delta\rho \cdot V \cdot g} \quad [33]$$

$$= \frac{k_B \cdot 298.15K}{1.03\rho_f \cdot \frac{4}{3}\pi \cdot (3\mu m)^3 \cdot g} \quad [34]$$

$$= 12.5\mu m\,. \quad [35]$$

This means that sedimentation can be observable if the chamber is smaller than 12.5$\mu m$, because thermal fluctuations would dominate otherwise sedimentation of aggregates caused by density differences. For chamber heights on length scales below the characteristic sedimentation length $l$, sedimentation would be negligible. Since our chamber height of about 10mm is multiple orders of magnitude larger, sedimentation plays a key role in our geometry. Even if the relative density difference between water and the aggregates would be 10 to 1000 times smaller (1.003 or 1.00003 $\cdot\rho_f$) the characteristic sedimentation length would not exceed the chamber height. Since the radius of particles influences the sedimentation length to the power of 3, the particles cannot be a lot smaller than the estimated value of 3$\mu m/s$. Since the chamber is 10mm high, a 10x smaller particle radius would result in a 1000x larger sedimentation length of about 10mm. Based on these calculations, we estimate the minimal particle radius for this geometry to be around 0.3$\mu m$.

## 10. Mass-balance over refeeding cycles

We use mass balance to determine the concentration of the bottom fraction for each cycle $n$. Using the concentration of the removed top fraction through HPLC analysis, and the concentration of the initial pool $c_0$, we calculate all the intermediate bottom fractions as plotted in Fig.3b. At step 1, the system phase separates which conserves the total mass (Fig.3a):

$$\bar{c}_n \, 2V_{\frac{1}{2}} = c_{\mathrm{top},n} \, V_{\frac{1}{2}} + c_{\mathrm{bottom},n} \, V_{\frac{1}{2}} \,, \qquad \text{and thus} \quad c_{\mathrm{bottom},n} = 2\bar{c}_n - c_{\mathrm{top},n} \,, \tag{36}$$

where $V_{\frac{1}{2}} = V/2$ denotes half the total volume ($V$ is the total volume of the system), $c_{\mathrm{top},n}$ and $c_{\mathrm{bottom},n}$ the concentrations of material in the top and bottom fraction in the $n$-th cycle, respectively. Note that the bottom fraction includes the sedimented dense phase. Moreover, $\bar{c}_n$ describes the concentration in the whole volume during the $n$-th cycle, which is conserved during phase separation (see Fig. S12).

At step 2 the top half of the system is removed and fed at step 3 by the pool of concentration $c_{\mathrm{pool}}$ (Fig.3a). The corresponding mass balance reads:

$$c_{\mathrm{pool}} \, V_{\frac{1}{2}} + c_{\mathrm{bottom},n} \, V_{\frac{1}{2}} = \bar{c}_{n+1} \, 2V \,. \tag{37}$$

Using Eq. (36) to substitute Eq. (37), $\bar{c}_{n+1}$ can be written as a function of $c_{\mathrm{top},n}$, $c_n$ and $c_{\mathrm{pool}}$. This relation can be used to obtain an equation for $c_{\mathrm{bottom},n+1}$, through substitution in Eq. (36):

$$\bar{c}_{n+1} = \bar{c}_n + \frac{c_{\mathrm{pool}} - c_{\mathrm{top},n}}{2} \,, \tag{38}$$

$$c_{\mathrm{bottom},n+1} = 2\bar{c}_n + c_{\mathrm{pool}} - c_{\mathrm{top},n} - c_{\mathrm{top},n+1} \,. \tag{39}$$

Since for $n = 0$, $\bar{c}_n = c_{\mathrm{pool}}$ and by measuring the top fraction concentrations $c_{\mathrm{top},n+1}$, we can recursively calculate all the bottom fraction concentrations using Eq. (39).

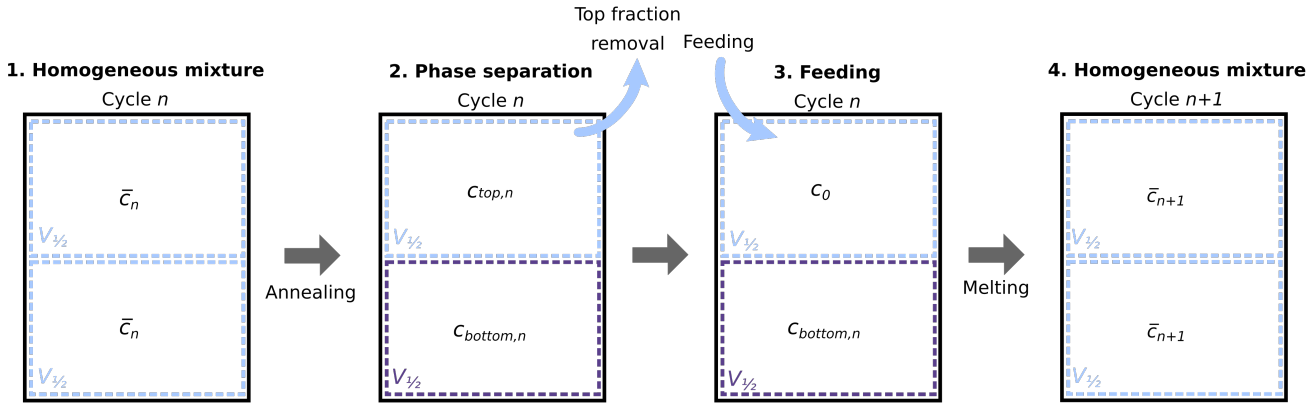


**Fig. S12. Schematic of a feeding cycle and the corresponding concentrations for top and bottom volume fractions.**

## 11. Feeding cycles in the absence of SYBR Green I

In order to understand the potential influence of SYBR Green I on the phase separation and sedimentation of the sequence pair 1, we repeated the feeding cycle experiment presented in Fig3b without including SYBR Green I 5X. The concentration of each of the sequence pairs over cycles of refeeding, for a mixture of sequence pair 1 and 2 is plotted in Fig. S13 and for sequence pairs 2 and 3 in the corresponding inset. The quantification was performed with HPLC at 260nm UV detection. Similarly to the original experiment containing SYBR Green I, the concentration of sequence pair 1 increases linearly in the bottom fraction and is depleted in the top fraction. Additionally, sequence pair 2 also does not partition between the top and bottom fractions, maintaining an approximately constant concentration around 25µM, the initial concentration. For the control experiment, with sequence pairs 2 and 3, the concentration of both systems stays approximately constant over cycles.

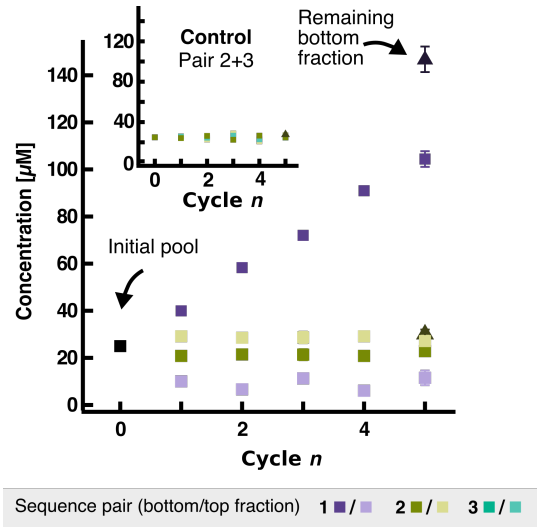


**Fig. S13. Concentration of the sequence pairs over five cycles of refeeding in the absence of SYBR green I.** The initial pool contained 25µM of sequence pairs 1 and 2. The sequences were slowly cooled to 15℃ at 6K/min similarly to the remaining sedimentation experiments (see Fig.3). The volume removed for the top fraction was half of the total volume ($\alpha$ = 0.5). The quantification was performed with HPLC through the measurement of absorbance at 260nm. For the mixture of sequence pairs 1 and 2, the concentration of sequence pair 1 increased linearly with a slope of approximately (16.1 $\pm$ 0.3) µM/cycle . Data are the average of three independent repeats and error bars correspond to one standard deviation of the mean.

The general behaviour of the pools is similar in the presence and absence of SYBR Green I at 5X concentration. However, the linear increase in concentration of sequence pair 1 in the bottom fraction is about 1.6 times higher without it, with an approximate slope of (16.1 $\pm$ 0.3) µM/cycle (compared to 10.2 $\pm$ 0.4 with SYBR Green I). This difference could be explained by the fact that SYBR Green I is an intercalating agent, which has been described to bind to the minor groove of a DNA helix (5). This could interfere with the structure of the dsDNA, as well as changing its melting temperature (8). This has also been reported for other intercalating dyes (9), and could influence the phase separation. However, for consistency with the remaining fluorescence microscopy experiments in Fig.2 and Fig.4 that require a fluorophore, we have kept the concentration of SYBR Green I constant over cycles in Fig.3.

## 12. Finite element calculations

The experimental chamber was designed using an Autodesk CAD software (Inventor 2019). The 2D drawing was then exported to COMSOL Multiphysics 5.4 using the Inventor live-link plugin. The geometry is therefore matching the experimental chamber in the two dimensions $x$ and $y$. Since the flows and the sedimentation of material occur only in $x$- and $y$-direction, omitting the $z$-axis allowed us to effectively screen parameters focusing on the key dynamics of the system.

The inlet and outlet of the well were emulated defining a constant normal inflow speed of 2µm/s as well as an outlet with a pressure boundary condition. Assuming a constant 20°C across the whole geometry, stationary laminar flow was solved with Navier-Stokes equations for conservation of momentum and continuity equation for conservation of mass. Since the flow speed of 2µm/s is very low (Mach number M $< 0.3$), the flow can be considered as incompressible. Thus, the density is assumed to be constant and the continuity equation reduces to the condition:

$$\nabla \cdot \boldsymbol{u} = \boldsymbol{0}\,. \qquad [40]$$

with $\rho$ denoting the mass density and $\boldsymbol{u}$ the velocity field. The Navier-Stokes equation then reduces to

$$\rho(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot[-p\boldsymbol{I} + \eta(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^T] + \boldsymbol{F} = \boldsymbol{0}\,, \qquad [41]$$

with $p$ being pressure, $\boldsymbol{I}$ the unity tensor, $\eta$ the fluid dynamic viscosity and $\boldsymbol{F}$ the external forces applied to the liquid. The reference pressure was set to 1[atm], the reference temperature was 20°C and all surfaces are described as non-slip boundary conditions.

Assuming the laminar flow to be stationary, we described the transport of a diluted species $c_i$ by combining convection and diffusion effects:

$$\frac{\partial c_i}{\partial t} + \nabla(-D_i\nabla c_i + \boldsymbol{u}c_i) = 0\,, \qquad [42]$$

where $D_i$ denotes the diffusion coefficient of species $i$. Sedimentation of oligomers was simulated assuming a downwards oriented flow speed $v_{sed}$ affecting diluted species only. With increasing local relative concentration $\frac{c}{c_0}$, the flow speed decreases from 0.1µm/s to 0, using an inbuilt smoothened heavyside step function, which is shown in Fig. S14. The $y$-component of the velocity field $\boldsymbol{u}$ used to describe the convective movement of diluted species $c_i$ then reads:

$$u_y = v - v_{sed} \quad \text{with} \quad v_{sed} = v_0 \cdot step\left(\frac{c_i}{c_0}\right)\,. \qquad [43]$$

Here, $v_{sed}$ ensures that molecules cannot sediment into an area, where the local concentration has already reached $c_i/c_0 = 10$. This maximum density was estimated using the fluorescence data from the experiments. Choosing the center of the step function to be at $c/c_0 = 7$ gave best results matching the experiments. The sedimentation speed of maximum 0.1 µm/s was chosen through observation of the distance traveled in between images for the very first visible phase separation DNA aggregates. The aggregates sediment quicker, the larger they become. These kinetics were not simulated. Rather, the system was treated as if it was filled with those small aggregates from the beginning, sedimenting with $v_{\mathrm{sed}}$. Meshing was done using an automatic finer physics-controlled mesh (see Fig. S14. After setting the parameters as above, a complete numerical solution could be found. A complete set of all input parameters can be found in Table S1.

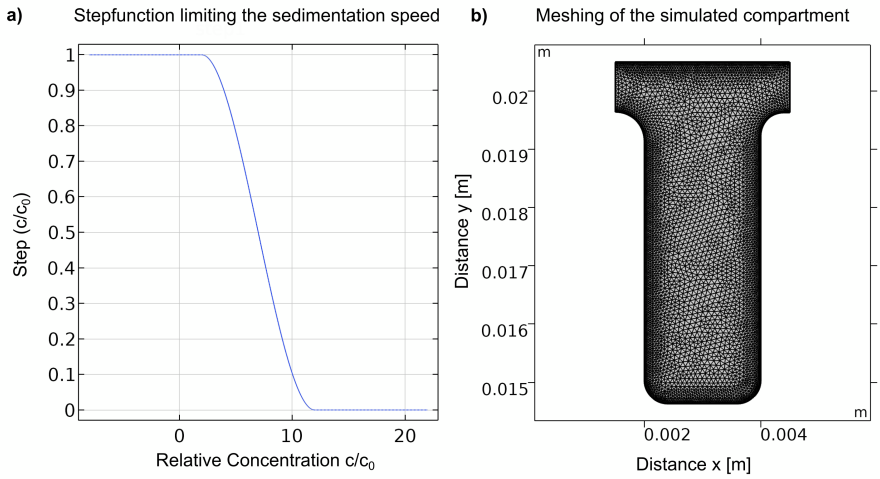


**Fig. S14. Sedimentation speed stepfunction and compartment meshing**. **a** Stepfunction dependent on local relative concentration $c/c_0$. This function is multiplied with the sedimentation velocity in each respective mesh triangular. Thus, molecule sedimentation arrests as they reach the bottom of the well. **b** Depiction of the mesh-size used for the simulation. The surfaces required more refined meshing to avoid problems related to boundary conditions.

| Parameter | Value | Description |
|---|---|---|
| $v_{\text{in}}$ | 2 $[\mu m/s]$ | Inflow velocity |
| $D$ | 5 $[\mu m^2/s]$ | Diffusion constant of species c |
| $v_0$ | 0.1$[\mu m/s]$ | Sedimentation velocity |
| $c_0$ | 25 $[\mu M]$ | Initial concentration of species c |
| well widht | 2 $[mm]$ | Width of the simulated pore |
| well height | 5 $[mm]$ | Height of the simulated pore |
| sed cutoff | 7 | Position of the center of the stepfunction |
| sed smoothing | 10 | width of the stepfunction |

**Table S1. Parameters used for the numerical calculation for the system with continuous feeding flow. Final set of parameters usd to simulate the flow-through experiment from Fig.4. Water-specific parameters such as dynamic viscosity or density were taken from inbuilt features of COMSOL Multiphysics 5.4..**

## 13. Experimental setup

The self-built experimental setup used to conduct and image sedimentation experiments is composed of:

1. Fluorescence microscope (Fig. S15)
2. Temperature control and chamber mount (Fig. S16 and Fig. S17)

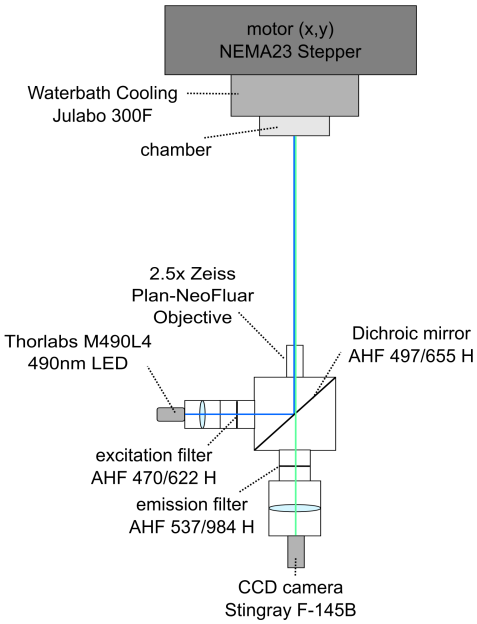


**Fig. S15. Schematic drawing of the measurement setup used to take fluorescent images for all experiments** Motors, LED as well as the camera were controlled using a self-written LABview software. The setup was surrounded by black curtains during measurements.

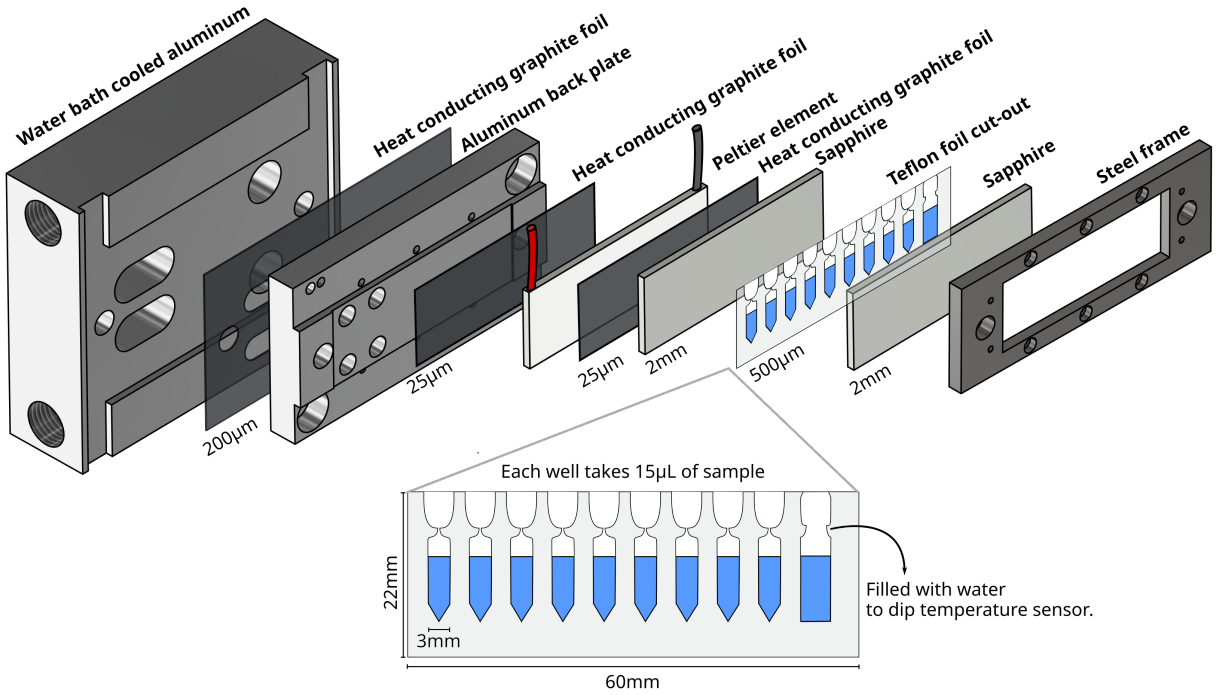


**Fig. S16. Schematic of sedimentation wells. a** The chamber is mounted from left to right. The temperature is controlled with the Peltier element, heat conducting graphite foil is used between all the parts to enhance heat conductivity. The water bath is used to dissipate the heat generated by the rear panel of the Peltier element back. The microfluidic Teflon cut-out is sandwiched between two sapphires which provide high heat conductivity when coupled to visible light transparency. The thickness is shown below each of the relevant parts. **b** Teflon cut-out has 9 each 3mm-wide wells and a temperature sensor is connected to each well. The shape of the chamber was designed to prevent the evaporation of the sample upon heating. The wide area above the narrow bottleneck is filled with parafilm, which seals the well.

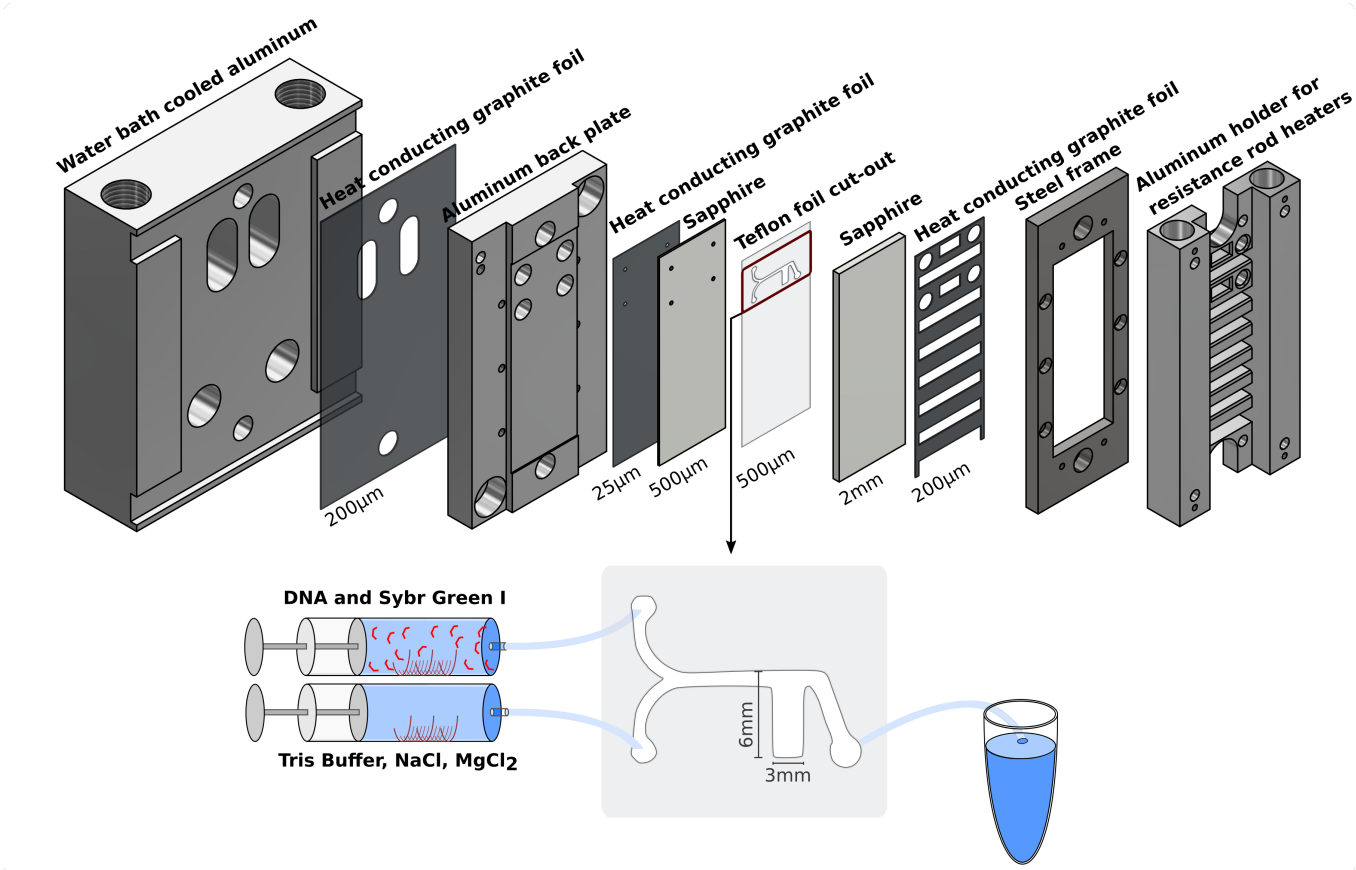


**Fig. S17. Schematic of the chamber used for the continuous feeding experiments. a** The chamber is mounted from left to right. The temperature is controlled with the water bath (set to 15℃), heat conducting graphite foil is used between all the parts to enhance heat conductivity. The microfluidic Teflon cut-out is sandwiched between two sapphires, used for their high heat conductivity coupled to visible light transparency. The thickness is written below each of the relevant parts. **b** Two 100µL syringes are used to inject the sample in the microfluidic well. DNA and Sybr Green I was injected through a separate inlet than the Tris Buffer and salts to prevent sedimentation to happen in the syringes and tubings.

**Movie S1. Timelapse video of Sequence pair 1 phase-separating and sedimenting inside the chamber.**

Chamber depth is 500$\mu$m. Temperature profile as described in methods. Contrast and brightness have been modified for qualitative reasons, the analysis of the readout was performed using the raw images. Sample contained 25$\mu$M of sequence pair 1 10mM TRIS pH 7, 5X Sybr Green I, 125mM NaCl and 10mM $MgCl_2$.

**Movie S2. Timelapse video of sequence pair 1 and 2 phase-separating and sedimenting inside the chamber.**

Chamber depth is 500$\mu$m. Temperature profile as described in methods. Contrast and brightness have been modified for qualitative reasons, the analysis of the readout was performed using the raw images. Sample contained 25$\mu$M of both sequence pairs 10mM TRIS pH 7, 5X Sybr Green I, 125mM NaCl and 10mM $MgCl_2$.

**Movie S3. Timelapse Video of the flow-through experiment shown in Fig.4.**

Sample contained 25$\mu$M of sequence pair 1 10mM TRIS pH 7, 5X Sybr Green I, 125mM NaCl and 10mM $MgCl_2$. Chamber depth is 500$\mu$m. Contrast and brightness have been modified for qualitative reasons, the analysis of the readout was performed using the raw images.

**Movie S4. Timelapse video of sequence pair 1 phase-separating and sedimenting inside the chamber (10mM TRIS pH 7, 5X Sybr Green I, 125mM NaCl and 10mM $MgCl_2$) at 25°C as an incubating temperature.**

Chamber depth is 500 $\mu$m. Framerate is 4 pictures per minute. Note how the aggregates do not grow in size as much as with 15°C.

**Movie S5. Video 5: 'Video5.avi' Timelapse video of sequence pair 1 phase-separating and sedimenting inside the chamber (10mM TRIS pH 7, 5X Sybr Green I, 125mM NaCl and 10mM $MgCl_2$) at 15°C as an incubating temperature.**

Chamber depth is 500 $\mu$m. Framerate is 4 pictures per minute. Note how the aggregates do grow way larger than in the equivalent experiment at 25°C.

## References


1. GH Fredrickson, ST Milner, L Leibler, Multicritical phenomena and microphase ordering in random block copolymers melts. *Macromolecules* **25**, 6341–6354 (1992).
2. JN Zadeh, et al., Nupack: Analysis and design of nucleic acid systems. *J. computational chemistry* **32**, 170–173 (2011).
3. DT Nguyen, OA Saleh, Tuning phase and aging of DNA hydrogels through molecular design. *Soft Matter* **13**, 5421–5427 (2017).
4. J SantaLucia Jr, A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci.* **95**, 1460–1465 (1998).
5. H Zipper, H Brunner, J Bernhagen, F Vitzthum, Investigations on dna intercalation and surface binding by sybr green i, its structure determination and methodological implications. *Nucleic acids research* **32**, e103–e103 (2004).
6. JL Mergny, L Lacroix, Analysis of thermal melting curves. *Oligonucleotides* **13**, 515–37 (2003).
7. N Qasem, MM Generous, B Qureshi, S Zubair, A comprehensive review of saline water correlations and data: Part ii—thermophysical properties. *ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING* (2021).
8. H Gudnason, M Dufva, DD Bang, A Wolff, Comparison of multiple dna dyes for real-time pcr: effects of dye concentration and sequence composition on dna amplification and melting temperature. *Nucleic acids research* **35**, e127 (2007).
9. PJ Kolbeck, et al., Molecular structure, dna binding mode, photophysical properties and recommendations for use of sybr gold. *Nucleic acids research* **49**, 5143–5158 (2021).

# Prebiotic gas flow environment enables isothermal nucleic acid replication

Philipp Schwintek,[1] Emre Eren,[1] Christof Mast,[1] and Dieter Braun[1,*]

[1]*Systems Biophysics, Physics department, Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstraße 54, 80799 Munich, Germany*

(Dated: August 19, 2024)

Nucleic acid replication is a central process at the origin of life. On early Earth, replication is challenged by the dilution of molecular building blocks and the difficulty of separating daughter from parent strands, a necessity for exponential replication. While thermal gradient systems have been shown to address these problems, elevated temperatures lead to degradation. Also, compared to constant temperature environments, such systems are rare. The isothermal system studied here models an abundant geological environment of the prebiotic Earth, in which water is continuously evaporated at the point of contact with the gas flows, inducing up-concentration and circular flow patterns at the gas-water interface through momentum transfer. We show experimentally that this setting drives a 30-fold accumulation of nucleic acids and their periodic separation by a 3-fold reduction in salt and product concentration. Fluid dynamic simulations agree with observations from tracking fluorescent beads. In this isothermal system, we were able to drive exponential DNA replication with Taq polymerase. The results provide a model for a ubiquitous non-equilibrium system to host early Darwinian molecular evolution at constant temperature.

Keywords: molecular evolution, DNA, nucleic acids, origins of life, non-equilibrium, isothermal, replication, gas flux

## Introduction

The emergence of life on Earth is still an unsolved puzzle to contemporary research. It is estimated that this event dates back approximately 3.7 - 4.5 billion years, with fossil carbon isotope signatures being the oldest evidence for life around 3.7 billion years ago [1, 2]. In order to reconstruct how early molecular life began before this time, it is crucial to identify and understand plausible geological environments, which support early prebiotic reaction networks that could have lead to the life we know today [3].

The common theory is that the Darwinian evolution of informational polymers was at the core of the origin of life [3]. Among these, nucleic acids, like RNA, stand out for their capability to both store genetic information and catalyze their own replication through transient formation of double-stranded helices [4]. These abilities allow them to mutate and evolve, enabling them to adapt to diverse environments and eventually encode, build and utilize proteins as the catalysts used in modern life.

Dilution, however, poses a significant obstacle, since such prebiotic reactions require sufficiently high concentrations of their reagents to work [5]. Large reservoirs, such as the ocean, cannot compensate for diffusion, because they lack local sources of energy to drive reaction pathways out of equilibrium [6]. The resulting homogeneity renders these environments unlikely to have harbored early molecular life [7].

Local physical non-equilibria, however, have shown the ability to up-concentrate molecules, such as nucleic acids, in a variety of different geological settings [8]. Examples range from thermal gradients in rock pores, local evaporation, re-hydration cycles of warm ponds, adsorption to mineral surfaces, heated gas bubbles in porous rocks, foams, or the eutectic phase in freeze-thaw cycles [9–17].

* dieter.braun@lmu.de

However, the accumulation of salts and molecules comes at a cost. Single-stranded nucleic acids replicate into double-stranded forms. These strands must separate again to complete a full replication cycle. But strand separation becomes increasingly difficult after accumulation, because the melting temperature of oligonucleotides is strongly dependent on the local salt concentration [18]. Despite high $Mg^{2+}$ concentrations being required for replication and catalytic activity [19], they can elevate the melting temperature of nucleic acid duplex structures to levels surpassing even the boiling point of water [20]. Oligonucleotides readily hydrolyze into nucleotide fragments under these conditions, rendering high temperature spikes as a primary strand separation mechanism more detrimental than beneficial [21].

Therefore, other mechanisms are required at the origin of life to separate nucleic acid strands with minimal thermal stress, and at best combined with an environment where supplied biomolecules are accumulated from the environment and trapped for long periods of time. Examples have used pH oscillations to drive nucleic acid strand separation, which can be caused either by differential thermophoresis of ionic species or by periodic freeze-thaw cycles [22–24]. Also, dew droplet cycles in a rock pore subjected to a temperature gradient can periodically melt strands by transiently lowering the salt concentration [25, 26]. Heated gas-water interfaces were also shown to promote many prebiotic synthesis reactions [14, 27, 28]. The above scenarios require temperature gradients or thermal cycling. This creates degradation stress for nucleic acids and limits the scenarios to geological settings with a thermal gradient. Here, we investigated a simple and ubiquitous scenario in which a water flux through a rock pore was dried by a gas flux at constant temperature (Fig.1). This can be found in the vicinity of underwater degassing events, where gases percolate through rocks to reach the surface, or in porous rocks at the surface exposed to atmospheric winds [29, 30]. Such a setting would be very common on volcanic islands on early

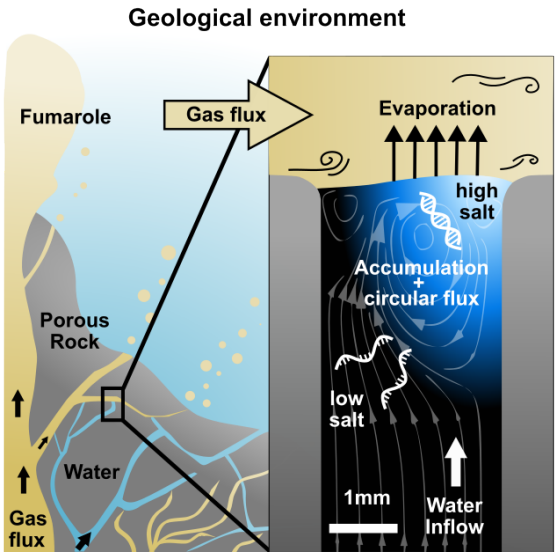


Figure 1. **Replication at the gas-water interface.** We considered a geological scenario in which water, containing biomolecules, is evaporated by a gas flow at the scale of millimeters. In volcanic porous rock, many of such settings can be imagined. The gas flow induces convective water currents and causes it to evaporate. Dissolved nucleic acids and salts accumulate at the gas-water interface due to the interfacial currents, even if the influx from below is pure water. Through the induced vortex, nucleic acids pass through different concentrations of salt, promoting strand separation and allowing them to replicate exponentially. Our experiments replicate this environment on the microscale, subjecting a defined sample volume to a continuous influx of pure water with an airflux brushing across.

Earth which also offered the necessary dry conditions for RNA synthesis[31].

We created an experimental model of such an evaporation pore, shown in Fig.1, and studied how combined gas and water fluxes can lead to early replication of nucleic acids. We first analyzed accumulation flow speeds at the interface in Fig.2, then monitored cyclic strand separation dynamics in Fig.3, and finally showed how both drive DNA-based replication under isothermal conditions in Fig.4.

### Results and Discussion

**Molecule Accumulation at the Gas-Water Interface.** We started off by constructing a laboratory model of the rock pore shown in Fig. 1. Here, we focused on the key properties of the system: An upward water flux evaporating at the intersection with the perpendicular gas flux. This leads to an accumulation of dissolved molecules at the interface since they cannot evaporate. Simultaneously, the momentum transfer of the gas flux induces circular currents in water, forcing molecules back into the bulk.

In the following, we analyse how these two effects act on dissolved nucleic acids. For simplicity, we used ambient air as the gas source, enabling us to focus solely on evaporation and the resulting currents. The velocity of the water flowing in was controlled by a syringe pump and chosen to match the velocity of the water evaporating in the given geometry. This ensured reliable and stable conditions in long lasting experiments. Fixed volumes of sample solutions (containing beads, labelled molecules, salts etc.) were always loaded ahead of an influx of pure water, simulating a continuous dilution scenario.

The micro scale gas-water evaporation interface consisted of a 1.5 mm wide and 250 µm thick channel that carried an upward pure water flow of 4 nl/s ≈ 10 µm/s perpendicular to an air flow of about 250 ml/min ≈ 10 m/s (Suppl. IV). The temperature of the chamber was controlled by a water bath at 45°C, while a self-built fluorescence microscope provided imaging (See Suppl. Fig.III.1). Two-dimensional finite element simulations were performed to model the diffusion of molecules in water, as well as the flow of water and gas.

First, using a particle tracking algorithm (See Suppl. Sec. V), we measured the flow velocities of individual fluorescent 0.5 µm beads to monitor the dynamics of the water flow as the air-flux streamed across the interface (Fig.2(a)). As expected, these velocities were dependent on their distance from the channel walls (see Suppl. Fig. VI.1(e)). The beads that were far from the interface were moving at water inflow velocities of 15 µm/s. Closer to the interface, the velocities increased to about 1 mm/s due to the momentum transfer of the gas flow (Supplementary Movie 1). This resulted in a circular flow pattern with the vortex center right below the interface. The flow lines in Fig. 2a) show how the upwards water flux reaches the interface on one side of the vortex, whereas on the opposite side, the beads are pushed back down into the bulk.

The extracted traces in Fig. 2(a) were compared with a finite element simulation. In a two-dimensional projection of the experimental geometry, we modeled laminar gas and water flow, diffusive nucleic acid mass transport in water, interfacial evaporation dynamics, and momentum transfer of gas flowing over the water surface (Suppl. Sec. VI). In agreement with the experimental results, the simulation showed a chamber-averaged water evaporation speed of 10.5 µm/s. The tangential velocity at the interface reached 0.9 mm/s. The modeled flow speed distributions agreed well with distribution of the experimental bead velocities, as shown in Fig.2(a,c).

To further test our understanding of the dynamics of the system, we imaged fluorescently labeled nucleic acids. The expectation was that the continuous evaporation would lead to an accumulation of the strands at the interface, while the gas flow would induce a vortex analogous to the beads. Both are found to be present in the experiment and agree qualitatively with the finite element model. In our experiment, 10 µl of 5 µM FAM-labeled 63mer DNA were introduced into the system, followed by a continuous diluting pure water inflow. Temperature, water flow and air flow were unchanged from the previous experiment.

Water continuously evaporated at the interface, but nucleic acids remained in the aqueous phase and could only escape downward either by diffusion or by the vortex, thus accumulating near the interface. The evaporation rate at the interface was set to be proportional to the vapor concentration gradient and varied spatially along the interface between 5 and 10.5 µm/s (See Suppl. Fig. VI.1d)). As the gas flow continuously removed excess vapor, the evaporation rate remained

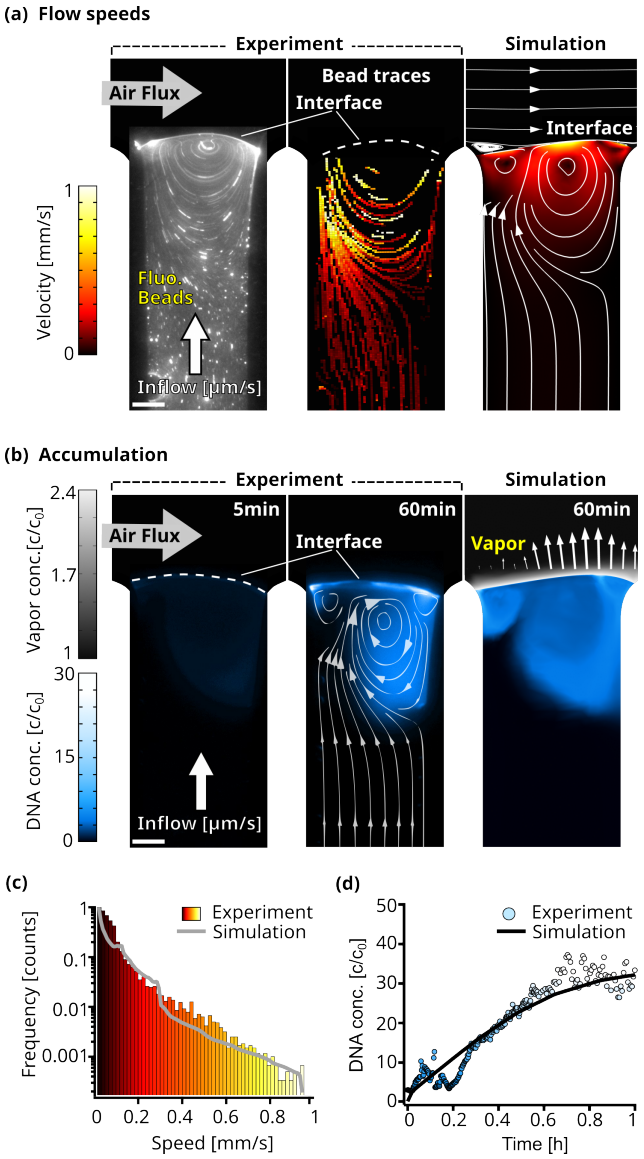


Figure 2. **Flow and accumulation dynamics.** **(a)** Imaging of fluorescent beads (0.5 µm) reveals a flow vortex right below the air-water interface, induced by the air flux across the interface (left panel). The bead movements were traced (middle panel) and the measured velocities were confirmed by a detailed finite element simulation (right panel). The color scale is equal for both simulation and experiment and the scale bar = 500 µm. **(b)** The accumulation of fluorescently labeled 63mer DNA was imaged and confirmed our understanding of the environment based on a diffusion model. Concentration reaches up to 30 times relative to the start $c_0$. The accumulation profile of the experiment (middle panel) and simulation (right panel) match well, showcased by overlaying the simulated flowlines. Blue colorscale represents DNA accumulation for experiment and simulation, while grey color scale shows the relative vapor concentration in the simulation. Arrows (right panel) proportionally show the evaporation speed along the interface. **(c)** The simulated and experimentally measured distribution of flow velocities of dissolved beads plotted in a histogram, showing a similar profile. Color scale is equal to (a). **(d)** The maximum relative concentration of DNA increased within an hour to ≈ 30 X the initial concentration, with values following the simulation.

constant. Thus, except for fluctuations, a stable interface shape was obtained. The vortex induced by the gas flowing across the interface pushed the molecules back deeper into the bulk (See the flow lines in Fig.2(b) taken from the simulation). 5 minutes after starting the experiment, the maximum DNA accumulation was 3-fold, while after one hour of evaporation, up to 30-fold accumulation was observed.

With the known diffusion coefficient of 95 µm$^2$/s for the 63mer[9], a good agreement between the simulation and the experiment was achieved (Fig.2(b), right panel). The DNA accumulation was located in the area where the flow induced by the gas flow created a vortex. In the simulation, the shape of the interface was kept constant, but in the experiment it fluctuated around a steady state shape, likely in response to small fluctuations in gas pressure and spatial variations in water surface tension (Supplementary Movie 2). The simulated maximum accumulation followed the experimental results and starts saturating after about one hour (Fig.2(d)).

**Strand Separation Dynamics.** As discussed earlier, strand separation is essential for the replication of nucleic acids. Only then can replication become exponential and compete with naturally exponential degradation kinetics. Usually, an elevation of temperature can separate strands but is accompanied with a higher risk for hydrolysis. The chosen isothermal setting requires changes in salt concentration for this process. More specifically, the circular fluid flow at the interface provided by the gas flux, together with Brownian motion, was expected to drive cyclic strand separation by forcing nucleic acid strands through areas of varying salt concentrations.

We used Förster resonance energy transfer (FRET) microscopy to optically measure the strand separation of DNA (Fig.3). A high FRET signal indicates that two DNA strands are bound, while a low FRET signal indicates that the strands are separated. In this way, FRET becomes an indirect measure of the salt concentration, since a low salt concentration will induce strand separation due to the reduced ionic shielding of the charged DNA or RNA backbones. Specifically, we chose a complementary 24mer DNA pair, with the FRET-pair fluorophores positioned centrally on opposite strands. 1 µL Sample (10 mM TRIS at pH 7, 50 µM $MgCl_2$, 3.9 mM NaCl, and 5 µM of each DNA strand) was injected into the chamber and flushed towards the interface by pure water with all other conditions equal to before.

Fig.3(a) shows micrographs of the recorded FRET values for each pixel (Supplementary Movie 3). Initially, the FRET signal increased near the interface (green), indicating areas where DNA is forming double stranded DNA. This area is localized around the vortex created by the gas flow across the interface. In the upward flow to the left of the vortex, DNA was found to be single-stranded (blue). During the course of the experiment, the low and high FRET regions remained stably separated (Fig.3(b)). This configuration suggests that the vortex could drive a cycle of replication and strand separation (see the scheme in Fig.3(a) - right panel).

To confirm this, we simulated the accumulation of $Mg^{2+}$ ions in the chamber (Suppl. VII), since divalent ions have a large effect on the melting temperature of nucleic acids [18]. We then used a Monte Carlo random walk model (Suppl. VIII) to

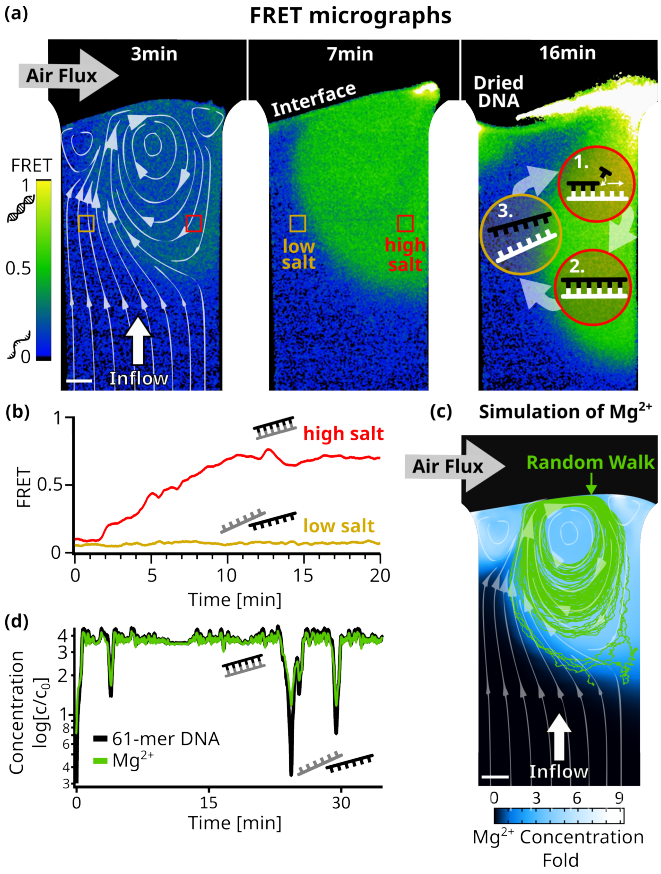


Figure 3. **Strand separation by salt cycling.** Fluorescence resonance energy transfer measurements revealed cycles of strand separation. **(a)** Micrographs of 24bp DNA FRET pair in the chamber at 45°C. 1 µl sample (5 µM DNA, 10 mM TRIS pH7, 50 µM MgCl2, 3.9 mM NaCl) was subjected to a 3 nl/s diluting upflow of pure water and a gas flow of 230 ml/min across. The induced vortex, shown by the simulated flow lines (left panel), overlays with regions of high FRET indicative of double-stranded DNA. The vortex flow was expected to enable replication reactions by (**1+2**) strand replication in the high salt region and (**3**) strand separation of template and replicate in the low salt region. **(b)** FRET signals confirmed strand separation in low salt regions and strand annealing in high salt regions in (a). After about 10 minutes, DNA and salt accumulated at the interface forming stable and clearly separated regions of low – where the influx from below reaches the interface – and high – located at the vortex – FRET signals. **(c)** Comsol simulation of $Mg^{2+}$ ions (D = 705 $\mu m^2/s$ in the chamber agreed with the FRET signal and showed up to 9-fold salt accumulation at the interface. The path of a 61mer DNA molecule from a random walk model is shown by the green lines and the white flowlines are taken from the simulation. **(d)** Concentrations along the DNA molecule path in (c) show oscillations relative to the initial concentration of up to 3-fold for $Mg^{2+}$ and 4-fold for 61mer DNA. This could enable replication cycles, as the vortex provides high salt concentrations for replication, while drops in salt and template concentrations regularly trigger strand separation.

simulate individual 61mer DNA molecules following the vortex and undergoing Brownian motion. Such a path is shown in Fig. 3(c), plotted over the simulated steady-state concentration of $Mg^{2+}$ along with the simulated flow lines. Starting in a region of low $Mg^{2+}$ concentration, the strand enters the vortex created by the gas flow. We have plotted the $Mg^{2+}$ concentration along its path, showing significant salt oscillations of up to 3X the initial salt concentration, capable of inducing strand separation (Fig. 3(d), and Supplementary Fig. VII.2). When plotting the simulated steady-state concentration of other dissolved – complementary – 61mer DNA molecules along its path, we observed even stronger oscillations of up to 4X the initial concentration. Together with significant drops in $Mg^{2+}$ concentration, this suggests the possibility of exponential replication by strand separation cycles.

**Isothermal Replication with PCR.** We saw that nucleic acids and salts accumulated near the interface, but far from the interface, in the bulk below, the concentrations remained vanishingly low due to the diluting inflow of pure water. The air flux induced an accumulation pattern of vortices in which molecules were trapped. The salt and DNA concentration changed cyclically, resulting in periodic strand separation of nucleic acids. Motivated by the above results, we used a model system to test whether nucleic acid replication could actually be implemented in this environment.

We chose to use Taq DNA Polymerase because it does not have a protein-based strand separating mechanism. Starting with a 51mer template and two 30mer primer strands, each with a 5'-AAAAA overhang for detection, the reaction is expected to form a 61mer replicate (Suppl. IX), the same length as the DNA used in the random walk model in Fig. 3(c)&(d). In contrast to standard PCR, which uses thermal cycling to separate the strands, we operated the experiment at isothermal conditions (68°C) and used 10 µl of the reaction mix (0.25 µM primers, 5 nM template, 200 µM dNTPs, 0.5 X PCR buffer, 2.5 U Taq polymerase, 2 X SYBR Green I). This reaction mixture was then exposed to a constant pure water influx of 5 nl/s towards the gas-water interface, matching the rate of evaporation at the interface.

Through the oscillations in salts and DNA observed along the random walk, we expected the 61mer product strand to be able to separate from its respective template strand, enabling exponential replication. The progress was monitored using the intercalating dye SYBR Green I, which binds preferentially to double-stranded DNA [32]. Fig.4(a) shows fluorescence micrographs of the reaction in the chamber. Initially, minimal fluorescence is seen. This indicates that the replicated templates are below the detection limit of SYBR Green.

The SYBR Green fluorescence increased after two hours, recording the increase of replicated DNA forming duplex structures. In an identical setting when the gas- and water flux were switched off, no fluorescence increase was found. Replication was confirmed under flux with the 61mer product being visible in gel electrophoresis with depleted primers (Fig.4(b)). The fluorescence signal over time is shown in Fig.4(c), recorded from a rectangular region of interest.

Fluorescence variations are caused by fluctuations in the position of the gas-water interface in addition to air bubbles caused by degassing of the liquid at this temperature (Supplementary Movie 4&5). With both gas flow and water influx turned off, no product band was found. We verified the replication reaction by repeating the experiments without the addition

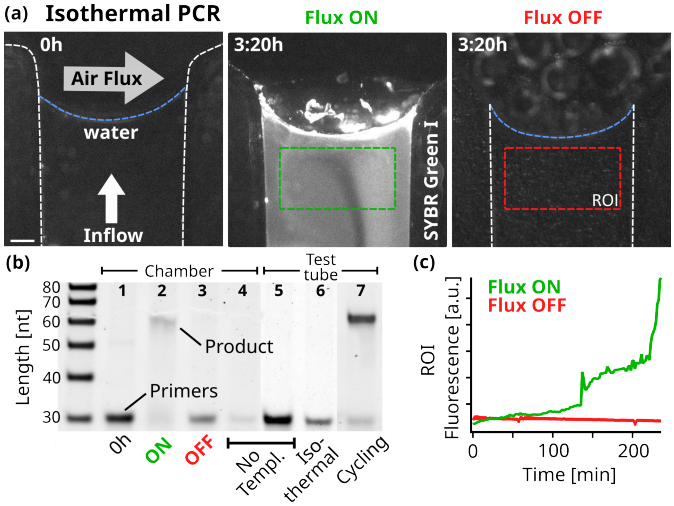


Figure 4. **Replication.** **(a)** Fluorescence micrographs of the PCR reaction in the chamber. At isothermal 68°C, 10 µl of reaction sample was subjected to a constant 5 nl/s pure water flow towards the interface where a 250ml/min gas flowed perpendicularly. The initial state on the left shows the background fluorescence. Fluorescence increased under flux (middle, after 3:20h), while without flux the fluorescence signal remained minimal (right). The reaction sample consisted of 0.25 µM primers, 5 nM template, 200 µM dNTPs, 0.5 X PCR buffer, 2.5 U Taq polymerase, 2 X SYBR Green I. Scale bar is 250 µm. **(b)** 15% Polyacrylamide Gel Electrophoresis of the reactions and neg. controls. After 4 hours in the reaction chamber with air- and water-flux ON, the 61mer product was formed under primer consumption **(2)**, unlike in the equivalent experiment with the fluxes turned OFF **(3)**. At the beginning of the experiment **(1)** or in the absence of template **(4)**, no replicated DNA was detected. The reaction mixture was tested by thermal cycling in a test tube **(5-7)**. As expected, replicated DNA was detected only with the addition of template: **(7)** shows the sample after 11 replication cycles. The sample was also incubated for 4 hours at the chamber temperature (68°C) yielding no product **(6)**. Primer band intensity variations are caused by material loss during extraction from the microfluidic chamber. **(c)** SYBR Green I fluorescence increased when gas and water flow were turned on, but remained at background levels without flow. Fluorescence was averaged over time from the green and red regions of interest shown in (a). SYBR Green I fluorescence indicates replication, as formed products are able to hybridize.

of the template, primer or DNA in the chamber as well as in a test tube (Suppl. Fig. IX.2). As expected, no product was found in any of these cases. We also compared the chamber experiment with a regular, temperature cycling based, PCR reaction in a test tube, revealing that in the chamber, about 10-11 cycles of PCR were finished after the 4 hours of experiment (Suppl. IX). The findings above confirm that the gas flow at the simulated rock opening was necessary for nucleic acid replication.

Prebiotic chemical reactions such as polymerization of imidazole[20] or 2′,3′-cyclic phosphate-activated nucleotides[28, 33] will benefit from the reduced RNA hydrolysis in the gentle isothermal replication environment. Most importantly, the combination of actively generated high concentrations by evaporation, dry-wet cycles at the interface caused by interface fluctuations, shielding from harmful UV, and the possibility of constant feeding by water influx makes the environment a compelling candidate for implementing the geophysical boundary condition of the early RNA world stage of emergent life.

**Conclusion.** In this work we investigated a prebiotically plausible and abundant geological environment to trigger replication of early Life. We considered an isothermal setting of gas flowing over an open rock pore filled with water. Previously, thermal gradients have been used to separate the strands of nucleic acids, risking its degradation. Now, the combined gas and water flow at an open pore trigger salt oscillations. We found that this condition supports oligonucleotide replication. We began by probing the system with fluorescent bead and DNA measurements, finding our results to agree with fluid dynamics theory using finite elements simulations. While DNA accumulates at the vortex close to the interface, oscillations in nucleic acid and salt concentration are created by a combination of molecular accumulation and interfacial flow, periodically separating nucleic acid strands under chemically gentle conditions. Due to the limitations of RNA-based replication, we probed the environment with protein-driven DNA replication and found isothermal replication in this common geological micro-environment, showing that it provides a setting for early nucleic acid replication chemistry. Physical non-equilibria, such as steep temperature gradients, pose many boundary conditions, decreasing the likelihood of readily finding such a setting. This isothermal environment, however, greatly extends the repertoire of prebiotic settings that enable replication on early planets.

## Methods

A microfluidic chamber was created between two sapphire plates (0.5mm thick on the front and 1mm thick on the back), sandwiching a 0.25mm thin Teflon sheet that defined the geometry created by a computer-controlled cutter. The plates were held together by a steel frame bolted to an aluminum back to ensure gas tightness. The back was connected to a water bath (Julabo) to control the temperature. Samples were injected into the chamber using syringe pumps (Nemesys) with tubings inserted into holes in the back sapphire. Gas flow was generated under pressure control using the AF1 dual pump system (Elveflow). Temperature was measured during the experiments with a thermal sensor attached to the back sapphire. A more detailed schematic of the microfluidic chamber can be found in Suppl. IV.

DNA fluorescence measurements were performed in a self-built tilted epi-fluorescence microscope setup using two M490L4 and M625L3 light-emitting diodes (Thorlabs), a 470/622 H dual-band excitation filter (AHF), a 497/655 H dual-band dichroic mirror, and a 537/694 H dual-band emission filter. A more detailed schematic of the setup can be found in Suppl. III. DNA strands were ordered from biomers.net including purification by high-performance liquid chromatography (Suppl. II.1). The strands used for fluorescence quantification of accumulation are (5'-3')-24bp DNA: *CY5*CGTAGTAAATATCTAGCTAAAGTG, 63bp DNA: *FAM*CCAGCCTCCAGTGCCTCGTATCATTGTG-CCAAAAGGCACAATGATACGAGGCACTGGAGGCTG diluted to 5 µM in Nuclease free water. Images were captured using a Stingray F145B camera (Allied Vision). Bead experiments used 0.5 µM fluorescent microspheres (Invitrogen) diluted 1/2000 in water (Suppl. Sec. V).

2D finite element simulations were performed using COMSOL Multiphysics 5.4. Fluid dynamics were simulated by solving the Navier-Stokes equation in two dimensions. Parameters used are available in table VI.1 in the supplementary Information. The complete description of the model can be found in Suppl. VI.

FRET imaging was performed using a second custom-built fluorescence microscopy setup consisting of light-emitting diodes (M470L2, M590L2;

Thorlabs) combined by a dichroic mirror on the excitation side, while an Optoplit II with a ratiometric filter set (DC 600LP, BP536/40, BP 630/50) and a Stingray-F145B ASG camera (Allied Vision Technologies) through a 1X objective (AC254 100-A-ML Achromatic Doublet; Thorlabs) detected and superimposed both fluorescence emission channels (Suppl. III). The DNA sequences used for FRET experiments were: strand 1 5'-CGTAGTAAATAT*FAM*CTAGCTAAAGTG-3', strand 2 5'-CACTTTAGCTAGAT*ROX*ATTTACTACG-3'. The two labeled complementary strands were diluted from stock solution (100 µM in nuclease-free water) and mixed together to a final concentration of 5 µM in buffer (10 mM TRIS, 50 µM $MgCl_2$, 3.9 mM NaCl, pH7). To promote annealing of the two complementary strands, the solution was heated and slowly cooled from 80°C to 4°C (ramp rate of -1°C per 5 s) in a standard thermocycler (Bio-Rad CFX96 Real-Time System) prior to each experiment.

Polymerase chain reaction (PCR) was performed using an AllTaq PCR Core Kit (QUIAGEN). Samples were mixed with 0.5 X AllTaq PCR Buffer, 5 nM template strand, 0.25 µM primers, 200 µM of each dNTP, 2 X SYBR Green I and AllTaq polymerase at 2.5 U/reaction. The reaction in the thermocycler was performed using a temperature protocol of 95°C for 2 minutes for heat activation of the enzyme, then annealing the primers to 52°C for 10 seconds, then 68°C for 10 seconds, and finally 10 seconds at 95°C. This cycle was repeated 40 times (See Suppl. Fig. IX.2b)). The reaction in the chamber was performed with 10 µl of the above mixture at 68°C. The solution was also heat activated at 95°C for 2 min followed by an annealing step to 52°C before loading into the chamber. The DNA sequences for the reaction were as follows Template (5'-3')-51bp DNA: TTAGCAGAGCGAGGTATGTAGGCGGGACGCTCAGTGGAACGAAAACTCACG, Reverse primer (5'-3')-30bp DNA: AAAAACGTGAGTTTTCGTTCCACTGAGCGT, forward primer (5'-3')-30bp DNA: AAAAATTAGCAGAGCGAGGTATGTAGGCGG.

For PAGE and gel imaging, a 15% denaturing (50% urea) polyacrylamide gel with an acrylamide:bis ratio of 29:1 was solidified with TEMED (tetramethylethylenediamine) and ammonium persulfate. 2 µl of sample was mixed with 7 µl of 2X loading buffer (Orange G, formamide, EDTA), of which 5 µl were loaded onto the gel. Staining was performed with 2X SYBR Gold in 1X TBE buffer for 5 minutes and the gel was imaged using the ChemiDOC MP imaging station (Bio-Rad).

**Data availability.** Supplementary data beyond the supplementary material will be given upon request.

**Author contributions.** Project conception: P.S., D.B., Research design: P.S., D.B., Methodology development: P.S., Experiments: P.S., E.E. Data analysis: P.S., Programming: P.S., Manuscript writing: P.S., Manuscript reviewing: P.S., D.B., Supervision: D.B. Funding acquisition: D.B.

**Competing interests.** The authors declare no competing interests.

**Additional information.** The online version contains supplementary material available at `https://...`

**Funding and acknowledgments.** We would like to acknowledge the following agencies for funding: Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 364653263 – CRC 235, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 521256690 – CRC 392, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 201269156 – SFB 1032, Volkswagen Initiative' Life? – A Fresh Scientific Approach to the Basic Principles of Life', Germany's Excellence Strategy EXC-2094-390783311, Simons Foundation 327125, European Research Council EvoTrap #787356, ERC-2017-ADG. This work was supported by the Center for Nanoscience Munich (CeNS).

[1] B. K. Pearce, A. S. Tupper, R. E. Pudritz, and P. G. Higgs, Astrobiology **18**, 343 (2018).
[2] M. T. Rosing, Science **283**, 674 (1999).
[3] L. E. Orgel, Scientific American **271**, 76 (1994).
[4] W. Gilbert, nature **319**, 618 (1986).
[5] P. L. Luisi, Israel Journal of Chemistry **55**, 906 (2015).
[6] N. Goldenfeld and C. Woese, Annu. Rev. Condens. Matter Phys. **2**, 375 (2011).
[7] N. Lane, J. F. Allen, and W. Martin, BioEssays **32**, 271 (2010).
[8] A. Ianeselli, A. Salditt, C. Mast, B. Ercolano, C. L. Kufner, B. Scheu, and D. Braun, Nature Reviews Physics **5**, 185 (2023).
[9] C. B. Mast, S. Schink, U. Gerland, and D. Braun, Proceedings of the National Academy of Sciences **110**, 8030 (2013).
[10] T. Matreux, P. Aikkila, B. Scheu, D. Braun, and C. B. Mast, Nature **628**, 110 (2024).
[11] B. K. Pearce, R. E. Pudritz, D. A. Semenov, and T. K. Henning, Proceedings of the National Academy of Sciences **114**, 11327 (2017).
[12] B. Damer and D. Deamer, Life **5**, 872 (2015).
[13] J. P. Ferris, A. R. Hill Jr, R. Liu, and L. E. Orgel, Nature **381**, 59 (1996).
[14] M. Morasch, J. Liu, C. F. Dirscherl, A. Ianeselli, A. Kühnlein, K. Le Vay, P. Schwintek, S. Islam, M. K. Corpinot, B. Scheu, *et al.*, Nature Chemistry **11**, 779 (2019).
[15] E. Tekin, A. Salditt, P. Schwintek, S. Wunnava, J. Langlais, J. Saenz, D. Tang, P. Schwille, C. Mast, and D. Braun, ChemBioChem **23**, e202200423 (2022).
[16] H. Trinks, W. Schröder, and C. K. Biebricher, Origins of life and evolution of biospheres **35**, 429 (2005).
[17] R. Stribling and S. L. Miller, Journal of molecular evolution **32**, 289 (1991).
[18] C. Schildkraut and S. Lifson, Biopolymers: Original Research on Biomolecules **3**, 195 (1965).
[19] A. Hampel and J. Cowan, Chemistry & biology **4**, 513 (1997).
[20] J. W. Szostak, Journal of Systems Chemistry **3**, 1 (2012).
[21] Y. Li and R. R. Breaker, Journal of the American Chemical society **121**, 5364 (1999).
[22] A. Mariani, C. Bonfio, C. M. Johnson, and J. D. Sutherland, Biochemistry **57**, 6382 (2018).
[23] L. M. Keil, F. M. Möller, M. Kieß, P. W. Kudella, and C. B. Mast, Nature communications **8**, 1897 (2017).
[24] N. Takenaka, M. Tanaka, K. Okitsu, and H. Bandow, The Journal of Physical Chemistry A **110**, 10628 (2006).
[25] A. Ianeselli, C. B. Mast, and D. Braun, Angewandte Chemie **131**, 13289 (2019).
[26] A. Salditt, L. Karr, E. Salibi, K. Le Vay, D. Braun, and H. Mutschler, Nature Communications **14**, 1495 (2023).
[27] A. M. Deal, R. J. Rapf, and V. Vaida, The Journal of Physical Chemistry A **125**, 4929 (2021).
[28] A. V. Dass, S. Wunnava, J. Langlais, B. von der Esch, M. Krusche, L. Ufer, N. Chrisam, R. C. Dubini, F. Gartner, S. Angerpointner, *et al.*, ChemSystemsChem **5**, e202200026 (2023).
[29] Y. Zhang, *Treatise on Geochemistry* (Elsevier, 2014) pp. 37–69.
[30] H. Kurnio, S. Lubis, and H. C. Widi, Bulletin of the Marine Geology **30**, 85 (2016).
[31] M. W. Powner, B. Gerland, and J. D. Sutherland, Nature **459**, 239 (2009).
[32] A. Dragan, R. Pavlovic, J. McGivney, J. Casas-Finet, E. Bishop, R. Strouse, M. Schenerman, and C. Geddes, Journal of fluorescence **22**, 1189 (2012).
[33] A. C. Serrão, S. Wunnava, A. V. Dass, L. Ufer, P. Schwintek, C. B. Mast, and D. Braun, JACS (2023).
[34] L. Yuan-Hui and S. Gregory, Geochimica et cosmochimica acta **38**, 703 (1974).

# Prebiotic gas flow environment enables isothermal nucleic acid replication

Supplementary information

Philipp Schwintek[1], Emre Eren[1], Christof Mast[1] and Dieter Braun[1,†]

[1]*Systems Biophysics, Physics department, Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstraße 54, 80799 Munich, Germany*

## Contents



† dieter.braun@lmu.de

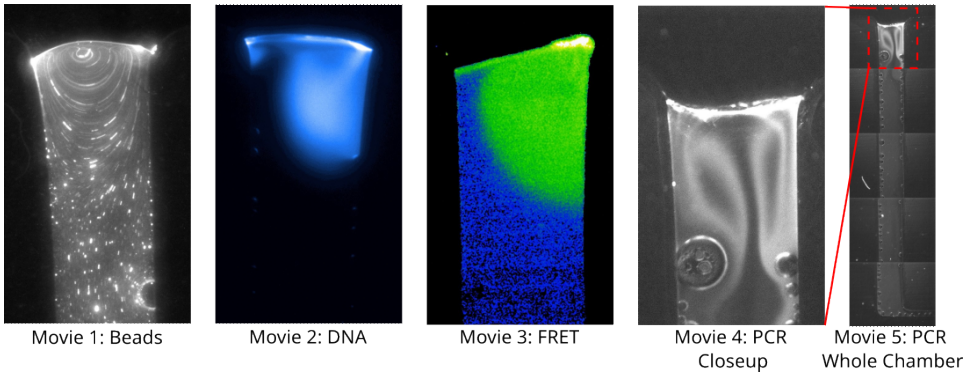


Figure I.1. Snapshots of each supplementary movie.

## Supplementary information I: Movies

The recorded fluorescence images movies which are the basis of the data shown in Fig.2(a), Fig.2(b), Fig.3(a) and Fig.4(a) are provided as supplementary movies.

**Supplementary Movie 1**
Fluorescence beads are used to track the fluid flow shown in Fig.2(a).

**Supplementary Movie 2**
Concentration of fluorescently labelled 63mer DNA is imaged to infer the accumulation at the interface in Fig.2(b).

**Supplementary Movie 3**
FRET imaging of dual-labaled DNA strands discriminate between for single stranded DNA in blue and double stranded DNA in green to yellow as detailed in Fig.3(a).

**Supplementary Movie 4**
Top fraction of the chamber used for the figures. SYBR green fluorescence shows the amount of DNA generated in the replication reaction, indicating by a raise of fluorescence how DNA becomes copied for Fig.4(a). In addition, the accumulation at the interface is also seen.

**Supplementary Movie 5**
Whole length of the chamber. SYBR green fluorescence shows the amount of DNA generated in the replication reaction, indicating by a raise of fluorescence how DNA becomes copied for Fig.4(a).

## Supplementary information II: DNA Strands

| Length | 5'- Sequence -3' | Label |
|---|---|---|
| 63mer | ccagcctccagtgcctcgtatcattgtgccaaaaggcacaatgatacgaggcactggaggctg | 5′ FAM |
| 24mer FRET strand 1 | CGTAGTAAATA8CTAGCTAAAGTG | 8 = FAM |
| 24mer FRET strand 2 | CACTTTAGCTAGA8ATTTACTACG | 8 = ROX |
| 51mer Template | TTAGCAGAGCGAGGTATGTAGGCGGGACGCTCAGTGGAACGAAAACTCACG | - |
| 30mer forward primer | AAAAA TTA GCA GAG CGA GGT ATG TAG GCG G | - |
| 30mer reverse primer | AAAAA CGT GAG TTT TCG TTC CAC TGA GCG T | - |

Table II.1. DNA sequences as ordered from biomers.net.

## Supplementary information III: Experimental Setups

All setups used in this paper were designed and assembled from components purchased from several companies. Figure III.1 shows schematics of both setups used. All specifications below are given in nanometers.

Setup A was built to capture fluorescence signals from beads and labeled DNA and was constructed using SM1 lens tubes (Thorlabs), excitation LEDs M490L4 and M625L3-C4 (Thorlabs), excitation filters (470/622H, AHF) and emission filters (497/655H, AHF), and a dual-band dichroic beamsplitter (497/655H, AHF). The camera (Stingray F-145 B/C) was purchased from Allied Vision. An achromatic doublet (AC254-100-A-ML, Thorlabs) was used as the objective.

Setup B was built to capture FRET. A Zeiss Axiotech Vario microscope body was used and equipped with excitation LEDs M590L4 (yellow) and M470L2 (blue) (Thorlabs) with excitation filters BP588/20 and BP482/29 (Thorlabs) coupled into the same beamline via a DC R 475/40 beamsplitter. A dual-band dichroic mirror (505/606 T) and an Optosplit II with a ratiometric filter set (DC600 LP, BP630/50 and BP 536/40) were coupled to the system to capture the FRET signal of the dyes ROX and FAM. Camera and objective are identical to setup A.

For both setups, fluorescence measurements were performed under curtains to ensure low background noise. Temperature was controlled by a JULABO Corio CD water bath connected to the microfluidic chamber (Suppl. IV) by thermally insulated tubing. The temperature was measured directly at the back sapphire surface of the chamber. Gas flow was generated using an AF1 Dual Pump (Elveflow) system with ambient air as the gas source, and gas flow rate was measured with a flow sensor (FS2000, from IDT). Syringes were driven by a Cetoni syringe pump (Nemesys).

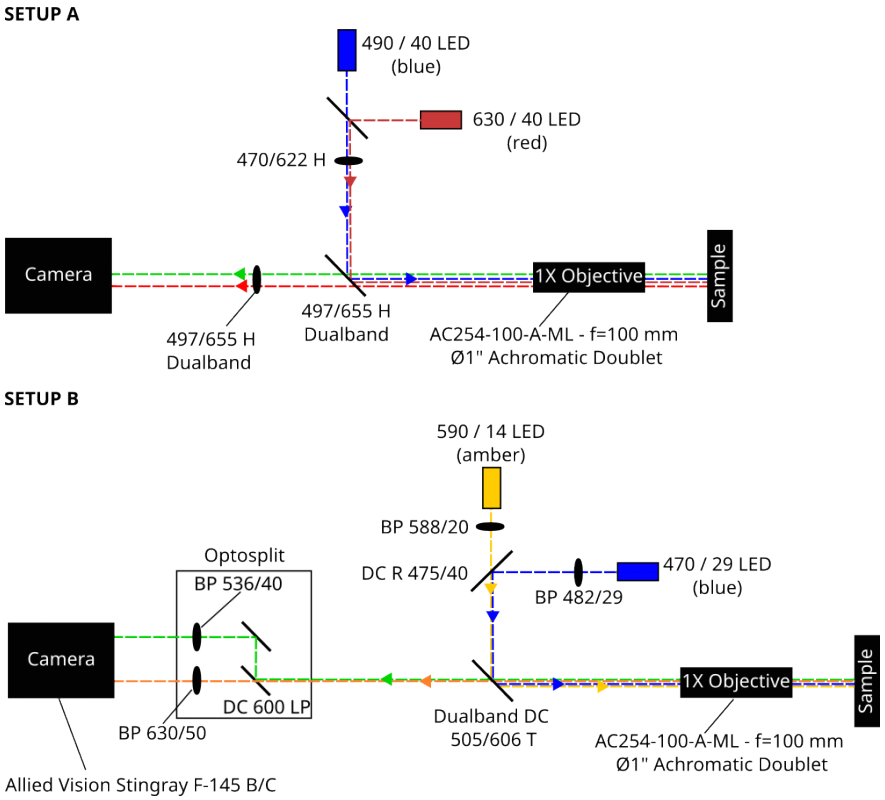


Figure III.1. Sketch of setup A and B used for all fluorescence imaging. Setup A was used for fluorescence measurements containing only fluorescent beads, DNA with a FAM/Cy5 label, and for those containing SYBR Green I. Setup B was used for FRET measurements using the FAM/ROX FRET pair.

## Supplementary information IV: Microfluidic Chamber

The microfluidic chamber used for the experiments in this work was constructed as follows:

Two 60x22mm sapphire plates (front 1mm thick, back 0.5mm thick) were used to sandwich a 250 µm thick Teflon foil from which the channels were cut using a Graphtec CE6000-40 Plus plotter. We used sapphire plates for their higher thermal conductivity compared to normal glass. The back sapphire had four holes to allow gas and water flow in and out of the chamber. The sandwich was held together by the steel frame and aluminum back, which were screwed together with a torque of 0.2 Nm. The aluminum back was held in place on the water bath fixture by magnets. A thin ($25\mu m$) graphite foil was placed between the sapphire back and the aluminum back to increase the thermal conductivity. Figure IV.1 shows an exploded view of the chamber.

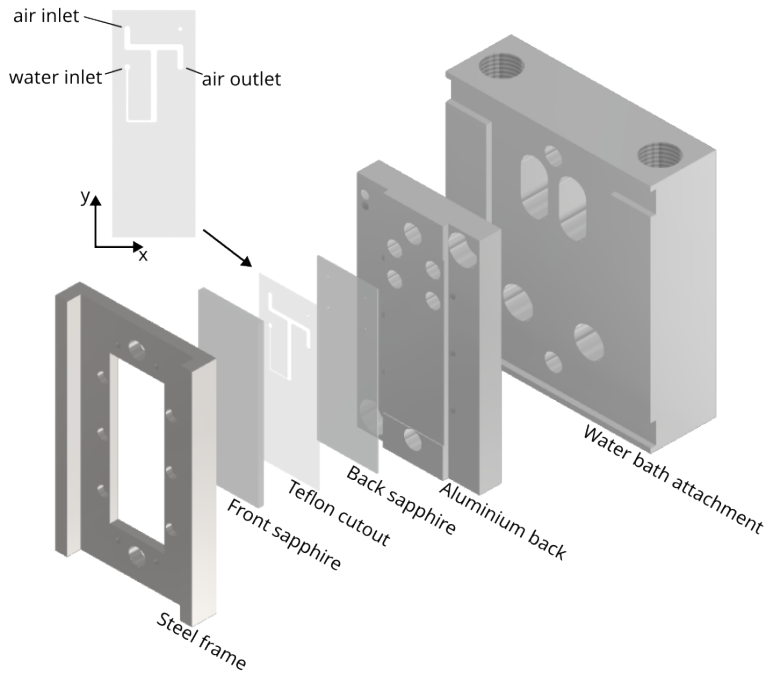


Figure IV.1. Sketch of the microfluidic chamber assembly. Between a steel frame front and an aluminium back attached to the waterbath attachment, a sapphire doublet sandwiches the teflon cutout. The 250µm thick cutout is connected to airflow and waterflow through holes in the back sapphire.

## Supplementary information V: Bead Measurements

A self-written Labview script (Fig. V.1) was used to track the movement of the fluorescent beads. 10µl solution containing 0.5 µm fluorescent beads (Invitrogen, Eugene, Oregon, USA, Lot: 31373W), diluted 1 to 2000, were loaded into the chamber and subjected to a pure water upward flow of 3 nl/s and a perpendicular gas flow of 230ml/min (Supplementary Movie 1). Images were captured using a 50ms exposure time, resulting in approximately 20 fps. Our setup did not allow very fast beads to be traced, as the maximum frame rate of 20fps did not capture particles faster than about 1mm/s. The 2D map of traced velocities therefore has some dark spots near the interface where the fastest beads are located.

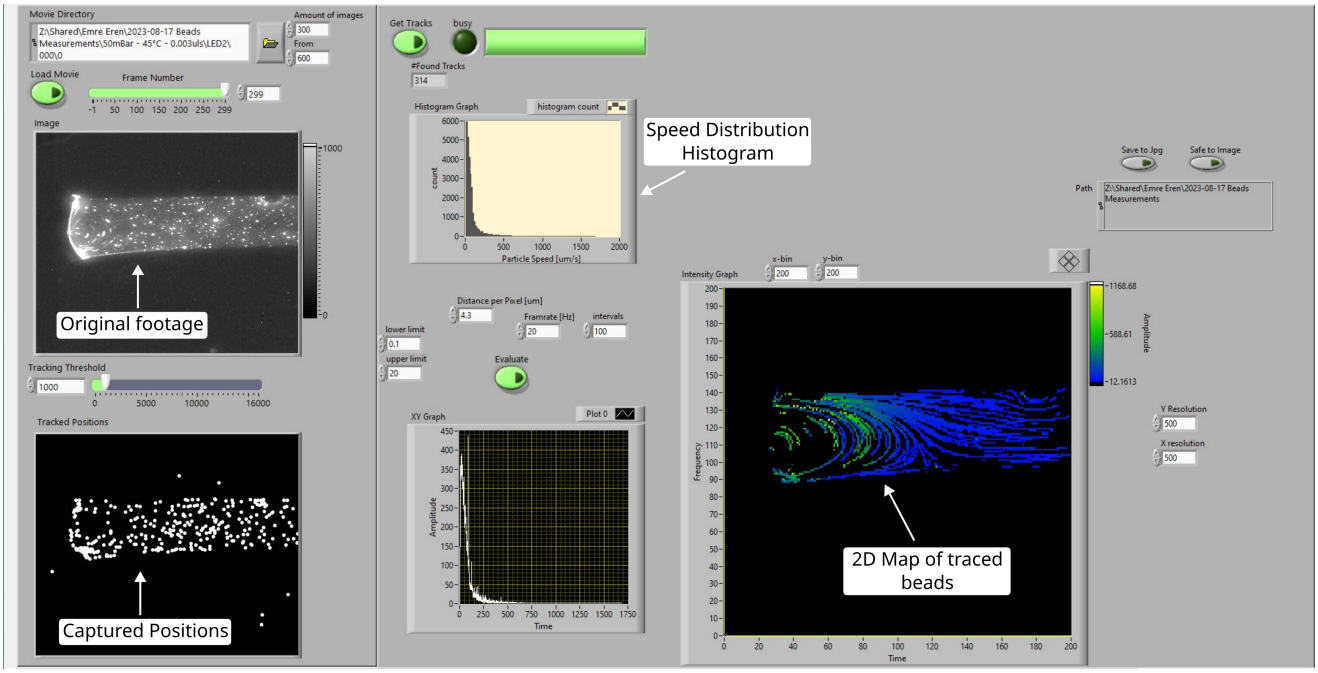


Figure V.1. Screenshot of the user interface of the self-written LabView script used for particle tracking. A series of microscopy images are loaded and beads are located and positional changes between images are traced. Statistics of particle speeds spanning hundreds of images are calculated using imaging framerate and previously measured length/pixel values. A 2D map of the obtained traces converted to the corresponding speed is generated as well.

## Supplementary information VI: Finite Elements Simulations

Finite element simulations were performed using COMSOL Multiphysics 5.4. A 2D geometry was designed using the same parameters as the Teflon cutout for the experimental chamber. Since the dynamics of the experiment take place mainly in the x-y plane, the system was simulated without the z dimension, focusing on the key properties. The geometry is coupled to a gas inlet and outlet as well as a water inlet from the bottom, analogous to the experiment (Suppl. VI). The normal gas inflow has been set to experimentally measured values (about 236 ml/min), resulting in velocities up to about 12 m/s. The gas outlet uses pressure as its boundary condition, releasing as much gas as necessary to maintain a constant pressure. The system is assumed to be laminar, since the velocities don't exceed Mach ¡ 0.3.The transport of water vapor in the gas is coupled at the top, using the stationary velocity field previously established. Simultaneously, the stationary velocity field of the laminar upward flow of water was calculated and coupled with the time-dependent transport of dilute species, in this case dissolved DNA or salts.

For both the gas and the water, Navier-Stokes-Equations were solved under the assumption that the flows are laminar due to their relatively low speed to viscosity ratio (Reynolds number). The flows can therefore be considered incompressible, the density constant and the continuity equation reduces to the condition:

$$\nabla \cdot \boldsymbol{u} = \mathbf{0}\,. \tag{1}$$

with $\rho$ denoting the mass density and $\boldsymbol{u}$ the velocity field. The Navier-Stokes equation then reduces to

$$\rho(\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = \nabla \cdot [-p\boldsymbol{I} + \eta(\nabla \boldsymbol{u} + (\nabla \boldsymbol{u})^T] + \boldsymbol{F} = \mathbf{0}\,, \tag{2}$$

with $p$ being pressure, $\boldsymbol{I}$ the unity tensor, $\eta$ the fluid dynamic viscosity and $\boldsymbol{F}$ the external forces applied to the liquid. The reference pressure was set to 1[atm], the reference temperature was 45°C and all surfaces, except the gas-water interface, are described as non-slip boundary conditions. The diffusion dependent transport of diluted species was simulated using Fick's law and convection due to the laminar flow fields:

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (-D_i \nabla c_i + \boldsymbol{u} c_i) = 0 \tag{3}$$

Equivalently, equation 2 and 3 are used for the dynamics of the gas channel. The boundary condition to combine the gas-flow with the water-flow is embedded in the gas-water interface: The velocity field components in x- and y-direction of the gas- as well as water-flow are required to be equal at the gas-water interface:

$$\boldsymbol{u} = \boldsymbol{u2} \tag{4}$$

where $\boldsymbol{u}$ describes the vector field of water, while $\boldsymbol{u2}$ denotes the vector field of the passing gas. The interface acts as a sliding wall, moving in the x-direction of the gas-flux to emulate the momentum transfer of the wind to the water surface. To simulate the evaporation of water into the gas phase, we used the August equation, describing the relation between saturation vapor pressure and temperature:

$$P_{sat} = \exp\left\{20.386 - \frac{5132K}{T}\right\}[mmHg] \tag{5}$$

The saturation concentration of water vapor therefore is

$$c_{sat} = \frac{P_{sat}}{R \cdot T} \tag{6}$$

,where R denotes the ideal gas constant and T the temperature. At the interface, the boundary condition reads:

$$c_{vapor} = c_{sat}, \tag{7}$$

while at the ceiling of the gas-channel, far away from the interface, the concentration is set to:

$$c_{vapor} = h \cdot c_{sat}, \tag{8}$$

where h denotes the relative humidity in percent. Furthermore, the speed of evaporation at the interface is proportional to the vapor concentration gradient:

$$\boldsymbol{v}_{\text{evap}} = -D_{vap} \cdot \frac{M}{\rho} \cdot \nabla c_{vap} \tag{9}$$

where M represents the molar mass of water, $\rho$ the density of water and $D_{vap}$ the diffusion coefficient of vapor.

At the inlet, the vapor concentration in the gas is set to a constant humidity dependent value (See table VI.1 for a detailed parameter list).

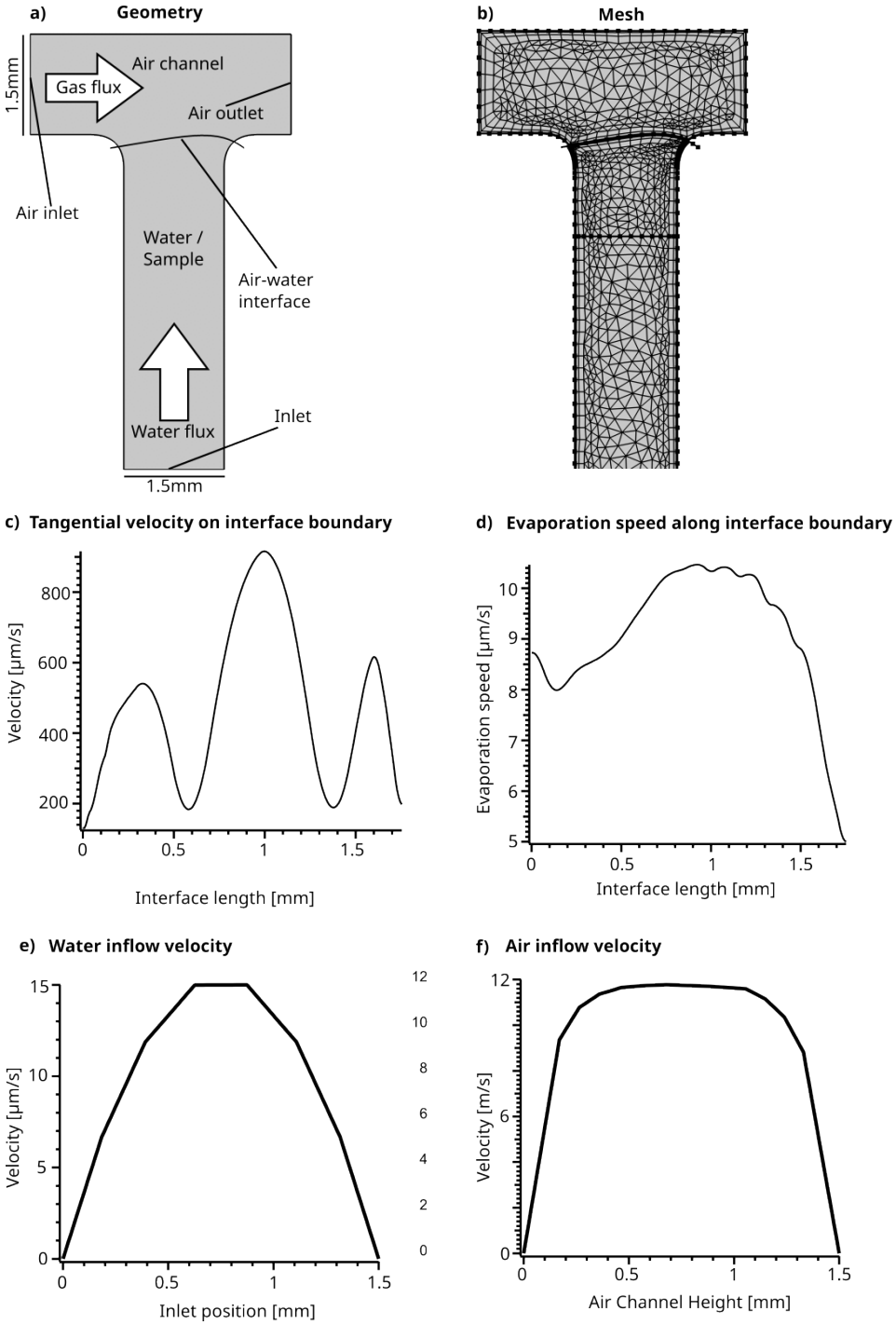


Figure VI.1. **a)** The geometry as it is used for the simulation. **b)** The geometry of the system after it has been meshed.**c)** Tangential velocity directly at the interface. x-Axis goes from the left-most point of the interface (see a)) to the right-most. The velocity is induced by the momentum transfer of the gas brushing across the interface (boundary condition equation 4). **d)** Evaporation speed along the interface. **e)** Parabolic flow profile at the inlet of the chamber in a). **f)** Parabolic flow profile of the gas flow measured at the inlet

| Parameter | Value | Description |
|---|---|---|
| $D_{vapor}$ | (21.2E-6)*(1/1[K])*(1 + (0.0071*(T - 273))) $[m^2/s]$ | Diffusion Vapor |
| $D_{63mer}$ | 643*$n^{-0.46}$[µ$m^2$/s]= 95.6 µ$m^2$/s | Diffusion Coefficient of a 63mer DNA Strand [9] |
| $D_{Mg^{2+}}$ | 705 µ$m^2$/s | Diffusion Coefficient of $Mg^{2+}$ [34] |
| $c(vapor)_0$ | humidity*0.01*(exp(20.386- (5132[K]/T))[mmHg]) / (R * T) | Initial Vapor concetration |
| Humidity | 40 % | Ambient relative humidity |
| $p_{sat}$ | (exp(20.386 - (5132[K]/T))[mmHg]) | Vapor saturation pressure |
| $c_{sat}$ | $p_{sat}$/(R*T) | Vapor saturation concentration |
| $M_{vapor}$ | 0.0180 [kg/mol] | Molar mass of vapor |
| $M_{water}$ | 18.01528[g/mol] | Molar mass of water |
| T | 45°C | Temperature |

Table VI.1. **Parameters used for the finite elements simulation.** Final set of parameters used to simulate the system. Water-specific parameters such as dynamic viscosity or density were taken from inbuilt features of COMSOL Multiphysics 5.4..

## Supplementary information VII: Förster Resonance Energy Transfer (FRET)

To measure the FRET signal in our microscope, we used an alternating illumination protocol (Supplementary Movie 3). The FRET pair FAM-ROX was excited by two LEDs in rapid succession. The blue LED excited the donor dye (FAM), while the acceptor (ROX) can only be excited indirectly while both dyes are in the FRET region. The yellow LED excited only the acceptor dye (ROX). Individual images of each illumination were captured using an Optosplit II to separate the individual emission wavelengths of FAM and ROX before they reached the camera. This allowed the emission of FAM and ROX to be captured simultaneously for each of the two illuminations, providing four images for each time point: DD, DA, AA, and AD (see table VII.1 for details). The spatially averaged, temperature-dependent, crosstalk- and artifact-corrected FRET signal was calculated using the equation 10[9]. Crosstalk between the two channels (aa(T) and dd(T)) was calculated in separate experiments using the same setup parameters with the equation 11 and 12, respectively. The data used for the crosstalk calculations are shown in Fig.VII.1. To test how different salt concentrations affect the FRET signal, we performed melting curves of different salt concentrations (See Supplementary Figure VII.2). We found that the sodium concentration has little effect on the melting temperature, while $Mg^{2+}$ strongly influences the hybridization state. In the FRET experiment in Figure 3, the initial $Mg^{2+}$ concentration was 50 µM at 45 °C. In this state the double stranded fraction is about 0.3. When the salts accumulated at the interface, salt concentrations increased up to 9 fold, strongly changing the double stranded fraction to around 0.8.

| Channel | Excitation | Emission | Label |
|---|---|---|---|
| DD | FAM - 470nm | FAM - 536nm | FAM/ROX |
| DA | FAM - 470nm | ROX - 630nm | FAM/ROX |
| AA | ROX - 590nm | ROX - 630nm | FAM/ROX |
| AD | ROX - 590nm | FAM - 536nm | FAM/ROX |
| $AA_A$ | ROX - 590nm | ROX - 630nm | ROX |
| $DA_A$ | FAM - 470nm | ROX - 630nm | ROX |
| $DD_D$ | FAM - 470nm | FAM - 536nm | FAM |
| $DA_D$ | FAM - 470nm | ROX - 630nm | FAM |

Table VII.1. Channel definitions for FRET calculation. First capital letter denotes the excitation wavelength (D = Donor, A = Acceptor), second the measured emission wavelength and the subscript stands for the label used in a separate experiment to determine crosstalk related artifacts.

$$FRET(T) = \frac{DA(T) - dd(T) \cdot DD(T) - aa(T) \cdot AA(T)}{AA(T)} \tag{10}$$

where dd(T) and aa(T) represent the non-FRET artifacts (crosstalk) in the DA and AA channels and are defined as:

$$dd(T) = \frac{DA_D(T)}{DD_D(T)} \tag{11}$$

and

$$aa(T) = \frac{DA_A(T)}{AA_A(T)} \tag{12}$$

Before each experiment, a melting curve of the FRET strands was performed inside the experimental setup chamber. The melting curve was used to normalize the FRET signal to 0 and 1 using the following equation:

$$FRET_{norm}(T) = \frac{FRET(T) - \alpha}{\beta} \tag{13}$$

where $\alpha = min(FRET(T))$ and $\beta = max(FRET(T))$.

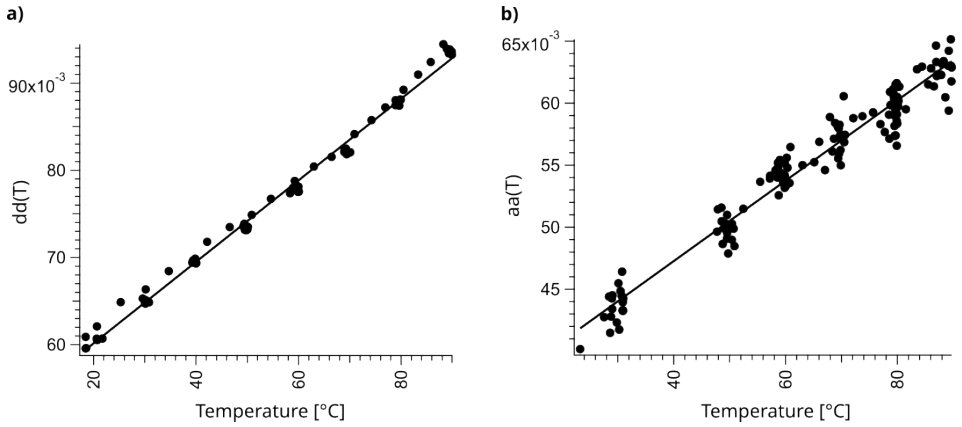


Figure VII.1. Crosstalk between donor and acceptor channel, **a)**: dd(T) and **b)**: aa(T), plotted as a function of temperature and fitted linearly.

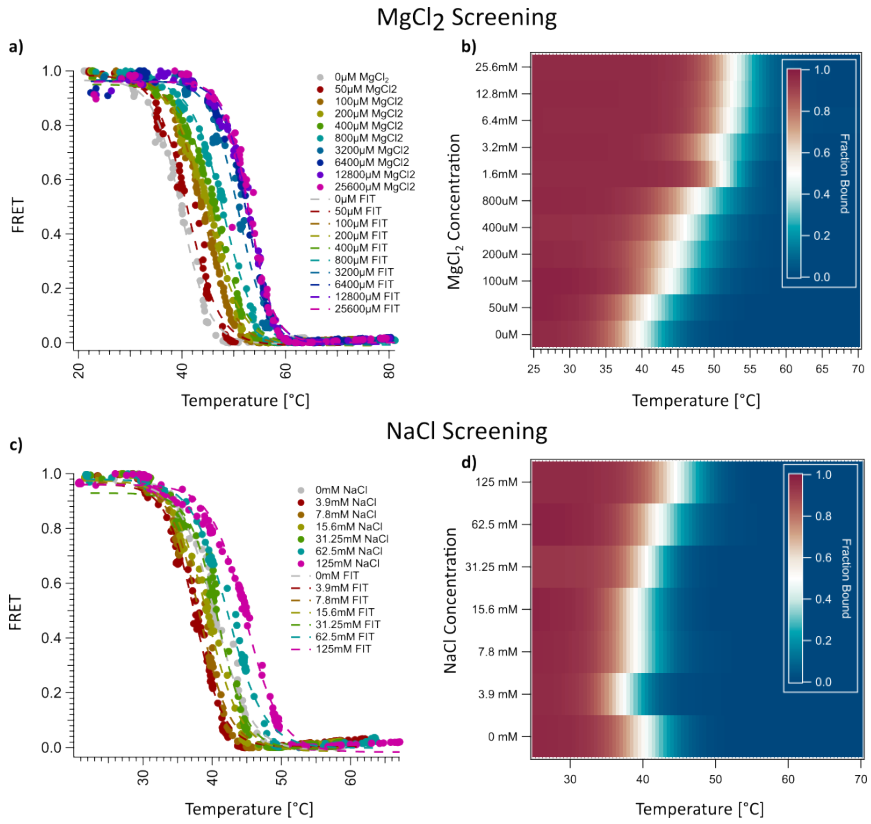


Figure VII.2. Melting curves performed in the FRET setup using the 24mer strands labelled with ROX and FAM respectively. a) Raw data of the melting curves of different $MgCl_2$ concentrations with $\alpha$ and $\beta$ already applied for normalization. b) Data from a displayed as a heatmap. The white areas display a fraction bound of 0.5 corresponding to the melting temperature $T_m$. Note that small oscillations in $Mg^{2+}$ strongly influence the melting temperature, which can enable strand separation at isothermal settings. c) Raw data of the melting curves of different NaCl concentrations with $\alpha$ and $\beta$ already applied for normalization. d) Data from c displayed as a heatmap. The white areas display a fraction bound of 0.5 corresponding to the melting temperature $T_m$.

## Supplementary information VIII: Random Walk Model

For the random walk model, we first used the existing Comsol simulation and ran the simulation with $Mg^{2+}$ ions and a 61mer DNA as the dilute species (the diffusion constant of $705\frac{\mu m^2}{s}$ for $Mg^{2+}$ at 25°C was taken from [34] and the diffusion constant of 97.04 $\frac{\mu m^2}{s}$ for a 61mer DNA strand from [9]). The simulation was performed in the same chamber, with the same characteristics and settings as in Suppl. Sec. VI. The resulting stationary salt and DNA concentration fields after 2 hours were exported as a 2D table with 200 values in x-direction and 200 values in y-direction, representing the whole simulated geometry. The same was done for the stationary laminar flow field in x-direction and y-direction induced by the air flow across the gas-water interface. Values were linearly interpolated for points between values from the grid.

Then, a self-written LabView script was used to simulate the Brownian motion of a particle with a chosen diffusion constant starting at a random position in the chamber and propagating along the flow vector field in 10ms time steps. To do this, we look at the random square displacement of a particle with diffusion constant D:

$$x^2 = Dt \tag{14}$$

As this particle can move in two directions, left and right this becomes

$$x^2 = 2Dt \tag{15}$$

Expanding this into two dimensions, we get:

$$x_{2D}^2 = x_x^2 + x_y^2 \text{ and therefore: } x_{x,y} = \sqrt{x_x^2 + x_y^2} = \sqrt{2Dt + 2Dt} = \sqrt{4Dt} \tag{16}$$

We then inserted the diffusion constant of a 61mer DNA[9] of $97.04 \cdot 10^{(-12)}\frac{m^2}{s}$ and a random unit vector phi with values between [-1,1]. $\vec{u}$ and $\vec{v}$ represent the exported laminar flow field data from Comsol:

$$\text{x displacements:} \sqrt{4 \cdot 97.04 \cdot 10^{(-12)} \cdot dt} \cdot phi + dt \cdot \vec{u} \tag{17}$$

$$\text{y displacements:} \sqrt{4 \cdot 97.04 \cdot 10^{(-12)} \cdot dt} \cdot phi + dt \cdot \vec{v} \tag{18}$$

The particle was then displaced according to equations 17 and 18 with a timestep of 10ms for a total of 35 minutes. Along its path, the respective local $Mg^{2+}$ and 61mer DNA concentration were plotted, yielding the graph displayed in main text Fig.3(d). The path was overlayed with the original Comsol simulation graphic of the salt concentration distribution in Figure 3(c).

## Supplementary information IX: PCR using Taq Polymerase

Replication reactions were performed using the AllTaq PCR Core Kit (QIAGEN). Each reaction contained 2.5U of AllTaq polymerase, 2X SYBR Green I, 5nM template, 0.25µM of each primer, 200µM of each dNTP and 0.5X PCR buffer (contains Tris HCl, KCl, $NH_4SO_4$ and $MgCl_2$).

To distinguish the template from the product strand on a PAGE image, we have added an overhang of 5A's to the 5' end of each primer. The 51mer template will be extended with the product strand then becoming a 61mer. Figure IX.1 shows a scheme of the replication cycle. Steps A to E represent the phase of replication in which the original 51mer template is extended step by step, first to a 56mer and finally to the 61mer product. Once the template is extended, the reduced replication cycle F to G (Fig.IX.1 green boxes) begins, in which the concentration of template and product strands increases exponentially. Once the initial amount of template (5nM) is consumed, the reaction can only be represented by the reduced scheme.

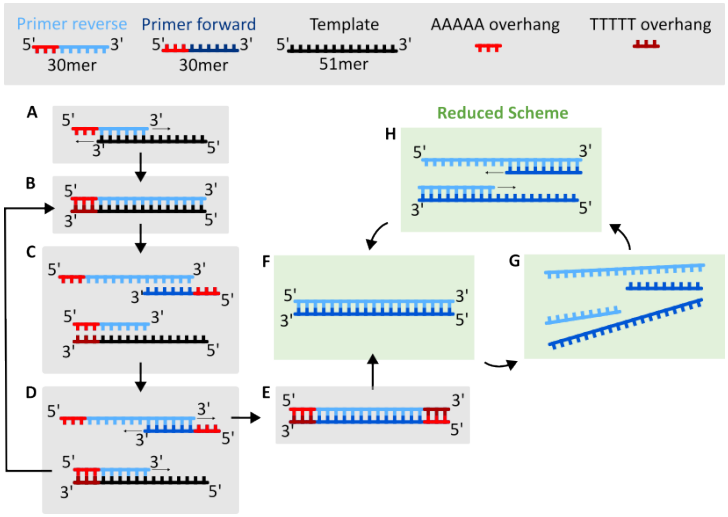


Figure IX.1. Complete schematic of the replication reaction using Taq polymerase. **A** The 30mer reverse primer binds to the 51mer template. The 5-A overhang remains unbound. **B** Taq polymerase adds nucleotides from 3' to 5' and a complete double strand is formed, now containing 56 base pairs. **C** After strand separation, the newly formed 56mer intermediate product with a 5-A overhang at the 5' end and the intermediate 56mer template with a 5-T overhang at the 3' end are bound by the primers. **D** In one case, Taq elongates the forward primer bound to the intermediate product and proceeds to step E. In the other case, elongation of the reverse primer bound to the intermediate template leads back to step B. **E** The result of the extension is a new product and a new template of 61 bases each. From here, the cycle enters the "reduced scheme". The newly formed products, together with the original primers, now replicate exponentially: **F** The double strand formed in E can now be considered as product and template for the reduced scheme. **G** After de-hybridization, both primers can anneal to template and product. **H** Taq extends again from 3' to 5', forming two new double strands of template and product, doubling their amount and completing one cycle of exponential amplification.

To confirm that the reaction in the chamber followed the predicted scheme, we performed a series of control experiments. Fig.IX.2 shows the resulting PAGE gels and the temperature protocol. The experiment was performed either in the microfluidic chamber (Supplementary Movies 4 & 5) or in a test tube inside a thermocycler (Fig.IX.2a)). After a heat activation step of 95°C for Taq polymerase, the temperature was kept constant at 68°C in the microfluidic chamber, while in the thermocycler we followed the PCR protocol for Taq, in which the sample underwent multiple cycles of replication (Fig.IX.2b)). After 95°C, the primers are given time to anneal by cooling the sample to 52°C, followed by a replication step at 68°C, where Taq is at its peak performance. This temperature protocol is then repeated an additional 39 times to complete the 40 cycles, and the sample is then cooled to 4°C, extracted, and then stored at -20°C until PAGE analysis.

Fig.IX.2(c) shows the PAGE results for the test tube samples in the thermocycler. On the left, the triplicate of the full sample (conditions shown in Fig.IX.2(a)) shows primer consumption and the formation of a product band in all cases. To be sure that Taq is not forming an unwanted side product, such as primer dimers, we repeated this experiment without adding the template strand, and indeed no product strand can be detected. Furthermore, the reaction cannot proceed to generate product without having both the reverse and forward primers. In the PAGE gel on the right, experiment triplicates are shown without the forward primer, without the reverse primer, or without primers at all. As expected, no product was formed in any of these cases. Without the addition of DNA (no primers and no template), Taq polymerase does not form a new strand. As a further negative control, we kept a complete sample in the chamber at isothermal 68 degrees C, analogous to the experiment, and observed no product formation. This is particularly interesting because it shows that without the microfluidic chamber environment, the replication cycle cannot be completed and the reaction is halted, further emphasizing the need for salt cycling in the chamber experiment.

Fig.IX.2(d) shows the PAGE gels of all experiments performed in the microfluidic chamber. While the full sample replicates 2 and 3 show product formation, no product is observed without the addition of the template strand to the reaction mix. When the reaction is run without the reverse primer, forward primer, primers in general, or no DNA at all, no product formation is observed. To show that the replication reaction is only possible when both the gas flow and the water flow are turned on, we repeated the experiment without any fluxes turned on. Here, the full sample was placed in exactly the same microfluidic chamber and kept at isothermal 68 degrees C as in the other chamber experiments. However, without upconcentration at the interface and without continuous stirring by the gas flow, no product formation can be observed.

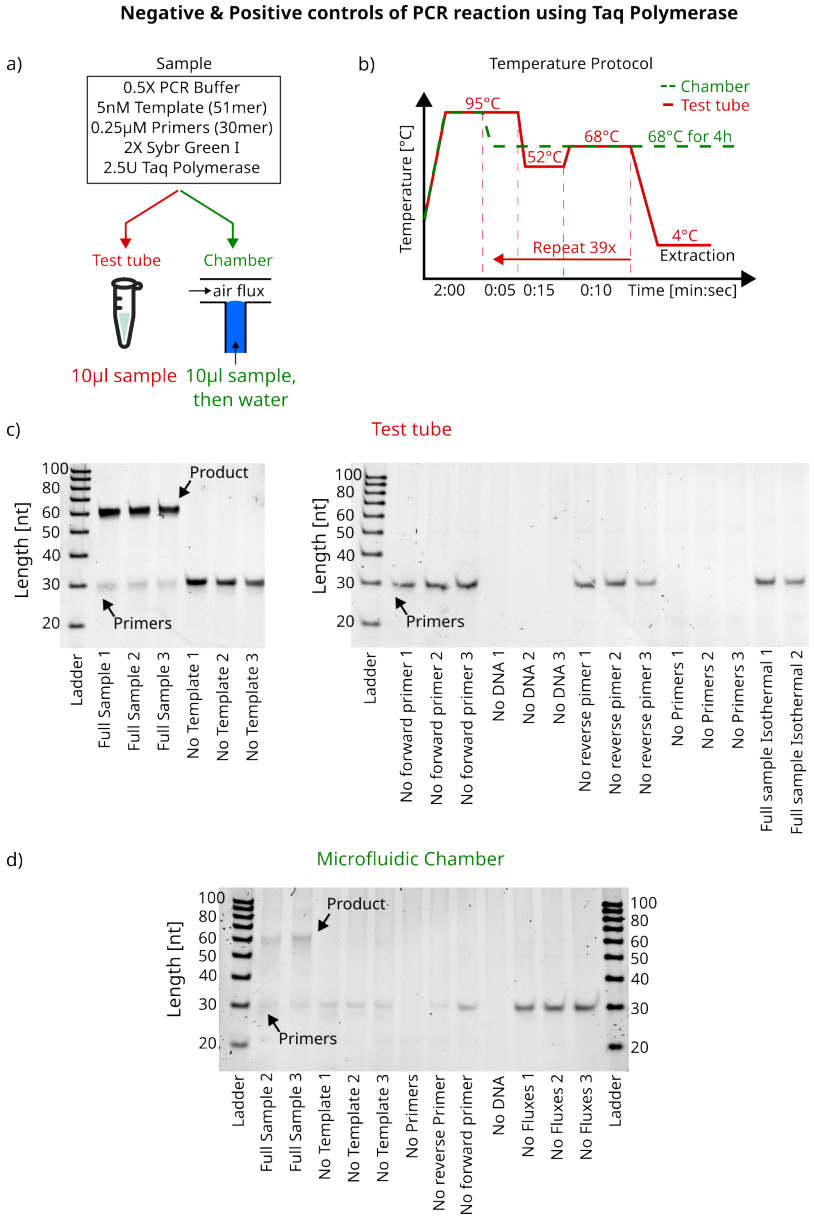


Figure IX.2. PAGE analysis of Taq PCR reactions. **a)**Sample composition used for the PCR reaction. The sample is either placed in a test tube (Eppendorf tube) and the temperature is controlled in a thermocycler (red text), or 10 µl of the sample is loaded into the microfluidic chamber prior to dilution water inflow (green text). **b)** Temperature protocol for the two individual experiments. In the thermocycler (red line), the sample undergoes a heat activation step at 95°C, followed by an annealing step at 52°C for 15s, then a replication step at 68°C for 10s. This is repeated 40 times before the sample is extracted and stored at -20°C prior to loading on a gel. Chamber experiments are performed at isothermal 68°C after the same heat activation step and extracted after 4 hours. **c)** PAGE images for the test tube samples. "Full sample isothermal 1&2" samples have the same temperature protocol as the chamber samples. A slight band may be visible around 51nt caused by the 5nM of the 51mer template. The numbers indicate the nth replicate of the experiment. **d)** PAGE image of the chamber control experiments. Primer fluorescence varies between samples, which is caused by variability during sample extraction or primer consumption by Taq (in case of "Full Sample").

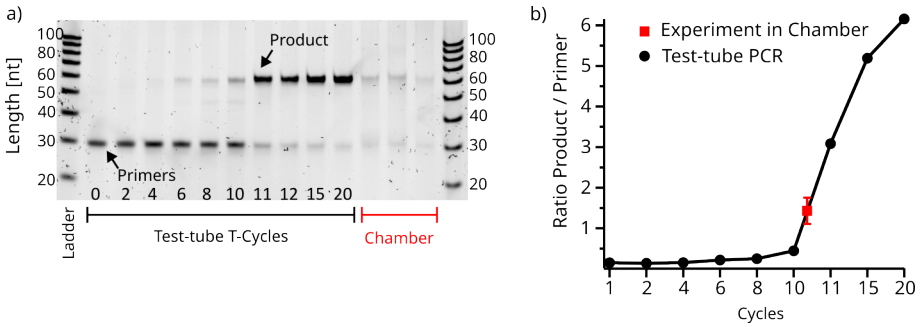


Figure IX.3. Comparison of test-tube temperature cycling vs. chamber experiments. **a)** Full Sample (10 µL of 2.5U of AllTaq polymerase, 2X SYBR Green I, 5nM template, 0.25µM of each primer, 200µM of each dNTP and 0.5X PCR buffer) were subjected to various amounts of temperature cycles, as displayed in IX.2b). The more cycles are performed, the more primers are consumed to form the product strand. The three experiments performed in the chamber show a generally lower gel intensity, which is due to losses during sample extraction from the microfluidic chamber. **b)** Comparison of Product/Primer intensity-ratio of the test-tube sample to the extracted chamber samples. This reveals that in the 4h chamber experiment about 10-11 cycles were performed. Errorbar of the experiment data point is the standard deviation of the 3 chamber samples in b).

Furthermore, we were interested in how many full PCR cycles the sample underwent in the chamber compared to regular PCR using temperature cycling in a test-tube. Therefore we performed the experiment in a test-tube with 10 different amount of temperature cycles as displayd in Figure IX.2b). We then compared these results with the samples extracted from the air-flux chamber after 4 hours. Figure IX.3a) shows the corresponding 15% PAGE image. However, due to losses during the extraction of the sample from the microfluidic chamber, only comparing the gel band intensity of the product from chamber to test-tube is not representative, because the corresponding primer band intensity does not match any of the test-tube samples. To account for this, we calculated the ratio of product to primer intensity for all lanes (Fig. IX.3b)). Since losses during extraction are the same for primer as well as for product strands, the ratio of product to primer strands stays unaffected. This reveals that inside the microfluidic chamber, with air- and water-fluxes turned on, after 4 hours, 10-11 full cycles of replication were performed. Gel intensities were extracted using a self-written LabView tool.

# Acknowledgements

First of all I want to thank Dieter for offering me an opportunity to pursue science in his lab, to supervise me for all the years I've been here and for his helping advice in times of need. Also for the amazing Kaiserschmarrn and the way too cheesy Käsespätzle. For purchasing such a damn good coffee machine for our lab, so that I could replace morning sleepiness with an addiction to coffee. And finally for the never ending enthusiasm to hike in Lenggries. I will never forget the hike to the top of the mountain, where he just pulled out a paraglider from his backpack and waved goodbye, flying off the cliff. SIC.

Secondly, I'd like to thank Christof for the infinite amount of tech knowledge he brought to the table. For offering help in any situation and then apologizing for not helping well enough. Also for making me feel like a total noob when programming LabView, which is true, but man, did you have to solve the problem I could not do in 1 week in five minutes?

Then of course I'd like to thank Zhenya for showing me how to pipette like an absolute gigachad, Matze and Lorenz for their never ending dispute, entertaining me to no end. Matze, I will never forget the "Ver***** dich mit deinen Rittern aus meiner Basis, du...". Good times.

I want to thank Adriana for being the perfect collaborator on a project, for making the lab a better place, for her endless moral support and for making me laugh on a daily basis. I don't think I could have lived through all these years without you. I will never forget the moment you walked into our lab... and how fast you raise your hand when someone asks for a coffee. Thank you for always casting a smile on my face and for making this lab an actual pleasure to go to every day. To many more years not only as colleagues ;).

Sree, thank you for taking care of our HPLC, teaching me how to make proper cappuccino, and googling things faster than your own shadow. Juliette, thanks for the many many crepes over the years, all those mispronunciations of english words containing the letter h: made me laugh every time. Thanks to Martina for being the real G and taking care of the lab organisation, without you, I don't know how it would've worked out. Sree, Juliette, Adriana and you are like family to me.

Thank you Saroj for the amazing indian food and all these amyloid references. Thanks Annalena for the help on COMSOL stuff and for bringing Annika along, as she is the definition of cuteness. Thanks to Paula and Theresa for the great gossip sessions in the coffee corner, Alan for showing me how not to give a f*** and still giving support even after having left the country. Thanks!

Thanks to Alex for being one of the only people to actually restock drinks in the fridge alongside me and for that core memory involving the trash bin outside McDonalds. Thanks to Noel a.k.a. Chaval for flossing with me whenever required. Thank you Thomas for ruining my spotify algorithm and for being emperor trappatine in the backlab. Thanks to Felix and Arthur for the amazing chess games.

Special thanks to Christoph Weber and Giacomo for the lovely zoom discussion sessions and for being the best collaborators on a paper a PhD student could wish for.

Also thanks to my DOTA comrades Rabaukenraph, Atomjürgen, StGB-Stephan and SeegurkenSam for having my back in the most dire but also radiant times. You have shown me

the art of maintaining the same level of skill over many years. Spears shall be missed and our shackles broken, but we shall never forfeit the fight for the ancient.

Last but not least I want to thank my family for the continuous support in any situation, I could not have done this without you.