Autonomous non-equilibrium mechanisms for molecular evolution

Lorenz Magnus Richard Keil

München 2017

Autonomous non-equilibrium mechanisms for molecular evolution

Lorenz Magnus Richard Keil

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

vorgelegt von
Lorenz Magnus Richard Keil
aus Bad Aibling

München, den 30.10.2017

Erstgutachter: Prof. Dr. Dieter Braun
Zweitgutachter: Prof. Dr. Ulrich Gerland
Tag der mündlichen Prüfung: 14.12.2017

Zusammenfassung

Grundlegende Prozesse in biologischen Systemen beruhen auf der Kopplung physikalischer und chemischer Ungleichgewichte. Die Untersuchung solcher Systeme sowie deren Wechselwirkungen untereinander liefern dabei wichtige Erkenntnisse über lebende Systeme und geben Aufschluss über deren Entstehung. Vereinfachte Modellsysteme bieten dabei die Möglichkeit, die Grundprinzipien der molekularen Evolution näher zu untersuchen und notwendige Randbedingungen für den Ursprung des Lebens abzuleiten.

Im ersten Teil dieser Arbeit wird ein thermisches Nichtgleichgewichtssystem beschrieben, welches einen natürlichen Selektionsdruck auf replizierende Polymere ausübt. In dem hier beschriebenen System führt ein transversaler Temperaturgradient innerhalb einer Kapillare zu einer gerichteten Bewegung von gelösten Nukleinsäuren, was zu einer lokalen Akkumulation der Teilchen führt. Ein auf Proteinen basierender Replikationsprozess nutzt dabei konvektiv erzeugte Temperaturschwankungen aus, um die Nukleinsäuren exponentiell zu vervielfältigen. Da kürzere Nukleinsäuren schneller repliziert werden, würden diese auf Dauer die längeren, mehr Information enthaltenden Polymere verdrängen. Durch das Anlegen eines äußeren Flusses wird daher zeitgleich ein Selektionsdruck auf die akkumulierten Moleküle ausgeübt, der die kürzeren Stränge verdrängt. Dieses Nichtgleichgewichtssystem begünstigt eine stabile Replikation langkettiger Nukleinsäuren und schützt die darin gespeicherte Information. Kürzere Nukleinsäuren werden, trotz eines inhärenten Replikationsvorteils, aus dem System verdrängt.

Der zweite Teil dieser Arbeit behandelt die Replikation von Ribonukleinsäuren (RNS) durch ein Polymerase Ribozym in einem durch thermische Konvektion getriebenen System. Die Replikation von Information wird, anstelle des Protein-basierten Replikationsprozesses, durch ein präbiotisch plausibles RNS-Replikationssystem durchgeführt. Die thermische Umgebung induziert dabei, basierend auf Thermophorese, Konvektion und Diffusion, einen Anreicherungsmechanismus, welcher sich ausschließlich auf das RNS-Ribozym auswirkt. Das Ribozym wird dabei in den kälteren Regionen des Systems aufkonzentriert und ist dadurch besser vor Degradation geschützt. Gleichzeitig katalysiert es eine exponentielle Replikation kürzerer Ribonukleinsäuren, welche sich im Konvektionsfluss bewegen. Die stochastischen Temperaturschwankungen dieses Systems erfüllen dabei die stringenten Bedienungen für den Verlängerungs- und Denaturierungsprozess einer ribozymatischen Replikation.

Im dritten Teil dieser Arbeit wird ein Mechanismus beschrieben, der zur Bildung eines stabilen pH Gradienten mittels geladener Puffermoleküle führt. Ein Temperaturunterschied von einigen Kelvin bewirkt dabei eine Trennung von OH^- und H^+ Ionen, wodurch pH Unterschiede von bis zu zwei pH-Werten entlang der Kammer erzeugt werden. Dieser Effekt tritt bei allen Molekülen mit einer Säure-Base-Reaktion auf, beispielsweise bei Aminosäuren, Nukleinsäuren und Phosphaten. Das System koppelt zudem den robusten pH Gradienten mit einer gerichteten Bewegung von Molekülen. Die thermische Konvektion transportiert diese Moleküle - abhängig von ihrer Größe - auf unterschiedlichen Zeitskalen zwischen den verschiedenen pH Regionen. Auftretende periodische pH Schwankungen können beispielsweise den Ausgang pH-abhängiger Reaktionen beeinflussen oder zur Erzeugung von pH Gradienten an einer Membran eingesetzt werden.

Abstract

Central processes in biological systems rely on the coupling between physical and chemical disequilibria. To understand living systems and how they evolved on earth, it is essential to study possible implications and interplays of non-equilibrium systems. Laboratory models can provide a versatile toolbox to explore basic principles of molecular evolution, allowing to set boundary conditions on the origin of living systems.

The first part of this thesis describes a thermal non-equilibrium system that exerts a natural selection pressure on replicating polymers. Here, a temperature gradient across submerged porous rocks drives the accumulation of nucleic acids. In addition, convectively driven temperature oscillations prompt a protein-mediated replication mechanism to exponentially amplify the accumulated polymers. Simultaneously, a diluting flow acts as a length selective filter on the accumulated molecules, favoring the replication of longer nucleic acids over shorter ones. This environment protects replicating systems from the kinetically favored loss of genomic information, the usual destiny of in-vitro replicators first explained by Sol Spiegelman. A non-equilibrium setting can therefore continuously support chemical replicators and liberate them from central entropic constraints, a fundamental prerequisite for Darwinian evolution.

The second part of the thesis describes the replication of RNA by an RNA polymerase ribozyme in a thermal micro-environment. Here, more rigid environmental constraints are necessary to extend the previously discussed protein-based model systems to a prebiotically plausible RNA-only replicating system. The thermal habitat features an accumulation mechanism that solely accumulates the RNA ribozymes, stabilizing them from spontaneous, temperature-dependent cleavage. These ribozymes then catalyze the addition of single nucleotides to a template-bound RNA primer. The ribozyme catalysis thereby utilizes recurrent temperature rises, given by thermal convection. These oscillations meet the delicate requirements for the elongation and denaturation process of the ribozymatic replication. It is therefore shown that the thermal non-equilibrium system can drive a prebiotically plausible, RNA-only replicating system. The mechanism delivers not only the energy necessary to for the replication, but also protects the system from molecular degradation.

The third part of the thesis reports on the accumulation of charged molecules that can form persistent pH gradients. A moderate temperature difference of several Kelvin across a sheet of water is sufficient to drive the separation of water into OH^- and H^+ to create a pH difference of more than two pH values. The mechanism is demonstrated for molecules engaging in acid-base reactions, such as amino acids, phosphates, and RNA. In addition, the system also features an interesting coupling between a directed movement of molecules by thermal convection and a robust and persistent pH gradient. The convection shuttles molecules - depending on their size - with different time scales between regimes of high and low pH, prompting them to undergo continuous pH oscillations. These pH oscillations open an unique way to drive pH-dependent reactions between molecules in the solution. This findings provide a novel mechanism to create a self-sustained proton gradient to drive biochemical reactions.

Each of the three chapters in this thesis are self contained and comprise of a separate conclusion and materials & methods section, making it accessible without knowledge of the prior chapters. The results presented in this thesis were published in form of three peer-reviewed publications and also include one manuscript in preparation:

1. **Lorenz Keil**, Michael Hartmann, Simon Lanzmich and Dieter Braun, **"Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations"**, *Physical Chemistry Chemical Physics* (2016), DOI: 10.1039/c6cp00577b

2. Moritz Kreysing*, **Lorenz Keil***, Simon Lanzmich* and Dieter Braun, **"Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length"**, *Nature Chemistry* (2015), DOI: 10.1038/nchem.2155
*contributed equally to this work

3. **Lorenz Keil**, Friederike Möller, Michael Kieß, Patrick Kudella and Christof Mast, **"Proton gradients and pH oscillations emerge from heat flow at the microscale"**, *Nature Communications* (2017), DOI: 10.1038/s41467-017-02065-3

Manuscript in preparation:

4. **Lorenz Keil**, David Horning, Christof Mast, Gerald Joyce and Dieter Braun, **"A thermal habitat drives the retention and ribozymatic replication of RNA"**

Contents

Chapter 1

Introduction

1.1 Evolution of life

The question of how life emerged in the universe has been driving modern research for decades. This field of research nowadays comprises interdisciplinary teams of many natural sciences - biology, chemistry, geology, astronomy, and physics. Its purpose is not only to show how life on earth originated but to depict potential pathways for the transition from non-living to living matter. The definition of living systems has, however, proven to be inherently difficult, even when common principles are deduced by living systems on earth. A generally recognized working definition for life was given by the NASA,[2] where G. Joyce summarized the definition of life as

"a self-sustained chemical system capable of undergoing Darwinian evolution."

Here, self-sustainability implies that the system does not need interventions from higher entities to continue as a living system. However, the formation and preservation of increasingly complex biomolecules solely occurs in a far-from-equilibrium driven setting. As a result, living systems depend on external energy sources to maintain local low-entropy states.[3–6] In addition, the term "chemical system" denotes the necessity of complex chemical reaction networks, e.g. to drive the synthesis of purine and pyrimidine ribonucleotides or chemiosmotic coupling.[7, 8] The central constraint for life, however, is its capability to undergo Darwinian evolution.[9] Darwinian evolution involves a molecular genetic system that is imperfectly replicated. These random mutations in the geno- and phenotype affect their ability to survive and reproduce against natural selection. Selective pressure on living systems is either given intrinsically or from environmental constraints. Darwinian evolution is therefore the key mechanism to explain how life has developed starting from the first cells to today's life forms.

The interdisciplinary work on the origins of life also made substantial progress in understanding what chemical compounds were available prebiotically,[10] before the first cells have emerged. These studies include the synthesis of sugars and purine nucleobases to Miller-Urey's

experiment, which yields diverse compounds such as amino acids.[11, 12] In an astrophysical and planetary context, amino acids have also been found in meteorites, while complex cyanides were detected in the protoplanetary disk around a young star.[13, 14] Given this diversity in starting materials, the development of more complex aspects of prebiotic chemistry then led to the synthesis of peptides, lipids, and nucleic acids,[15, 16] and therefore the formation of information encoding polymers. This is the starting point for the research presented in this thesis, addressing fundamental challenges for the onset of molecular evolution in the context of a non-equilibrium setting.

1.2 Non-equilibrium settings for molecular evolution

The formation of RNA-like biopolymers that exhibit both catalytic functions and information storage capabilities is central to the origin of life. However, geochemical evidence points towards very low molecular concentrations in prebiotic oceans, termed the ”concentration problem of the origin of life”.[17–19] All approaches to generate molecules such as amino acids,[12] purines,[11, 20, 21] pyrimidines, and oligonucleotides, as well as alternative early replicators, are limited to high initial concentrations of precursor molecules.[15, 22–26] While present-day cells run elaborate systems to maintain spatial compartmentalization, and feed their interiors by complex protein-based transport machineries,[27, 28] only a few settings on the primordial Earth are predicted to have featured comparable segmentation and accumulation of molecules from aqueous solutions. Therefore, the formation of complex informational molecules that require a number of molecular precursors is severely hindered.[17–19]

The concentration problem of the origin of life can be approached by searching for geological non-equilibrium conditions that make an origin of life possible, given suitable boundary conditions. Such a search focuses on experimentally testable conditions that create an evolutionary machine for protobiomolecules. The aim is to achieve the first steps of Darwinian evolution by a combination of physico-chemical effects. Favorable conditions for the synthesis or preservation of protobiomolecules are for example given by evaporated terrestrial ponds,[29] self-assembled lipid bilayers, or coacervates as precursors of protocells and catalyzing inorganic surfaces.[30–34] Some of these systems however rely on periodic changes in the environment or change the concentrations of ionic species, e.g. divalent metal ions, which considerably reduces the stability of biomolecules such as RNA.[35–37]

This thesis studies a non-equilibrium setting for molecular evolution, driven solely by a natural temperature gradient spanning across submerged porous rocks. Such rocks emitting geothermally heated water into the ocean were an abundant setting on the early Earth.[38, 39] and have been advocated as a possible source for the origin of life (Fig. 1.1a-c).[40] The dissipation of heat forms a temperature gradient across sub-millimeter sized pores inside these rocks. This type of heat flux drives a highly efficient accumulation mechanism which is based on the interplay of thermal convection and thermophoresis. Temperature gradients across artificial pores lead to the accumulation of dilute lipids and nucleotides,[40–42] enable the polymerization of long DNA/RNA strands,[43] and select, feed, and replicate nucleic acids towards increasing length.[44] Numerical approaches and experiments have shown that the mechanism is robust with respect to a large variety of geometries and works within artificial pores of different sizes (Fig. 1.1d).[40, 45]

A supply of protobiomolecules inside such pores, however, is not only limited to dif-

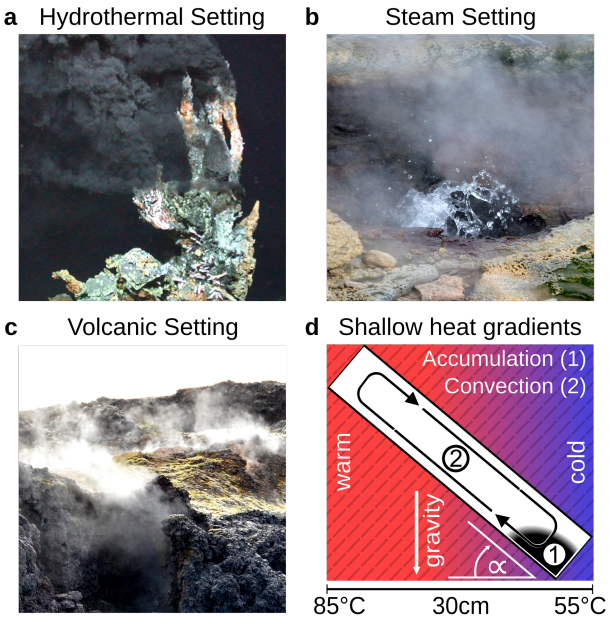


Figure 1.1: Possible microthermal habitats for the origins of life. (a) Heat dissipation across submerged porous rocks and the cold ocean form steep temperature gradients which drive an efficient accumulation mechanism.[38, 39] (b, c) Temperature gradients of 100 K per metre enable an efficient accumulation, extending the range of possible microthermal habitats from hydrothermal settings to volcanic and steam settings. These surface-based microhabitats provide wet-dry cycles and UV illumination for trapped molecules, facilitating the generation and polymerization of nucleotides.[15] (d) Numerical approaches show that elongated pore systems within shallow temperature gradients efficiently accumulate molecules such as DNA and RNA (1) and enable a heat-driven replication reaction due to cyclic temperature changes induced by the laminar thermal convection (2). (a) adapted from MARUM, University of Bremen/Germany.

fusive coupling with the ocean and Fischer-Tropsch-type synthesis.[46] Porous rocks near the surface could have also been supplied with feedstock molecules synthesized by surface chemistry.[15, 16] Numerical simulations have shown that shallow temperature gradients of 0.1 K/mm still achieve at least a million-fold accumulation within elongated pore systems. Those pores are formed by a stacking of shorter pores that are connected by mass diffusion. They behave analogously to a single, elongated pore, thereby increasing the effective pore length. Shallow temperature gradients can be found not only in hydrothermal settings but also in near-surface volcanic and steam settings (Fig. 1.1,1.2).[1] Here, molecules such as precursors of ribonucleotides, lipids, or amino acids are synthesized on the surface and subsequently leached into the pores e.g. by downhill streams from rainfall. This makes reaction products, particularly those of wet/dry and UV-reactions, available in the pores. Surface directed ends of the pores can also be directly struck by sunlight and include the case of partially dried pores, e.g. based on moisture changes in steamy environments.

Temperature gradients have shown to accumulate dilute lipids and polynucleotides.[40–42] Given a suitable geometry of the pore, simulations have also shown to enable the accumulation of a variety of small molecules, including monovalent ions like Li^+ and divalent ions such as Mg^{2+} and Ca^{2+}.[47] To calculate the particle's accumulation efficiency, only its diffusion and Soret coefficients D and S_T, respectively, are necessary.

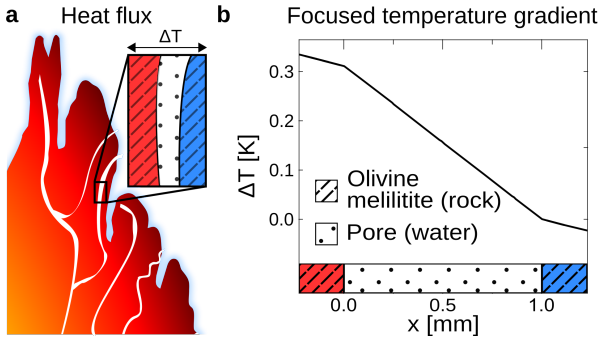


Figure 1.2: Heat flux across porous rocks. (a) Formation of temperature gradients across porous rocks, heated by volcanic activity against the cold surrounding. Shallow temperature gradient settings can be found in various environments such as volcanic, water-vapour, or hydrothermal settings. (b) Simulations show that a difference in heat conductivity between the rock (olivine-melilitite $\kappa = 2\,\mathrm{W/mK}$) and the water filled pore ($\kappa = 0.6\,\mathrm{W/mK}$) results in a local increase in temperature gradient across the pore by at least a factor of 3. Taking other naturally occurring minerals such as quartz, peridotite, or pyrite-silica into account, the temperature gradient could further increase e.g. by a factor of 33 for pyrite. Therefore, a minor temperature difference between both heat reservoirs suffices to enable efficient molecular accumulation.

Besides the accumulation of molecules, heated pores offer an additional feature relevant to the origin of life: Convective motion drives accumulated molecules through varying temperature regions. Stochastic particle simulations show that biomolecules frequently shuttle between hot and cold parts of the pore (Fig. 1.3).[1] The accumulated molecules are therefore subjected to temperature oscillations in a laminar convective flow, allowing e.g. for Watson-Crick-type replication mechanisms.[26, 48–53] As a result, the concentration mechanism could also offer a pathway to the long-term storage, propagation, and mutation of the first self-replicating molecules. The RNA-world hypothesis posits that RNA played a crucial role in the origin of life due to its catalytic function and information storage capabilities. On prebiotic earth, thermal habitats could have provided the energy necessary to support the formation of self-replicating ribozymes containing at least 200 nt and continuously drive an RNA-based replication machinery.

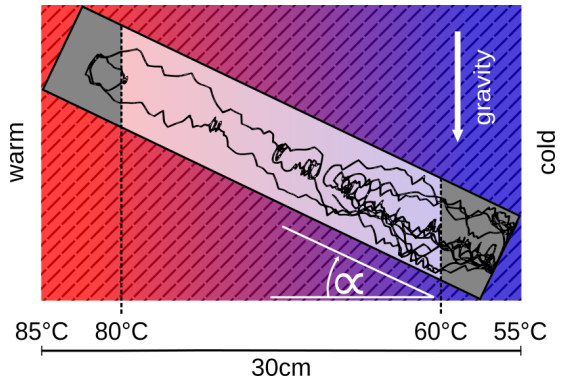


Figure 1.3: Temperature cycle statistics of individual molecules derived by numerical simulations. The trajectory of a 100mer oligonucleotide inside an elongated, 45 ° tilted pore within a 0.1 K/mm temperature gradient. A convectional flow profile with a maximum velocity of ~2 µm/s shuttles molecules between warm and cold areas, enabling replication reactions by cyclic temperature changes.

1.3 Theory on thermogravitational columns

The accumulation of molecules can be described by Debye's approach, originally used to characterize separation columns.[54] The basic principle of molecular accumulation is given by the superposition of gravitational convection and thermophoresis. Both mechanisms result from heat fluxes across water-filled compartments. The thermophoretic effect moves molecules along the thermal gradient, resulting in a net movement of

$$\mathbf{v}_{\mathrm{TD}} = -D \cdot S_{\mathrm{T}} \cdot \nabla T \tag{1.1}$$

Hereby, D and S_{T} denote the diffusion and Soret coefficients and ∇T the temperature gradient, respectively. Thermophoresis is still subject to active research;[55, 56] however, it has been found that the effect on charged molecules, e.g. short oligonucleotides, is dominated by ion shielding and Seebeck effects.[57] The latter is induced by the fact that each ionic species has a different Soret coefficient, generating a global electrical field that moves charged molecules. Thereby, the Seebeck effect highly depends on the ionic composition of the solution. In conjunction with convection, molecules erratically move within the temperature gradient while slowly accumulating at the bottom of the compartment. The superposition of thermophoresis, convection and diffusion has emerged to be highly potent in gaining extreme concentrations in distinct regions. The resulting molecular flux given by:

$$\mathbf{j} = -D \cdot \nabla c - D \cdot S_{\mathrm{T}} \cdot \nabla T \cdot c + \mathbf{v}(\alpha, w) \cdot c \tag{1.2}$$

with the molecule's diffusion coefficient D and Soret coefficient S_{T}, temperature T, the convective flow profile $\mathbf{v}(\alpha, w)$, pore width w, and the angle α between the pore and the direction of gravity. The tilting angle α and pore width w affect the gravitationally induced convective flow $\mathbf{v}(\alpha, w)$. Here, 90 ° and 0 ° denote a vertically and horizontally aligned pore, respectively. The Soret coefficients for oligonucleotides were experimentally measured by Reineck *et al.* [58] and Reichl *et al.* [57]. Following Debye's approach, perturbations at the ends of the pore are neglected.

$$c \cdot \nabla \cdot (\nabla T) = 0; c \cdot \nabla \cdot \mathbf{v}(\alpha, w) = 0 \tag{1.3}$$

Those assumptions have previously been shown to be in good agreement with experimental data, e.g. Duhr and Braun [59], Mast *et al.* [43], Reichl *et al.* [57], and Kreysing *et al.* [44]. A change in molecular concentration over time (see Equation 1.4,1.5) is derived from the molecular flux equation 1.2 by using the continuity equation:

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) + D \cdot S_{\mathrm{T}} \cdot \nabla \cdot (\nabla T \cdot c) - \nabla \cdot (\mathbf{v}(\alpha, w) \cdot c) \tag{1.4}$$

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2} \right) + D \cdot S_{\mathrm{T}} \cdot \nabla T \cdot \nabla c - \mathbf{v}(\alpha, w) \cdot \nabla c \tag{1.5}$$

Here, x and y denote the coordinates along the pore length and perpendicular to the pore in the direction of the temperature gradient shown in Fig. 1.1d. Given equilibrium buoyancy

forces, Debye derived a differential equation for the velocity in y-direction

$$\frac{\partial^3 v}{\partial x^3} = \frac{-\beta \cdot \rho \cdot g \cdot \Delta T}{\mu \cdot w} \tag{1.6}$$

Here, g denotes the gravitational acceleration, μ the viscosity, ρ the fluid density and β the volumetric temperature expansion coefficient. In case of a vertical aligned pore, a horizontal temperature gradient and non slip boundary conditions, the velocity profile in y-direction is given by:

$$v(x) = \frac{-\beta \cdot \rho \cdot g \cdot \Delta T \cdot w^2}{6\mu} \cdot \left(\frac{x}{2w} - \frac{3x^2}{2w^2} + \frac{x^3}{w^3} \right) = -v_0 \cdot \left(\frac{x}{2w} - \frac{3x^2}{2w^2} + \frac{x^3}{w^3} \right) \tag{1.7}$$

The velocity profile is then applied to 1.5. In order to solve the equation, non-dimensional variables are introduced by using $q = v_0 w/D, p = S_T \Delta T, \eta = y/w, \xi = x/w$ and $\tau = Dt/w^2$. The separation of variables is given by $c(\eta, \xi) = U(\xi) \cdot V(\eta)$ whereas V is assumed to be exponential $V(\eta) = \exp(\kappa \cdot p \cdot \eta)$ and p to be small ($p \ll 1$). As a result, the concentration $c(\xi, \eta)$ is given by

$$c(\xi, \eta) = \left(1 + p \left(-\xi + \frac{84p \cdot q}{10080 + q^2} \left(\frac{\xi^5}{20} - \frac{\xi^4}{8} + \frac{\xi^3}{12} \right) \right) \right) \cdot exp \left(\frac{84p \cdot q}{10080 + q^2} \eta \right) \tag{1.8}$$

The maximal concentration is given for $q_{opt} = \sqrt{10080}$, deriving an exponential factor of $\kappa_{opt} = 0.42$. For a given width ($w_{opt} = q_{opt} D/v_0$) and trap height h, the maximum concentration can be written by

$$\frac{c_{bottom}}{c_{top}} = \exp \left(0.42 \cdot S_T \cdot \Delta T \cdot h \cdot w_{opt}^{-1} \right) \tag{1.9}$$

As a result of 1.9, the accumulation efficiency of a specific molecule is defined by its diffusion coefficient D and Soret coefficient S_T (Fig. 1.4). While the height h of the pore exponentially increases the accumulation efficiency, the optimal width w_{opt} of the pore varies for each molecule.[1, 60] Both parameters are strongly affected by salt concentration and ambient temperature, which is shown for 1-200mer oligonucleotides (Fig. 1.5). The diffusion and Soret coefficients can be experimentally measured for a given Debye length. A Debye length for example of λ_D of 0.79 nm – 5.6 nm corresponds to salt concentrations in physiological solutions and more diluted solutions, respectively.[57]

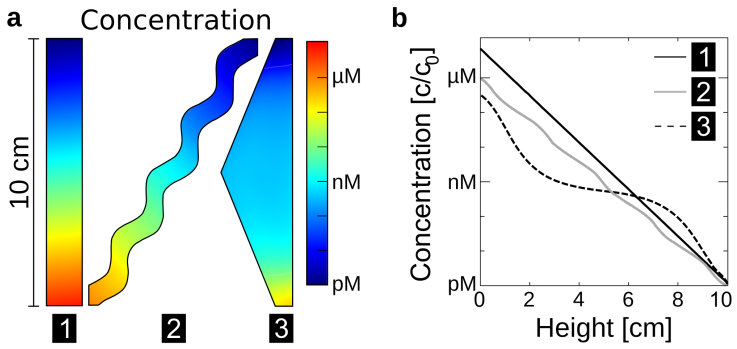


Figure 1.4: Extreme accumulations of 100mer oligonucleotides for various pore geometries via finite-element analysis, assuming shallow temperature gradients of 0.1 K/mm. Diffusion and Soret coefficients are based on experimental data.[57, 58] (a) The accumulation mechanism is found to be highly robust with respect to the shape of the pore system. Pore systems of 10 cm in length achieve accumulations in concentration from pM to near µM. (b) The shape has only a minor impact on the magnitude of accumulation. The exact distribution of concentration is shown for the geometries in (a). A rectangular geometry shows the highest efficiency, accumulating molecules by a factor of $7 \cdot 10^6$. All pore geometries achieve at least a 10^5 fold accumulation.

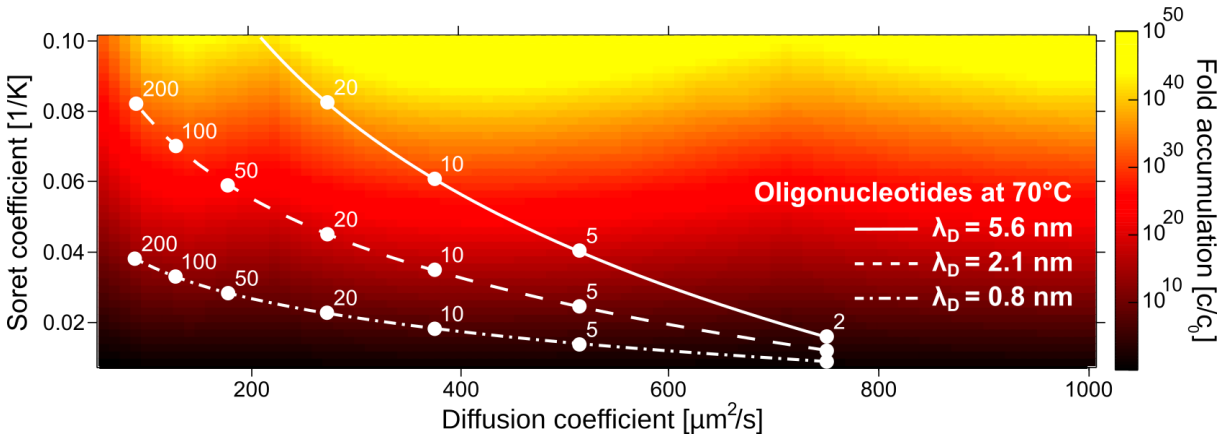


Figure 1.5: Accumulation of various biomolecules in shallow temperature gradients. A temperature gradient of 0.1 K/mm across a 1 m high pore accumulates various molecules with a given diffusion coefficient and Soret coefficient, including monovalent ions like Li^+ (D = 1029 µm^2/s, S_T = 0.0007 K^{-1}) and divalent ions such as Mg^{2+} (D = 706 µm^2/s, S_T = 0.012 K^{-1}) and Ca^{2+} (D = 792 µm^2/s, S_T = 0.013 K^{-1}).[47]. The curves show scenarios for the accumulation of 1-200mer oligonucleotides at various salt concentrations. The Debye lengths of 0.79 nm, 2.1 nm, and 5.6 nm denote high, medium, and low salt concentrations, respectively.

Chapter 2

Continuous replication and selection of oligonucleotides towards increasing length

Summary

The replication of nucleic acids is central to the origin of life. On the early Earth, suitable non-equilibrium boundary conditions would have been required to surmount the effects of thermodynamic equilibrium such as dilution and degradation of oligonucleotides. One particularly intractable experimental finding is that short genetic polymers replicate faster and outcompete longer ones, leading to ever shorter sequences and the loss of genetic information. This chapter shows that a heat flux across an open pore in submerged rocks concentrate replicating oligonucleotides from a constant feeding flow and select for longer strands. The experiments utilize the interplay of molecular thermophoresis and laminar convection, the latter driving strand separation and exponential replication. Strands of 75 nucleotides survive while strands half as long die out, inverting above dilemma of the survival of the shortest. The combined feeding, thermal cycling and positive length selection opens the door for stable molecular evolution in the long-term micro-habitat of heated porous rock.

This Chapter was published by Kreysing *et al.* [44] in *Nature Chemistry* and adapted with permission from the Nature Publishing Group.

2.1 Introduction

From a wide range of exploratory experiments, much is known about the capabilities and limitations of chemical replication systems.[15, 26, 48–51] It has become increasingly clear that such replicators are delicate systems that require a suitable supportive micro-environment to host non-equilibrium conditions. These conditions permit to sustain molecular evolution and the synthesis of molecules against equilibrating forces.[7, 15, 27, 61] To the same end, modern cells provide active compartments of reduced entropy that protect genetic information against its thermodynamically favored decay.[27, 28] This is facilitated by sophisticated membrane trafficking machinery and a metabolism that feeds on chemical low entropy sources or light energy (Fig. 2.1a).

It has been known since Sol Spiegelman's experiments in the late 1960s that,[62] even if humans assist with the assembly of an extracellular evolution system, genetic information from long nucleic acids is quickly lost. This is because shorter nucleic acids are replicated with faster kinetics and outcompete longer sequences. If mutations in the replication process can change the sequence length, the result is an evolutionary race towards ever shorter sequences.

In the following experiments, a counterexample is given. Here, heat dissipation across an open rock pore, a common setting on the early Earth (Fig. 2.1b),[63] provides a promising non-equilibrium habitat for the autonomous feeding, replication and positive length selection of genetic polymers. Previously, it has been argued that a temperature gradient spanning across a sub-millimeter wide, closed compartment is able to accumulate dilute nucleotides, to enforce their polymerization, or to concentrate lipids to form vesicles.[40, 41, 43]

This study extends the concept for the geologically realistic case of an open pore with a slow flow passing through. Theoretical and experimental observations demonstrate continuous, localized replication of DNA together with inherent non-linear selection for long strands. With

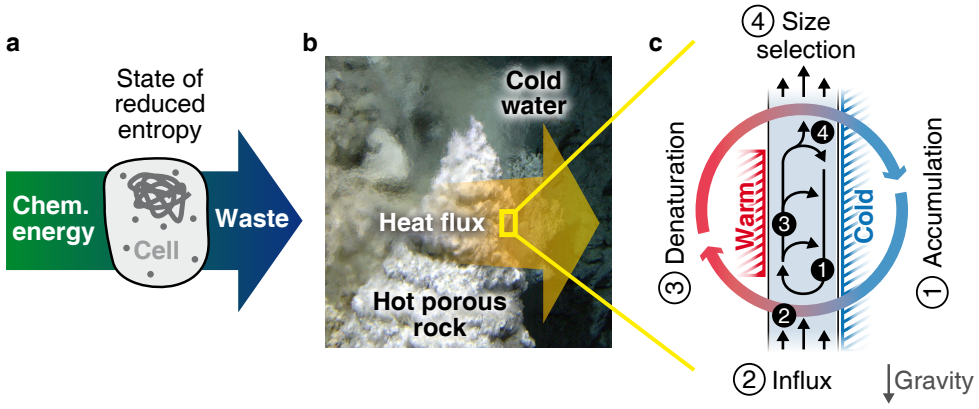


Figure 2.1: Reduction of local entropy is the key for living systems and can be caused by the flux of thermal energy. (a) Modern cells feed on chemical energy, enabling them to host, maintain and replicate information coding polymers, processes necessary for Darwinian evolution. (b) Flux of thermal energy across geological cracks near a heat source (image of white smoker adapted from Kelley *et al.* [64]). (c) A thermal gradient across a millimeter-sized crack induces the accumulation of molecules by thermophoresis and convection (1). A global through-flow imports nutrients into the open pore (2). Exponential replication is facilitated by the local convection, which shuttles the molecules repetitively between warm and cold, thus inducing the cyclic denaturation of nucleotides (3). The combination of influx, thermophoresis and convection selectively traps long molecules and flushes out short ones (4). The inflow speed determines the cut-off size of the resulting length selection.

an added mutation process, the shown system would bode well for autonomous Darwinian evolution based on chemical replicators with a built-in selection for increasing sequence length. The complex interplay of thermal and fluid dynamic effects, which lead to a length selective replication (Fig. 2.1c, 1-4), will be introduced in a stepwise manner in the following.

2.2 Thermophoretic accumulation of nucleic acids

The accumulation mechanism responsible to counterbalance the mixing entropy relies on the interplay of thermophoresis and gravitationally driven convection (Fig. 2.2a). In the presence of a temperature difference, thermophoresis drives the molecules horizontally from the warm left to the cold right side. On a similar time scale, the fluid moves vertically by convection and carries the molecules with it. Convection deflects the horizontal thermophoretic depletion and amplifies it to a strong vertical molecule accumulation (refs. [54, 65], Methods). This interplay of molecular movement and fluid flow therefore results in an efficient net transport of oligonucleotides to the bottom of the compartment, visualized experimentally in Fig. 2.2b.

For oligonucleotides with a length of 75 bases, concentrations increase by a factor of 10 per millimeter pore length, resulting in a million-fold concentration increase for a six millimeter high pore. Larger nucleic acids are exponentially better trapped, because their higher charge contributes quadratically to the achievable accumulation.[66, 67] This length-selective accumulation bias can be directly detected experimentally (Supplementary Fig. 2.7). The accumulation counterbalances diffusional dilution and offers a solution to the concentration problem associated with the origin of life.

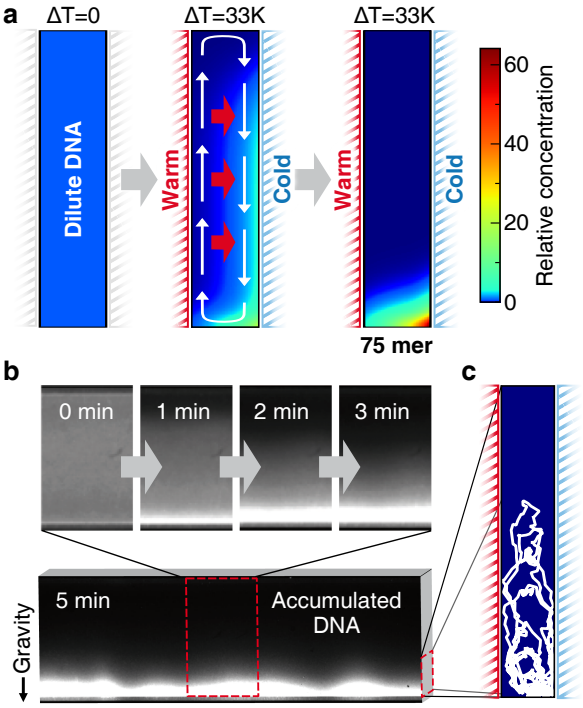


Figure 2.2: Accumulation of oligonucleotides. (a) The temperature gradient drives oligonucleotides horizontally from warm to cold by thermophoresis and triggers the vertical thermal convection of water at the same time. Its combination results in a length dependent accumulation at the bottom of an elongated pore within minutes. (b) The accumulation of dilute double-stranded oligonucleotides (100–1000mer) at the bottom is monitored within a 100 µm thin and 2 mm high capillary via SYBR Green I fluorescence. (c) The accumulation is dynamic: the nucleotides cycle between the warm and cold side, visualized in white for a single 500mer of DNA.

2.3 Size-selective trapping from feeding flow

To establish efficient feeding with replication-relevant monomers, the pore was opened at both ends. This permitted an upward feeding flow through the pore, originating from the overall large-scale upward flow in a hydrothermal situation. Interestingly, this lead to all-or-nothing trapping characteristics depending on the strand length. An oligonucleotide ladder comprising of 20 bp fragments was loaded (20-200 bp dsDNA) in a 3.5 mm high and 70 µm wide pore and subjected to an upward flow with a velocity of 6 µm/s. The analysis was performed by gel electrophoresis. Solely nucleic acids above a certain length threshold were trapped inside the pore, while shorter strands followed the upward flow and were washed out of the pore (Fig. 2.3a).

For the given velocity, this sharp length fractionation had a transition between 80 and 100 base pairs and can be understood by the interaction of the flow profile inside the trap with the thermophoretic concentration profile. The upward feeding flow superimposes on the internal convection pattern, generating an asymmetrical flow profile inside the trap (Fig. 2.3b). Long strands are pushed by thermophoresis into the descending flow at the cold side, transporting the molecules downwards. These are then localized against the upward feeding flow at the bottom end of the heated section. Shorter strands experience weaker thermophoresis and the overall upward flow drags them out of the trap. The flow rate, at which the solute nucleic acids start to move upwards and leave the pore, depends monotonically on the strand length. Consequently, the gradual increase of the flow rate with time results in the sequential release of longer strands (Fig. 2.3c). The existence of the observed length threshold might come as a surprise, but a finite element model combining flow, diffusion and thermophoresis reproduces the behavior of the trap in detail (Fig. 2.3d, Methods).

2.4 Exponential replication by convective thermal cycling

Besides continuous feeding and length selective trapping, the asymmetrically heated pore offers another important feature relevant to the origin of life: laminar, convective temperature cycling of the accumulated nucleic acids.[52, 68] This opens the door to Watson-Crick-type replication mechanisms, which are otherwise hindered by the considerable energy costs required to separate double-stranded oligonucleotides.[53] The thermal cycling can be predicted from the fluid dynamic model, including thermophoresis and diffusion (Fig. 2.4a). It is sufficient to cyclically separate double-stranded DNA (dsDNA) in order to drive exponential base-by-base replication with duplication times on the order of minutes, as documented by SYBR Green I fluorescence (Fig. 2.4b). The focus was to study the boundary conditions which enable early chemical systems for oligonucleotide replication. For this reason, the polymerase chain reaction (PCR) was chosen as a fast and well characterized placeholder for the large family of template-directed replication mechanisms that depend on temperature oscillations for long substrates.[26, 48–51]

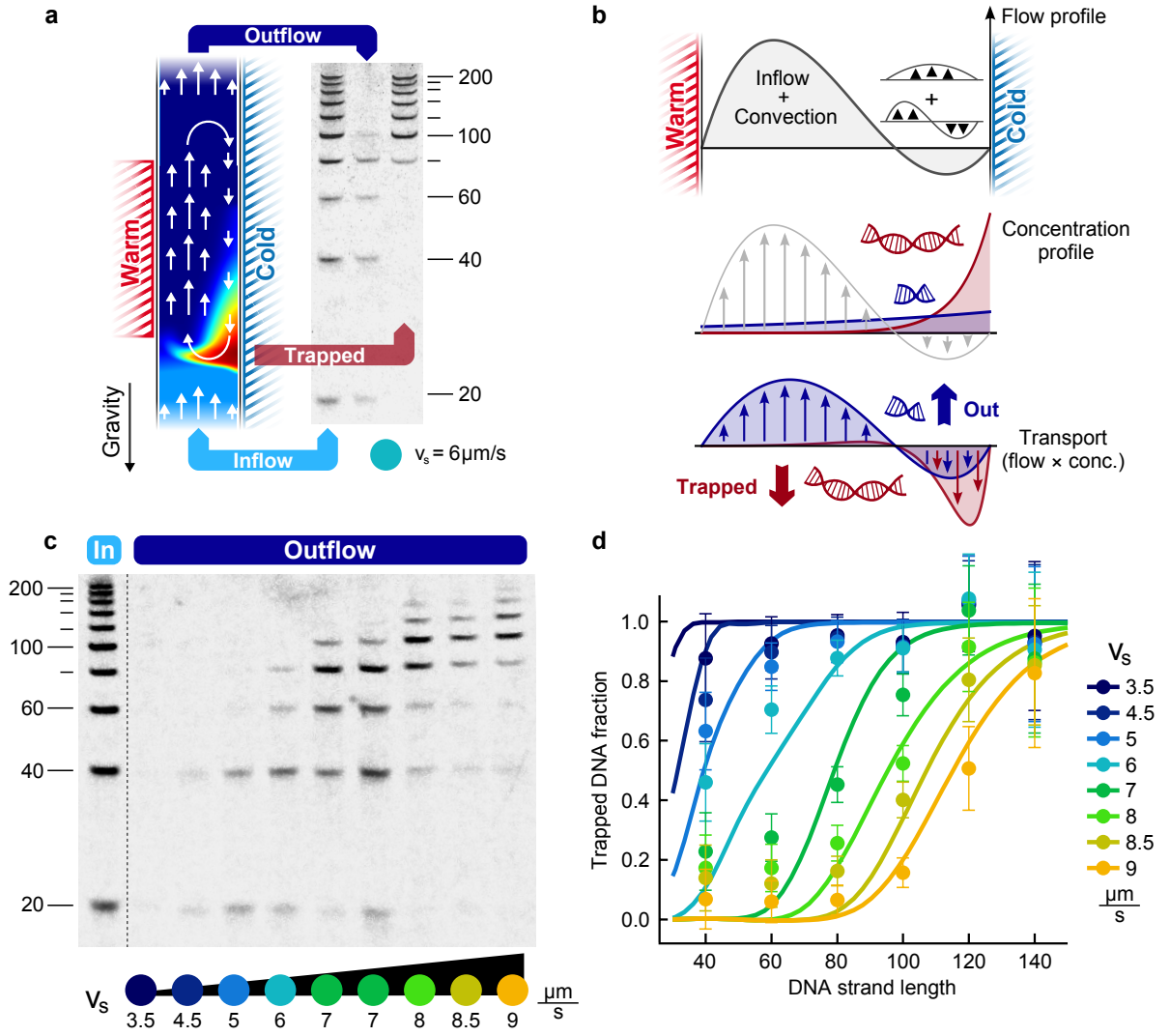


Figure 2.3: Heat-driven filter selecting for strand length. (a) A steady upward feeding flow is triggered by opening the asymmetrically heated pore. A ladder of double-stranded DNA (20-200 bp, 20 bp steps) was injected into the trap. Subsequent flushing of the capillary with pure buffer at a single velocity (6 µm/s) reveals the filter's thresholding characteristics: lengths ≤ 80 bp flow through the pore while longer strands get trapped. (b) An asymmetric flow pattern is generated by the superposition of the upward flow and the convection. Thermophoresis pushes the long strands into the downward flow and traps them. Short strands are subjected to the overall upward flow and leave the pore. The trapping is a function of the feeding flow speed. (c) The velocity of the external flow tunes the fractionation of nucleic acids. As in the experiment before, a DNA ladder was initially introduced at low flow velocity, which is then sequentially increased. The released DNA was measured using gel electrophoresis. (d) The fraction of trapped DNA obtained from the electrophoresis gel constitutes a selection landscape of this thermal habitat in favor of long oligonucleotides. The velocity dependent trapped fraction is described by the fluid-dynamic model (see Methods). Error bars reflect the signal-to-noise ratio of the gel images (cf. Supplementary Fig. 2.15 for details).

2.5 Differential survival of replicating strands

Combining all of the mechanisms, the joined thermally induced trapping and replication allows to overcome Spiegelman's evolutionary dilemma of the degeneration of strand length and therefore loss of genomic information.[62] For this reason, an experimental study followed the composition of a heterogeneous DNA population that continuously replicates inside the open pore. A 2.5 mm short capillary was seeded with a population of unlabeled template DNA strands with identical primer binding sites and a binary length distribution of 36 bp / 75 bp at a concentration of 1 nM each.

While applying a temperature gradient from 61 °C to 94 °C, a continuous upward flux of template-free PCR buffer containing nucleotides, polymerase, and 7 nM of fluorescently labeled primers was run through the system at a speed of 5.5 µm/s. Over the course of the experiment (7 hours), the trapping volume was exchanged approximately 150 times with the template-free feeding buffer. Aliquots containing the product of the continuously running reaction were taken from the outflow and analyzed using gel electrophoresis. Since the primers carried the labels, only replicated DNA strands were detected (Fig. 2.4c).

The experiments show that only the long strands were able to replicate sufficiently to withstand the diluting flow through the pore. This determined the increase of the relative concentration of the long, viable strands with respect to the total amount of DNA (Fig. 2.4d, yellow graph). The twofold shorter strands got diluted and became extinct. This competitive replication and selection of two genetic polymers in favor of larger molecular length can easily be understood by a simple model. The determinants of the growth kinetics $dc_i/dt = (rep_i - dil_i) \cdot c_i$ for either the short or the long species $i = \{S, L\}$ are given by the replication rates rep_i and the dilution rates dil_i. Expressing the relative concentration of the long strands yields $c_L/(c_S + c_L) = \left(1 + A \cdot e^{\Delta k \cdot t}\right)^{-1}$. $A = c_S^0/c_L^0$ is the initial concentration ratio, $\Delta k = (rep_L - rep_S) - (dil_L - dil_S)$ the differential growth rate. The experiments show that, inside the pore, long strands (L) outcompete shorter ones (S) with $\Delta k = 0.55\,\mathrm{h}^{-1}$ (yellow fit curve). The length selective fractionation model (Fig. 2.3c) confirmed that the shorter strands suffer from a fourfold higher dilution rate as compared to the trapped long strands. Notably, this selection of the longer replicating strand works best if the mechanism of replication is inefficient, such that the dilution of the short strand occurs before it can be replicated efficiently.

On the other hand, in a well mixed situation and hence in the absence of the selection pressure of the pore, experimental observations recovered Spiegelman's dilemma of the tyranny of the short. In a serial dilution experiment using a conventional thermal cycler with dilution rates reproducing the pore conditions, the long strands died out rapidly with a differential growth rate of Δk = -2.5 h^{-1} (blue fit curve).

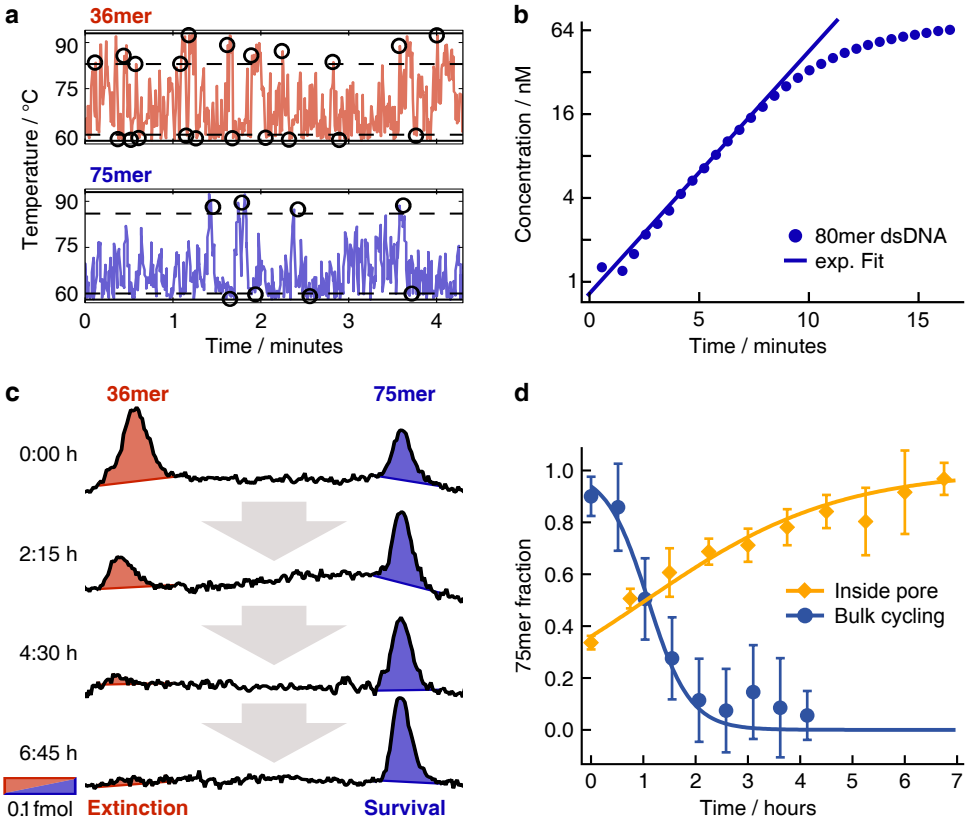


Figure 2.4: Selection of a replicating DNA population occupying the thermal habitat. (a) Strands are subjected to temperature oscillations by the combination of thermophoresis, convection, feeding flow, and diffusion. Simulations of stochastic molecule traces show that strands of 75 base pairs cycle inside the system for 18 minutes on average. In comparison, 36mers, due to their enhanced diffusion, show faster temperature cycles, but are flushed out of the system after 5 minutes. (b) Taq DNA polymerase assisted replication of 80mer dsDNA by convective temperature cycling. Quantitative SYBR Green I fluorescence measurements show exponential replication with a doubling time of 102 s. (c) An open pore (cf. Fig. 2.1c) was seeded with a binary population of nucleic acids. Quantitative gel electrophoresis revealed sustainable replication for only the long strand. Short strands got diluted and became extinct despite their faster replication. (d) Relative concentrations of the two competing species inside the thermal habitat. The selection pressure of the thermal gradient altered the composition of the binary population with time (yellow symbols) in good agreement with an analytical replication model. The absolute fitness values were 1.03 and 0.87 for long and short strands, respectively. Without the thermal gradient, the short oligonucleotides won over the long strands (blue symbols), analogous to the Spiegelman experiment. Error bars reflect the signal-to-noise ratio of the gel images (cf. Supplementary Fig. 2.15 for details).

2.6 Discussion

The experimental findings conclusively show that, at the expense of dissipating free thermal energy, a habitat is created that drives and sustains the replication of long oligonucleotides by exploiting both, convective temperature cycling and a selection pressure that supports long over short sequences. Therefore, heat dissipation allows the pore to overcome Spiegelman's classic problem for *in vitro* replication systems to create ever shorter genetic polymers, resulting in the loss of genetic information.

On the early hot Earth, the pore system described here was likely widespread because of porous, partially metallic volcanic rock, both near the surface and at submarine sites. As metals have a more than 100-fold larger thermal conductivity than water, metallic inhomogeneities near the pores can focus the thermal gradient from centimeters down to a micrometer-sized cleft (Supplementary Fig. 2.5). The kinetics of replication and selection were realized in the most simple geometrical setting of a single pore section with dimensions of 0.07 mm x 3.5 mm. Metallic inclusions do allow to focus thermal gradients up to 100-fold to reach the thermal gradients from realistic geological settings (Supplementary Fig. 2.5). It is however important to note that the steepness of the thermal gradient can be further relaxed by at least one order of magnitude by separating replication and selection into two adjacent pores (Supplementary Fig. 2.6). At the bottom, a wide pore could provide the necessary temperature difference for replication.[69] On its top, the outflow would be constricted through one or more thin, but longer selecting pores. Their increased length of several centimeters instead of 3.5 mm compensates for the reduced temperature difference linearly.[40]

While the demonstrated length selective trapping requires a temperature difference to work, the average temperature of the trap is not a critical parameter and can be tuned easily to fit the replication reaction. Therefore, the core mechanism of temperature cycling and selection studied here will also work for replication systems that require colder temperatures, including for example ribozymes or Q-beta replicase. However, many early replication systems are likely to rely on high temperatures for temperature induced strand separation.

For the PCR reaction used in the experiment, the strand lengths are highly controlled by the primers. In comparison, reactions involving ligations have a tendency to extend the strands with partial templating and initiate the length extension of the genetic polymers.[70] To extend this work in order to achieve Darwinian evolution in the demonstrated system, the replication process would require a significant mutation rate, including changes of the sequence length. The use of error-prone PCR with deep sequencing is therefore an interesting prospect for future experiments. At this point, the amount of less than 1 pg inside the pore is preventing such an approach: The necessary strong pre-amplification would highly bias the obtained sequences and obscure their analysis. Importantly, the thermophoretic selection pressure applies to each individual molecule of the population. Since it is ultimately sensitive to the thermophoretic strength, the selection does not only favor the survival of long strands over short strands. It is possible that the found mechanism could even be tuned to select for the formation of macromolecular complexes or even for binding of aptamers.[71]

2.7 Conclusion

The experiments reveal how temperature gradients, the most simple out-of-equilibrium setting, can give rise to local environments that stabilize molecular replication against entropic tendencies of dilution, degradation, and negative length selection. A thermal gradient drives replication oligonucleotides with an inherent directional selection of long over short sequence lengths. Interestingly, when the replication and trapping inside the pore reach their steady state, the newly replicated molecules will leave the trap with the feeding flow. This ensures an efficient transfer of the genetic polymers to neighboring pore systems. Heat dissipation across porous rock was likely placed in close proximity to other non-equilibrium settings of pH, UV, and electrical potential gradients, all of which are able to drive upstream synthesis reactions producing molecular building blocks. An exciting prospect of the presented experiments is the possible addition of mutation processes in order to achieve sustained Darwinian evolution of the molecular population inside the thermal gradients of the early Earth. Accordingly, the onset of molecular evolution could have been facilitated by the natural thermal selection of rare, long nucleic acids in this geologically ubiquitous non-equilibrium environment.

2.8 Materials and methods

2.8.1 Temperature gradients

Temperature gradients were generated across rectangular borosilicate glass capillaries (VitroTubes, VitroCom, USA) with a cross-sectional aspect ratio of 1:20, and a thermal conductivity of 1.2 W/m·K. To this end, two different approaches were followed. 1.) For direct observation of the accumulation effect, glass capillaries were coated with a transparent conducting oxide layer that allowed for one-sided heating at constant electric power while cooling from the other side. 2.) Fractionation and replication experiments were performed in capillaries sandwiched between, and thermally connected to temperature controlled metal surfaces.

2.8.2 Accumulation-only experiments

Double-stranded DNA was diluted in 1× Taq reaction buffer (New England Biolabs, Germany) containing 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 % Tween20, with a pH of 8.3 at room temperature. 10 µg/ml dsDNA ladder (100–1000 bp, 10 equidistant bands, weight equalized) was used in combination with 0.5× SYBR Green I. The applied temperature gradient from 22 °C to 88 °C resulted in temperatures from 38 °C to 71 °C inside the capillary (inner dimensions: 100 µm × 2000 µm and 70 µm × 1400 µm, as specified).

2.8.3 Fractionation experiments

A DNA ladder (20 bp–200 bp, 10 equidistant bands) was suspended in 1× PCR buffer including 0.1 % Tween20. Fractionation was carried out in a vertically oriented capillary (inner dimensions: 70 µm × 1400 µm) with an internal temperature gradient from 39 °C to 73 °C present over a capillary length of 3.5 mm (see SI for details). Threshold trapping characterization was determined using a constant flow speed. Gradual fractionation was achieved by increasing the flow rate with time using a feedback controlled syringe pump (neMESYS, Cetoni, Germany, see SI for detailed protocol).

2.8.4 *In vitro* selection and replication

Extracellular selection of replicating DNA strands was studied in a temperature gradient from 61 °C to 94 °C inside a thoroughly cleaned (DNA Away, Molecular BioProducts, USA) capillary (inner dimensions: 70 µm × 1400 µm) at a mean solvent velocity of 6 µm/s. DNA replication was facilitated in a commercially available, glycerol-free master mix (fast cycling PCR Kit, Qiagen, Germany) containing Taq polymerase, free nucleotides and standard concentrations of mono- and bivalent salts. The overall efficiency of DNA replication was reduced to less than 8 % by means of low concentration (7 nM) of each 14mer primer (primers FWD-Cy5 and RV, see Supplementary Fig. 2.9 for sequences) in the feeding buffer. Unlabeled DNA templates (36mer, 75mer) were seeded into the region of replication through the system's output, leaving the feeding buffer template free. Reaction products containing the incorporated Cy5 primer from the feeding buffer were extracted from the output of the artificial pore in 1.5 µl aliquots. Controls were performed in a conventional real-time PCR cycler (CFX96, Bio-Rad, Germany). A serial dilution experiment was performed to derive the replication efficiencies of the 36mer

and 75mer DNA. Temperature cycles emulated the mean temperature cycle of 75mer DNA inside the pore, consisting of 3 s at 94 °C and 14 s at 60 °C (Supplementary Fig. 2.13). Including transition times, the total cycle time was 46.5 s. Initial concentrations were 2 pM (36mer) and 18 pM (75mer) for the PCR templates and 7 nM for the common primers. Every 40 cycles, the sample was diluted by a factor of 20, yielding a dilution rate of $dil_S = dil_L = 5.8\,\mathrm{h}^{-1}$ that counterbalanced the concentration increase of the 36mer DNA within 40 cycles. This scheme prevented a depletion of the primer concentration and ensured that the efficiencies of the PCR reaction stayed constant over all 320 cycles. Replication rates were determined by comparing the amount of DNA before each dilution using gel electrophoresis. Mean replication rates were determined to be $rep_S = (5.8 \pm 0.6)\,\mathrm{h}^{-1}$ (36mer) and $rep_L = (3.3 \pm 0.4)\,\mathrm{h}^{-1}$ (75mer).

2.9 Supplementary material

2.9.1 Temperature gradients i) Ohmic heating with a transparent electrode

Glass capillaries were plasma cleaned and coated with a thin layer of indium tin oxide (ITO, see Supplementary Fig. 2.7) in a radio frequency sputtering chamber (LS320, Von Ardenne, Germany), equipped with a custom-built translation stage. Sputtering under an argon atmosphere for 40 minutes at 30 W and subsequent heat treatment for 30 minutes at 250 °C resulted in a typical sheet resistance of 12 Ohms per square and high optical quality. The one-side coated capillaries were glued onto a commercially available water CPU cooler (Innovatek, Germany) using a thin film of silver-filled thermally conducting epoxy (Arctic Silver, Arctic Silver, USA). The ITO-coated side of the capillary was electrically connected to a digitally controlled power supply (6010A, Agilent, USA) via copper wires and conductive paint (Busch5900, Busch, Germany). Temperature gradients across the capillary were established through electric heating at constant power, with the cooled side of the capillary being controlled by a water bath (F31-C, Julabo, USA) operating at constant temperature. The thermal response of the capillary was calibrated with a thermochromic dye (70C black, Sintal Chemie, Germany) that was put on top of the capillary.

2.9.2 Temperature gradients ii) Ohmic heating at monitored temperatures

Glass capillaries were sandwiched between plan sapphire windows (thickness: 100 µm, Sappro, Germany) on copper substrates using a thermally coupling adhesive (TC-2707, 3M, USA, cf. Supplementary Fig. 2.8). This ensured accurate temperature conditions in experiments requiring intra-capillary temperatures as high as 90 °C to 100 °C. Heating was achieved by an Ohmic resistor connected to a computer controlled power supply. A Peltier element on a water-based CPU cooler was used for cooling. During the experiments, temperatures were measured on both copper surfaces by thermocouples in a LabVIEW-based computer environment, and stabilized by a PID-controlled feedback loop acting on the cooling side (±50 mK), resulting in stable temperature conditions also on the heated side (±1 K long term drift). Heating was applied along a 2.5–3.5 mm long section of the capillary.

2.9.3 Microfluidics

Glass capillaries were connected to a feedback controlled syringe pump (neMESYS, Cetoni, Germany), via high purity PFA tubing (HPFA+, Upchurch Scientific, USA), and tightly matched silicon seals. Microfluidic distances between the heated region of the capillary and its accessible output measured 3 µl to 5 µl, and were determined with a precision higher than 0.2 µl prior to fractionation experiments and the seeding of the pore in the selection and replication experiment. Degassing of this microfluidic system was done by flushing isopropyl alcohol followed by degassed PCR reaction buffer (Standard Taq Reaction Buffer, New England Biolabs, Germany) using an overpressure of several bars. Crucially, all assays also had to be degassed carefully prior to loading into the system in order to avoid air bubble formation during the experiments. This was achieved by heating 200 µl sample tubes to 88 °C. After one minute, a mechanical shock was applied to induce to the formation of gas bubbles. Consequently, the temperature was kept at 88 °C for five minutes, followed by an increase to 94 °C for four minutes. Finally, gas bubbles were released from the tube walls by vortexing for three seconds. In order to avoid re-saturation of the samples with oxygen from the air, tubes were maintained at 90 °C during injection of the assay into the system.

2.9.4 Imaging of fluorescently labelled oligonucleotides

Fluorescent imaging of DNA was realized with a 90°-tilted upright microscope (Axioscope A1, Zeiss), using a 2.5× objective (Plan-Neofluar 2.5× 0.075 NA, Zeiss, Germany), equipped with a CCD camera (1400, PCO, Germany) and two alternating light-emitting diodes (LED 470 nm, LED 625 nm, Thorlabs, USA) in combination with a dual band filter set (Dual band FITC / Cy5, AHF, Germany). Gel electrophoresis and documentation Native gel electrophoresis was performed in 12.5 % polyacrylamide gels inside a 1× TBE buffer at electric field strengths of 60 V/cm and 30 °C for 13 minutes. After running, the gels were stained by incubation in fresh 1× solutions of SYBR Green I (Invitrogen, Germany) in TBE buffer for four minutes followed by a one minute washing step in pure TBE. Imaging of SYBR Green I stained gels was done by CCD photography through a green bandpass filter (520 nm, 10 nm FWHM, Newport, Germany) under spectrally filtered (470 nm, 10 nm FWHM, Thorlabs, USA) light emitting diode excitation (LED 470 nm, Thorlabs, Germany). Denaturing gel electrophoresis was performed after a standard protocol.[72] In short, DNA samples were denatured in a formamide glycerol buffer at 95 °C for 2 minutes, followed by shock cooling on ice. Then, the samples were loaded into 12.5 % polyacrylamide gels containing 50 % urea. After a 5 minutes pre-run at 7.5 V/cm, the samples were separated by an electric field of 60 V/cm in 45 °C – 50 °C warm TBE buffer for 13 minutes. For analysis, the gels containing the Cy5 labeled reaction products from the extra-cellular selection and replication experiment were illuminated with two spectrally filtered light emitting diodes (LED 625 nm, filter 630 nm, 10 nm FWHM, Thorlabs, Germany). Detection was done through a pair of high quality interference filters (bandpass 692 ± 20 nm, OD6 blocking, Edmund Optics, USA, and bandpass 700 ± 35 nm, OD 2 blocking, Newport, USA, resulting in an excitation rejection of OD8+) by an actively cooled CCD camera (Orca 03-G, Hamamatsu, Japan).

2.9.5 Fractionation experiments

A weight equalised double stranded DNA ladder (10 equidistant bands, 20–200 bp, Carl Roth, Germany) was separated from its loading dye by ethanol precipitation and resuspended at a final concentration of 0.25 µg/µl in 1× PCR buffer containing 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.1 % Tween20 with a pH of 8.3 at room temperature. After degassing the DNA-free microfluidic system, the DNA ladder was sucked into a reservoir with an inlet just before the region of the temperature gradient (cf. Supplementary Fig. 2.9). After flushing the main channel with DNA free buffer again, a 1.5 µl pulse of the DNA ladder was injected into the trapping region, followed by the constant flow of pure buffer driving the fractionation. Experiments were carried out in a capillary (internal dimensions: 70 µm × 1400 µm) with a heated region of 3.5 mm length and a temperature gradient ranging from 39 °C to 73 °C. A fractionation run with higher terminal velocities than shown in Fig. 2.3c is depicted in Supplementary Fig. 2.10c.

2.9.6 Quantitative gel analysis

Gel image quantification (as shown in Supplementary Fig. 2.15) was done with a custom LabVIEW program, after point-like outliers have been removed using NIH ImageJ. Before integrating the intensities, images were corrected for inhomogeneous illumination. To improve the signal to noise ratio, the intensity of each gel lane was then integrated along the horizontal axis. Further, a local linear background was subtracted from each gel band (shaded regions in Supplementary Fig. 2.15d). The uncertainty of this integral was estimated using the standard deviation of the values around the base points of the linear background.

2.9.7 Replication model

The amplification of a target DNA sequence using PCR is described by $c(n) = c_0 \cdot (1+E)^n$, where n denotes the number of cycles and E the PCR efficiency. Under the replication conditions in the experiment, molecules are also subjected to a continuous outflow. The latter is modeled by a dilution rate per unit time dil. In a continuous time description, the replication rate per unit time is $rep = ln\,(1+E)\,/(t_c)$ with the temperature cycle time t_c of the PCR reaction. This leads to a combined growth equation for each species i given by $dc_i/dt = c_i \cdot (rep_i - dil_i)$. Its solution is

$$c(t) = c_i^0 \cdot e^{(rep_i - dil_i)\cdot t} \tag{2.1}$$

and the relative concentration of 75mer DNA is given by

$$\frac{c_L}{(c_S + c_L)} = \frac{c_L^0 \cdot e^{k_L \cdot t}}{\left(c_L^0 \cdot e^{k_L \cdot t} + c_S^0 \cdot e^{k_S \cdot t}\right)} \tag{2.2}$$

with $k_i = (rep_i - dil_i)$. Defining $A = c_S^0/c_L^0$ as the ratio of the initial concentrations of short versus long strands, and the differential growth rate $\Delta k = (rep_S - rep_L) - (dil_S - dil_L)$, this can be simplified to

$$\frac{c_L}{(c_S + c_L)} = \frac{1}{\left(1 + A \cdot e^{\Delta k \cdot t}\right)} \tag{2.3}$$

Taking into account separately determined parameters for temperature gradients, Soret coefficients, diffusion coefficients, and inflow velocities, the fluid-dynamic model yields dilution rates of $dil_L = 3.2\,\mathrm{h}^{-1}$ and $dil_S = 12.5\,\mathrm{h}^{-1}$ for the selection and replication experiment. Replication rates were experimentally determined in a PCR cycler set up to match temperature cycling rates inside the pore (compare Supplementary Fig. 2.13 for their determination). Replication rates of $rep_L = (3.3 \pm 0.4)\,\mathrm{h}^{-1}$ for the long strands and $rep_L = (12.05 \pm 0.06)\,\mathrm{h}^{-1}$ for the short strands were found. Interestingly, in the absence of dilution, the short strand replicates 3.7 times faster inside the pore than the longer strand, partly reflecting its a priori evolutionary advantage of a higher replication efficiency E (as described by Spiegelman). In addition, the short strand experiences faster temperature cycles (19 s) than the long strand (44 s) due to its higher diffusive mobility.

The absolute fitness for a specific genotype is defined by the ratio of individual (strands) before and after selection. The fitness is a binary distribution (zero or one) when evaluated on time scale that the flow needs to run through the pore. Long strands exhibit a lifetime of $\tau = 18.8\,\mathrm{min}$ in the pore, defined by on/off-rate decay of $c/c_0 = e^{-1}$. The population of long strand growths around 3 %, and the population of short strands shrinks by 13 %, leading to absolute fitness values of 1.03 and 0.87 for the long and short strands respectively. If the time axis is scaled by the lifetime of the short strands, these numbers would read 1.01 and 0.96, respectively.

Taken together, the higher replication rate of the short strand is overcompensated by the length selective dilution. Using the here determined rates in the exponential growth model yields relative growths of $c_L(7\mathrm{h})/c_L(0) \approx 2.0$ and $c_S(7\mathrm{h})/c_S(0) \approx 0.04$ for short and long strands, respectively. This is in good agreement with the experimental results of the selection and replication experiment of $c_L(7\mathrm{h})/c_L(0) = 1.7 \pm 0.3$ and $c_S(7\mathrm{h})/c_S(0) = 0.1 \pm 0.1$.

2.9.8 Diffusion and screening length parameters of DNA

The Debye length is a major determining parameter for the strength of thermophoresis. It was estimated as $\lambda_{\mathrm{DH}} = 1.30\,\mathrm{nm}$ for the used PCR buffer (10 mM TRIS, 50 mM KCl, 1.5 mM $MgCl_2$) at 75 °C. Measurements of DNA at this Debye length were interpolated from the previously measured two-dimensional data set. As seen in Supplementary Fig. 2.11a, the Soret coefficients of single and double stranded DNA do not significantly differ. Interestingly, RNA shows very similar thermophoretic properties. The measured values of S_{T} were used and interpolated with a square root function, resulting in $S_{\mathrm{T}} = -0.0063 + 0.0115 \cdot (\#bases)^{0.5}$. Here, $\#bases$ is the number of bases on a single strand. A very similar approach of fitting was used previously. To infer the diffusion coefficient from the same measurements, the radius of dsDNA was fitted from the same data set as plotted in Supplementary Fig. 2.11b. The radius shows good agreement with a line fit according to $R = (0.8 + \#bases \cdot 0.059)\,\mathrm{nm}$. Based on the same measurements, the hydrodynamic radius is then translated into the diffusion coefficient by the Einstein relation with the viscosity taken at the temperature of 75 °C, resulting in a diffusion coefficient for the COMSOL simulation given by $D = 6.69 \cdot 10^{-19}/(8 \cdot 10^{-10} + (\#bases) \cdot 5.9 \cdot 10^{-11})\,\mathrm{m}^2/\mathrm{s}$.

2.9.9 Computer simulations and analytical trapping model

Non-stochastic computer simulations of the nucleotide accumulation in a thermal gradient were performed in COMSOL Multiphysics, similar to simulations published before. Additionally, a custom computer program was used to access stochastic information of particle motion which is relevant to calculate cycling rates. Here, individual particles were traced on a biased random walk trajectory inside the combined temperature and velocity fields. As a basis for this stochastic simulation, COMSOL provided the temperature field (conductive heat transfer module) and the velocity field (incompressible Navier-Stokes module).

2.9.10 Calculation of temperature cycling times

Typical results of these stochastic particle tracing simulations are visualised in Supplementary Fig. 2.12, showing individual trajectories for 36mer and 75mer DNA in the trapping geometry relevant to the selection and replication experiment of Fig. 2.4d. Using these trajectories, statistical data on thermal cycles and particle lifetimes can be obtained. For each DNA species, a temperature cycle is defined using two threshold temperatures: an annealing temperature T_A and a denaturation temperature T_D. A temperature cycle is defined as the time it requires for a particle to go from T_A to T_D and back. Supplementary Fig. 2.13 shows length resolved DNA strand trajectories and cycling parameter extracted from it. The thermal cycling in the pore is comparable to that of a standard PCR protocol with short denaturation times and a longer annealing/elongation step. Compared to the 75mer DNA, the 36mer cycles faster between the warm and cold sides of the pore, which is due to its higher diffusion and lower Soret coefficients. Thermal cycling statistics for the two DNA species and different influx velocities is presented in Supplementary Fig. 2.14. Notably, the cycling time depends only weakly on the influx velocity, whereas the total number of cycles is determined by time the particles residence time inside the pore before being flushed out.

2.9.11 Modelling of fractionation experiments

Simulations of the fractionation of a DNA ladder, as presented in Fig. 2.3, required multiple steps. First, simulations derived the length-dependent propagation and trapping of a mixed-length DNA pulse through the trap for different influx velocities. This data was then used to calculated the length distribution of strands that have been flushed out after a given time. The time steps were chosen to match the increases of inflow velocities presented in Supplementary Fig. 2.10a,b. For each velocity step, the concentration of the DNA leaving the trap was normalized to the step duration.

2.10 Supplementary figures

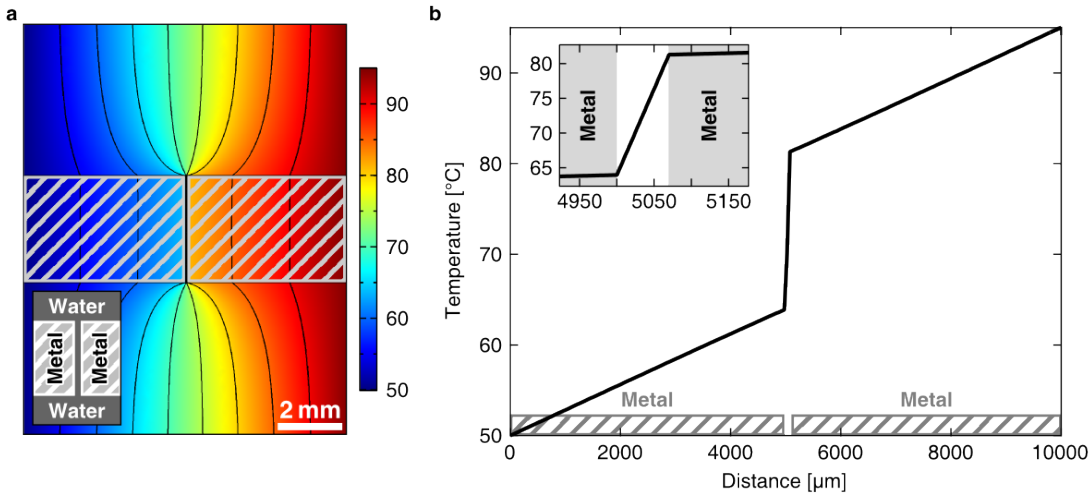


Figure 2.5: Focussing of a temperature gradient in a millimeter-sized pore. (a) Metal inclusions (grey) focus a temperature gradient across a micrometer-sized pore. As metals show about 100-fold higher thermal conductivities than water, the temperature gradient is strongly focused to the 70 µm gap between the two inclusions. (b) Horizontal cut along the metal inclusions in panel a. Inside the gap, the temperature gradient is increased to 250 K/mm, compared to 4.5 K/mm in the bulk water. For the calculation, thermal conductivities of 0.58 W/m·K for water and 50 W/m·K for the metal inclusion have been used. The latter is on the lower end of the range for metals, with copper having the highest (400 W/m·K).

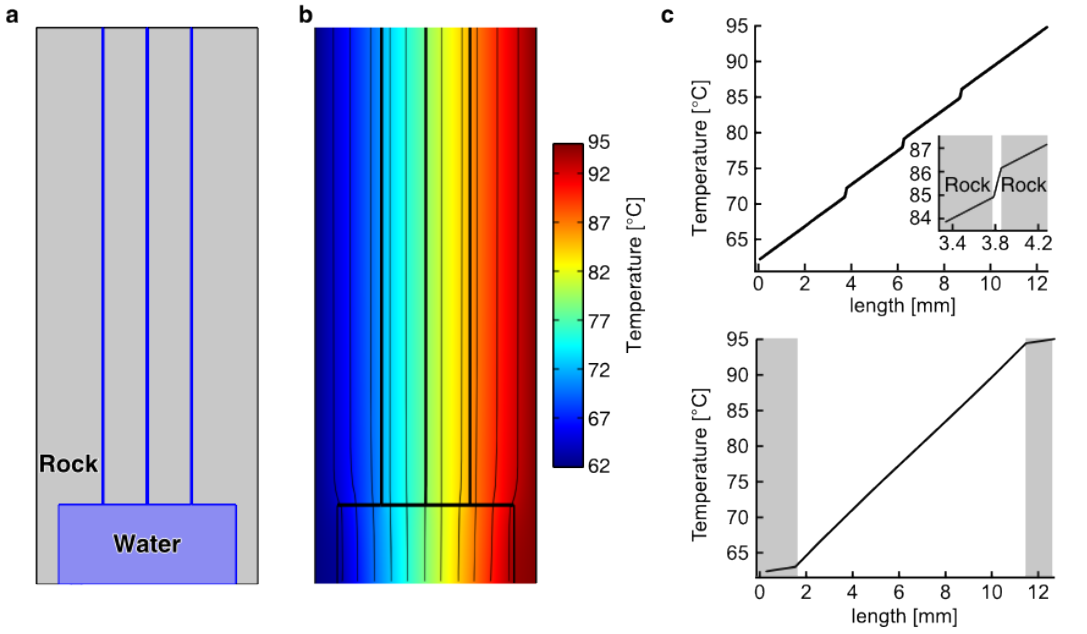


Figure 2.6: Separation of accumulation and thermal cycling in multipore system. (a) A multipore system combining a large 1 cm sized pore with multiple 70 µm small sized pores. The small pores with a length of 3 cm are able to accumulate DNA 103-fold and the accumulation grows exponentially with the length of the pore system. The large pore exhibits a strong convection flow which shuttles the molecules between warm and cold with velocities up to 4.8 cm/s. (b) By assuming a thermal conductivity of 5 W/m K for porous rocks, a temperature gradient of 31 K and 1.2 K is formed along the large and small pore, respectively.

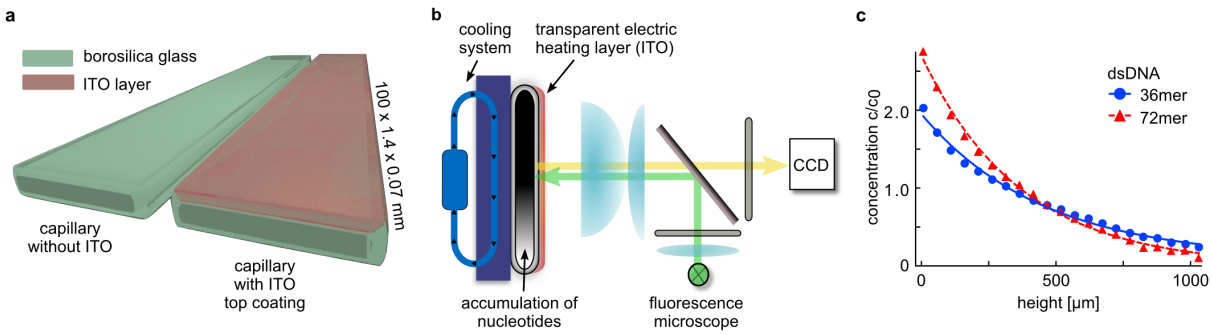


Figure 2.7: Experimental set-up to generate temperature gradients across rectangular borosilicate capillaries by electrically heating a transparent electrode. (a) Schematics of the capillary before and after coating with a transparent layer of indium tin oxide (ITO, here red). (b) Capillary thermally connected to a vertically oriented heat sink (blue). The electrically generated heat from the ITO layer (red) flows through the capillary, giving rise to the accumulation of nucleotides at its bottom. Time-resolved accumulation is recorded by a standard wide-field fluorescent microscope. (c) Height-resolved concentration profiles of FAM-labeled dsDNA templates 36mer and 72mer reveal exponential accumulation characteristics with a stronger spatial confinement of longer oligonucleotides. The experiment was carried out in a capillary with internal dimensions of 70 µm × 1400 µm and a temperature gradient ranging from 23 °C to 58 °C at the outer walls and 31 °C to 50 °C at the inner walls of the capillary.

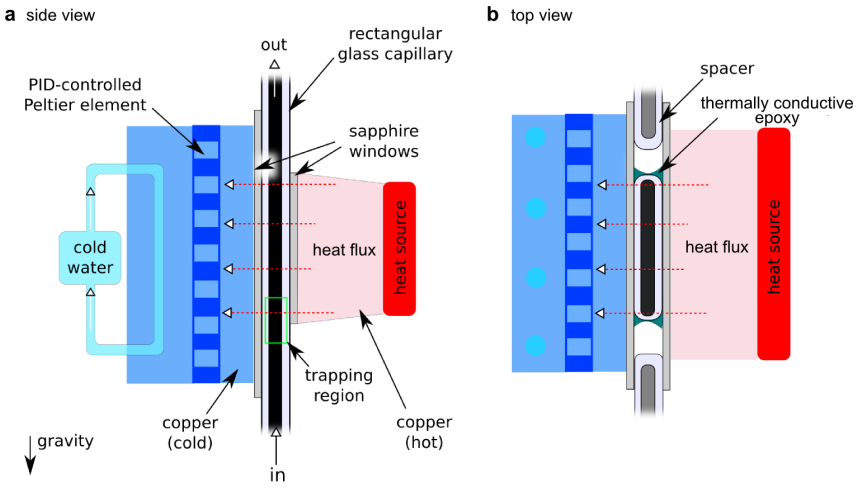


Figure 2.8: Cross-sectional drawings of flow-through set-up employed for fractionation and selection & replication experiments, side and top view. A rectangular glass capillary is sandwiched between two sapphire windows. Thermal coupling is achieved by thermally conductive epoxy adhesives. The heat created by a constantly powered resistor on the red side flows through the capillary into a PID-regulated heat sink. Spacer capillaries ensure a homogeneous temperature gradient.

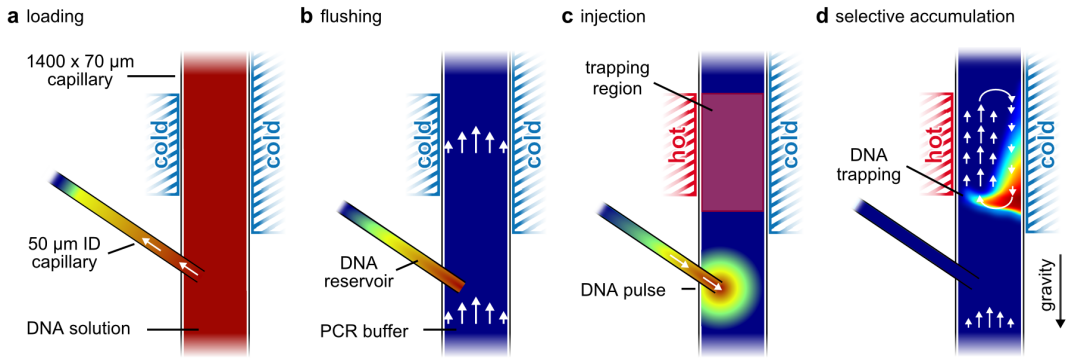


Figure 2.9: Injection of a DNA pulse for the fractionation experiments. (a) Concentrated DNA was pulled from the main channel into a 50 µm capillary, serving as a DNA reservoir. (b) The DNA solution inside the main capillary is replaced by PCR buffer solution. (c) Injection of a 1.5 µl pulse of DNA below the trapping region after applying a temperature gradient. (d) External inflow exerts selection pressure on oligonucleotides.

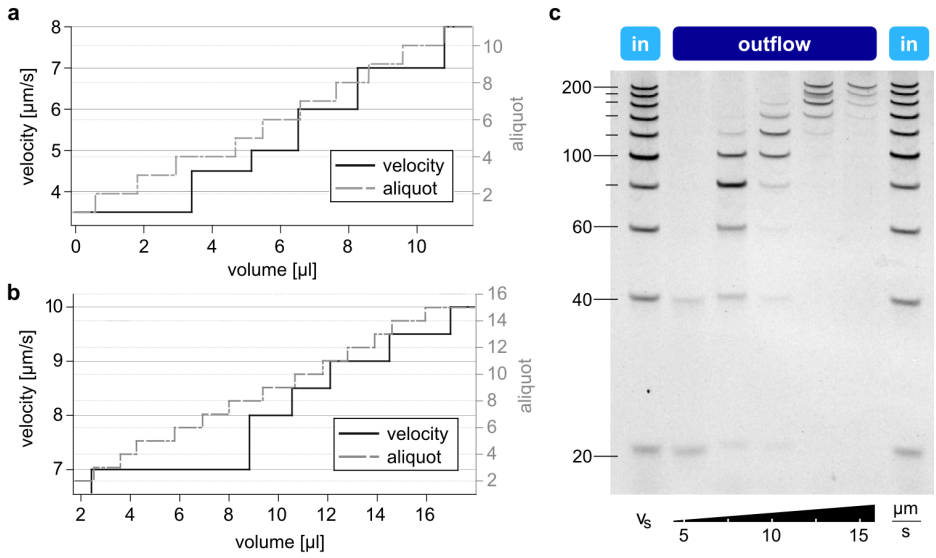
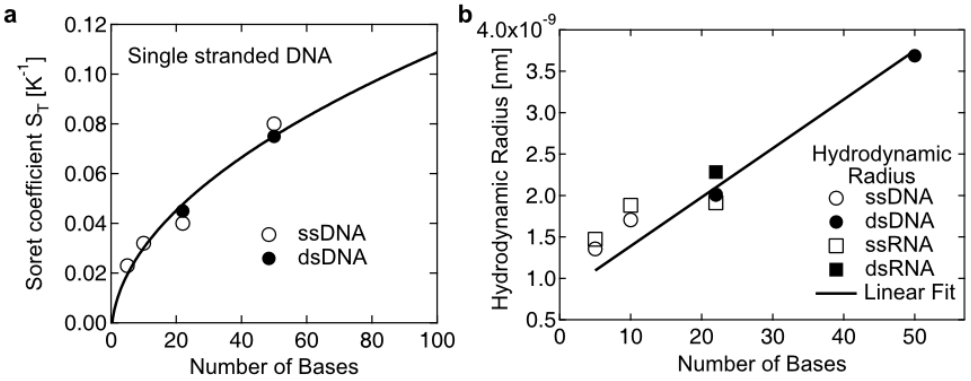


Figure 2.10: Determination of critical trapping velocities for each 20 bp ladder fragment. Fractionated DNA was obtained from the output of the microfluidic system in volumes between 0.9-1.5 µl and assigned to the step-wise increased velocities. (a) The influx was gradually increased from 3.5 nl/s to 10 nl/s. At an inflow of 3.5 µm/s, 4.5 µm/s, 5 µm/s, 6 µm/s and 7 µm/s, DNA fragments with a length of 20 bp, 40 bp, 60 bp, 80 bp and 100 bp started to be flushed out of the trapping region, respectively. (b) At an inflow of 7 µm/s, 8 µm/s, 8.5 µm/s and 9 µm/s, DNA fragments with a length of 100 bp, 120 bp, 140 bp and 160 bp were flushed out, respectively. (c) Gel electrophoresis of thermally fractionated double stranded DNA ladder (20-200 bp) under same conditions as data presented in main text (Supplementary Fig. 2.3c), but at coarser velocity steps and with a wider range of inflow velocities v_s.



c

Name	Length	DNA Sequence 5'- to 3'-end
72mer	72bp	GGCAATCCATAAAACCCCAACCCCAACCCCAACCCCAA CCCCAACCCCAACCCCAACCCCGATTTGACACGA
36mer	36bp	TCATCCATAGTTGCACGAGAGCACTTTATCCGCCTC
75mer	75bp	TCATCCATAGTTGCGAGTGTAGATACGAGGGCTTACCA GCGAGACCCCAGTAACCAAGTCTACTTTATCCGCCTC
forward primer	14bp	TCATCCATAGTTGC
reverse primer	14bp	GAGGCGGATAAAGT

Figure 2.11: Analytic fit functions for the Soret and diffusion coefficients of DNA and RNA based on published measurements of short DNA. (a) The fit of the Soret coefficient is used as input parameter for the fluid dynamic computer simulation of the thermal trap. (b) Hydrodynamic radius of single and double-stranded DNA/RNA. (c) Sequences of 72mer, 36mer, 75mer, and forward and reverse primers.

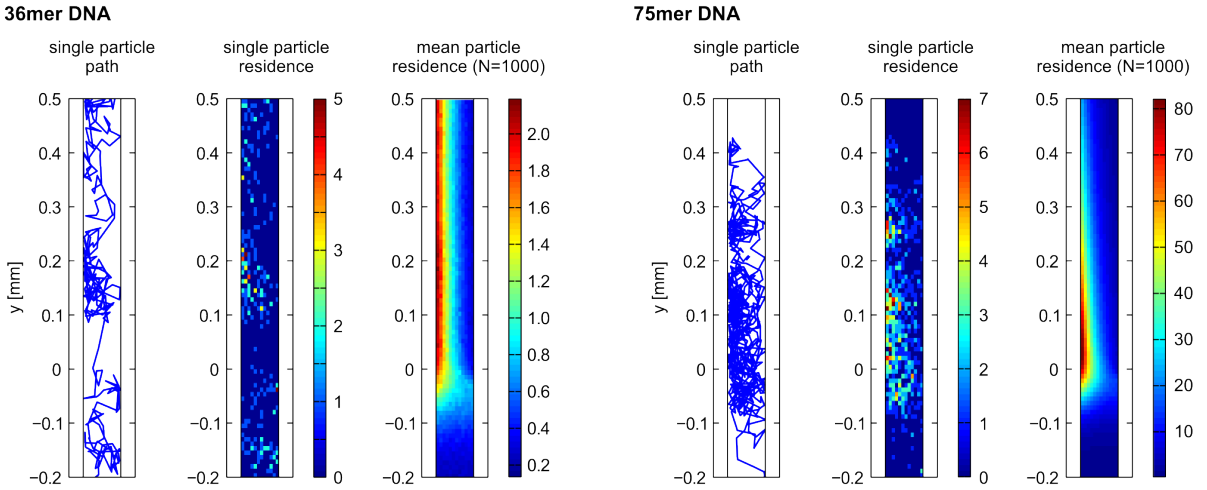


Figure 2.12: Visualization of random walk simulations of 36mer and 75mer DNA inside a 70 µm wide, asymmetrically heated pore. A mean influx velocity of 4 µm/s is applied from below. Individual panels show a single particle trajectory (left), the corresponding single particle density function (middle), and the mean concentration profile from 1000 independent simulations (right). The trajectories cover the first 500 s of the simulations. Note the different color scales for the short and the long strand. For the depicted influx velocity, the longer DNA species is trapped, while the shorter strand is not.

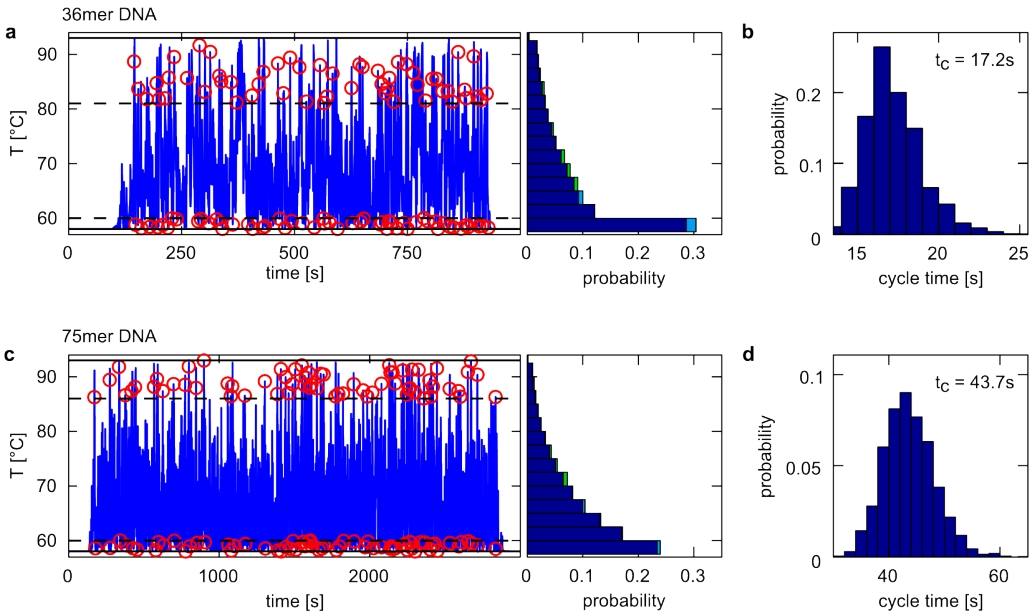


Figure 2.13: Thermal cycle times of individual nucleotide particles inferred from trajectories. The underlying simulation has been performed for N = 1000 particles at an influx velocity of 5.5 µm/s. (a,c) Temperature experienced during the lifetime of an individual particle. The dashed lines indicate the temperatures T_A and T_D. Red circles mark all entries into the regions colder than T_A and hotter than T_D, respectively. The temperature histograms summarize the time-resolved data of the full set of particles. Light blue (green) areas indicate negative (positive) deviations of the histogram for the single particle from that of the full set. (b,d) Cycle time histograms. The mean cycle time is indicated in the top right corner.

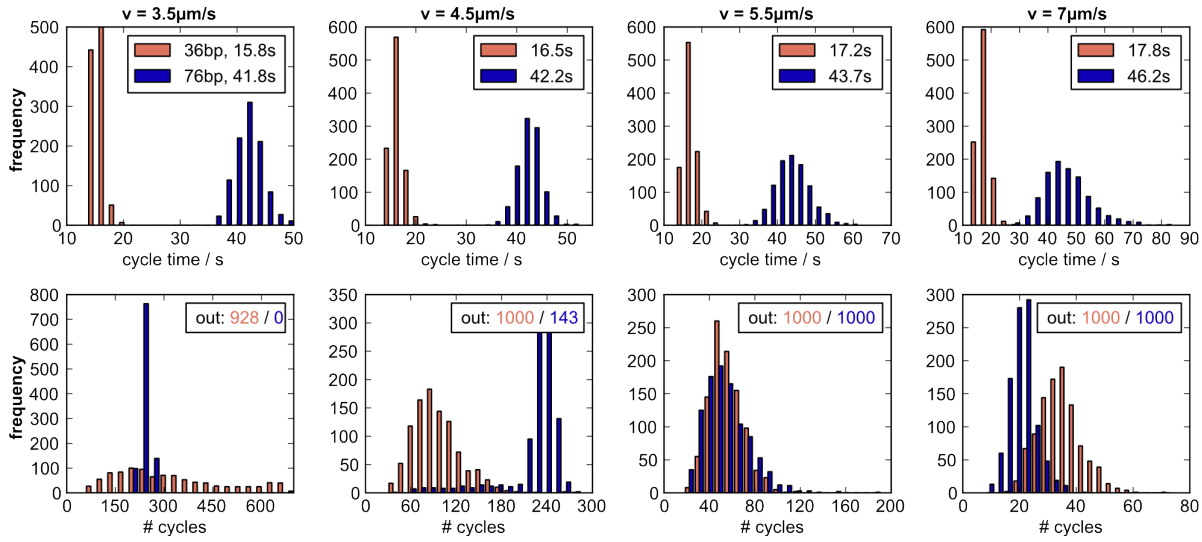


Figure 2.14: Temperature cycle statistics obtained from random walk simulations. N = 1000 particles have been simulated for 10^5 s. Top row: Particle cycle time distributions. The cycling times depend only weakly on the inflow rate. In all cases, short strands (red) cycle faster than long strands (blue) due to their higher mobility and lower affinity to the cold side. Bottom row: Particle cycle counts. The numbers in the top right of the diagrams indicate the number of particles that left the simulation volume, color coded for the oligonucleotide length. For low flow rates through the system, the higher cycling frequencies of short strands lead to higher numbers of temperature cycles before a strands leave the system. In the intermediate velocity range, long strands benefit from their significantly higher abidance time, allowing them to cycle more often than the short ones.

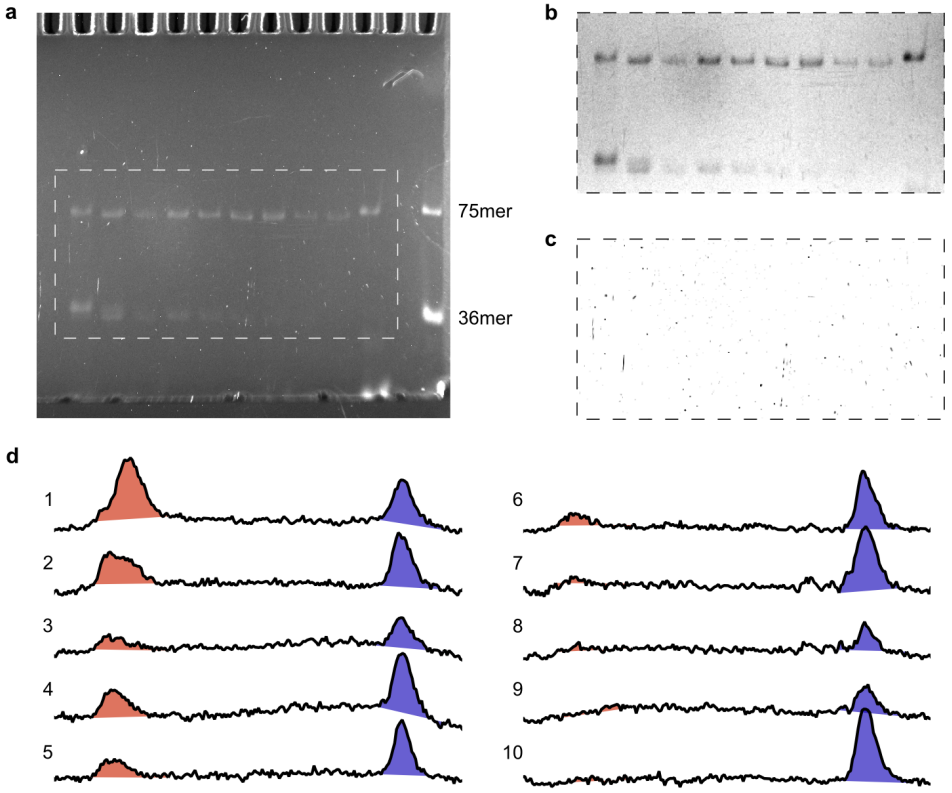


Figure 2.15: Quantification of native polyacrylamide gel data from the selection and replication experiment. (a) Raw gel image and selection of the region of interest. (b) Outliers which are due to residual dust grains are visible due to the long exposure time. They were removed before further analysis. (c) Removed outliers. (d) Quantitative data of all 10 gel lanes. The integrated intensities of the 36mer (red, left) and 75mer DNA (blue, right) were evaluated as indicated. The error bars in Fig. 2.3d,2.4d were calculated from the contribution of noise at the end points of the local baselines to the integrated intensities.

Chapter 3

A thermal habitat drives the retention and ribozymatic replication of RNA

Summary

The replication of information is central for the origin of life. In the RNA-World hypothesis, RNA polymers encode and transmit information while simultaneously catalyzing their own replication. At elevated temperatures, RNA is however inherently fragile due to hydrolysis, making a thermal separation of RNA strands for replication difficult. This chapter describes a natural micro-environment of a focused temperature field that both protects and drives replication with a 210 nucleotide long functional ribozyme. In the experiments, the heat flux selectively accumulates 210 nt long oligonucleotides into colder areas of the thermal habitat by thermophoresis, stabilizing them against enhanced RNA cleavage. The replication of shorter 35 nt RNA is catalyzed exponentially by the ribozyme from recurrent laminar thermal convection. The temperature oscillations meet the delicate requirements for the elongation and denaturation process of the ribozymatic replication. Autonomous replication, accumulation and thermal protection of RNA can be accomplished in a single, thermally driven environment.

3.1 Introduction

The evolution of protein-based enzymes are the basis of the complex machinery necessary to drive modern-day living systems. Here, DNA serves as the predominant information storage molecule that is translated to proteins via an RNA intermediate. This allows both the replication of informational polymers and expression of functional enzymes. In a proposed RNA-world on the early earth, however, RNA oligonucleotides could have been an early proxy for this central dogma of microbiology.[10, 73, 74] They are able to store genetic information and also fold into enzymatically active structures.[75–78]. While protein mediated replication requires a sophisticated supporting machinery, the replication process is much easier for an RNA-only system. The RNA-world hypothesis is backed by the fact that RNA ribozymes are an active part of the large ribosomal subunit that catalyzes the synthesis of proteins in modern biology.[79]

One of the central problems of an RNA-world is its dependence on long oligonucleotides that feature enzymatic activities. Longer ribozymes come with the payoff of faster reaction rates: Ribozymes with a length of at least 50-100 nucleotides were shown to achieve turnover rates (k_{cat}) of 1 min^{-1}.[80, 81] However, the spontaneous cleavage of RNA limits the formation of longer nucleic acids, resulting in an exponentially decaying length distribution that is dominated by short, nonfunctional strands. The limiting factor are the slow reaction rates during polymer synthesis. This can be overcome by utilizing a suitable non-equilibrium driving force.[27, 43, 61] A temperature gradient e.g. can locally increase nucleotide concentrations by several orders of magnitude, shifting the length distribution towards longer polymers.[40, 43]

The ribozyme-catalyzed copying of RNA requires energy for the continuous separation of Watson-Crick base pairs after the templated product elongation. Base pairing is essential for all RNA-only copying mechanisms based on ligation or replication. However, these processes require delicate temperature cycles as high temperatures facilitate spontaneous RNA cleavage.[82, 83] Evaporation also changes the concentrations of ionic species, e.g. divalent metal ions, which account for the stability of the RNA.[35–37]

The experiments described here present a thermal micro-environment that naturally drives functional ribozymes. It demonstrates that a heat flux across a water-filled pore selectively accumulates long oligonucleotides in the order of 210-nt in colder areas of the habitat, stabilizing them from enhanced RNA cleavage (Fig. 3.1a,b). Previous studies have shown that heat fluxes across closed compartments accumulate nucleotides and enforce their elongation process towards longer DNA/RNA strands.[40, 43] In open pores of submerged rocks, the same mechanism has been demonstrated to selectively replicate long oligonucleotides by using a protein based replication mechanism, inverting Spiegelman's dilemma.[44, 62]

In this study, the concept of replication is extended to a prebiotically plausible, RNA-only replication scenario. The results demonstrate an exponential replication of RNA by an RNA polymerase in a natural thermal habitat. Here, the polymerase ribozyme includes a sequence tag that binds to a complementary template region via Watson-Crick base pairing. The RNA polymerase then catalyzes the addition of single nucleotides to a template-bound RNA primer. The ribozyme catalysis thereby utilizes recurrent temperature rises, given by thermal convection. These stochastic temperature oscillations meet the delicate requirements for the elongation and denaturation process of the ribozymatic replication (Fig. 3.1b). Therefore, replication, preservation and expressions of genetic information can now be accomplished in a single, thermally driven environment. On the prebiotic earth, those thermal hatcheries might have been essential to drive an RNA-based origin of life.

3.2 Accumulation of RNA polymerase ribozymes

Heat fluxes, for example occurring near sites of volcanic or hydrothermal activity, implement a natural temperature difference across submerged porous rocks. These temperature differences give rise to two effects, the thermophoretic movement of molecules (Thermophoresis) and gravitational convection (Fig. 3.1b). Thermophoresis drives biomolecules along a temperature gradient, in case of nucleic acids from warmer to cooler areas. The thermophoretic movement has been found to depend on non-ionic interactions, ionic shielding and Seebeck effects.[57] For thermal habitats that exhibit a punctual heat source, the thermophoretic velocity field points outward from the given heat spot. Convection triggers the upwards movement of water in warmer regions and a resulting back flow in colder regions of the chamber that continuously drag biomolecules along (see Supplementary Fig. 3.4). The interplay of both effects, thermophoresis and convection, results in an efficient net transport to regions beneath the warm temperature spot. Here, the natural dissipation of heat provides a supporting and protective environment for the replication and storage of functional ribozymes against equilibrating dilution effects.[27]

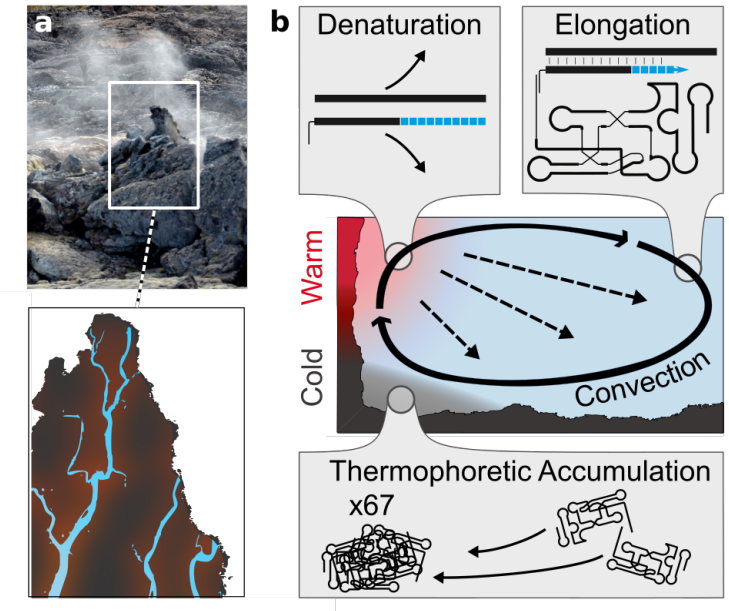


Figure 3.1: Heat fluxes across water filled pores enable the replication of RNA by an RNA polymerase ribozyme. (a) The dissipation of heat across porous rocks forms natural temperature differences such as volcanic or steam activity. (b) The temperature difference induces thermophoresis (dashed arrows), a directed movement of molecules in a thermal field, and convection (solid arrows). In combination, both effects locally concentrate the RNA ribozyme up to 67-fold inside the chamber. In addition, RNA oligonucleotides are subjected to continuous temperature oscillations that drive the denaturation and replication of an RNA template in warm and cold areas, respectively. A natural setting thereby mimics the delicate thermal oscillation conditions necessary for an RNA polymerase ribozyme.

In the experiments, these conditions were implemented using an IR-laser that is absorbed in a water-filled chamber, thereby creating a radial symmetric heat profile across the cylindrical shaped compartment (see Supplementary Fig. 3.4). Within this compartment, nucleic acids with a length of 210-nt are concentrated by a factor of 55 within 60 min (Fig. 3.2a). This length corresponds to the "24-3" RNA polymerase ribozyme which was evolved by Horning and Joyce [75]. The RNA accumulation site is a highly confined, cold region at the bottom center of the chamber, which depletes the RNA from the warmer sections of the reaction vessel (Fig. 3.2b,c). Detailed finite-element simulations describe the accumulation kinetics and take into account diffusion, thermophoresis and convection. Accumulation reaches a steady state after 90 min with a maximum concentration of 67-fold for 210mer polynucleotides compared to the initial conditions (Fig. 3.2d). This steady state is defined by the point in time at which accumulation and diffusive transport against the implemented concentration gradient are equal. Interestingly, the thermophoretic movement of short molecules is significantly lower compared to longer strands.[40] As a result, these molecules are not accumulated in the chamber, here shown for a 35mer oligonucleotide (Fig. 3.2c,d). This size selectivity locally enhances the concentration of rare, long nucleic acids and thus the selection for complex genetic information as well as potential enzymatic activities.

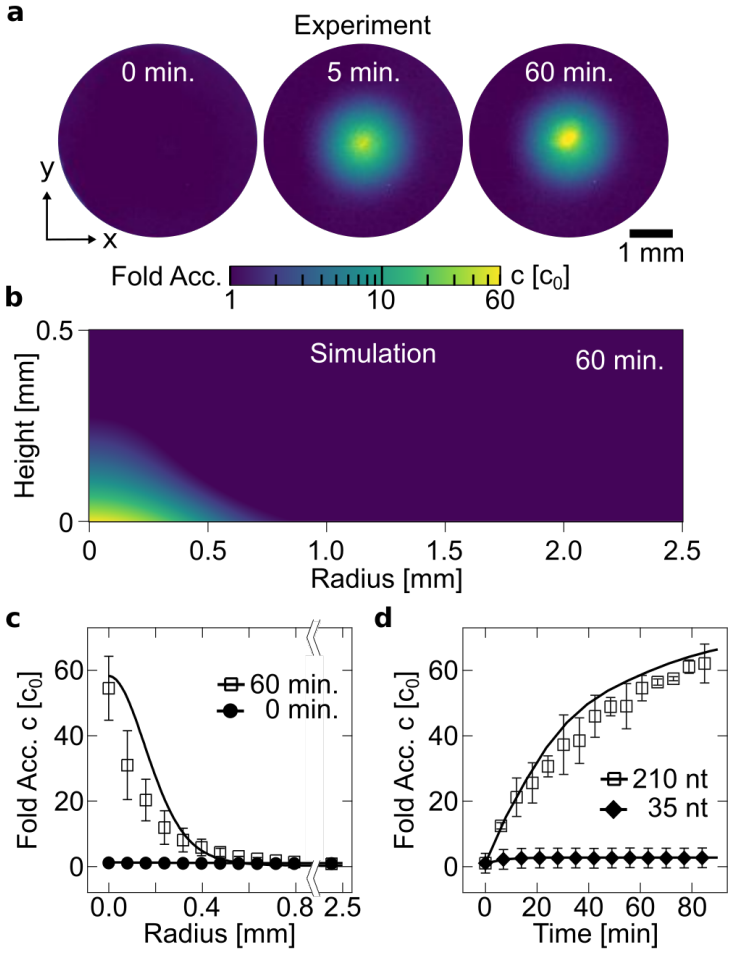


Figure 3.2: Accumulation of nucleic acids in a convective chamber. (a) A radial symmetric heat profile found in naturally heated porous rocks is experimentally implemented using the absorption of an IR-Laser beam in a water-filled compartment. In these habitats, 210mer oligonucleotides accumulated beneath the heated regions in the center up to 55-fold within 60 min. The movement of molecules is defined by drift, diffusion and thermophoresis. (b) Detailed accumulation dynamics are derived by a finite-element simulation and resolve the concentration profile along the height of the chamber. (c) The radial accumulation profiles from both experiment (markers) and simulation (lines) are in good agreement. (d) The accumulation dynamics for 210-mers polynucleotides equilibrate within 90 min, achieving a maximum accumulation of 67-fold. 35mers are not accumulated in the same setting. The error bars depict the standard deviation for duplicate measurements.

3.3 Convectively driven temperature oscillations of RNA templates

The heat driven molecular movement not only accounts for the accumulation but also for repetitive temperature cycling of nucleic acids. Brownian motion and convection shuttle the nucleic acids through regions of varying temperature (Fig 3.3a). As a result, the thermal habitat continuously provides the energy necessary to separate double-stranded nucleic acids. Supposing suitable temperature conditions for the template directed polymerization/ligation, the system also allows Watson-Crick-type replication mechanisms.

Potential natural habitats for an RNA-catalyzed replication of RNA, however, require delicate temperature cycle conditions. The stability of nucleotides, for example, is given at low ambient temperatures. These temperatures reduce non-enzymatic RNA hydrolysis,[82, 83] in particular in the presence of divalent metal ions such as Mg^{2+}.[35–37] While the ribozymatic polymerization operates best at low temperatures (< 20 °C),[75] the denaturation process for RNA requires considerably higher temperatures (> 60 °C). Therefore, to reduce hydrolysis, the time that RNA polymers spend at high temperatures should be minimal. In addition, the replication process also requires a longer period of time at low temperature regions, in order to match time scales for the polymerization process. Primer annealing and ribozyme-catalyzed nucleotide addition times for a 10-nt RNA sequence are in the order of 20 min. These elongation times result from an average nucleotide addition rate of up to 1.2 nt/min, catalyzed by the 24-3 ribozyme.[75]

Simulation of stochastic trajectories in the convection chamber show that both criteria, long elongation and short denaturation times, are met (Fig. 3.3b). Polynucleotides are mostly kept in low temperature regions and therefore match the extended nucleotide addition times of the ribozyme. A 210-nt RNA strand remains at temperatures above 37 °C with a probability of 3 %, a temperature where RNA degradation begins.[83] The probability for a 35mer to be above 37 °C is three times higher compared to a 210mer. Due to the toroidal flow profile, molecules are subjected to a short denaturation phase in the center of the reaction vessel. This is necessary to release the RNA template and extended primer from the ribozyme, which prepares the ribozymes for a next cycle of replication. High flux velocities at the temperature source lead to fast heating/cooling rates of up to 3 K/s (see Supplementary Fig. 3.3). Stochastic simulation derives a mean cycle time of 31 min for a 35mer RNA strand that oscillates between two threshold temperatures, 20 °C and 60 °C. In comparison, 20 min cycling time with 2 s at 72 °C and 20 min at 17 °C were shown to be optimal for bulk cycle experiments.[75]

3.4 Thermal habitat drives exponential amplification of RNA

The polymerase chain reaction (PCR) is a commonly used method in biology for the amplification of DNA. Here, the templated polymerization of DNA is mediated by sophisticated polymerase proteins. These proteins enable an enzyme-driven DNA replication, given the products are exposed to cycles of repeated heating and cooling. The ribozymatic amplification of RNA (rPCR) operates analog. It does, however, not rely on enzymatic activity of proteins but of folded RNA strands. Thus, the replication of informational polymers and expression of functional RNA is possible in a RNA-only system.

The convection chamber allows for the ribozymatic amplification of a 20-nt template that is composed of two 10-nt primer binding sites. Its reaction buffer includes PEG8000 as a molecular crowding agent and tetrapropylammonium chloride (TPA) for lowering the melting temperature of nucleic acids. A reduced Mg^{2+} concentration of 50 mM also minimizes non-enzymatic hydrolysis. Under these conditions, the convection chamber enables the exponential replication of RNA template concentrations as low as 100 fM (Fig. 3.3c,d). The 24-3 polymerase faces difficulties in the extension of purines, forming intermediate products at their positions. The polymerization kinetics of the ribozyme requires the experiment to last at least 18 hours. In comparison, thermal cycling of a bulk reaction solution with 50 temperature cycles under optimized conditions achieves similar amounts of product concentration.

For long lasting experiments, replication driven by a traditional thermal cycler exhibits slightly higher product yield compared to the convective chamber. In these habitats, RNA ribozymes are longer exposed to high temperatures and therefore exhibit higher amount of spontaneous RNA cleavage at later stages of the amplification process (Fig. 3.3d). The exponential replication exhibits product yields of $2{\cdot}10^5$ and $4{\cdot}10^5$ for bulk cycle and convectively driven, respectively, given an initial template concentration of 100 fM.

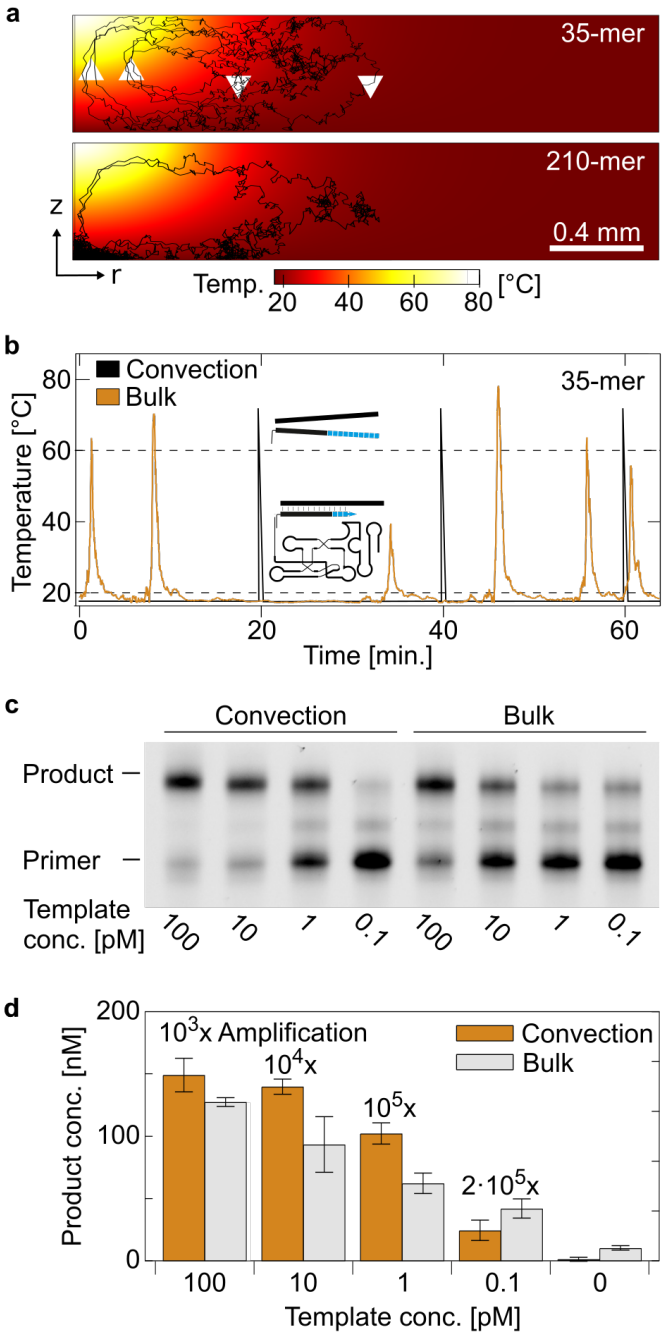


Figure 3.3: Thermally driven replication of RNA templates by an RNA ribozyme. (a) Stochastic trajectory simulations of 35mer (top) and 210mer (bottom) RNA depicts their movement inside the convection chamber. Here, short molecules are subjected to frequent temperature oscillations while longer molecules accumulate in the center. (b) In the convection chamber the 35mer mostly occupies low temperature regions, where primer extension can take place. It passes through high temperature regions for short periods of time, allowing for a denaturation process. Bulk cycling experiments were performed using a conventional thermal cycler with temperature cycles of 17 °C for 20 min and 72 °C for 2 s. (c) The convection chamber and bulk experiments both show a primer extension of a 20mer RNA template. Here, the RNA polymerase exponentially amplifies initial template concentrations as low as 100 fM within 24 hours. Intermediate bands can be attributed to temporary stalling of the RNA ribozyme by purines. (d) The convection chamber achieves product amplification yields comparable to optimized bulk cycling experiments. A maximum amplification of $2{\cdot}10^5$ was observed in the convection chamber. The error bars depict the standard deviation for duplicate measurements.

3.5 Discussion

Many efforts have been made to investigate and optimize the RNA replication by an RNA polymerase.[75, 84–87] These replicators, however, have not been successfully shown to operate in a prebiotically plausible environment, given their delicate temperature requirements. The experiments conclusively show that thermal habitats, an abundant scenario on early earth, allow the replication and sustainment of nucleic acids in an RNA-only system. Here, the thermally induced accumulation mechanism exhibits a length selection towards longer oligonucleotides, enhancing their concentration in colder regions of the thermal habitat. These molecules then exhibit a higher stability for spontaneous cleavage, whereas shorter molecules undergo faster degradation.

The convective movement also shuttles biomolecules between regions of varying temperatures, meeting the requirements for the elongation and denaturation processes of RNA. Importantly, RNA molecules are subjected to steep temperature changes, further decreasing spontaneous cleavage while simultaneously enabling a Watson-crick type replicator. The prolonged elongation times in the order of several minutes also compensate for slow ribozyme-catalyzed nucleotide addition rates. As a result, the system prompts the RNA polymerase to exponentially amplify 10-nt sequences up to $2{\cdot}10^5$-fold within 24 hours. At low initial concentrations, the natural thermal setting achieves merely 32 % less product amplification compared to optimized temperature cycle conditions. Here, RNA ribozymes have an advantage over proteins, since the denaturation at elevated temperatures is always followed by a reversible refolding in colder regions. The given temperature oscillations equally suit other replication systems such as RNA ligation and Q-beta replicase.

The length dependent cleavage in a natural convection chamber could indeed help to overcome the threat of short, parasitic sequences to overwhelm the replication mechanism. Short sequences are generally copied more quickly, exhibiting an intrinsic advantage to survive natural selection. In the thermal habitat, however, short sequence motifs exhibit three times higher probability to remain at elevated temperatures where RNA degradation is present (see Supplementary Fig. 3.5d).[83] As a result, the thermophoretic size-selectivity of the thermal habitat might act as a length selective pressure on replicating nucleic acids.

3.6 Conclusion

Searching for an answer to the question of the origin of life implies finding a location for the molecules to replicate and select according to Darwinian evolution. The experiments demonstrate how heat flow across water filled porous rocks could form a suitable habitat for the preservation and replication of nucleic acids. The energy necessary to continuously drive the strand separation and accumulation of an RNA-based origin of life would be given by the dissipation of thermal energy, a likely abundant geological scenario on the prebiotic earth. If the replicating ribozyme could be assembled from shorter strands, similar to experiments in eutectic ice, shown thermal scenario could be a viable thermal habitat for the first replicating RNA molecules.

3.7 Methods

3.7.1 Template-directed replication of RNA

The RNA replication was carried out in 10 mM TE Buffer (pH 8, sterilized filtered), 0.05 % Tween20 (sterile filtered, Millex-GP Filter Unit 0.22 µm, Merck Chemicals GmbH, Germany), 50 mM $MgCl_2$, 6 % w/v PEG8000, 0.9 M TPA, 400 nM of 24-3 polymerase ribozyme, 200 nM primer, 4 mM NTPs and varying amounts of template. The annealing of the polymerase was performed at 95 °C for 30 s, followed by a temperature ramp to 4 °C at 0.2 °C/s before adding it to the mixture. For purification processes and gel analysis, primers were 5‘-biotinylated and labeled with FAM fluorophores. Following the replication process, the samples were quenched by adding 500 mM of EDTA. Streptavidin C1 Dynabeads (Thermo Fischer Scientific, USA) were used to capture biotinylated primers and extended products. The purification process consisted of six washing steps, four times with an alkali buffer (25 mM NaOH, 0.05 % TWEEN20, 1 mM EDTA) and two times with TEU buffer (1 mM EDTA, 0.05 % TWEEN20, 10 mM Tris, 8 M urea pH 8.0). For PAGE gel analysis, samples were eluted with 98 % formamide and 10 mM EDTA (pH 8.0) and heated for 10 min at 95 °C. The reaction products were analyzed by denaturing gel electrophoresis in 12.5 % polyacrylamide gels at electric field strengths of 60 V/cm and 30 °C inside a 1× TBE buffer for 16 minutes. FAM labeled reaction products, the imaging of PAGE gels was done by CCD photography (Orca 03-G, Hamamatsu, Japan) through a green bandpass filter (520 nm, 10 nm FWHM, Newport, Germany) and spectrally filtered (470 nm, 10 nm FWHM, Thorlabs, USA) light emitting diode excitation (LED 470 nm, Thorlabs, Germany). The detection of Cy5 labeled products was done through spectrally filtered light emitting diodes (LED 625 nm, filter 630 nm, 10 nm FWHM, Thorlabs, Germany) in combination with high quality interference filters (bandpass 692 ± 20 nm, OD6 blocking, Edmund Optics, USA, and bandpass 700 ± 35 nm, OD 2 blocking, Newport, USA). The convectively driven RNA replication was compared to bulk cycling experiments, carried out in a thermocycler with improved heating/cooling rates (Analytik Jena AG, Germany).

3.7.2 Thermally driven convection habitat

The convective chamber is made of a 500 µm thermally conductive soft-silicone film, comprising of blanket out holes of 5 mm in diameter (KU-TCS, Aavid Kunze GmbH, Germany). The chamber is closed at the bottom using a silica wafer (Si-Mat-Silicon Materials e.K., Germany), having a thickness of 525 ± 25 µm, a 100 nm SiO_2 coating and a <100> orientation. The upper enclosure consists of a 170 µm borosilicate glass coverslip (Carl Roth, Germany), allowing for the permeability of IR-light. For the generation of a stable temperature gradient, an IR-Laser (TLR-30-Y12, IPG Laser GmbH, Germany) and a peltier-element (Uwe Electronic Vertriebs GmbH, Germany) were used to act as a heat source and heat sink, respectively. The temperature of the peltier-element was regulated by an PID-controlled feedback loop, where the excess heat was dissipated through a water bath (CF41 Kryo-Kompakt-Thermostat, Julabo, Germany). Fluorescent imaging of the accumulation was realised with optomechanical components (Thorlabs GmbH, Germany), using a long working distance 2× objective (Mitutoyo Plan Apo Infinity 2x, 0.055 NA, Mitutoyo Corporation, Japan), equipped with a CCD camera (Stingray F-145B, Allied Vision Technologies GmbH, Germany) and two alternating light-emitting diodes (LED 625 nm, LED 470 nm, Thorlabs, USA) in combination with a dual band

filter set (Dual band FITC / Cy5, AHF, Germany). The heating is provided by an IR laser (1940 nm, 30 Q, IPG, Burbach) and directed by scanning mirrors (6200-XY, Cambride Technology, England) on four samples simultaneously. The IR laser is coupled into the optical path between the sample and the objective by a special designed dichroic (Transmission 400-700 nm, Reflection 633/1940 nm, AHF, Germany). The temperature profile was measured using the fluorescence of 50 µM BCECF (2',7'-Bis-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein), diluted in TRIS buffer (10 mM).

3.8 Supplementary Information

3.8.1 Finite-element simulations

The finite-element simulations are performed in a two-dimensional, radial symmetric chamber. A punctual heat source in the cylindrical compartment forms a temperature gradient, pointing outward from the given heat spot. Following steps are solved consecutively: (i) The temperature profile is calculated by using partial differential equations for transient heat transfer. The Gaussian profile shaped IR laser is assumed to be absorbed based on Beer-Lambert law. The absorption coefficient for the IR light of 1940 nm is given by 129.6 cm^{-1}. (ii) The numerical solution of the incompressible Navier-Stokes equation derives the convective flow profile inside the thermal convection chamber. The viscosity of the rPCR buffer is experimentally measured via Brownian motion of fluorescently labeled, 1 µm sized, beads. At a temperature of 20 °C, the viscosity is given by 28 mPas·s. (iii) The accumulation of the 35mer and 210mer is solved following Debye's approach. Here, the simulation neglects perturbations at the reversal points of the chamber by assuming $c \cdot \nabla \cdot (\nabla T) = 0; c \cdot \nabla \cdot \mathbf{v} = 0$, where $\mathbf{v}$ denotes the flow velocity. The molecular flux $\mathbf{j}_i$ for a nucleic acid species i is defined by of diffusion, thermophoresis, and convection. It can be described by

$$\mathbf{j}_i = -D_i \cdot \nabla c_i - S_{\mathrm{T}i} \cdot D_i \cdot \nabla T \cdot c_i + \mathbf{v} \cdot c_i \tag{3.1}$$

where c_i, D_i, $S_{\mathrm{T}i}$ and denote the concentration of molecule, diffusion, Soret coefficient and convective transport. The Soret and Diffusion coefficient are given by $S_{\mathrm{T}} = -0.0063 + 0.0115 \cdot (n)^{0.5}\ \mathrm{K}^{-1}$, and $D = 6.69{\cdot}10^{-19}/(8{\cdot}10^{-10}{+}(\mathrm{n}){\cdot}5.9{\cdot}10^{-11})\ \mathrm{m}^2/\mathrm{s}$ adapted from Kreysing *et al.* [44], where n denotes the length of the strand.

3.8.2 Random walk model

Random walk simulations are used to derive statistical data on thermal cycles of RNA oligonucleotides inside the thermal habitat. The temperature cycle statistics consider the movement of 10^4 particles for 4 h experimental time, given time-steps of $\Delta t = 1$ ms. The displacement of particles is thereby given by

$$\mathbf{s} = \Delta t \cdot (\mathbf{v} + D \cdot S_T \cdot \nabla T) + \sqrt{4 \cdot D \cdot \Delta t} \cdot \eta(\mathbf{t}) \tag{3.2}$$

where Δt denotes the time-step, T the local temperatures, and $\mathbf{v}$ the convective flow-profile. The convective flow profile is thereby calculated by finite element simulations (see chapter finite-element simulations). Brownian motion is implemented by a randomly directed movement $\eta(t)$.

Molecules complete a full temperature cycle when moving form a high temperature region (> 60 °C for denaturation) to a low temperature region (< 20 °C for elongation) and back to a high temperature region. Partial elongation scenarios, however, are not considered in the random walk model. Sequences of RNA molecules Both primers poses a 5‘ tag to improve the polymerase ribozyme processivity and a FAM modification.

Type	RNA Sequence (5'-3')
Fwd. Primer	GACAAUGAC AAAAAA GCAACUUUUC
Rev. Primer	GACAAUGAC AAAAAA CACUCCACAC
rPCR Template	GACAAUGAC AAAAAA CACUCCACAC GAAAAGUUGC
24-3 Polymerase	AGT CAT TGC CGC ACG AAA GAC AAA TCT GCC CTC AGA GCT TGA GAA CAT CTT CGG ATG CAG AGG AGG CAG CCT TCG GTG GAA CGA TCG TGC CAC CGT TCT CAA CAC GTA CCC GAA CGA AAA AGA CCT GAC AAA AAG GCG TTG TTA GAC ACG CCC AGG TGC CAT ACC CAA CAC ATG GCT GAC

Table 3.1: RNA sequences for the ribozymatic replication of RNA.

3.9 Supplementary Figures

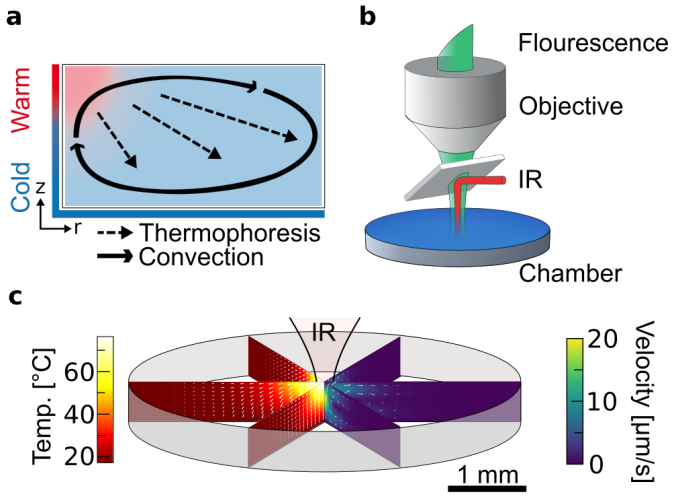


Figure 3.4: Thermally driven convective habitat. (a) A temperature gradient induces two effects in solution, thermophoresis and convection. The superposition of both effects result in an efficient accumulation mechanism for longer polymers. (b) A radial symmetric heat profile is applied using an IR-Laser that mimics a naturally heated pore. The accumulation kinetics are observed using a fluorescence microscope. (c) The absorption of an IR laser in the upper layers of a cylindrical compartment induces a convective motion of the fluid.

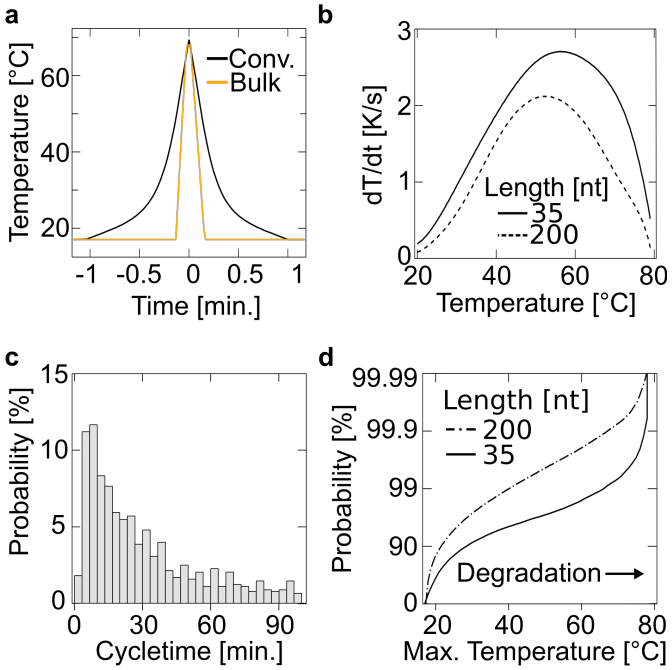


Figure 3.5: Heating/cooling rates derived by stochastic trajectories. (a) The convective motions is the main driving force to subsequently shuttle molecules between regions of varying temperatures. High flux velocities in the center account for heating/cooling rates, necessary for the ribozymatic RNA amplification. Here, the convection motion achieves heating/cooling rates of up to 3 K/s. These rates differ for various oligonucleotides, where short molecules are more prone to undergo faster heating/cooling rates. (b) The convectively driven denaturation process takes on average 39 s, which includes a temperature increase from 30 °C to 60 °C and vice versa. In comparison, bulk cycling experiments exhibit faster heating and cooling rates of 8 °C/s and 6 °C/s, respectively. As a result, the denaturation process takes only 16 s in total. (c) Cycle time histogram of 10^4 particles simulated for 4 hours gives a mean cycle time of 31 minutes for temperature oscillations between 20 °C and 60 °C. (d) The accumulation of long polymers considerably reduces their time to remain at high temperatures. As a result, the 210-nt polymerase ribozyme is sheltered from temperature dependent hydrolysis.

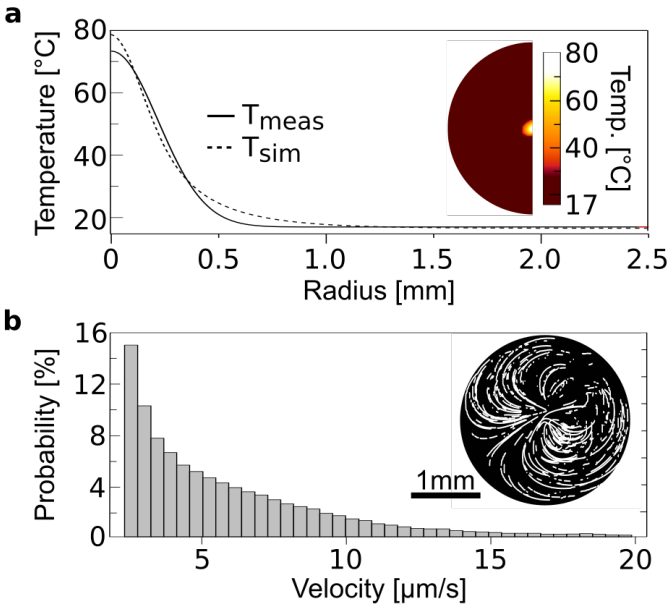


Figure 3.6: IR laser induces convective motion of macromolecules. (a) The Gaussian profile of the IR laser locally heats the compartment in the center. Here, the laser induces a temperature difference between heated regions (80 °C) and the bottom layer, which is actively cooled to 17 °C. The radial temperature profile is derived from the temperature dependent fluorescence of BCECF (50 µM) diluted in TRIS buffer (10 mM). Experimentally derived temperatures are in good agreement with a 2-D finite-element simulation. (b) The temperature gradient induces a convective flow which shuttles 1 µm sized, fluorescently labeled beads between the inner and outer parts of the chamber. The maximum velocities of 20 µm/s occur in the heated regions of the chamber.

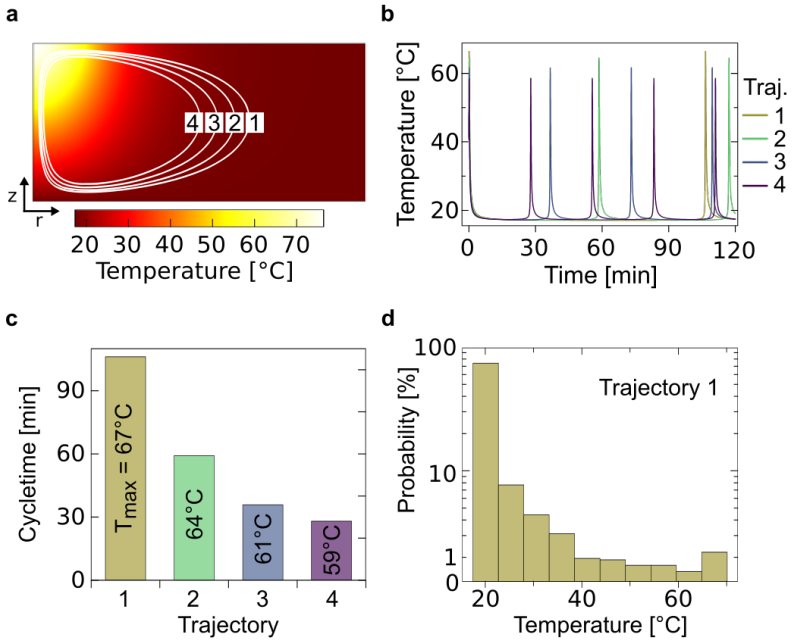


Figure 3.7: Finite-element simulation derives temperature cycle conditions. (a) Exemplary trajectories of macro molecules solely driven by thermal convection. Various thermal scenarios were screened to find a scenario which matches the delicate temperature conditions necessary for RNA amplification. For this purpose, a finite-element simulation derives temperature cycles of molecules with respect to its initial starting position, neglecting thermophoretic and Brownian motion. (b) The temperature oscillation inside a chamber of 0.5 mm in height and 2.5 mm match the temperature cycle requirement of a RNA-only replication system. Here, molecules undergo fast temperatures changes while remaining for most of the time at low temperatures. (c) The trajectories derive cycle times of at least 31 min for temperature cycles between 20 °C and 60 °C. Excessive temperature oscillations, such as temperature differences between 20 °C and 67 °C, require extended cycle times. (d) In these thermal habitats, the molecules are convectively shuttled between regions of varying temperatures, however remain for most of the time at low temperatures.

Chapter 4

Formation of proton gradients and pH oscillations

Summary

Proton gradients are essential for biological systems. They not only drive the synthesis of ATP, but also initiate molecule degradation and recycling inside lysosomes. However, the high mobility and permeability of protons through membranes make pH gradients very hard to sustain in vitro. This chapter reports on a heat flow across a water-filled chamber that forms and sustains stable pH gradients. Charged molecules accumulate by convection and thermophoresis better than uncharged species. In a dissociation reaction, this imbalances the reaction equilibrium and creates a difference in pH. Solutions of amino acids, phosphate, or nucleotides, achieve pH differences of up to 2 pH units. The same mechanism cycles biomolecules by convection between the different pH regions. This implements a feedback between biomolecules and a cyclic variation of the pH. The finding provides a novel mechanism to create a self-sustained proton gradient to drive biochemical reactions.

This Chapter was published by Keil *et al.* [88] in *Nature Communications* under a Creative Commons CC-BY licence.

4.1 Introduction

The establishment of ion gradients holds an important place for the sustainment of all current life forms. Regions of different pH are essential for various cellular processes such as the synthesis of DNA, RNA, and proteins as well as their degradation and recycling.[89–92] These processes are accommodated in specialized compartments separated by biological membranes and important in lysosomes, mitochondria, and cell vacuoles. To form complex reaction networks, the variation of pH is essential to create different reaction conditions.[93, 94] In the process of chemiosmosis, for example, the channeled movement of ions across a membrane drives the phosphorylation of ADP to ATP, life's most commonly used energy currency.[95] Since ancient microbes, such as the last universal common ancestor (LUCA), are assumed to be chemiosmotic,[8] it has been argued towards the significance of proton gradients at the origin of life.[96–98] In addition, the synthesis of precursor molecules for early molecular evolution requires differing pH values for the synthesis of purine and pyrimidine ribonucleotides,[7] aminonitriles,[99] amino acids,[16] and phosphoenol pyruvate.[100]

For living systems, the creation of pH differences is tedious and requires complex protein-mediated transport mechanisms. On prebiotic earth, natural pH gradients would form by laminar mixing of fluids with different pH values, e.g. at the interface between disequilibria at submarine hydrothermal vents and the Hadean ocean.[64, 101, 102] However, these systems rely on a continuous influx of mass and energy. Molecules would experience this gradient only for a short time and are washed out into a pH-equilibrated reservoir, keeping the mechanism transient. Repetitive pH oscillations inside a closed system would be desirable and keep molecules at one location without being lost in a flow or by diffusion.

Here, a mechanism is described that forms stable pH gradients and enables continuous cycling of pH in a closed system. Thermal energy, the form of energy necessary for the mechanism, triggers an accumulation mechanism for individual ionic species. The heat flow forms a temperature gradient, spanning across sub-millimeter sized water-filled compartments. This is likely a common setting on early earth, found for example in geothermally heated porous rocks such as hydrothermal vents or cooling volcanic sites.[38, 39] This elementary setting has previously been shown to concentrate dilute nucleotides,[40] accumulate lipids to facilitate the formation of vesicles,[41] shift polymerization reactions towards longer DNA/RNA strands,[43] and selectively replicate longer polymers.[44] This study now extends these characteristics to include the formation of a stable pH gradient, providing another mechanism for a promising long-term microhabitat for the onset of molecular evolution.

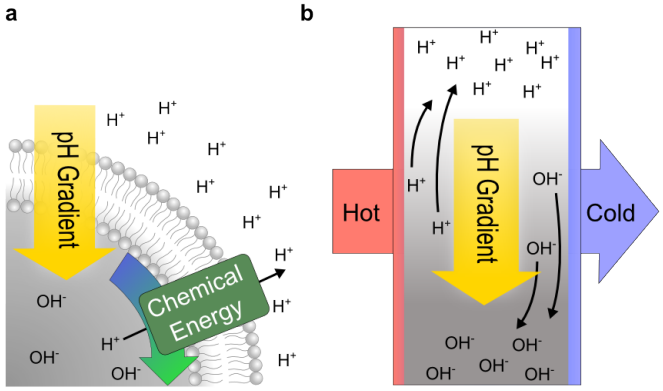


Figure 4.1: Formation of pH gradients by chemical or thermal energy. (a) Modern cells run on an elaborate protein system to maintain proton gradients across a membrane. Chemical energy is used to pump protons against their concentration gradient, which is then harnessed by chemiosmosis. (b) Heat fluxes across confined solutions, a common geological setting, induce a movement of ions that results in a stable pH gradient of up to 2 pH units. The charge selective thermophoretic accumulation of buffer molecules such as amino acids, phosphates, or nucleotides keeps the water self-ionization persistently out of equilibrium. The formation of a pH gradient is achieved in a closed system and biomolecules are repetitively exposed to differences in pH.

The following experiments demonstrate the formation of proton gradients in various buffer solutions. These include prebiotically plausible solutions of phosphate buffers, amino acids, and nucleotides.[12, 15, 103–105] In a temperature gradient, the thermophoretic movement of each ionic species is different. As a result, the concentration of ionic species is locally shifted, prompting the formation of proton gradients (Fig. 4.1). As indicated by the modeling, the formation of proton gradients is not affected by smaller concentrations of larger biomolecules or vesicles. These are shuttled inside the already established pH gradient by laminar convection, prompting them to move regularly between regions of varying pH. Thereby, each molecule undergoes individual pH oscillations with respect to their molecular properties, mostly their diffusion and Soret coefficients. The pH oscillations enable chemical reaction pathways that rely on pH fluctuations. The comparably fast shuttling of vesicles could trigger proton gradients across protocellular membranes without the need for active proton pumps.

4.2 Thermally driven accumulation of ionic species

The formation of pH gradients is facilitated by a thermally driven accumulation mechanism inside a water-filled compartment. These flow chambers are subjected to a spatially confined horizontal temperature gradient. Based on the second law of thermodynamics, energy fluxes are necessary to maintain states of low entropies, in this case accumulated regions of molecules. The accumulation mechanism relies on the superposition of two effects, (i) thermal convection drives molecules up- and downwards in the flow chamber, and (ii) thermophoresis – the movement of molecules in a temperature gradient (Fig. 4.2a). In combination, the two effects result in an efficient net transport of molecules, as small as single ions, to the bottom of the flow chamber and a simultaneous depletion at the top. The flux $\mathbf{j}_i$ for an ionic species i inside the

flow chamber is composed of diffusion, thermophoresis, and convection. It can be described by

$$\mathbf{j}_i = -D_i \cdot \nabla c_i - S_{\mathrm{T}i} \cdot D_i \nabla T \cdot c_i + \mathbf{v} \cdot c_i \tag{4.1}$$

where v denotes the convection flow, c_i, D_i and $S_{\mathrm{T}i}$ the concentration of the ionic species, diffusion coefficient, and Soret coefficient, respectively.[43] The Soret coefficient $S_{\mathrm{T}i} = D_{\mathrm{T}i} \cdot D_i^{-1}$ is defined as the quotient of the thermal and collective diffusion coefficient. Using Debye's approach, an analytical solution for the accumulation mechanism in a flow chamber at equilibrium is given by [65, 68]

$$\frac{c}{c_0} = exp\left(\kappa_i \cdot S_{\mathrm{T}i} \cdot \Delta T \cdot r\right) \tag{4.2}$$

where c/c_0 denotes the relative concentration of a species at the bottom of the chamber, c_0 the initial concentration, the temperature difference ΔT, the aspect ratio of the chamber $r = h \cdot w^{-1}$ with the height h and width w, and the experimental prefactor κ_i ranging between 0 and 0.42 (see Supplementary equation 4.11,4.12). The experimental prefactor κ_i is calculated for each ionic species and depends on experimental parameters such as the temperature difference ΔT, the chamber width w and the diffusion coefficient of the species D_i.

For a given buffer reaction in solution, each buffer subspecies accumulates differently. Its respective accumulation efficiency, as shown in equation (4.2), strongly depends on molecular properties which define its diffusion and Soret coefficients. As a result, different ratios in concentrations of proton acceptor and donor species are present along the height of the flow chamber, which then form a pH gradient (Fig. 4.2b,c). Chamber widths above 150 µm for Earth's gravity field are optimal for the accumulation of fast diffusing, small molecules to achieve a maximum pH gradient. These pores exhibit high convection velocities which are necessary to counterbalance the fast diffusion of the molecules. As a result, the accumulation mechanism concentrates small molecules at the bottom of the pore. Longer oligonucleotides cannot compensate for the convection, reducing their accumulation efficiency while simultaneously being more prone to shuttle through the thermal habitat. The Soret coefficients of the phosphate buffer species are estimated with respect to their net charge (see Supplementary table 4.2).[57]

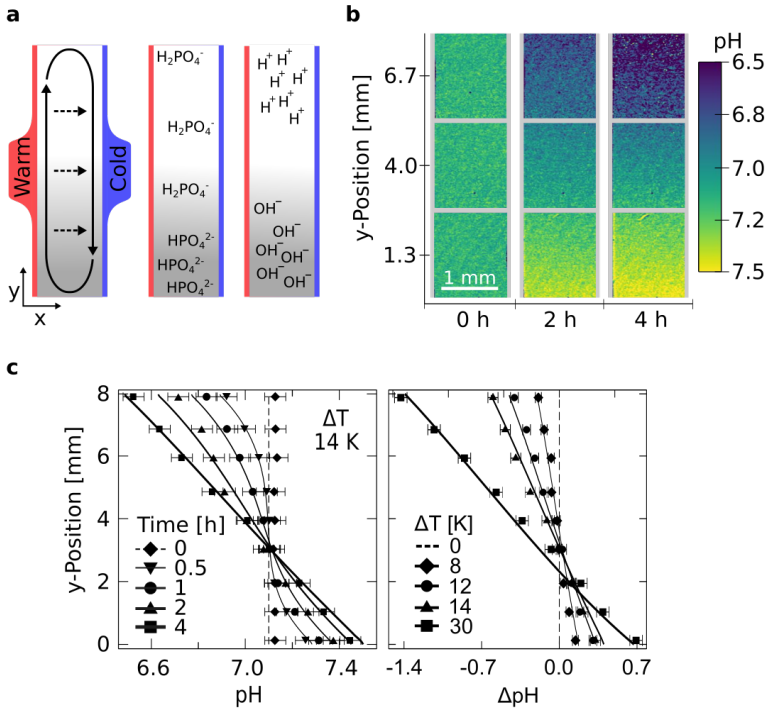


Figure 4.2: Formation of pH gradients in phosphate solutions. (a) The temperature gradient creates a convective flow in the solution while simultaneously inducing a movement of molecules along the temperature gradient based on thermophoresis. The combination of both effects results in an accumulation of molecules at the bottom of the chamber. For the case of the phosphate buffer, HPO_4^{2-} accumulates much stronger compared to $H_2PO_4^-$. Based on the concentration of both buffer species, the pH is locally shifted. As a result, the pH increases in the lower regions and decreases in the upper regions of the chamber. (b) The formation of a pH gradient is monitored in an 8 mm high chamber via the ratiometric dye SNARF-1. (c) For a given temperature difference of ΔT = 14 K, a stable proton gradient of ΔpH = 0.93 develops within 4 hours. The steepness of the proton gradient changes with respect to ΔT and is in good agreement with theoretical predictions. Differences in pH are achieved after 4 h for a ΔT = 8, 12, 14 K and after 1 h for a ΔT = 30 K. The error bars give the standard deviation of the SNARF-1 detection for duplicate measurements.

4.3 Phosphate mediated formation of stable pH gradients

At a pH of 7, phosphate buffer consists mostly of two ionic species, hydrogen phosphate (HPO_4^{2-}) and dihydrogen phosphate ($H_2PO_4^-$). Both ionic species obtain different vertical accumulation profiles along the flow chamber, as the dominant contribution to the thermophoretic force approximately scales with the charge of the molecules squared (Fig. 4.3a). Based on their net charge, HPO_4^{2-} accumulates tenfold stronger compared to $H_2PO_4^-$, locally shifting the acid-base equilibrium in the buffer solution. The acid-base reaction in general and for the phosphate buffer solution in particular is defined by

$$HX + H_2O \rightleftharpoons X^- + H_3O^+ \tag{4.3}$$

$$H_2PO_4^- + H_2O \rightleftharpoons HPO_4^{2-} + H_3O^+ \tag{4.4}$$

respectively, where HX and $H_2PO_4^-$ denote the proton donors and X^- and HPO_4^{2-} the proton acceptors. High concentrations of HPO_4^{2-} at the bottom of the chamber drive a protonation

reaction that locally reduces the hydronium ion concentration (Fig. 4.3b). Simultaneously, a deprotonation reaction of highly concentrated $H_2PO_4^-$ releases hydronium ions at the top of the chamber. The acid-base reaction locally shifts the pH along the chamber. The pH of a solution is derived by the Henderson-Hasselbalch equation

$$\mathrm{pH} = \mathrm{pK}_a + \log\left(\frac{c_{\mathrm{acceptor}}}{c_{\mathrm{donor}}}\right) \tag{4.5}$$

where c_{acceptor} and c_{donor} denote local proton acceptor and proton donor concentrations, respectively. Given sufficiently high buffer concentrations, the pH is only affected by the ratio of proton acceptors to donors, and not by their total concentrations. The pH at the bottom of the chamber $\mathrm{pH}_{\mathrm{bottom}}$ is therefore derived by calculating the corresponding proton acceptor/donor concentration at the bottom, given by equation (4.2) at equilibrium.

$$\mathrm{pH}_{\mathrm{bottom}} = \mathrm{pK}_a + \log\left(\frac{c_{0\ \mathrm{acceptor}}}{c_{0\ \mathrm{donor}}} \cdot \exp\left(\Delta(\kappa_i \cdot S_{\mathrm{T}i}) \cdot \Delta T \cdot r\right)\right) \tag{4.6}$$

$$= \mathrm{pK}_a + \log\left(\frac{c_{0\ \mathrm{acceptor}}}{c_{0\ \mathrm{donor}}}\right) + \frac{\Delta(\kappa_i \cdot S_{\mathrm{T}i}) \cdot \Delta T \cdot r}{ln(10)} \tag{4.7}$$

with $\Delta(\kappa_i \cdot S_{\mathrm{T}i}) = \kappa_{\mathrm{acceptor}} \cdot S_{\mathrm{T\ acceptor}} - \kappa_{\mathrm{donor}} \cdot S_{\mathrm{T\ donor}}$ and $r = h \cdot w^{-1}$ the aspect ratio of the chamber. By assuming $\kappa_{\mathrm{acceptor}} \approx \kappa_{\mathrm{donor}}$ and $\ln(10)^{-1} \approx 0.43$, $\mathrm{pH}_{\mathrm{bottom}}$ is given by

$$\mathrm{pH}_{\mathrm{bottom}} \approx \mathrm{pH}_{\mathrm{init}} + 0.43 \cdot \kappa \cdot \left(S_{\mathrm{T\ acceptor}} - S_{\mathrm{T\ donor}}\right) \cdot \Delta T \cdot r \tag{4.8}$$

The maximum pH difference ΔpH inside the flow chamber, arising between the lowest and highest regions of the chamber, is given by

$$\Delta\mathrm{pH} \approx 0.86 \cdot \kappa \cdot (\Delta S_{\mathrm{T}}) \cdot \Delta T \cdot r \tag{4.9}$$

with $\Delta S_{\mathrm{T}} = S_{\mathrm{T\ acceptor}} - S_{\mathrm{T\ donor}}$. The steepness of the pH gradient is affected by the temperature difference ΔT, Soret coefficient S_{T}, height h, and width w of the chamber (Fig. 4.3c, Supplementary fig. 4.9). An elongated flow chamber can compensate for lower temperature differences, achieving similar or higher pH gradients (Fig. 4.3d). The theoretical model is applicable for shallow proton gradients and is in good agreement with experimental data such as the formation of pH gradients in solutions of phosphate. Here, the model predicts a proton gradient of ΔpH = 0.40, 0.90, and 1.2 which describes comparably well the experimental values of ΔpH = 0.34, 0.69, and 0.93 for a given temperature difference of ΔT = 8, 12, and 14 K, respectively (Fig. 4.3c). The pH model (Eq. (4.9)), however, neglects the build-up of an electrical field that leads to an electrophoretic movement of ions, inhibiting the formation of a pH gradient. As a result, the model is not applicable for steep proton gradients, for example formed by large differences in temperature or by large differences in Soret-coefficients of proton donors/acceptors. In addition, above formula is subjected to the limitations of the Henderson-Hasselbalch equation (4.5), i.e. limited to acid/base concentration in the millimolar regime and pH values of about 7. Better descriptions of the experimental results are obtained using time dependent finite element simulations (Fig. 4.3c), including explicit reaction kinetics for the self-ionization of water, acid-base reaction kinetics, and electrostatics (see Supplementary table tables 4.1 to 4.9). Simulation results are derived using a time dependent finite-element model with a commercial solver.

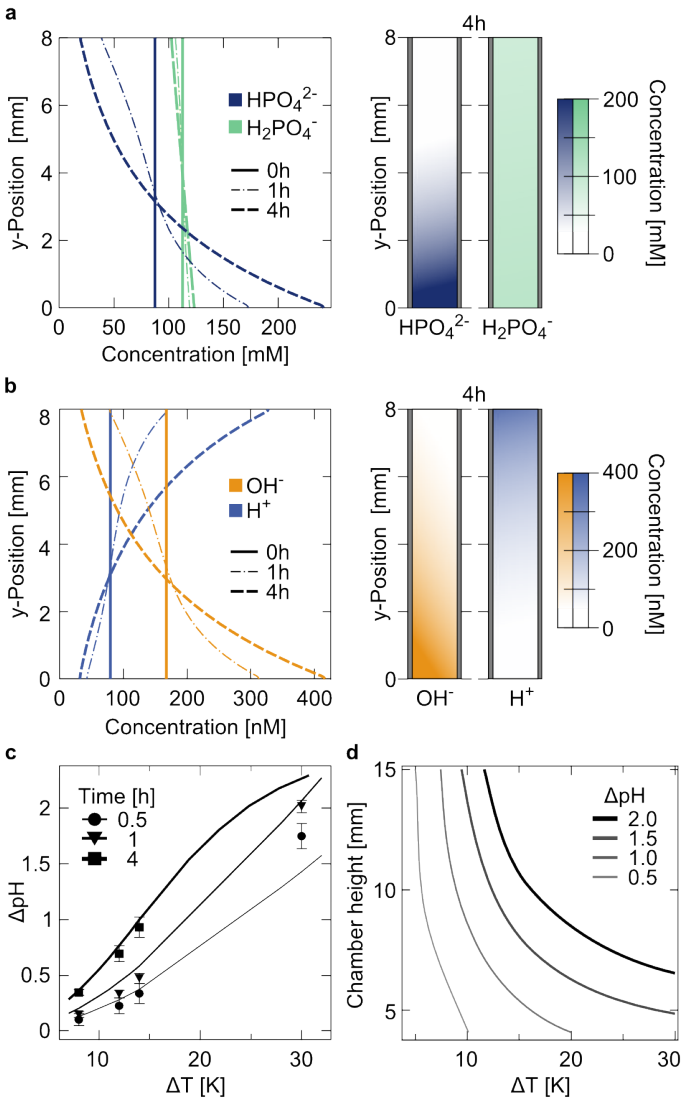


Figure 4.3: Accumulation of phosphate buffer species imbalances the reaction equilibrium. (a) Each phosphate species accumulates at the bottom of the chamber, depending on its thermophoretic force. At near neutral pH, HPO_4^{2-} and $H_2PO_4^-$ are predominately in the solution. As the thermophoretic force scales with the charge of the molecules squared, HPO_4^{2-} achieves a tenfold higher accumulation compared to $H_2PO_4^-$. (b) The protonation reaction of HPO_4^{2-} reduces the hydronium ion concentration at the bottom of the chamber while the deprotonation reaction of $H_2PO_4^-$ simultaneously releases hydronium ions at the top. The system reaches its steady state on the order of hours. (c) The pH gradient scales approximately linearly with the applied temperature difference. A maximum proton gradient of ΔpH = 2.1 is formed for ΔT = 30 K. (d) An elongated flow chamber compensates for lower temperature differences, achieving similar or higher pH gradients. The error bars give the standard deviation of the SNARF-1 detection.

4.4 Ionic solutions mediate formation of pH gradients

The formation of a proton gradient was not limited to phosphate buffer solutions but resulted also from various ionic and amino acid solutions (see Supplementary figs. 4.10 to 4.16). The primary requirement is an acid-base reaction and therefore buffer solutions of HEPES, TRIS, and single nucleotides formed stable pH gradients (Fig. 4.4a). The fluorescent pH indicator SNARF-1 could also act as a buffer, however, due to the low concentrations of 50 µM, only shallow pH gradients emerged as shown in the control measurement. Solutions of TRIS and phosphate achieve the highest pH gradients of ΔpH = 0.47 and 0.79 at concentrations of 200 mM after 60 min. Amino acids have the capability to react with both bases and acids. Prebiotically plausible amino acids such as alanine, arginine, glycine, proline, and serine at concentrations of 200 mM formed proton gradients ranging between ΔpH = 0.1-1.1 within 60 min (Fig. 4.4b). Finite element simulations describe the formation of the pH gradient in detail by assuming a difference in Soret coefficient ΔS_T of 0.022, 0.019, 0.016, 0.012 and 0.003 K^{-1} for the acceptor and donor species of alanine, glycine, arginine, serine and proline, respectively. These Soret coefficients are in the range for unevenly charged ionic species.[57] The models indicate that the thermophoretic properties of each buffer species are the major source for differences in the proton gradients. However, the thermophoretic properties of small molecules are rather tedious to measure and not all available, making it difficult to give a detailed prediction for individual buffers.[106] Mixtures of various buffers or amino acids are expected to form an averaged pH gradient, derived from (4.2) and seen on the example for phosphate/arginine mixtures (see Supplementary fig. 4.13). This pH gradient, however, is shifted by the exact ionic composition of the solution (see Supplementary eqs. (4.13) to (4.15)). Interestingly, short chains of amino acid monomers, shown for tetra-glycine and penta-alanine, form an inverted proton gradient of ΔpH = -0.33 and -0.06, respectively (Fig. 4.4c). Here, the accumulation of proton donors is much stronger compared to proton acceptors, resulting in a deprotonation reaction of proton donors that releases hydronium ions at the bottom of the chamber. The inversion of the pH gradient is explained by a negative difference in the Soret coefficient $\Delta S_T = \Delta S_{T\,\text{acceptor}} - \Delta S_{T\,\text{donor}}$, which is the result of additional non-ionic contributions to thermophoresis. The non-ionic contribution S_T^{NI} of the Soret coefficient S_T (see Supplementary eq. 4.13) scales linearly with the polymer length, as a result of local molecule-solvent interaction around the tube-like polymer.[57, 107, 108] In the case of longer peptides, the contribution of the Soret coefficient, which depends on the protonated state S_T^{CM}, becomes comparatively smaller. As a result, the difference in Soret coefficient ΔS_T between the proton acceptor and donor decreases or is even inverted for longer peptides.

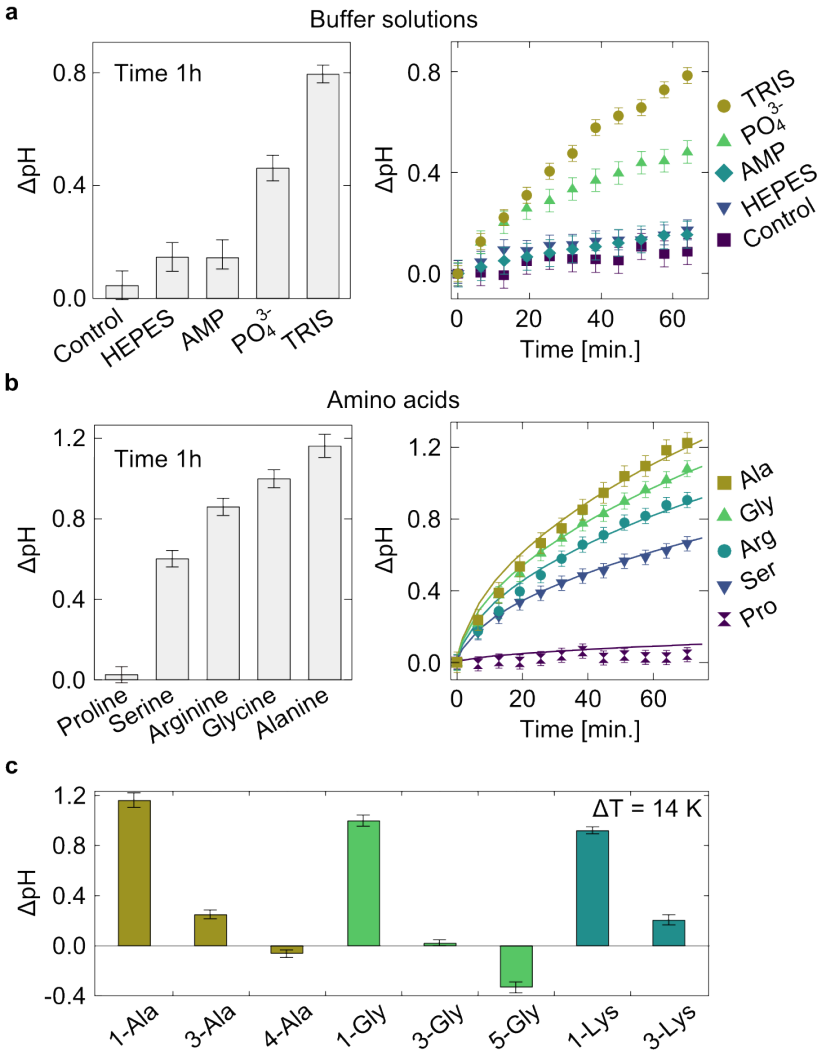


Figure 4.4: Formation of proton gradients in ionic and amino acid solutions. (a) Various ionic solutions form steep proton gradients within 60 min, achieving ΔpH = 0.14, 0.14, 0.47 and 0.79 for HEPES, AMP, Phosphate, and TRIS, respectively. As a control, the fluorescent pH indicator SNARF-1 is dissolved in pure water, forming a pH gradient of ΔpH = 0.09. (b) Prebiotically plausible amino acids form proton gradients of up to ΔpH = 1.1 within 60 min. Finite-element simulations are in good agreement using differences in Soret coefficients ΔS_T of 0.022, 0.019, 0.016, 0.012 and 0.003 K^{-1} as fitting parameters for the acceptor and donor species of alanine, glycine, arginine, serine and proline, respectively. (c) Measured peptides have the tendency to form an inversed proton gradient, showing a decrease of pH at the bottom. An inversion of pH gradients is attributed to a negative difference in Soret coefficient ΔS_T from non-ionic thermophoresis. Error bars give the standard deviation of the SNARF-1 detection.

4.5 Convectively driven pH oscillations

Molecules and particles accumulate and undergo laminar convection. The molecular motion inside the chamber is described by convection, Brownian motion, and thermophoresis (Fig. 4.5a). As a result, molecules shuttle in the arising pH gradient and are subjected to subsequent pH oscillations. These pH oscillations strongly depend on the molecular properties. Larger macromolecules are more prone to shuttle in the convective flow and exhibit faster pH cycle times compared to smaller molecules, which tend to solely accumulate at the bottom of the chamber (Fig. 4.5b). For example, 10mer strands of RNA complete pH oscillations of at least ΔpH = 0.5, 1.0, and 1.5 in 11, 38, and 73 min, respectively (Fig. 4.5c). Longer DNA/RNA strands such as 200mers undergo faster pH oscillations of at least ΔpH = 0.5, 1.0, and 1.5 in 7, 14, and 26 min, respectively. The accumulation mechanism not only defines the magnitude of pH oscillations for a species but also affects the probability for a species to remain in a specific pH region (Fig. 4.5d). The flow chamber efficiently accumulates DNA/RNA strands in the region of 3-10 nt, while smaller and larger molecules accumulate significantly weaker. A 10mer oligonucleotide inside an already formed pH gradient ranging from 7.7 to 5.6 has probabilities to stay in pH regions pH > 7.2 and pH > 6.7 of 0.81 and 0.98, respectively. This probability decreases for shorter and longer oligonucleotides. The accumulation kinetics hereby strongly depends on the flow chamber's width; smaller chambers achieve larger accumulation for longer molecules (see Supplementary fig. 4.18). The inversion of the pH gradients, facilitated by longer peptides, allows for a microscopic chemical feedback from polymerization back to the pH gradient in these chambers. Now, long molecules, which are concentrated at the bottom of the chamber, remain at low pH values for prolonged periods if the pH gradient is inverted.

4.6 Discussion

The experimental and theoretical findings show that heat fluxes across confined solutions form and sustain stable pH gradients. The mechanism solely requires a dissolved acid-base reaction buffer that is given by prebiotically plausible biomolecules such as phosphate, amino acids, or RNA. Phosphate for example is assumed to play an important role in prebiotic chemistry because it functions as a nucleophilic catalyst, pH buffer, and chemical buffer.[15] The formation of pH gradients is hereby achieved in a closed system that allows molecules to undergo repetitive pH oscillations without being washed away into an equilibrated reservoir.

The dissipation of free thermal energy, the sole form of energy necessary, is a geologically abundant scenario on early Earth. Although the formation of pH gradients depends on the thermal gradient that drives the accumulation mechanism, shallow thermal gradients can be compensated by longer flow chambers.[1] The geometry of the flow chamber is realized to mimic the most simple geometrical setting with dimensions of 8.0 x 0.17 mm. However, the shape of the flow chamber is found to have a minor impact on the accumulation mechanism.[40]

The thermally driven formation of pH gradients can be incorporated into a microfluidic device that forms tunable pH gradients and subjects molecules to pH oscillations defined in range and frequency. This device potentially allows for size selection and certain replication mechanisms to be functional.[44]

Interestingly, the existing pH gradient can simultaneously be utilized by actively driven biomolecules, not only based on Brownian motion. The accumulation mechanism thereby

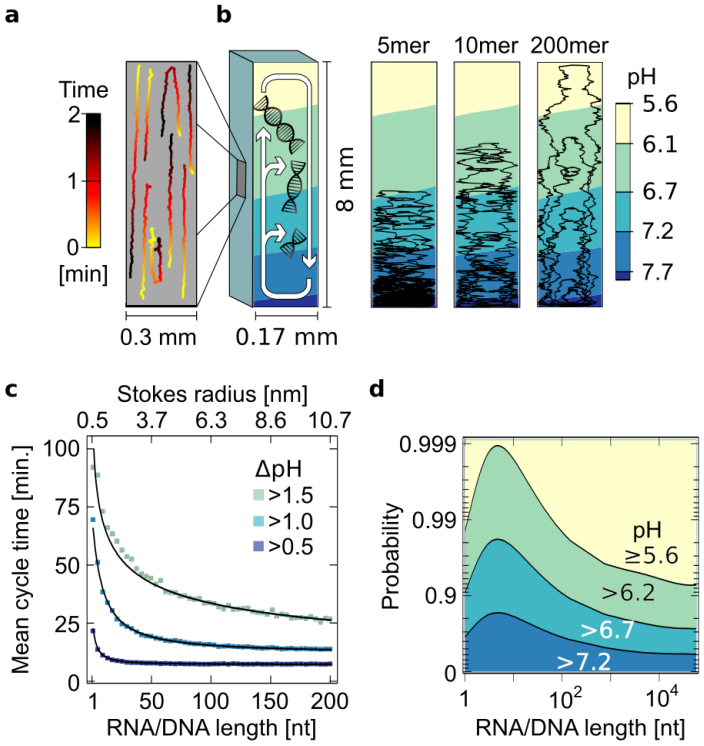


Figure 4.5: pH oscillations of individual particles derived by numerical simulations. (a) Laminar convection induces an upwards and downwards movement of molecules inside the chamber. The convection flow is visualized with 1 µm polystyrene beads over the time of 2 min. (b) Single molecules are shuttled in the existing pH gradient, undergoing subsequent pH oscillations. The more stochastic trajectories of a 5, 10 and 200mer RNA or DNA strand are depicted in black. The probability for a molecule to remain in a specific pH region strongly varies for each molecule. (c) Stochastic simulations of 1-200mer oligomers of RNA/DNA are used to derive mean cycle times for completing pH oscillation of $\Delta pH > 0.5$, 1.0 and 1.5. Longer strands of RNA/DNA are more prone to complete a full pH cycle and therefore exhibit faster cycle times compared to shorter strands. (d) Short DNA/RNA strands accumulate more strongly at the bottom, remaining mostly in a distinct pH region. Longer strands, however, are more likely to be shuttled between varying pH regions. The thickness of the flow chamber thereby determines which molecules remain at defined pH regions and which undergo frequent pH oscillations (see Supplementary fig. 4.18 for details).

subjects molecules to convectively driven pH cycles, resulting in permanent pH oscillations. Importantly, the oscillatory behavior applies for each individual species of the solution and strongly depends on its molecular properties. Hereby, specific molecular species harness the pH gradient by undergoing extensive pH oscillations, while others strongly accumulate by remaining in a limited pH region. This enables the formation of biomolecules that rely on fluctuations in pH during their assembly. The yield of pH-dependent multicomponent reaction systems such as a three-component purine/pyrimidine reaction system could be shifted along the height of the chamber. Here, purine ribonucleotides precursors would be synthesized at the top of the chamber at pH 5 while pyrimidine precursors would be synthesized at the bottom at pH 7.[7]

4.7 Conclusion

The results conclusively show in both theory and experiment how heat flows form stable pH gradients. Proton gradients hold an important place for the sustainment of all current life forms and have also been argued to play a central role in powering the origin of life.[96, 97] The simplicity of this novel mechanism opens various possibilities for rebuilding biological systems bottom up – to either understand them from a mechanistic perspective or to elucidate how cellular life might have originated.

The mechanism that forms pH differences here also drives the accumulation of dilute biomolecules,[40, 44] overcoming the concentration problem of the origin of life. The combination of such an accumulation mechanism that simultaneously subjects biomolecules to pH oscillations prompts interesting future experiments. In the origin of life, the abundance of geological environments that were capable of forming stable pH gradients in a closed system could have opened vast chemical possibilities for the formation and sustainment of the first replication cycles of prebiotic biomolecules.

An exciting prospect of this work is the periodical formation of proton gradients across vesicles. The leakiness of the first membranes,[109] made of fatty acids or phosphorylated isoprenoids,[48, 110] ensures a short-term pH equilibration with its surrounding. Based on the convective flow, vesicles are transported into different pH regions while preserving the equilibrated pH value at the inside, forming pH gradients across the membrane. These pH gradients could be harnessed by primordial forms of the ATP synthase to drive an energy storage system.[98, 111]

4.8 Methods

4.8.1 Microfluidic flow chamber

The temperature gradient is applied across a 3-D printed flow chamber, in between a silica wafer and a transparent sapphire window (Sappro, Germany, see Supplementary fig. 4.6). Hereby, a water bath (CF41 Kryo-Kompakt-Thermostat, Julabo, Germany) cools down the silica wafer while simultaneously heating up the sapphire with temperature PID-controlled hotend heater cartridges (12 V, 6/20 mm, RepRap, France, see Supplementary fig. 4.7). These heater cartridges are inserted into copper blocks that are thermally connected to the sapphire window. The silica wafer (Si-Mat-Silicon Materials e.K., Germany) has a diameter of 100 mm, a thickness of 525 ± 25 µm, a 100 nm SiO_2 coating and a <100> orientation. The boundaries of the flow chamber are fabricated by a 3-D printer using polylactic acid (PLA) material (Filamentworld, Germany). The flow chamber is positioned using a translation stage (Thorlabs, Germany). For in- and outflow of the chamber, capillaries with a height and width of 50 x 500 µm (vitrotubes 5003, CMScientific, UK) connected to a BTP E-60 tubing (0.76 x 1.22 mm, Instech Laboratories, Inc., USA) are used. A high precision syringe pump (neMESYS, Cetoni, Germany) and 250 µL microsyringes (Hamilton, USA) ensure an accurate loading of the flow chamber at flow rates ranging from 1000 nL·s^{-1} to 1 nL·s^{-1}.

4.8.2 Ratiometric pH detection

The pH in the solution is derived using the ratiometric dye SNARF-1 (5-[and-6]-Carboxy SNARF-1, Invitrogen AG, CA) at concentrations of 50 µM (see Supplementary fig. 4.8, Supplementary table 4.10). The starting pH of each solution was adjusted close to the pK of the SNARF-1 dye to ensure maximum accuracy of the pH measurement. For recording the fluorescence images at λ_1 = 580 nm and at λ_2 = 640 nm simultaneously, a self-built fluorescence microscope was used with a CMOS Stingray 145-b camera (AVT, Germany), a 2x objective (LD, NA 0.055, Mitutoyo, Japan) and an Optosplit 2 (Cairn Research, UK). The ratiometric dye is excited using a 470 nm LED (M470L2-C4, Thorlabs, Germany) in combination with a ratiometric filterset (BrighLine HC 482/35, HCBS506, Brightline HC 580/23, H606LP, BrightLineXF 643/20, AHF Analysentechnik AG, Germany). Depicted error bars are given from duplicate measurements (n = 2), and in particular cases from triplicates.

4.9 Supplementary Methods

4.9.1 Finite-element simulations

The finite-element simulations consist of three consecutive calculations of a two-dimensional flow chamber. The flow chamber is mimicked by using a rectangularly shaped compartment with a horizontal aligned temperature gradient.

(i) The temperature profile of the flow chamber is calculated using partial differential equations (PDE) for transient heat transfer. Thereby, a temperature of T_{max} and T_{min} is applied to the left and right vertical boundaries, respectively.

(ii) The convective flow profile inside the flow chamber is derived by numerically solving the incompressible Navier-Stokes equation. The µm-sized width and a moderate temperature gradient lead to a Rayleigh number of $Ra_L = 2$. As a result, the heat flow as well as the temperature profile almost entirely corresponds to heat transfer by conduction. The heat transfer caused by convection are negligible. Finite-element simulation of the heat transfer and fluid flow were preformed consecutively to decrease the computational time, however, yielded the same results compared to a coupled simulations of heat transfer/fluid.

(iii) Molecular concentrations for each buffer species are derived by simultaneously solving acid-base reaction kinetics, electrostatics and molecular flux equations. Following Debye's approach, the simulation neglects perturbations at the upper and lower reversal points by assuming $c \cdot \nabla \cdot (\nabla T) = 0; c \cdot \nabla \cdot \mathbf{v} = o$, where v denotes the flow velocity. Explicit reaction rates and kinetics are listed below for both cases, solutions of phosphate buffer (see Supplementary tables 4.2 to 4.6) and amino acids (see Supplementary tables 4.7 to 4.9).

Temp. [K]	pK_w
273.13	14.943
298.13	13.997
333.13	13.262

Table 4.1: Temperature dependence of pK_w.[112]

Variable	Formula	D [µm^2/s]	S_T 10^{-3} [K^{-1}]	pK_a	µ[10^{-7} m^2/(V·s)]
Phosphate	PO_4^{3-}	824	24.0	2.12	-2.89
Hydrogen phosphate	HPO_4^{2-}	760	12.0	2.12; 7.21	-1.18
Dihydrogen phosphate	$H_2PO_4^-$	959	0.7	7.21; 12.67	-0.373
ortho-Phosphoric acid	H_3PO_4	959	0.3	12.67	
Oxonium ion	H_3O^+	9311	18.0		3.62
Hydroxide ion	OH^-	5273	23.3		-2.05
Sodium ion	Na^+	1334	8.7		0.519

Table 4.2: Input parameters for finite-element simulation. The Soret coefficients for the oxonium ion, hydroxide ion and sodium ion are taken from literature,[113, 114] while the Soret coefficients of the phosphate buffer are estimated with respect to their net charge. The diffusion coefficients and mobilities are taken from Lide et al.[115]

Variable	Concentration [mol/l]
$c(H_3PO_4)$	$c(H_2PO_4^-)\cdot 10^{-pH+pK1}$
$c(H_2PO_4^-)$	$c(HPO_4^{2-})\cdot 10^{-pH+pK2}$
$c(HPO_4^{2-})$	$c(PO_4^{3-})\cdot 10^{-pH+pK3}$
$c(PO_4^{3-})$	$c_0/(1+10^{-pH+pK3})$
c_0	$200\cdot 10^{-3}$
c(HX)	$c(X^-)\cdot 10^{-pH+pK_a}$
$c(X^-)$	$c0/\left(1+10^{-pH+pK_a}\right)$

Table 4.3: Initial concentration for the phosphate buffer and amino acids. The concentrations of each species are calculated based on their pK_a and the starting pH

	pK_a
pK1	2.12
pK2	7.21
pK3	12.67

Table 4.4: pK_a values of phosphate buffer.

Variable	Rate constant
k_{offP1} [l/(mol·s)]	10^{13}
k_{offP2} [l/(mol·s)]	10^{13}
k_{offP3} [l/(mol·s)]	10^{13}
K_{Si} [mol/l]	10^{-pKi}
k_{onP1} [1/s]	$k_{offP1} \cdot K_{S1}$
k_{onP2} [1/s]	$k_{offP2} \cdot K_{S2}$
k_{onP3} [1/s]	$k_{offP3} \cdot K_{S3}$
k_A [l/(mol·s)]	$1.3 \cdot 10^{10}$
K_W [mol/l]	10^{-pK_w}
k_D [1/s]	$k_A \cdot K_w$

Table 4.5: Reaction rates of the phosphate buffer and water. The pK_i denotes the pK_a of a specific buffer species.[116]

Rate/Conc.	c(H_3O^+)	c(OH^-)	c(PO_4^{3-})	c(HPO_4^{2-})	c($H_2PO_4^-$)	c(H_3PO_4)	c(H_20)
d(H_3O^+)/dt		$-k_A \cdot$cH	$-k_{offP3} \cdot$cH	k_{onP3} - $k_{offP2} \cdot$cH	k_{onP2} -$k_{offP1} \cdot$cH	k_{onP1}	k_D
d(OH^-)/dt	- $k_A \cdot$cH						k_D
d(PO_4^{3-})/dt			- $k_{offP3} \cdot$cH	k_{onP3}			
d(HPO_4^{2-})/dt			$k_{offP3} \cdot$cH	-$k_{offP2} \cdot$cH -k_{onP3}	k_{onP2}		
d($H_2PO_4^-$)/dt				$k_{offP2} \cdot$cH	- $k_{offP1} \cdot$cH -k_{onP2}	k_{onP1}	
d(H_3PO_4)/dt					$k_{offP1} \cdot$cH	-k_{onP1}	

Table 4.6: Reaction kinetics of phosphate buffer.

Aminoacid	Initial pH	pK_a	$S_{T\,acceptor}$ [10^{-3} K^{-1}]	S_{Tdonor} [10^{-3} K^{-1}]	ΔS_T [10^{-3} K^{-1}]
Alanine	8.3	9.87	22.0	0.1	21.9
Glycine	8.2	9.60	22.0	3.0	19.0
Argenine	7,9	9.15	18.0	2.0	16.0
Serine	7.9	9.09	15.0	3.0	12.0
Proline	7.0	10.6	3.0	0.1	2.9

Table 4.7: Soret coefficients, pK_a and initial pH of amino acids. The initial pH is adjusted to be situated in the SNARF-1 detection range of $5 \leq pH \leq 9$. The diffusion coefficients for amino acids are set to 760 and 960 $\mu m^2/s$ for the proton donor and acceptor, respectively.[117] The difference in Soret coefficient ΔS_T is hereby defined by $\Delta S_T = S_{T\,acceptor} - S_{T\,donor}$.

Variable	Rate constant
k_{offAA} [l/(mol·s)]	10^{13}
K_{AA} [mol/l]	10^{-pK_a}
k_{onAA} [1/s]	$k_{offAA} \cdot K_{AA}$

Table 4.8: Reaction rates of amino acids.

Rate/Conc.	c(H_3O^+)	c(OH^-)	c(X^-)	c(HX)	c(H_2O)
d(H_3O^+)/dt		-k_A·cH	-k_{offAA}·cH	K_{onAA}	k_D
d(OH^-)/dt	-k_A·cH				k_D
d(X^-)/dt			-k_{offAA}·cH	k_{onAA}	
d(HX)/dt			+k_{offAA}·cH	-k_{onAA}	

Table 4.9: Reaction kinetics of amino acids. Hereby X^- and HX denote the proton acceptor and donor, respectively.

4.9.2 pH detection/calibration method

A spatially resolved pH is derived using the ratiometric dye SNARF-1. Following a modified Henderson-Hasselbalch equation (see Supplementary eq. 4.10),[118] the fluorescence intensity ratio R = $F^{\lambda 1}/F^{\lambda 2}$ provides a measure for the pH. The wavelengths λ_1 and λ_2 denote the emission wavelength of SNARF-1 at λ_1 = 580 nm and λ_2 = 640 nm. The fluorescent images are recorded simultaneously using a beam splitting device and background correction.

$$\mathrm{pH} = a + b \cdot \log\left(\frac{R - Ra}{Rb - R}\right) \tag{4.10}$$

The modified Henderson-Hasselbalch equation comprises of four parameters (Ra, Rb, a, b), derived from the ratio-to-pH calibration curve (see Supplementary fig. 4.8). The parameters a and b denote for the pK_a of SNARF-1 and the weighting of the spectra parameters while Ra and Rb account for the maximum and minimum fluorescence rations. Temperature dependent shifts of the pH calibration curve are corrected by linearly fitted parameters a, b, Ra, Rb (see Supplementary fig. 4.8, Supplementary table 4.10).

Parameter	Offset t	Slope m [1/°C]
a	7.2293 ± 0.0407	-0.021194 ± 0.00113
b	1.3941 ± 0.0564	-0.0010164 ± 0.00165
Ra	1.4917 ± 0.0351	0.0036805 ± 0.00102
Rb	0.13895 ±0.00986	-0.00077193 ± 0.000272

Table 4.10: Temperature dependence of SNARF-1 calibration curve. The temperature dependent correction of the SNARF-1 parameters is fitted linearly: y = t + m·x.

If not mentioned separately, shown measurements were conducted in 50 µM SNARF-1, a buffer/amino acid concentration of 200 mM and a temperature gradient of ΔT = 14 K.

subsectionExperimental prefactor κ The experimental prefactor for the accumulation ratio is defined by Mast and Braun [68]

$$\kappa = \frac{q/120}{1 + q^2/10080} \tag{4.11}$$

with q defined as

$$q = \frac{\Delta T \alpha g \rho_0 w^3}{6 \eta D} \tag{4.12}$$

where ΔT denotes the temperature gradient, α the volume expansion of the solvent, g standard gravity, w the width of the chamber, ρ_0 the density of the solvent, η the viscosity of the solvent and D the diffusion coefficient of the species. The accumulation prefactor strongly affects the accumulation efficiency. Therefore, the width of the flow chamber is specifically designed for certain molecules.

4.9.3 Thermophoresis

The thermophoretic effect moves molecules along a temperature gradient, mostly from the warm to the cold side. However, thermophoresis is still subject to active research. The Soret coefficient gives a measure for a molecules thermophoretic movement. It can thereby be described by four contributions, the capacitor S_{T}^{CM}, Seebeck effect S_{T}^{EL},[119, 120] non-ionic S_{T}^{NI}, and the temperature contribution.11

$$S_{\mathrm{T}} = S_{\mathrm{T}}^{CM} + S_{\mathrm{T}}^{EL} + S_{\mathrm{T}}^{NI} + 1/T \tag{4.13}$$

Recently, it has been suggested that thermophoresis on charged molecules is dominated by Seebeck effects and ion shielding.[57] The Seebeck effect solely depends on the ionic composition of the solution. Here, each ion in solution is subjected to thermophoresis and generates a global electric field. Within the electric field, molecules are subjected to electrophoresis. The electrical field is given by

$$E = \frac{k_B T \nabla T}{e} \frac{\sum_i Z_i c_i S_{Ti}}{\sum_j Z_j^2 c_j} \tag{4.14}$$

where E denotes the electric field, Z_i, c_i and $S_{\mathrm{T}i}$ denotes the charge number, concentration and Soret coefficient of an ionic species. The Seebeck contribution can be approximated by

$$S_{\mathrm{T}}^{EL} = -\frac{k_B T \mu_{DNA}}{e D_{DNA}} \frac{\sum_i Z_i c_i S_{Ti}}{\sum_j Z_j^2 c_j} \tag{4.15}$$

where $\mu_D NA$ denotes the electric mobility. As shown in Supplementary fig. 4.12, the Seebeck contribution strongly affects the formation of pH gradients.

4.9.4 Charge of amino acids and peptides

The net charge of a peptide sequence or single amino acid depends on the pH and is given by Paselk [121]

$$Z = \sum_i N_i \frac{10^{\mathrm{pK_{ai}}}}{10^{\mathrm{pH}} + 10^{\mathrm{pK_{ai}}}} - \sum_j N_j \frac{10^{\mathrm{pH}}}{10^{\mathrm{pH}} + 10^{\mathrm{pK_{aj}}}} \tag{4.16}$$

where Z denotes the net charge of the peptide sequence, N_i the number of histidine, lysine and arginine residues and the N-terminus with their respective $\mathrm{pK_{ai}}$ values, and N_j the number of cysteine and the C-terminus with their respective $\mathrm{pK_{aj}}$ values.

The net charge of an amino acid changes with respect to the alpha-amino group (pK ≈ 9), the carboxylic acid group (pK ≈ 2) and in some cases the side chain. The initial pH is adjusted to be situated in the detection range of SNARF-1 (pH 5-9), therefore solely the dissociation of the alpha-amino group and side-chain are detectable (see Supplementary fig. 4.13).

4.9.5 Random walk model

Random walk simulations investigate the pH cycling statistics of RNA/DNA strands. These molecules are placed inside a 0.17 x 8.0 mm flow chamber filled with 200 mM of phosphate solution. In theory and experiment, a temperature difference of ΔT = 30 K forms a pH gradient of up to ΔpH = 2.1 (see Fig. 4.2c,4.5a). The particle movement (N = 1000) inside the already established pH gradient is simulated for a given time-step of $\Delta t = 1 ms$ for a total simulation time of 3 h. The displacement $\Delta\mathbf{s}(x, y)$ of these particles is given by

$$\Delta\mathbf{s}(x, y) = \sqrt{4 \cdot D \cdot \Delta t} \cdot \eta(t) + \Delta t \cdot (\mathbf{v}(x, y) + D \cdot S_{\mathrm{T}} \cdot \nabla T(x, y)) \tag{4.17}$$

where Δt denotes the time-step, $\mathbf{v}(x, y)$ the convective flow-profile and T the local temperatures. Finite element simulations calculate the convective flow profile (see chapter finite-element simulations) while the Brownian motion is implemented by a randomly directed movement $\eta(t)$. Diffusion and Soret coefficients for RNA/DNA strands in the range of 1-200 nt are given by $D = 470 \cdot n^{-0.53}$ µm^2/s and $S_{\mathrm{T}} = 0.1 \cdot n^{0.5} K^{-1}$, where n denotes the length of the strand. The experimental scaling law of the Soret coefficient S_{T} is adapted from Mast *et al.* [43].

4.10 Supplementary Figures

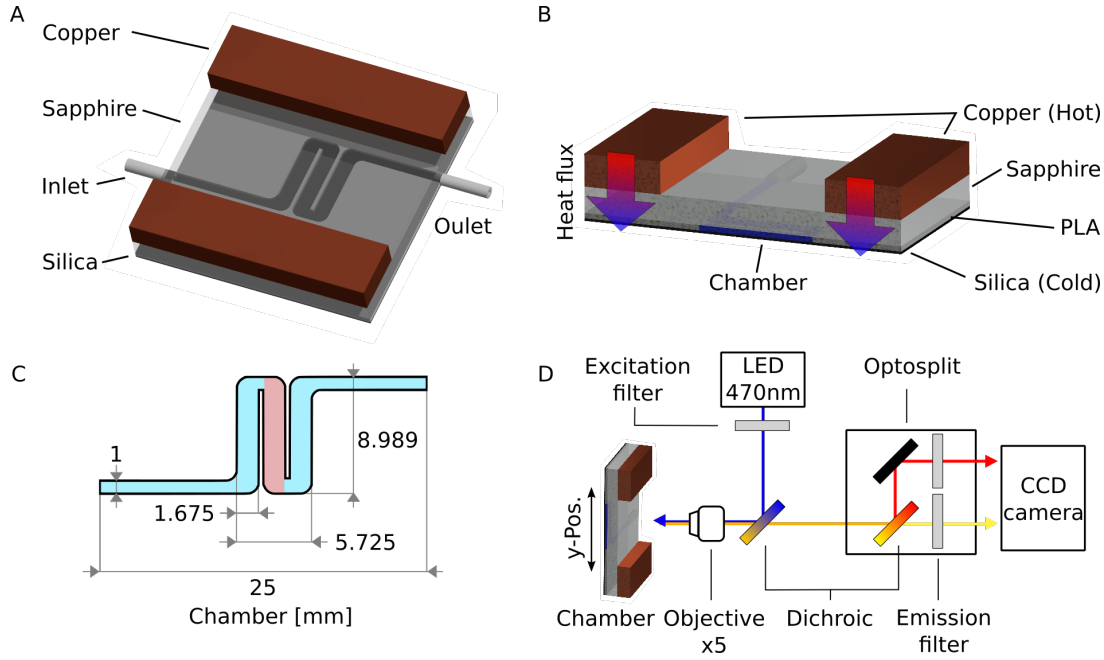


Figure 4.6: Experimental set-up for the generation and detection of pH gradients across water filled compartments. (A, B) Schematics of the experimental device. An S-shaped flow chamber is sandwiched between a transparent sapphire and a silica wafer. The copper/sapphire is heated up by an PID-controlled resitor while simultaneously cooling the silica wafer by a water bath. As a result, a fully tunable temperature gradient is formed across the flow chamber. (C) The boundaries of the flow chamber are fabricated by a 3-D printer using polylactic acid (PLA) material. The inner dimensions of the chamber (blue, red) are reduced from 1.675 x 0.2 x 8.989 mm to 1.2 x 0.17 x 8.0 mm during the enclosing process, where the sapphire windows is placed on top of the printed structure at 150 °C for approx. 10 min. The pH is measured in the marked area of the chamber (red). (D) The ratiometric detection of the pH is achieved by a fluorescence microscope equipped with an optosplit. The y-position of the chamber can be shifted by a translation stage to simultaneously detect pH changes along the height of the chamber.

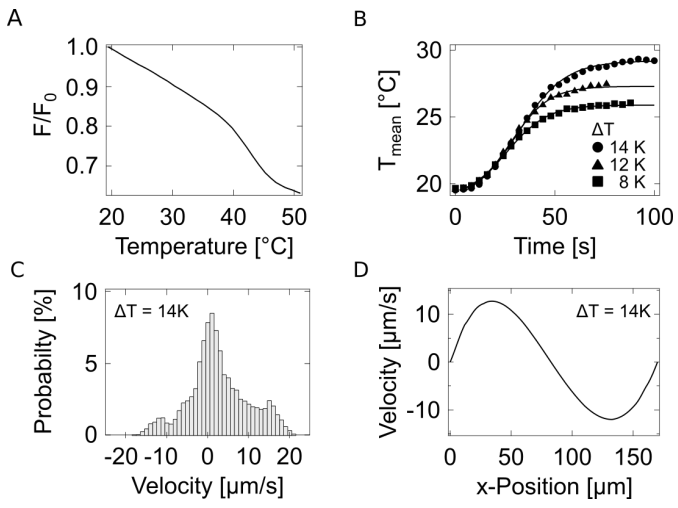


Figure 4.7: Calibration of the temperature gradient across the flow chamber. (A) Reference measurement for the temperature dependent fluorescence of (2',7'-Bis-(2-Carboxyethyl)-5-(and-6)-Carboxyfluorescein) (BCECF, 50 µM) diluted in TRIS buffer (10 mM). (B) The thermal equilibrium of a flow chamber was reached after maximum 90 s, exhibiting a mean temperature of 25.9 °C, 27.3 °C and 29.2 °C by applying a temperature gradient of ΔT = 8, 12 and 14 K, respectively. (C) The convection speeds inside a flow chamber were measured by using fluorescently labeled beads in a thermal gradient of ΔT = 14 K. (D) The experimental convection speeds are in good agreement with a 2-D finite-element simulations.

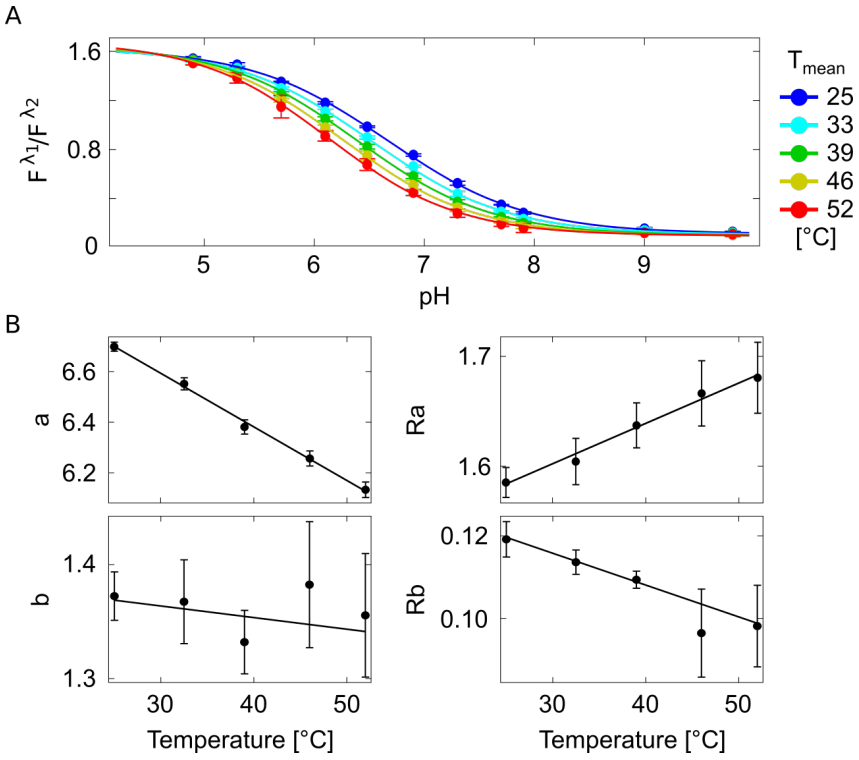


Figure 4.8: pH calibration curve for SNARF-1. (A) The fluorescence intensity ratio $R = F^{\lambda_1}/F^{\lambda_2}$ was measured for acidic and basic pH values in the range of pH 4.9 to 9.8 at different temperatures. F^{λ_1} and F^{λ_2} denote the fluorescence intensity of SNARF-1 at λ_1 = 580 nm and λ_2 = 640 nm. A modified Henderson-Hasselbalch equation (Supplementary eq. 4.10) comprising four parameters (Ra, Rb, a, b) was fitted to the ratio-to-pH calibration curve. The parameters Ra and Rb account for the maximum and minimum fluorescence rations while a and b denote for the pK_a of SNARF-1 and the weighting of the spectra. The pH calibration curve shifts with respect to the mean temperature. (B) The pH calibration curve was adjusted to the mean temperature of each measurement. Hereby, the parameters a, b, Ra, Rb are fitted linearly. The error bars depicts the standard deviation of the SNARF-1 fluorescence.

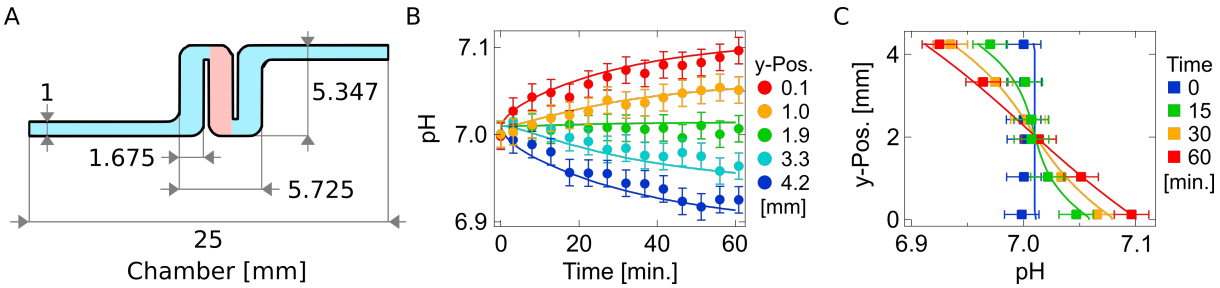


Figure 4.9: pH gradient in a shortened flow chamber. (A) Dimensions of the shortened flow chamber in mm. Reducing the height of the chamber from 8.9 to 5.347 mm results in an effective height of 4.3 mm in the measurement area (red) after the enclosure step (see Fig. 4.6). (B, C) The height of the flow chamber increases the equilibration time and the steepness of the proton gradient. By reducing the height of the flow chamber, the pH gradient of a phosphate buffer solution decreased from ΔpH = 0.47 to ΔpH = 0.17 after 60 min (see Supplementary fig. 4.13). However, the shortened chamber reached its equilibrium after 60 min, approximately four times faster compared to the 8 mm flow chamber. Finite-element simulations are in good agreement with experimental data. The error bars depicts the standard deviation of the SNARF-1 detection method.

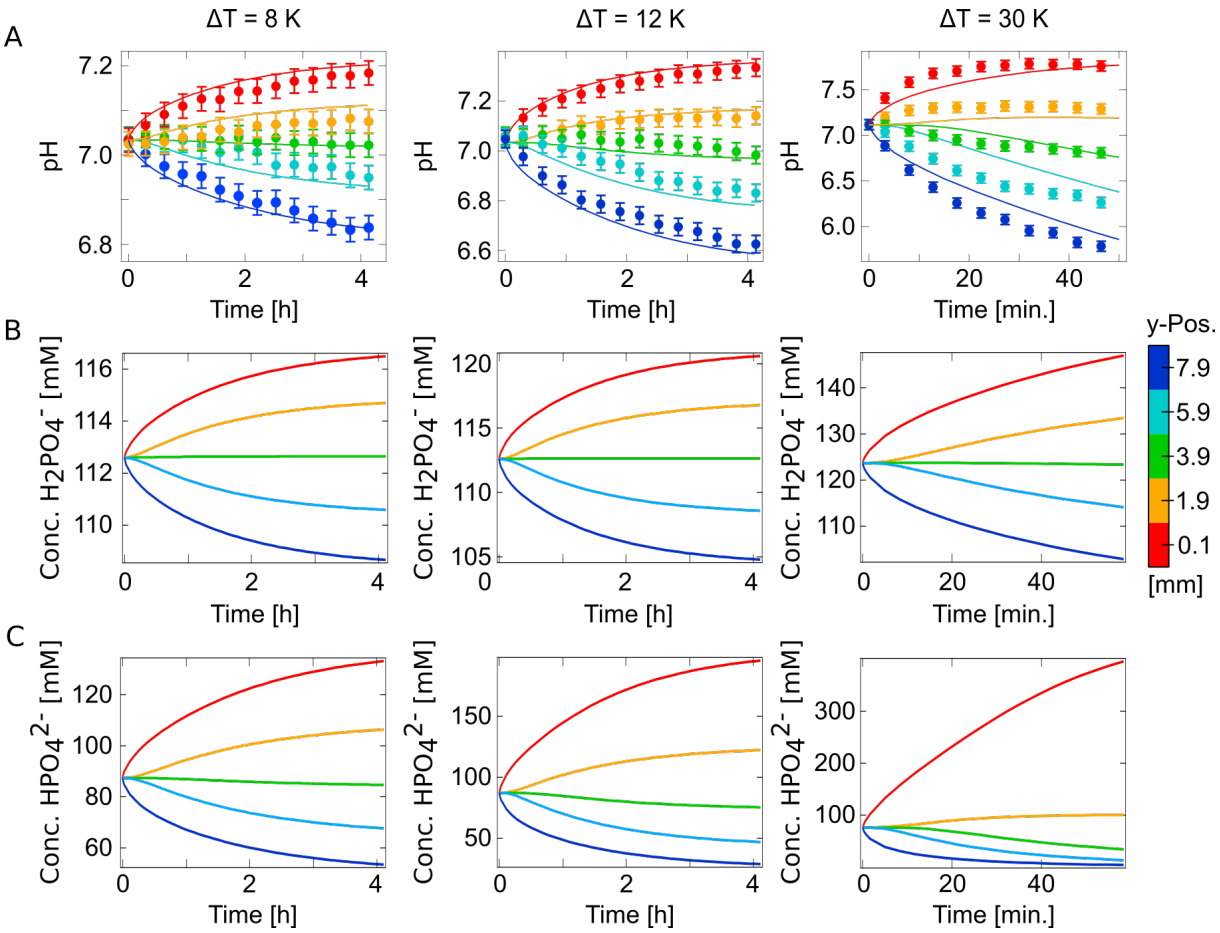


Figure 4.10: pH gradient increases with temperature difference. (A) The pH gradient was measured for temperature differences of ΔT = 8, 12, 30 K in 200 mM phosphate buffer, reaching an average temperature of 25.9, 27.3, 37.1 °C, respectively. The maximum pH gradient was formed between the upper and lower regions of the chamber and increases with the steepness of the thermal gradient. The formation of a pH gradient is modulated by a finite element simulation and fits well the experimental data. (B, C) *Insilico* accumulation of hydrogen phosphate and dihydrogen phosphate at specific positions in the chamber. Hydrogen phosphate accumulates more strongly at the bottom of the chamber compared to dihydrogen phosphate due to its higher Soret coefficient. As a result, a conjugate acid, dihydrogen phosphate, is formed from the reception of a proton by hydrogen phosphate. The protonation reaction locally reduces the oxonium ion concentration and therefore increase the pH at the bottom of the chamber. Phosphate and *ortho*-phosphoric acid are also included in the simulation, however the concentration of both molecules is 10^4-times lower and are therefore not depicted. The error bars depicts the standard deviation of the SNARF-1 detection method.

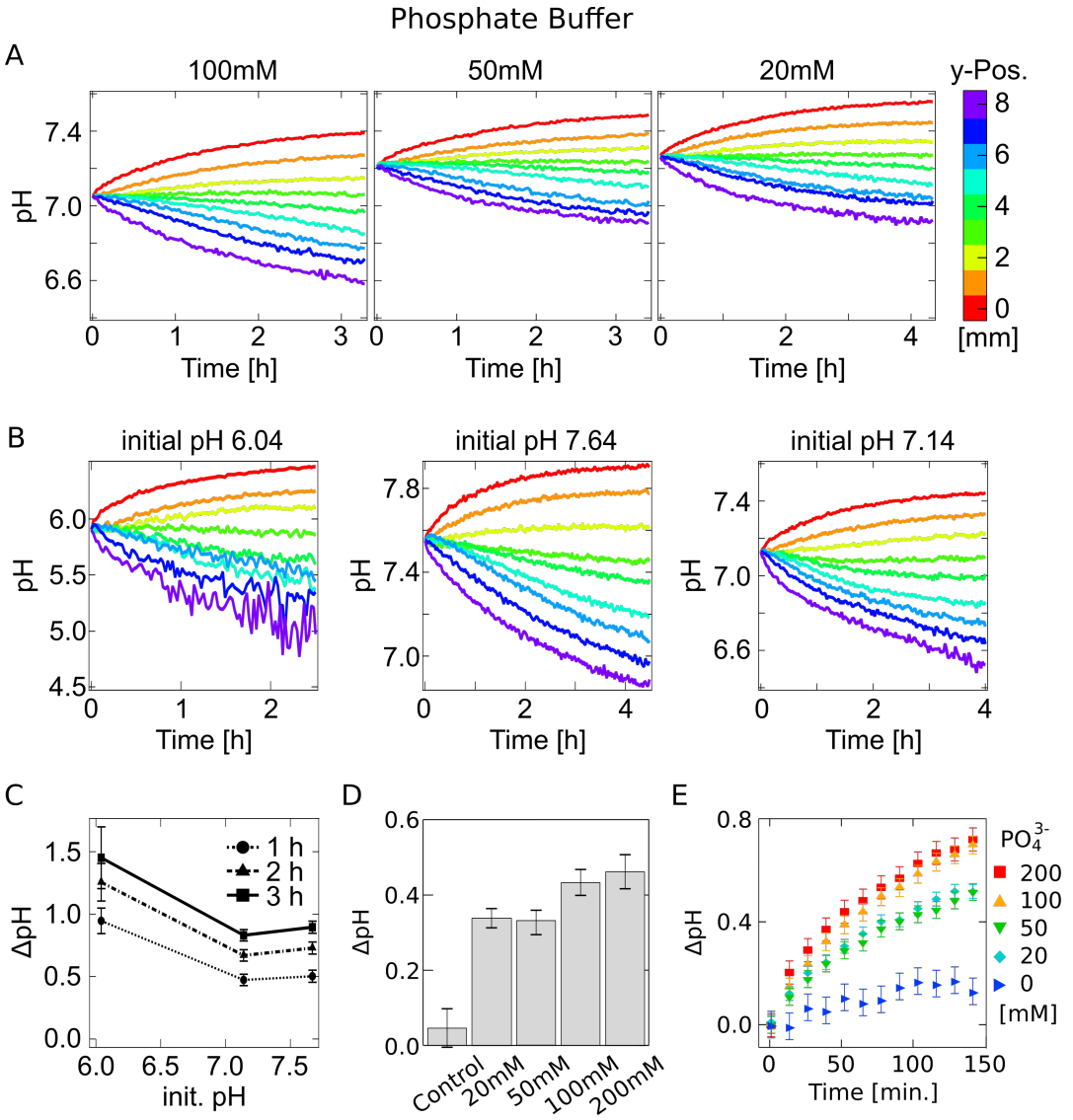


Figure 4.11: Effects of phosphate buffer concentration and initial pH on the pH gradient. (A) The pH gradient slightly increases with the phosphate buffer concentration, shown for 20, 50 and 100 mM for different y-positions in the chamber. (B, C) The initial pH affects the pH gradient, achieving the highest pH gradient for a low initial pH. Large pH fluctuation during the pH 6.04 measurement can be attributed to the detection limit of SNARF-1 from pH 5 to 9. (D) A pH gradient of ΔpH 0.47, 0.44, 0.34, 0.33 and 0.05 was formed for phosphate buffer concentrations of 100, 50, 20 and 0 mM after 60 min. (E) The maximum pH gradient increases for longer experimental times. The error bars depict the standard deviation of the SNARF-1 detection method.

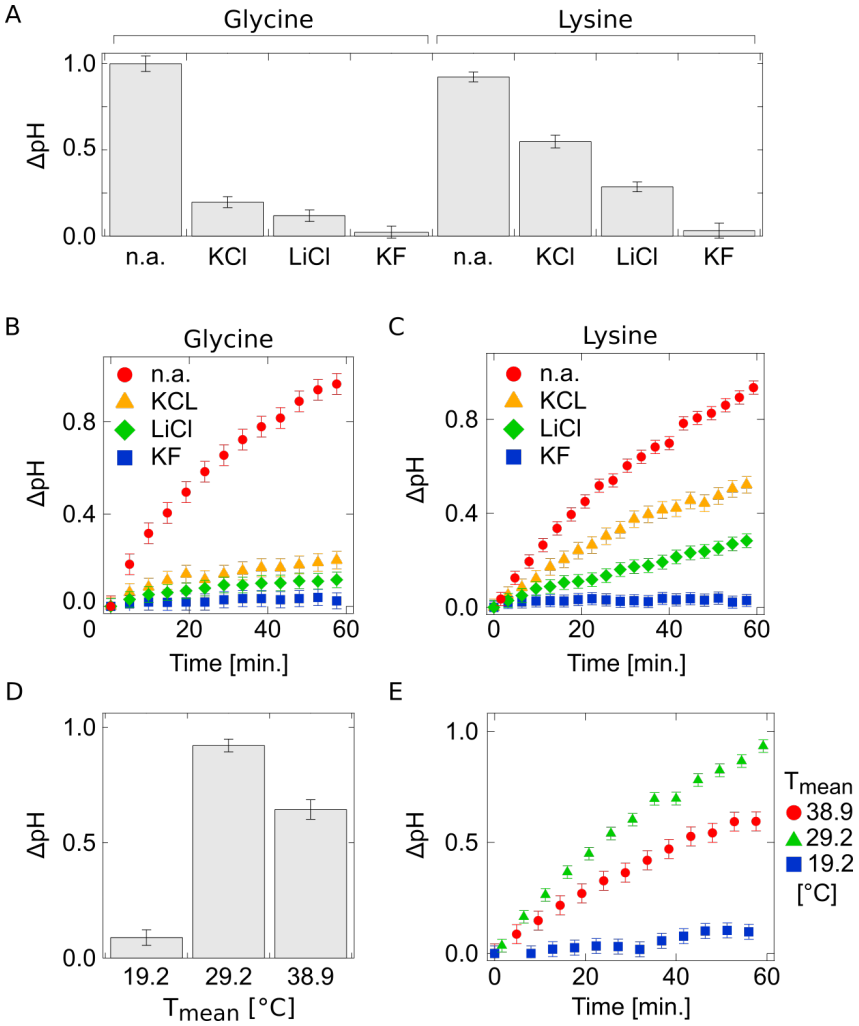


Figure 4.12: Ionic composition and temperature affect the formation of pH gradients. (A - C) The Seebeck contribution S_T^{EL} (see Supplementary eq. 4.13) affects the formation of the pH gradient. In accordance with Reichl *et al.* [57], the addition of the salt species LiCl or KF decrease the Seebeck contribution S_T^{EL} while the addition of KCL results in an increase of S_T^{EL}. The overall lower pH gradient can be attributed to the high ionic strength, which decreases the Debye length and therefore the accumulation efficiency. The measurements were performed using 50 µm SNARF-1, 200 mM phosphate buffer, 500 mM salt concentrations and a temperature gradient of 14 K. (D, E) The pH gradient can also be modulated by the overall temperature based on the temperature dependence of the Soret coefficients.[57] The maximum pH gradient was formed for a mean temperature of 29.2 °C while the pH gradient decreases for higher and lower temperatures. The pH gradient strongly increases between T_{mean} = 19.2 °C and 29.2 °C based on the temperature dependence of the Soret coefficient S_T. The decrease in ΔpH for T_{mean} = 38.9 °C can be attributed to the accumulation behavior of the thermal trap. Here, the accumulation prefactor (see Eq. (4.2), (4.3)) strongly decrease since the width of the thermal trap is specifically designed for a certain diffusion coefficient (see Fig S12).

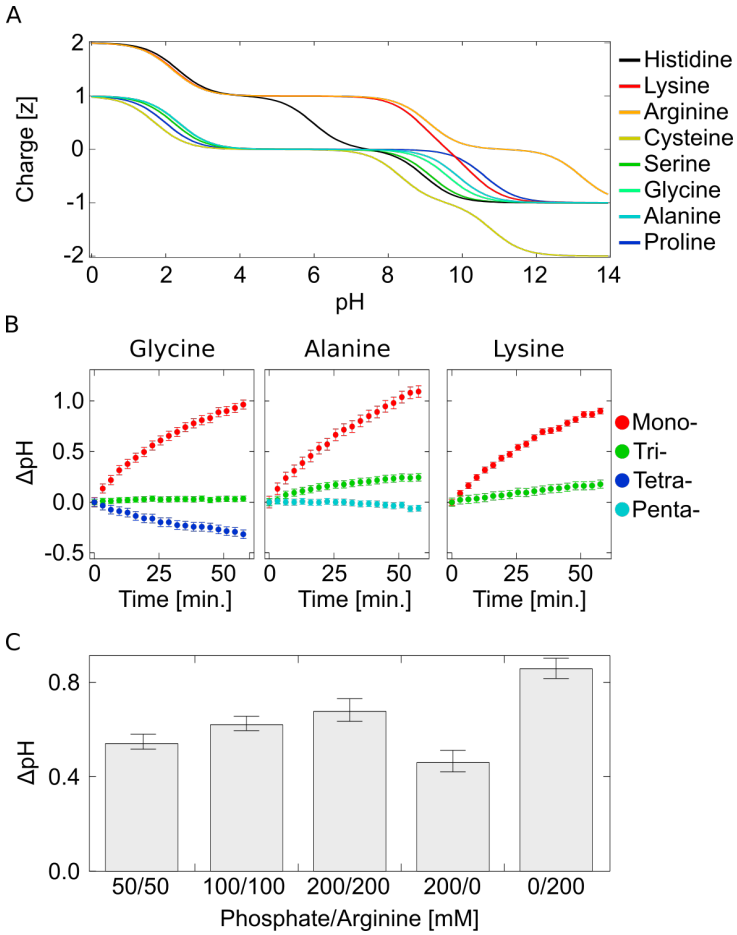


Figure 4.13: Formation of pH gradients with respect to charge, peptide length and amino acid/buffer mixtures. (A) The net charge of amino acids decreases with respect to the surrounding pH as a result of the alpha-amino group (pK ≈ 9), the carboxylic acid group (pK ≈ 2) and the side chain (see Supplementary eq. 4.16). Since the initial pH was adjusted to be situated in the detection range of SNARF-1 (pH 5-9), solely the dissociation of the alpha-amino group and side-chain are detectable. (B) The formation of a pH gradient depends on the length of the peptide. For Alanine, Glycine and Lysine, single amino acid achieved the highest pH gradient, while Tetra- and Pentapetides resulted in an inversed pH gradient. An inversion of the pH gradient can be attributed to negative difference in Soret coefficient $\Delta S_T = S_{T\,\text{acceptor}} - S_{T\,\text{donor}}$. (C) The mixture of amino acid and phosphate buffer strongly affects the formation of pH gradients. Single-component buffers of phosphate and arginine formed a pH gradient of ΔpH = 0.47 and 0.86, respectively, whereas a combination of both buffers formed intermediate pH gradients. The decrease in ΔpH for 200 mM of the two component mixture can be attributed to an increase in Debye length λ_D, which in turn results in a decrease in S_T. The error bars depicts the standard deviation of the SNARF-1 detection method.

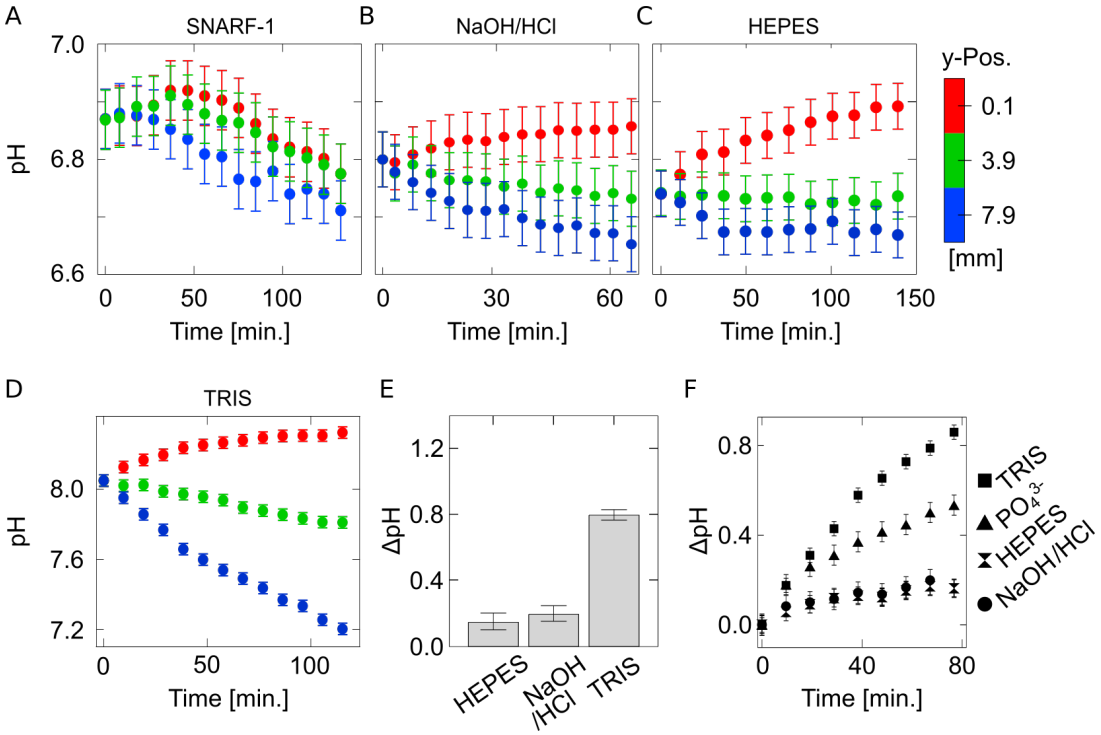


Figure 4.14: Effects of buffers and ions on the pH gradient. (A) The pH indicator SNARF-1 also acts as a buffer (pK ≈ 7.5), forming a maximum pH gradient of ΔpH = 0.12 for 50 µM SNARF-1 concentration. This concentration is used for all pH measurements. (B) The addition of 0.2 M NaOH/HCl resulted in an increase in the pH gradient. (C, D) The magnitude of the pH difference depends on the buffer. HEPES formed a shallow pH gradient while TRIS formed a relative steep pH gradient, both at a concentration of 0.2 M. (E, F) A pH gradient of ΔpH = 0.14, 0.19 and 0.79 was formed for HEPES, NaCl/HCl and TRIS after 60 minutes, respectively. The error bars depicts the standard deviation of the SNARF-1 detection method.

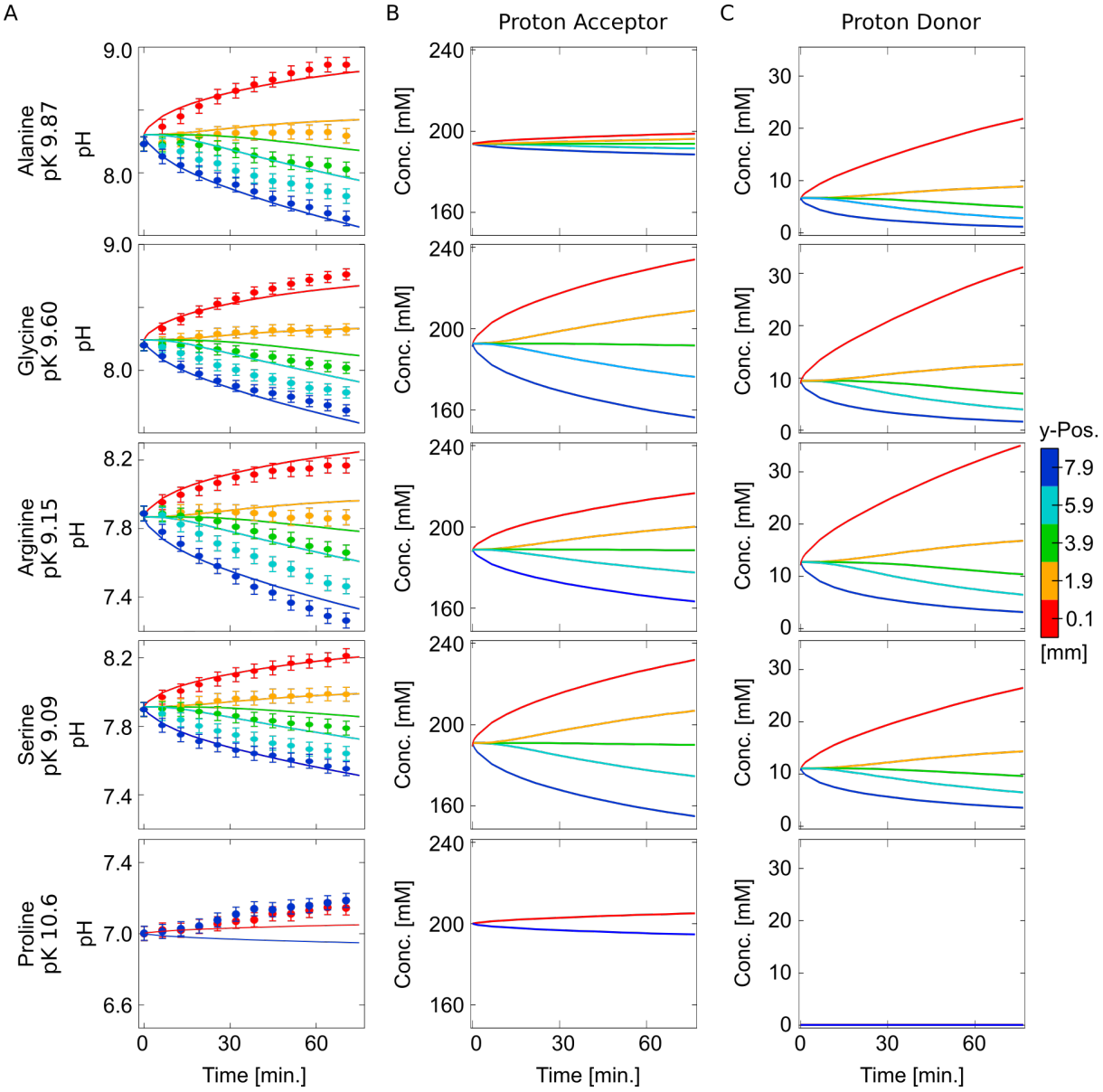


Figure 4.15: Formation of pH gradients for prebiotic plausible amino acids. (A) The pH gradient was measured for different amino acids with respect to the y-position in the chamber. The measurements were performed at an amino acid concentration of 200 mM. The starting pH was adjusted to be situated close to the pK of the amine group to ensure a sufficient quantity of both molecules, proton acceptor and donor. Finite-element simulation reproduce the increase in pH gradient by assuming a $\Delta S_T = S_{T\,\text{acceptor}} - S_{T\,\text{donor}}$ of 0.022, 0.019, 0.016, 0.012 and 0.003 K^{-1} for alanine, glycine, arginine, serine and proline, respectively. (B, C) Both species, proton acceptor and donor, exhibit an increase in concentration at the bottom and a depletion at the top of the chamber. However, the accumulation of the proton acceptor is much stronger compared to the proton donor. The error bars depicts the standard deviation of the SNARF-1 detection method.

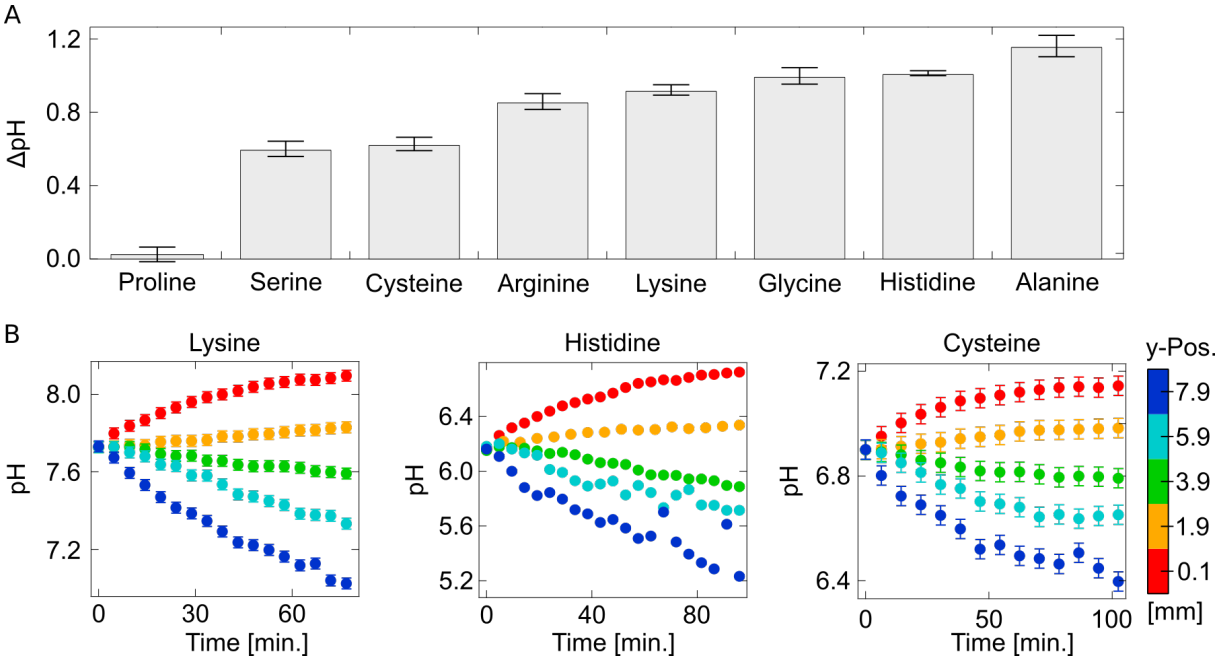


Figure 4.16: pH gradients for various amino acids in a thermal gradient. (A) The magnitude of the pH gradient strongly depends on the amino acid. After 60 min, alanine formed the highest pH gradient of ΔpH = 1.16 compared to the lowest ΔpH = 0.03 formed by proline. The measurements were performed at an amino acid concentration of 200 mM and 50 µM SNARF-1. (B) The pH was measured for lysine, histidine and cysteine with respect to the y-position in the chamber. In the course of the experiment, the pH decreased in the lower regions and increased in the upper regions of the chamber. The error bars depicts the standard deviation of the SNARF-1 detection method.

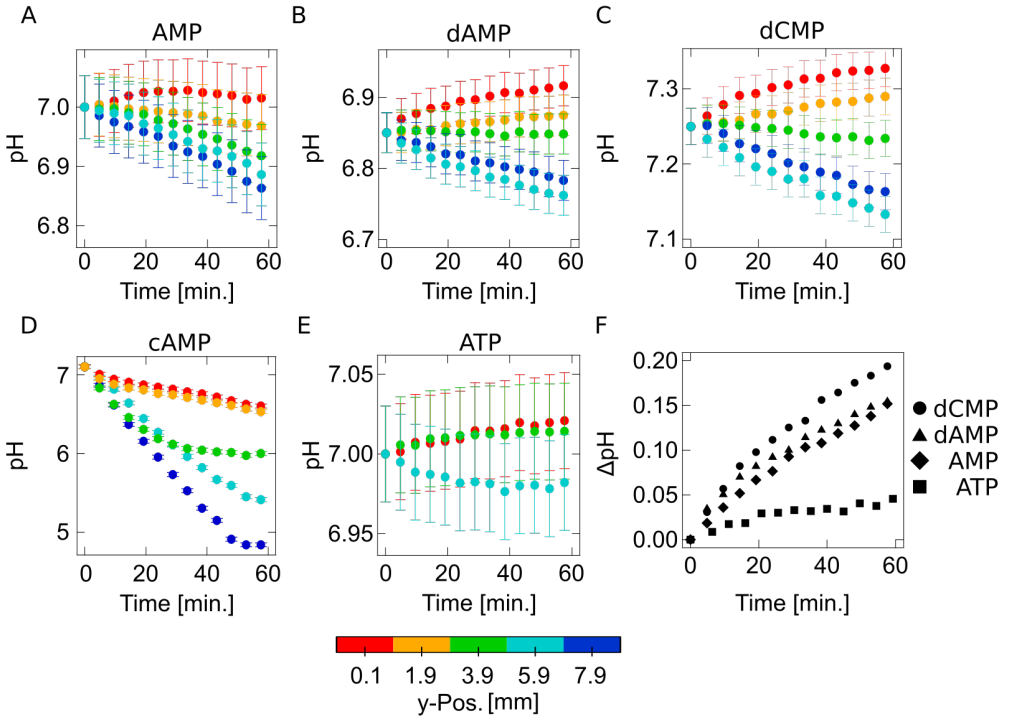


Figure 4.17: Formation of a pH gradients in solutions of RNA/DNA nucleotides. (A-E) Single nucleotides in solution with a concentration of 200 mM formed pH gradients of up to ΔpH = 0.2 within 60 min. The lowest pH gradient was achieved for ATP, having a pH gradient of ΔpH = 0.05. In the case of cAMP, a global shift in pH of several units was measured that cannot be described by the standard accumulation model. (F) dAMP, AMP and dCMP formed comparable pH gradients between upper and lower regions of the pore. The error bars give the standard deviation of the SNARF-1 detection.

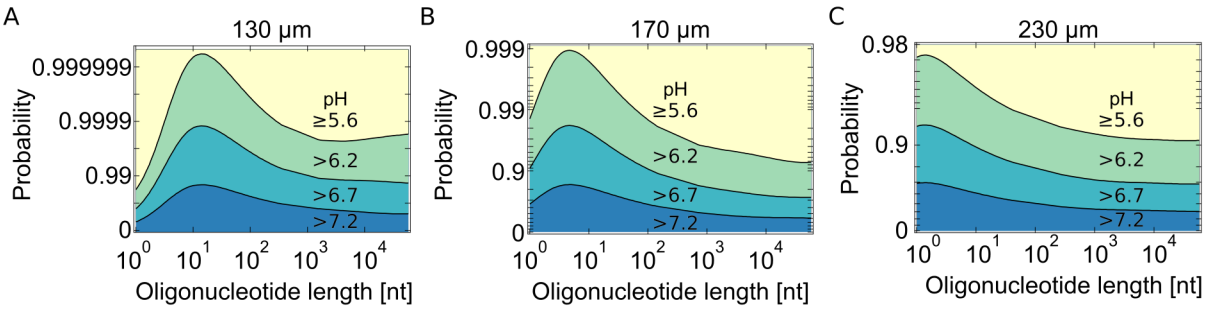


Figure 4.18: Accumulation kinetics of DNA/RNA strands depends on flow chamber thickness. (A) A flow chamber with a thickness of 130 µm strongly accumulates larger molecules such as long DNA/RNA strands at the bottom, while the concentration distribution along the y-axis of short molecules is mostly given by Brownian motion. (B,C) Short molecules accumulate more efficiently in wider chambers such as 170 µm and 230 µm. Here, high convection speeds can counterbalance the fast diffusion of small molecules. The flow chamber can be specifically adjusted that certain molecules remain at defined pH regions or undergo frequent pH oscillations.

Bibliography

[1] Lorenz Keil, Michael Hartmann, Simon Lanzmich, and Dieter Braun. Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations. *Phys. Chem. Chem. Phys.*, 18(30):20153–20159, 2016. doi: 10.1039/c6cp00577b.

[2] Steven A. Benner. Defining life. *Astrobiology*, 10(10):1021–1030, 2010. ISSN 1531-1074, 1557-8070. doi: 10.1089/ast.2010.0524.

[3] H. J. Morowitz. *Energy flow in biology*. Ox Bow, S.l., 1979. ISBN 0918024137.

[4] C. Huber and G. Wachtershauser. Activated acetic acid by carbon fixation on (fe,ni)s under primordial conditions. *Science*, 276(5310):245–247, 1997. ISSN 0036-8075, 1095-9203.

[5] Joshua Jortner. Conditions for the emergence of life on the early earth: summary and reflections. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 361(1474):1877–1891, 2006. doi: 10.1098/rstb.2006.1909.

[6] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin, and E. V. Koonin. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Natl. Acad. Sci. U.S.A.*, 109(14): E821–E830, 2012. doi: 10.1073/pnas.1117774109.

[7] Matthew W. Powner and John D. Sutherland. Phosphate-mediated interconversion of ribo- and arabino-configured prebiotic nucleotide intermediates. *Angew. Chem. Int. Ed.*, 49(27):4641–4643, 2010. ISSN 14337851. doi: 10.1002/anie.201001662.

[8] Armen Y. Mulkidjanian, Kira S. Makarova, Michael Y. Galperin, and Eugene V. Koonin. Inventing the dynamo machine: the evolution of the f-type and v-type atpases. *Nat. Rev. Microbiol.*, 5(11):892–899, 2007. doi: 10.1038/nrmicro1767.

[9] Charles Darwin. *On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life*. Dover Publications, Mineola, NY, dover giant thrift ed edition, 2006. ISBN 0486450066.

[10] Leslie E. Orgel. Prebiotic chemistry and the origin of the rna world. *Crit. Rev. Biochem. Mol. Biol.*, 39(2):99–123, 2004. ISSN 1040-9238. doi: 10.1080/10409230490460765.

[11] J. Oró and A. P. Kimball. Synthesis of purines under possible primitive earth conditions. i. adenine from hydrogen cyanide. *Arch. Biochem. Biophys.*, 94(2):217–227, 1961. ISSN 0003-9861. doi: 10.1016/0003-9861(61)90033-9.

[12] S. L. Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117(3046):528–529, 1953. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.117.3046.528.

[13] Arthur L Weber and Stanley L Miller. Reasons for the occurrence of the twenty coded protein amino acids. *J. Mol. Evol.*, 17(5):273–284, 1981. doi: 10.1007/bf01795749.

[14] Eric T Parker, Henderson J Cleaves, Jason P Dworkin, Daniel P Glavin, Michael Callahan, Andrew Aubrey, Antonio Lazcano, and Jeffrey L Bada. Primordial synthesis of amines and amino acids in a 1958 miller h2s-rich spark discharge experiment. *Proc. Natl. Sci. U.S.A.*, 108(14):5526–5531, 2011. doi: 10.1073/pnas.1019191108.

[15] Matthew W. Powner, Béatrice Gerland, and John D. Sutherland. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, 459(7244): 239–242, 2009. ISSN 0028-0836. doi: 10.1038/nature08013.

[16] Bhavesh H. Patel, Claudia Percivalle, Dougal J. Ritson, Colm D. Duffy, and John D. Sutherland. Common origins of rna, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.*, 7(4):301–307, 2015. doi: 10.1038/nchem.2202.

[17] K. Dose. Peptides and amino acids in the primordial hydrosphere. *Bio Systems*, 6(4): 224–228, 1975. ISSN 0303-2647. doi: 10.1016/0303-2647(75)90064-7.

[18] S. J. Mojzsis, T. M. Harrison, and R. T. Pidgeon. Oxygen-isotope evidence from ancient zircons for liquid water at the earth's surface 4,300 myr ago. *Nature*, 409(6817):178–181, 2001. ISSN 0028-0836. doi: 10.1038/35051557.

[19] Nick Lane. *Life ascending: the ten great inventions of evolution.* Profile books, 2010. ISBN 1861978189.

[20] John Oró. Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun.*, 2(6):407–412, 1960. ISSN 0006-291X. doi: 10.1016/0006-291x(60)90138-8.

[21] J. Oró and A. P. Kimball. Synthesis of purines under possible primitive earth conditions. *Arch. Biochem. Biophys.*, 96(2):293–313, 1962. ISSN 0003-9861. doi: 10.1016/0003-9861(62)90412-5.

[22] Bradley T. Burcar, Laura M. Barge, Dustin Trail, E. Bruce Watson, Michael J. Russell, and Linda B. McGown. Rna oligomerization in laboratory analogues of alkaline hydrothermal vent systems. *Astrobiology*, 15(7):509–522, 2015. ISSN 1531-1074, 1557-8070. doi: 10.1089/ast.2014.1280.

[23] H. Kuhn and J. Waser. Evolution of early mechanisms of translation of genetic information into polypeptides. *Nature*, 298(5874):585–586, 1982. ISSN 0028-0836. doi: 10.1038/298585a0.

[24] Manfred Eigen. Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften*, 58(10):465–523, 1971. ISSN 0028-1042, 1432-1904. doi: 10.1007/BF00623322.

[25] G. F. Joyce. Rna evolution and the origins of life. *Nature*, 338(6212):217–224, 1989. ISSN 0028-0836. doi: 10.1038/338217a0.

[26] D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369(6477):221–224, 1994. ISSN 0028-0836. doi: 10.1038/369221a0.

[27] Robert Pascal, Addy Pross, and John D. Sutherland. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open biology*, 3(11):130156, 2013. ISSN 2046-2441. doi: 10.1098/rsob.130156.

[28] E. Schrödinger. *What is life?* Cambridge Univ. Press, 1944.

[29] K. E. Nelson, M. P. Robertson, M. Levy, and S. L. Miller. Concentration by evaporation and the prebiotic synthesis of cytosine. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 31(3):221–229, 2001. ISSN 0169-6149.

[30] J. Oro and A. Lazcano. A minimal living system and the origin of a protocell. *Advances in space research : the official journal of the Committee on Space Research (COSPAR)*, 4(12): 167–176, 1984. ISSN 0273-1177. doi: 10.1016/0273-1177(84)90559-3.

[31] T. Oberholzer and P. L. Luisi. The use of liposomes for constructing cell models. *J. Biol. Phys.*, 28(4):733–744, 2002. ISSN 0092-0606. doi: 10.1023/A:1021267512805.

[32] Shogo Koga, David S. Williams, Adam W. Perriman, and Stephen Mann. Peptide-nucleotide microdroplets as a step towards a membrane-free protocell model. *Nat. Chem.*, 3(9):720–724, 2011. ISSN 1755-4330, 1755-4349. doi: 10.1038/nchem.1110.

[33] J. P. Ferris, A. R. Hill, JR, R. Liu, and L. E. Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61, 1996. ISSN 0028-0836. doi: 10.1038/381059a0.

[34] C. Huber. Peptides by activation of amino acids with co on (ni,fe)s surfaces: Implications for the origin of life. *Science*, 281(5377):670–672, 1998. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.281.5377.670.

[35] S. C. Dahm, W. B. Derrick, and O. C. Uhlenbeck. Evidence for the role of solvated metal hydroxide in the hammerhead cleavage mechanism. *Biochemistry (Mosc.)*, 32(48): 13040–13045, 1993. ISSN 0006-2960. doi: 10.1021/bi00211a013.

[36] S. C. Dahm and O. C. Uhlenbeck. Role of divalent metal ions in the hammerhead rna cleavage reaction. *Biochemistry*, 30(39):9464–9469, 1991. ISSN 0006-2960.

[37] R. B. van Atta and S. M. Hecht. A ribozyme model: Site-specific cleavage of an rna substrate by mn2+. *Adv. Inorg. Biochem.*, 9:1–40, 1994. ISSN 0190-0218.

[38] Michael J. Russell, Allan J. Hall, Adrian J. Boyce, and Anthony E. Fallick. On hydrothermal convection systems and the emergence of life. *Econ. Geol.*, 100(3):419–438, 2005.

[39] Deborah S. Kelley, Jeffrey A. Karson, Gretchen L. Früh-Green, Dana R. Yoerger, Timothy M. Shank, David A. Butterfield, John M. Hayes, Matthew O. Schrenk, Eric J. Olson, Giora Proskurowski, Mike Jakuba, Al Bradley, Ben Larson, Kristin Ludwig, Deborah Glickson, Kate Buckman, Alexander S. Bradley, William J. Brazelton, Kevin Roe, Mitch J. Elend, Adélie Delacour, Stefano M. Bernasconi, Marvin D. Lilley, John A. Baross, Roger E. Summons, and Sean P. Sylva. A serpentinite-hosted ecosystem: The lost city hydrothermal field. *Science*, 307(5714):1428–1434, 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1102556.

[40] Philipp Baaske, Stefan Duhr, and Dieter Braun. Melting curve analysis in a snapshot. *Appl. Phys. Lett.*, 91(13):133901, 2007. doi: 10.1063/1.2790806.

[41] Itay Budin, Raphael J. Bruckner, and Jack W. Szostak. Formation of protocell-like vesicles in a thermal diffusion column. *J. Am. Chem. Soc.*, 131(28):9628–9629, 2009. doi: 10.1021/ja9029818.

[42] Dieter Braun and Albert Libchaber. Computer-based photon-counting lock-in for phase detection at the shot-noise limit. *Opt. Lett.*, 27(16):1418–1420, 2002. ISSN 0146-9592. doi: 10.1364/ol.27.001418.

[43] C. B. Mast, S. Schink, U. Gerland, and D. Braun. Escalation of polymerization in a thermal gradient. *Proc. Natl. Acad. Sci. U.S.A.*, 110(20):8030–8035, 2013. doi: 10.1073/pnas.1303222110.

[44] Moritz Kreysing, Lorenz Keil, Simon Lanzmich, and Dieter Braun. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nat. Chem.*, 7(3):203–208, 2015. doi: 10.1038/nchem.2155.

[45] Barry Herschy, Alexandra Whicher, Eloi Camprubi, Cameron Watson, Lewis Dartnell, John Ward, Julian R. G. Evans, and Nick Lane. An origin-of-life reactor to simulate alkaline hydrothermal vents. *J. Mol. Evol.*, 79(5-6):213–227, 2014. ISSN 0022-2844, 1432-1432. doi: 10.1007/s00239-014-9658-4.

[46] T. M. McCollom, G. Ritter, and B. R. Simoneit. Lipid synthesis under hydrothermal conditions by fischer-tropsch-type reactions. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 29(2):153–166, 1999. ISSN 0169-6149.

[47] R. Büttner, B. Zimanowski, J. Blumm, and L. Hagemann. Thermal conductivity of a volcanic rock material (olivine-melilitite) in the temperature range between 288 and 1470 k. *J. Volcanol. Geotherm. Res.*, 80(3-4):293–302, 1998. ISSN 03770273. doi: 10.1016/S0377-0273(97)00050-4.

[48] S. S. Mansy and J. W. Szostak. Thermostability of model protocell membranes. *Proc. Natl. Acad. Sci. U.S.A.*, 105, 2008. doi: 10.1073/pnas.0805086105.

[49] A. Wochner, J. Attwater, A. Coulson, and P. Holliger. Ribozyme-catalyzed transcription of an active ribozyme. *Science*, 332(6026):209–212, 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1200752.

[50] N. Paul and G. F. Joyce. A self-replicating ligase ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12733–12740, 2002. ISSN 0027-8424. doi: 10.1073/pnas.202471099.

[51] Zunyi Yang, Fei Chen, J. Brian Alvarado, and Steven A. Benner. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.*, 133(38):15105–15112, 2011. ISSN 0002-7863. doi: 10.1021/ja204910n.

[52] M. Krishnan. Pcr in a rayleigh-benard convection cell. *Science*, 298(5594):793, 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.298.5594.793.

[53] Sudha Rajamani, Justin K. Ichida, Tibor Antal, Douglas A. Treco, Kevin Leu, Martin A. Nowak, Jack W. Szostak, and Irene A. Chen. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.*, 132(16): 5880–5885, 2010. ISSN 0002-7863, 1520-5126. doi: 10.1021/ja100780p.

[54] Klaus Clusius and Gerhad Dickel. Neues verfahren zur gasentmischung und isotopentrennung. *Die Naturwissenschaften*, 26(33):546, 1938. ISSN 0028-1042, 1432-1904. doi: 10.1007/BF01675498.

[55] Alois Würger. Hydrodynamic boundary effects on thermophoresis of confined colloids. *Phys. Rev. Lett.*, 116(13):138302, 2016. ISSN 0031-9007. doi: 10.1103/physrevlett.116.138302.

[56] Jan K. G. Dhont, S. Wiegand, S. Duhr, and D. Braun. Thermodiffusion of charged colloids: single-particle diffusion. *Langmuir : the ACS journal of surfaces and colloids*, 23(4): 1674–1683, 2007. ISSN 0743-7463. doi: 10.1021/la062184m.

[57] Maren Reichl, Mario Herzog, Alexandra Götz, and Dieter Braun. Why charged molecules move across a temperature gradient: The role of electric fields. *Phys. Rev. Lett.*, 112(19), 2014. ISSN 0031-9007. doi: 10.1103/PhysRevLett.112.198101.

[58] Philipp Reineck, Christoph J. Wienken, and Dieter Braun. Thermophoresis of single stranded dna. *Electrophoresis*, 31(2):279–286, 2010. ISSN 0173-0835. doi: 10.1002/elps.200900505.

[59] Stefan Duhr and Dieter Braun. Optothermal molecule trapping by opposing fluid flow with thermophoretic drift. *Phys. Rev. Lett.*, 97(3):038103, 2006. ISSN 0031-9007.

[60] Philipp Baaske, Franz M. Weinert, Stefan Duhr, Kono H. Lemke, Michael J. Russell, and Dieter Braun. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl. Acad. Sci. U.S.A.*, 104(22):9346–9351, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0609592104.

[61] Jack W. Szostak. The eightfold path to non-enzymatic rna replication. *J. Syst. Chem.*, 3 (1):2, 2012. ISSN 1759-2208. doi: 10.1186/1759-2208-3-2.

[62] Mills, R. L. Peterson, and Sol Spiegelman. An extracellular darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl. Acad. Sci. U.S.A.*, 58(1):217–224, 1967. ISSN 0027-8424. doi: 10.1073/pnas.58.1.217.

[63] Thorne Lay, John Hernlund, and Bruce A. Buffett. Core–mantle boundary heat flow. *Nat. Geosci.*, 1(1):25–32, 2008. ISSN 1752-0894. doi: 10.1038/ngeo.2007.44.

[64] Deborah S. Kelley, Jeffrey A. Karson, Donna K. Blackman, Gretchen L. Früh-Green, David A. Butterfield, Marvin D. Lilley, Eric J. Olson, Matthew O. Schrenk, Kevin K. Roe, and Geoff T. Lebon. An off-axis hydrothermal vent field near the mid-atlantic ridge at 30 n. *Nature*, 412(6843):145–149, 2001. ISSN 0028-0836. doi: 10.1038/35084000.

[65] P. Debye. Zur theorie des clusiusschen trennungsverfahrens. *Annalen der Physik*, 428 (3-4):284–294, 1939.

[66] Roberto Piazza and Andrea Guarino. Soret effect in interacting micellar solutions. *Phys. Rev. Lett.*, 88(20):208302, 2002. ISSN 0031-9007. doi: 10.1103/physrevlett.88.208302.

[67] S. Duhr and D. Braun. From the cover: Why molecules move along a temperature gradient. *Proc. Natl. Acad. Sci. U.S.A.*, 103(52):19678–19682, 2006. ISSN 0027-8424. doi: 10.1073/pnas.0603873103.

[68] Christof B. Mast and Dieter Braun. Thermal trap for dna replication. *Phys. Rev. Lett.*, 104 (18):188102, 2010. ISSN 0031-9007, 1079-7114. doi: 10.1103/physrevlett.104.188102.

[69] Dieter Braun, Noel Goddard, and Albert Libchaber. Exponential dna replication by laminar convection. *Phys. Rev. Lett.*, 91(15), 2003. ISSN 0031-9007, 1079-7114. doi: 10.1103/PhysRevLett.91.158103.

[70] Chrisantha Fernando, Günter von Kiedrowski, and Eörs Szathmáry. A stochastic model of nonenzymatic nucleic acid replication:"elongators" sequester replicators. *J. Mol. Evol.*, 64(5):572–585, 2007. ISSN 0022-2844, 1432-1432. doi: 10.1007/s00239-006-0218-4.

[71] Philipp Baaske, Christoph J. Wienken, Philipp Reineck, Stefan Duhr, and Dieter Braun. Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angew. Chem. Ed.*, 49(12):2238–2241, 2010. ISSN 1433-7851. doi: 10.1002/anie.200903998.

[72] Tom Maniatis, Andrea Jeffrey, and Hans Van deSande. Chain length determination of small double-and single-stranded dna molecules by polyacrylamide gel electrophoresis. *Biochemistry*, 14(17):3787–3794, 1975. doi: 10.1021/bi00688a010.

[73] Michael P. Robertson and Gerald F. Joyce. The origins of the rna world. *Cold Spring Harbor Perspect. Biol.*, 4(5), 2012. ISSN 1943-0264. doi: 10.1101/cshperspect.a003608.

[74] Walter Gilbert. Origin of life: The rna world. *Nature*, 319(6055):618, 1986. ISSN 0028-0836. doi: 10.1038/319618a0.

[75] David P. Horning and Gerald F. Joyce. Amplification of rna by an rna polymerase ribozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 113(35):9786–9791, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1610103113.

[76] Gerald F. Joyce. The antiquity of rna-based evolution. *Nature*, 418(6894):214–221, 2002. ISSN 0028-0836. doi: 10.1038/418214a.

[77] James Attwater, Aniela Wochner, Vitor B. Pinheiro, Alan Coulson, and Philipp Holliger. Ice as a protocellular medium for rna replication. *Nat. Commun.*, 1(6):1–8, 2010. ISSN 2041-1723. doi: 10.1038/ncomms1076.

[78] Abe Pressman, Celia Blanco, and Irene A. Chen. The rna world as a model system to study the origin of life. *Current biology : CB*, 25(19):R953–63, 2015. ISSN 0960-9822. doi: 10.1016/j.cub.2015.06.016.

[79] T. R. Cech. A model for the rna-catalyzed replication of rna. *Proc. Natl. Acad. Sci. U.S.A.*, 83(12):4360–4363, 1986. ISSN 0027-8424. doi: 10.1073/pnas.83.12.4360.

[80] J. A. Doudna, S. Couture, and J. W. Szostak. A multisubunit ribozyme that is a catalyst of and template for complementary strand rna synthesis. *Science*, 251(5001):1605, 1991. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1707185.

[81] Katarzyna Adamala, Aaron E. Engelhart, and Jack W. Szostak. Generation of functional rnas from inactive oligonucleotide complexes by non-enzymatic primer extension. *J. Am. Chem. Soc.*, 137(1):483–489, 2015. ISSN 0002-7863. doi: 10.1021/ja511564d.

[82] Matthew Levy and Stanley L. Miller. The stability of the rna bases: Implications for the origin of life. *Proc. Natl. Acad. Sci. U.S.A.*, 95(14):7933–7938, 1998. ISSN 0027-8424. doi: 10.1073/pnas.95.14.7933.

[83] M. G. AbouHaidar and I. G. Ivanov. Non-enzymatic rna hydrolysis promoted by the combined catalytic activity of buffers and magnesium ions. *Zeitschrift fur Naturforschung. C, Journal of biosciences*, 54(7-8):542–548, 1999. ISSN 0341-0382. doi: 10.1515/znc-1999-7-813.

[84] Michael D. Been and Thomas Cech. Rna as an rna polymerase: Net elongation of an rna primer catalyzed by the tetrahymena ribozyme. *Science*, 239(4846):1412–1417, 1988. ISSN 0036-8075. doi: 10.1126/science.2450400.

[85] Eric H. Ekland and David P. Bartel. Rna-catalysed rna polymerization using nucleoside triphosphates. *Nature*, 382(6589):373, 1996. ISSN 0028-0836.

[86] Jonathan T. Sczepanski and Gerald F. Joyce. A cross-chiral rna polymerase ribozyme. *Nature*, 515(7527):440–442, 2014. ISSN 0028-0836. doi: 10.1038/nature13900.

[87] W. K. Johnston. Rna-catalyzed rna polymerization: Accurate and general rna-templated primer extension. *Science*, 292(5520):1319–1325, 2001. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1060786.

[88] 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. *Nat. Commun.*, 8(1):1897, 2017. doi: 10.1038/s41467-017-02065-3.

[89] C. P. Burns. Extracellular na+ and initiation of dna synthesis: Role of intracellular ph and k+. *J. Cell Biol.*, 98(3):1082–1089, 1984. doi: 10.1083/jcb.98.3.1082.

[90] J. Pouysségur, A. Franchi, G. L'Allemain, and S. Paris. Cytoplasmic ph, a key determinant of growth factor-induced dna synthesis in quiescent fibroblasts. *FEBS Letters*, 190(1): 115–119, 1985. ISSN 00145793. doi: 10.1016/0014-5793(85)80439-7.

[91] Joseph A. Mindell. Lysosomal acidification mechanisms. *Annu. Rev. Physiol.*, 74:69–86, 2012. doi: 10.1146/annurev-physiol-012110-142317.

[92] I. H. Madshus. Regulation of intracellular ph in eukaryotic cells. *Biochem. J.*, 250(1):1–8, 1988. doi: 10.1042/bj2500001.

[93] D. E. Koshland, A. Goldbeter, and J. B. Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, 217(4556):220–225, 1982. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.7089556.

[94] Sergey N. Semenov, Albert S. Y. Wong, R. Martijn van der Made, Sjoerd G. J. Postma, Joost Groen, Hendrik W. H. van Roekel, Tom F. A. de Greef, and Wilhelm T. S. Huck. Rational design of functional and tunable oscillating enzymatic networks. *Nat. Chem.*, 7 (2):160–165, 2015. doi: 10.1038/nchem.2142.

[95] Peter Mitchell. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature*, 191(4784):144–148, 1961. ISSN 0028-0836. doi: 10.1038/191144a0.

[96] 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. doi: 10.1002/bies.200900131.

[97] Nick Lane and William F. Martin. The origin of membrane bioenergetics. *Cell*, 151(7): 1406–1416, 2012. ISSN 00928674. doi: 10.1016/j.cell.2012.11.050.

[98] Victor Sojo, Andrew Pomiankowski, and Nick Lane. A bioenergetic basis for membrane divergence in archaea and bacteria. *PLOS Biol.*, 12(8):e1001926, 2014. doi: 10.1371/journal.pbio.1001926.

[99] Saidul Islam, Dejan-Krešimir Bučar, and Matthew W. Powner. Prebiotic selection and assembly of proteinogenic amino acids and natural nucleotides from complex mixtures. *Nat. Chem.*, 2017. doi: 10.1038/nchem.2703.

[100] Adam J. Coggins and Matthew W. Powner. Prebiotic synthesis of phosphoenol pyruvate by α-phosphorylation-controlled triose glycolysis. *Nat. Chem.*, 2016. doi: 10.1038/nchem.2624.

[101] Michael J. Russell. The alkaline solution to the emergence of life: energy, entropy and early evolution. *Acta Biotheor.*, 55(2):133–179, 2007. doi: 10.1007/s10441-007-9018-5.

[102] Friederike M. Moller, Franziska Kriegel, Michael Kiess, Victor Sojo, and Dieter Braun. Steep ph gradients and directed colloid transport in a microfluidic alkaline hydrothermal pore. *Angew. Chem. Int. Ed.*, 56(9):2340–2344, 2017. doi: 10.1002/anie.201610781.

[103] Matthew A. Pasek and Dante S. Lauretta. Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early earth. *Astrobiol.*, 5(4):515–535, 2005. doi: 10.1089/ast.2005.5.515.

[104] David E. Bryant and Terence P. Kee. Direct evidence for the availability of reactive, water soluble phosphorus on the early earth. h-phosphinic acid from the nantan meteorite. *Chem. Commun.*, (22):2344–2346, 2006. doi: 10.1039/b602651f.

[105] S. L. Miller and H. C. Urey. Organic compound synthesis on the primitive earth. *Science*, 130(3370):245–251, 1959. ISSN 0036-8075, 1095-9203.

[106] Zilin Wang, Hartmut Kriegs, and Simone Wiegand. Thermal diffusion of nucleotides. *J. Phys. Chem. B*, 116(25):7463–7469, 2012. doi: 10.1021/jp3032644.

[107] D. Stadelmaier and W. Köhler. Thermal diffusion of dilute polymer solutions: The role of chain flexibility and the effective segment size. *Macromolecules*, 42(22):9147–9152, 2009. ISSN 0024-9297, 1520-5835. doi: 10.1021/ma901794k.

[108] Zilin Wang, Dzmitry Afanasenkau, Minjie Dong, Danni Huang, and Simone Wiegand. Molar mass and temperature dependence of the thermodiffusion of polyethylene oxide in water/ethanol mixtures. *The Journal of Chemical Physics*, 141(6):064904, 2014. ISSN 00219606. doi: 10.1063/1.4891720.

[109] David W. Deamer. Origins of life: How leaky were primitive cells? *Nature*, 454(7200): 37–38, 2008. ISSN 0028-0836. doi: 10.1038/454037a.

[110] Sheref S. Mansy, Jason P. Schrum, Mathangi Krishnamurthy, Sylvia Tobe, Douglas A. Treco, and Jack W. Szostak. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, 454(7200):122–125, 2008. ISSN 0028-0836. doi: 10.1038/nature07018.

[111] Thomas Le Saux, Raphael Plasson, and Ludovic Jullien. Energy propagation throughout chemical networks. *Chem. Commun.*, 50(47):6189–6195, 2014. doi: 10.1039/c4cc00392f.

[112] Herbert S. Harned and Benton B. Owen. *The Physical Chemistry of Electrolytic Solutions*. Reinhold, New York, 1958.

[113] Naokata Takeyama and Kimie Nakashima. Proportionality of intrinsic heat of transport to standard entropy of hydration for aqueous ions. *J. Solution Chem.*, 17(4):305–325, 1988. doi: 10.1007/BF00650412.

[114] C. J. Petit, M.-H Hwang, and J.-L Lin. The soret effect in dilute aqueous alkaline earth and nickel chloride solutions at 25°c. *Int. J. Thermophys.*, 7(3):687–697, 1986. doi: 10.1007/BF00502400.

[115] David R. Lide. *CRC handbook of chemistry and physics*. CRC Press, Boca Raton, 82nd, ed., 2001-2001. edition, 2001. ISBN 9780849304828.

[116] Ignacio Tinoco. *Physical chemistry: Principles and applications in biological sciences*. Pearson, Boston, 5th ed. edition, 2014. ISBN 032189104X.

[117] Youguang Ma, Chunying Zhu, Peisheng Ma, and K. T. Yu. Studies on the diffusion coefficients of amino acids in aqueous solutions. *Journal of Chemical & Engineering Data*, 50(4):1192–1196, 2005. ISSN 0021-9568. doi: 10.1021/je049582g.

[118] James E. Whitaker, Richard P. Haugland, and Franklyn G. Prendergast. Spectral and photophysical studies of benzo [c] xanthene dyes: dual emission ph sensors. *Anal. Biochem.*, 194(2):330–344, 1991. doi: 10.1016/0003-2697(91)90237-n.

[119] Alois Würger. Transport in charged colloids driven by thermoelectricity. *Phys. Rev. Lett.*, 101(10), 2008. ISSN 0031-9007. doi: 10.1103/PhysRevLett.101.108302.

[120] George Guthrie, J. Norton Wilson, and Verner Schomaker. Theory of the thermal diffusion of electrolytes in a clusius column. *The Journal of Chemical Physics*, 17(3):310, 1949. ISSN 00219606. doi: 10.1063/1.1747244.

[121] Richard A. Paselk. Principles of biochemistry. *J. Chem. Educ.*, 60(7):A201, 1983. ISSN 0021-9584. doi: 10.1021/ed060pA201.

List of Figures

List of Tables

Acknowledgments

An dieser Stelle möchte ich mich bei allen Menschen bedanken, welche mich im Laufe der Doktorarbeit unterstützt haben und zu dieser Arbeit beigetragen haben:

Dieter - für eine meist unbeschwerte, unvergessliche und äußert lehrreiche Zeit. Für die Unterstützung wenn sich Problem als äußerst hartnäckig erwiesen haben und für das Vertrauen in meine Arbeit.

Den alten wie auch neuen **Braunies, Liedls, Rädlers, und Gaubs** für ein entspanntes Arbeitsklima, der gegenseitigen Unterstützung, den vielen wissenschaftlichen Diskussionen und natürlich den Espresso-Pausen.

Ike, Simon, Christof, David und Moritz für eine tolle Zusammenarbeit an den facettenreichen und anspruchsvollen Projekten.

Fabian, Mathilde, Michi, Mike, und Thomas für deren Unterstützung beim Experimenten, Simulieren und Diskutieren. Ein besonderer Dank geht an **Matze**, der mir unschätzbar viel Hilfe beim Korrigieren der Arbeit und der Publikationen geleistet hat.

An meine **Freunde**, besonders den **Holis**, die jederzeit für Ablenkung außerhalb der Arbeit sorgten und mir mit Rat und Tat zur Seite standen.

Regina und meiner ganzen Familie - für Ihre selbstlose, geduldige und uneingeschränkte Unterstützung, die immer für mich da waren.

Article in Physical Chemistry Chemical Physics

Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations

by

Lorenz Keil, Michael Hartmann, Simon Lanzmich and Dieter Braun

published in

Physical Chemistry Chemical Physics doi: 10.1039/c6cp00577b (2016)
Reprinted from Keil *et al.* [1] with permission from the Royal Society of Chemistry

The publication in this Appendix was used as a basis for Chapter 2

PCCP

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2016, **18**, 20153

Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations†

Lorenz Keil, Michael Hartmann, Simon Lanzmich and Dieter Braun*

Received 26th January 2016,
Accepted 3rd May 2016

DOI: 10.1039/c6cp00577b

www.rsc.org/pccp

How can living matter arise from dead matter? All known living systems are built around information stored in RNA and DNA. To protect this information against molecular degradation and diffusion, the second law of thermodynamics imposes the need for a non-equilibrium driving force. Following a series of successful experiments using thermal gradients, we have shown that heat gradients across sub-millimetre pores can drive accumulation, replication, and selection of ever longer molecules, implementing all the necessary parts for Darwinian evolution. For these lab experiments to proceed with ample speed, however, the temperature gradients have to be quite steep, reaching up to 30 K per 100 μm. Here we use computer simulations based on experimental data to show that 2000-fold shallower temperature gradients – down to 100 K over one metre – can still drive the accumulation of protobiomolecules. This finding opens the door for various environments to potentially host the origins of life: volcanic, water-vapour, or hydrothermal settings. Following the trajectories of single molecules in simulation, we also find that they are subjected to frequent temperature oscillations inside these pores, facilitating *e.g.* template-directed replication mechanisms. The tilting of the pore configuration is the central strategy to achieve replication in a shallow temperature gradient. Our results suggest that shallow thermal gradients across porous rocks could have facilitated the formation of evolutionary machines, significantly increasing the number of potential sites for the origin of life on young rocky planets.

Introduction

The formation of RNA-like biopolymers that exhibit both catalytic functions and information storage capabilities is central to the origin of life. However, geochemical evidence points towards very low molecular concentrations in prebiotic oceans. Therefore, the formation of complex informational molecules that require a number of molecular precursors is severely hindered.[1–3] The problem can be approached by searching for geological non-equilibrium conditions that make an origin of life possible, if not highly likely or even imperative under certain boundary conditions. Such a search will focus on experimentally testable conditions that create an evolutionary machine for proto-biomolecules, achieving the first steps of Darwinian evolution naturally by a combination of physico-chemical effects. Here, we discuss such a machine, one driven solely by a natural temperature gradient in porous rocks.

One of the first problems to solve is the so-called "concentration problem of the origin of life".[1–3] All approaches to generate molecules such as amino acids,[4] purines,[5–7] pyrimidines, and oligonucleotides, as well as alternative early replicators, are limited to high initial concentrations of precursor molecules.[8–13] While present-day cells run elaborate systems to maintain spatial compartmentalization, and feed their interiors by complex protein-based transport machineries,[14,15] only a few settings on the primordial Earth are predicted to have featured comparable segmentation and accumulation of molecules from aqueous solutions. If periodic changes in the environment do not cause the degradation of the protobiological molecules, evaporated terrestrial ponds,[16] self-assembled lipid bilayers, or coacervates as precursors of protocells and catalysing inorganic surfaces can offer favourable conditions for the synthesis or preservation of protobiomolecules.[17–21]

On the other hand, porous rocks emitting geothermally heated water into the ocean were an abundant setting on the early Earth.[22,23] The dissipation of heat forms a temperature gradient across sub-millimetre sized pores inside these rocks. This type of heat fluxes drive a highly efficient accumulation mechanism which is based on the interplay of thermal convection and thermophoresis. Temperature gradients across artificial pores

Systems Biophysics, Physics Department, Nanosystems Initiative Munich and Center for NanoScience, Ludwig-Maximilians-Universität München, Amalienstraße 54, 80799 München, Germany. E-mail: dieter.braun@lmu.de

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c6cp00577b

lead to the accumulation of dilute lipids and nucleotides,[24–26] enable the polymerization of long DNA/RNA strands,[27] and select, feed, and replicate nucleic acids towards increasing length.[28] Previous numerical approaches and experiments have shown that the mechanism is robust with respect to a large variety of geometries and works within artificial pores of different sizes.[26,29]

This work demonstrates *in silico* that shallow temperature gradients of 100 K per metre in prebiotically abundant volcanic settings are sufficient to accumulate a variety of molecules at least a million-fold. Previously conducted studies used temperature gradients at least 2000-times higher,[26–28] which we show is not necessary to achieve high accumulation ratios. Here, we extend the concept of narrow, vertical pores to pores with variable orientation, proving their versatility to achieve extreme accumulations under all spatial orientations of the pore. We thereby expand the range of thermal gradients capable of driving prebiotic molecular evolution to various environments such as steam-, volcanic-, and hydrothermal settings (Fig. 1). The heat dissipated from these sites creates a temperature gradient across adjacent pore systems, irrespective of their being filled with water. The accumulation mechanism, however, occurs solely in water-filled parts of the pore. Our new theoretical findings provide an abundant, simple, and universally applicable scenario of accumulation and a possible solution to the concentration problem of the origin of life.

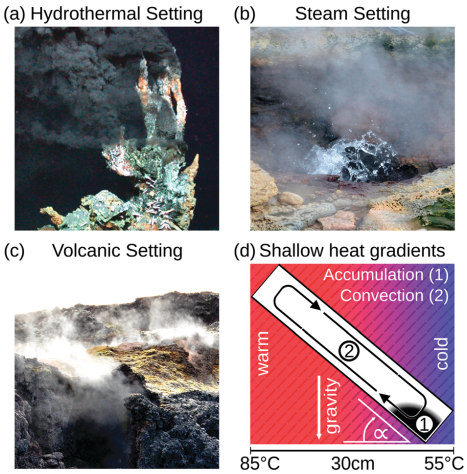


Fig. 1 Possible microthermal habitats for the origins of life. (a) Heat dissipation across submerged porous rocks and the cold ocean form steep temperature gradients which drive an efficient accumulation mechanism.[22,23] (b and c) Shallow temperature gradients, approximately 100–2000-fold weaker than previously assumed, still enable an efficient accumulation, extending the range of possible microthermal habitats from previously discussed hydrothermal settings to volcanic and steam settings. These surface-based microhabitats provide wet–dry cycles and UV illumination for trapped molecules, facilitating the generation and polymerization of nucleotides.[9] (d) Numerical approaches show that elongated pore systems within shallow temperature gradients efficiently accumulate molecules such as DNA and RNA (1) and enable a heat-driven replication reaction due to cyclic temperature changes induced by the laminar thermal convection (2). (a) Adapted from MARUM, University of Bremen/Germany.

In addition, we have simulated a large number of single, stochastic particles following unique trajectories that mimic the behaviour of nucleic acid strands inside the pores. The statistics derived from these numerical simulations suggest that particles frequently shuttle between hot and cold parts of the pore. The accumulated molecules are therefore subjected to temperature oscillations in a laminar convective flow, allowing for *e.g.* template-directed replication mechanisms.[13,30–33] Such mechanisms are central to the origin of life since they offer a pathway to the long-term storage, propagation, and mutation of information.

The concentration mechanism could also assist in the formation and selection of the first self-replicating molecules. The RNA-world hypothesis for example posits that RNA played a crucial role in the origin of life due to its catalytic function and information storage capabilities. The question remains of how a self-replicating ribozyme, containing at least 200 nt, could have emerged. Previous numerical and experimental studies have shown that the thermally driven accumulation mechanism concentrates oligonucleotides and thereby shifts a polymerization reaction towards longer polymers.[27] These polymers could then be selected for function and sequence, *e.g.* through a gelation process, providing an essential requirement for Darwinian evolution.[34] The numerical findings presented here suggest that pore systems subjected to shallow temperature gradients achieve comparable accumulation efficiencies and therefore have similar effects on the polymerization reaction. As a result, the number of potential sites for the formation of ribozymes vastly increases, assuming that enough feedstock molecules are available.

A supply of protobiomolecules inside such pores, however, is not only limited to *e.g.* diffusive coupling with the ocean and Fischer–Tropsch-type synthesis.[35] Porous rocks near the surface could have also been supplied with feedstock molecules synthesized by surface chemistry.[9,36] In such a regime, molecules such as precursors of ribonucleotides, lipids, or amino acids are synthesized on the surface and subsequently leached into the pores *e.g.* by downhill streams from rainfall. This makes reaction products particularly those of wet/dry and UV-reactions available in the pores. Surface directed ends of the pores can also be directly struck by sunlight and include the case of partially dried pores, *e.g.* based on moisture changes in steamy environments.

This work specifically studies the thermally driven accumulation mechanism inside porous rocks from the physical perspective, thereby neglecting chemical reactions.

Experimental

The accumulation of molecules can be described by Debye's approach, originally used to characterize separation columns.[37] The basic principle of molecular accumulation is given by the superposition of gravitational convection and thermophoresis. Both mechanisms result from heat fluxes across water-filled compartments. The thermophoretic effect moves molecules

along the thermal gradient, resulting in a net movement of $\vec{v}_{TD} = -D \cdot S_T \cdot \nabla T$. Hereby, D and S_T denote the diffusion and Soret coefficients and ∇T the temperature gradient, respectively. Thermophoresis is still subject to active research;[38,39] however, it has been found that the effect on charged molecules, *e.g.* short oligonucleotides, is dominated by ion shielding and Seebeck effects.[40] The latter is induced by the fact that each ionic species has a different Soret coefficient, generating a global electrical field that moves charged molecules. Thereby, the Seebeck effect highly depends on the ionic composition of the solution. In conjunction with convection, molecules erratically move within the temperature gradient while slowly accumulating at the bottom of the compartment. This effect has emerged to be highly potent in gaining extreme concentrations in distinct regions.

On the prebiotic Earth, such compartments could be found in volcanic rocks, highly porous minerals that exhibit micrometre sized pores. In this work, we approach molecular traps *in silico*, using a commercial finite-element solver (COMSOL Multiphysics 4.4). For calculating the temperature profile, rectangular compartments of 1 mm in width serve as pores, surrounded by 1 m of volcanic rock. Temperatures of 104 °C and 4 °C are applied to the left and right boundaries of the rock, respectively, spanning a temperature gradient of 0.1 K mm^{-1}. The linear temperature profile is calculated along a water filled pore for various minerals of hydrothermal-, steam-, and volcanic settings such as gabbro, peridotite, olivine-melilitite, and clay (see Fig. S1 in the ESI†).[41–44]

These minerals cover hydrothermal vent settings such as the Lost City (gabbro, peridotite), volcanic settings (olivine-melilitite), and the common material clay for comparison. Ancient versions of hydrothermal fields such as the Lost City existed on the early Earth; however, the exact composition is assumed to differ. The shape of the temperature profile for a mineral is mainly defined by its thermal conductivity, holding for different materials with equal thermal conductivity. Thermal conductivities may be even higher because of sediments within the rocks, *e.g.* sulfide sediments in black smokers.[45] The simulation, however, does not take into account interactions at the mineral/water boundary layers such as catalytic effects or surface induced polymerization reactions.[20,21]

The accumulation efficiencies are based on three consecutive steady state calculations of two-dimensional pore models:

(i) Partial differential equations (PDE) for transient heat transfer are employed to calculate the temperature profile within the pore. Here, a low temperature of T_{low} = 55 °C and a high temperature of $T_{high} = T_{low} + \Delta T$ on the right and left side are applied, respectively, while insulating the top and bottom. The temperature difference ΔT is calculated by assuming a temperature gradient of 0.1 K mm^{-1} across the porous rock.

(ii) The temperature profile is used to calculate a convectional flow profile by numerically solving the incompressible Navier–Stokes equations. Reciprocal effects upon the temperature profile are ignored since the laminar convection negligibly alters the heat-transfer of (i). The porous rock material is assumed to have a combined thermal conductivity of κ = 3 W mK^{-1}, a combination of quartz (κ = 6.6 W mK^{-1}) and olivine-melilitite (κ = 1.7–2.5 W mK^{-1}).[41,46]

(iii) The resulting flow profile is superimposed by thermophoretic and diffusive movement of the molecules, using PDEs (see eqn (1)). The top of the pores are assumed to have a constant molecule concentration $c = c_0$ (*i.e.* connected to a reservoir with the initial molecule concentration) and a closed bottom, resulting in a concentration distribution over the pore. The molecular flux is given by:

$$j = -\underbrace{(D \cdot \nabla c)}_{\text{Diffusion}} - \underbrace{D \cdot S_T \cdot \nabla T \cdot c}_{\text{Thermodiffusion}} + \underbrace{\vec{v}(\alpha, w) \cdot c}_{\text{Convection}} \quad (1)$$

with the molecule's diffusion coefficient D and Soret coefficient S_T, temperature T, the convective flow profile $\vec{v}(\alpha, w)$, pore width w, and the angle α between the pore and the direction of gravity. The tilting angle α and pore width w affect the gravitationally induced convective flow $\vec{v}(\alpha, w)$, which is solved in the preceding step of the simulation. Here, 90° and 0° denote a vertically and horizontally aligned pore, respectively. The Soret coefficients for oligonucleotides were experimentally measured by Reineck *et al.*[47] Our simulation follows Debye's approach by neglecting perturbations at the ends of the pore $c \cdot \nabla \cdot (\nabla T) = 0; c \cdot \nabla \cdot \vec{v}(\alpha, w) = 0$. Those assumptions have previously been shown to be in good agreement with experimental data, *e.g.* Duhr and Braun,[48] Mast *et al.*,[27] Reichl *et al.*,[40] and Kreysing *et al.*[28] A change in molecular concentration over time (eqn (2) and (3)) is derived from the molecular flux eqn (1) by using the continuity equation:

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) + D \cdot S_T \cdot \nabla \cdot (\nabla T \cdot c) - \nabla \cdot (\vec{v}(\alpha, w) \cdot c) \quad (2)$$

$$\frac{\partial c}{\partial t} = D \cdot \left(\frac{\partial^2 c}{\partial x^2} + \frac{\partial^2 c}{\partial y^2}\right) + D \cdot S_T \cdot \nabla T \cdot \nabla c - \vec{v}(\alpha, w) \cdot \nabla c \quad (3)$$

Here, x and y denote the coordinates along the pore length and perpendicular to the pore shown in Fig. 1d. Extending the simulation to the z-dimension was shown to have no effect on the accumulation efficiency.[26] For each molecule, the movement is defined by its diffusion coefficient D and Soret coefficient S_T. The accumulation efficiency is screened with respect to the Soret and diffusion coefficient, pore width, and tilting angle α. These conditions are tested in order to identify the angle and width for each molecule giving the highest accumulation (see Fig. S2 and S3 in the ESI†). The length of the pore L is adjusted to obtain a length to width ratio of 50 : 1, resulting in a maximum pore length of 50 mm given a width of 1 mm. The results from these simulations are then scaled according to Baaske *et al.*[26] to obtain concentrations for pore lengths of 1 m. Those pores are formed by a stacking of shorter pores that are connected by mass diffusion. They behave analogously to a single, elongated pore, thereby increasing the effective pore length. The large length allows both the accumulation and the thermal cycling of the molecules.

Random walk simulations investigate thermal cycling statistics of 100-mer oligonucleotides in shallow temperature gradients. The molecules are placed within an inclined, rectangular pore embedded in volcanic rock composed of olivine-melilitite.

The volcanic rock serves solely as a heat-conducting material, since surface chemistry is not included in the model. It is exposed to thermal gradients of 0.1 K mm^{-1} and 1 K mm^{-1}, applying a temperature difference of ΔT = 30 K over 30 cm and 3 cm, respectively. A temperature gradient in the range of 55 °C to 85 °C, as chosen here, enables *e.g.* replication reactions by allowing elongation and denaturation processes. In the steady state scenario, trapped molecules still shuttle between hot and cold areas, continuously undergoing temperature cycles. A temperature cycle is defined using two threshold temperatures for elongation and denaturation. Molecules complete a temperature cycle by first moving into a cold temperature region below 60 °C for elongation, followed by a high temperature region above 80 °C for denaturation, and finally back to the low temperature region. The thermal cycling statistics are inferred from 100 particle trajectories. These particles are randomly distributed along the pore and simulated for 3×10^5 h, corresponding to the steady-state reached after approximately 4.2×10^5 h, derived from L^2/D, where L denotes the pore length. The random walk model accounts for the superposition of laminar flow, thermophoresis, and Brownian motion for 100-mer oligonucleotides. The diffusion and Soret coefficients are given by D = 127 μm^2 s^{-1} and S_T = 0.07 K^{-1} for 100-mer RNA/DNA at a Debye length of λ_D = 2.1 nm at 70 °C, both measured experimentally.[40,47] Thermophoresis contributes with a net movement $\vec{v}_{DT}$ along the gradient, resulting in a thermophoretic velocity of $\left|\vec{v}_{DT}\right| = 2.5\,\mathrm{nm\,s^{-1}}$ for a 100-mer oligonucleotide. A convectional flow profile with a maximum of $v_{max} \approx 8$ μm s^{-1} shuttles molecules between warm and cold areas. The displacement $\Delta s(x,y)$ of these particles is given by:

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

with the time-step Δt, the temperature T, the convective flow-profile $v(\alpha)$, and the tilting angle α. Brownian motion is implemented by a randomly directed movement $\eta(t)$ for a given time-step of Δt = 1 ms.

Results and discussion

Heat dissipation across submerged porous rocks,[22,23] an abundant setting on the early Earth, has been advocated as a possible source for the origin of life (Fig. 1a).[26] Porous rocks comprise multiple branched pore systems which serve as water channels. We have proposed that a major temperature difference between the hot rock, generated by volcanic activity and the cold ocean, forms a temperature gradient, which drives a highly robust and efficient accumulation mechanism within the pore system. We have previously argued that these systems can thermophoretically accumulate, thermally cycle, and continuously feed the first prebiotic molecules for evolution, employing a temperature gradient of 400 K mm^{-1} and 200 K mm^{-1} in theory and experiment respectively.[27,28,49] Steep gradients of 10–400 K mm^{-1}, however, limit the scope of the approach to hydrothermal orifices and vapour heating with high flow rates. Here we show in simulations that 100–2000 fold shallower temperature gradients (0.1 K mm^{-1}) still achieve at least a million-fold accumulation within elongated pore systems. Such gradients can be found in near-surface volcanic and steam settings (Fig. 1b and c). Here, we further investigate the accumulation behaviour of tilted elongated pores in shallow gradients for oligonucleotides and small molecules. In addition, random walk simulations are performed to derive thermal cycling behaviour of single molecules within micro-thermal pores (Fig. 1d).

A minor difference in temperature between both heat reservoirs, *e.g.* volcanic rock and water/air, suffices to enable highly efficient accumulation of molecules (Fig. 2a). This is facilitated by a difference in heat conductivity between the rock (2 W mK^{-1} for olivine-melilite) and the water (0.6 W mK^{-1}), which results in a local increase of the temperature gradient across the water-filled pore by a factor of 3 (Fig. 2b). Steeper temperature gradients arise from different minerals, all naturally occurring in hydrothermal vents. The temperature gradient increases by a factor of 6.6, 20, or 33 for materials such as quartz, pyrite–silica, and pyrite,[46,50] respectively. A porous rock consisting of pyrite–silica is therefore able to locally create a temperature gradient of 2 K mm^{-1} across a single pore, despite having an average temperature gradient of 0.1 K mm^{-1} across the porous rock. Clay represents the worst case scenario for rock materials. Its heat conductivity of 0.9 W mK^{-1} increases the temperature gradient only by a factor of 1.5. Subsequent simulations are performed assuming a geological realistic temperature gradient of 0.1 K mm^{-1} along the porous rock.

The exact geometry of the pore system within rocks – be it *e.g.* triangular, rectangular, or curved – barely affects the accumulation efficiency, *i.e.* the maximum concentration (Fig. 3a). Based on these findings, subsequent simulations are carried out in rectangular pores only, assuming that the difference in amplification is insignificant for our main statement. However, extreme accumulation for 100-mer oligonucleotides up to 10^6-fold – concentrating

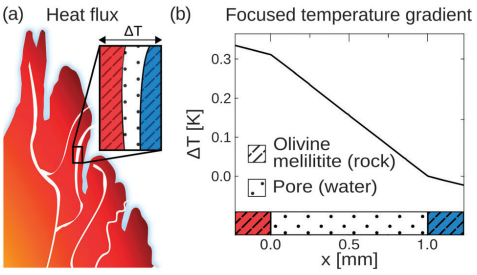


Fig. 2 Heat flux across porous rocks. (a) Formation of temperature gradients across porous rocks, heated by volcanic activity against the cold surrounding. Shallow temperature gradient settings can be found in various environments such as volcanic, water-vapour, or hydrothermal settings. (b) Simulations show that a difference in heat conductivity between the rock (olivine-melilitite κ= 2 W mK^{-1}) and the water filled pore (κ= 0.6 W mK^{-1}) results in a local increase in temperature gradient across the pore by at least a factor of 3. Taking other naturally occurring minerals like quartz, peridotite, or pyrite-silica into account, the temperature gradient could further increase *e.g.* by a factor of 33 for the case of pyrite. Therefore, a minor temperature difference between both heat reservoirs suffices to enable highly efficient molecular accumulation.

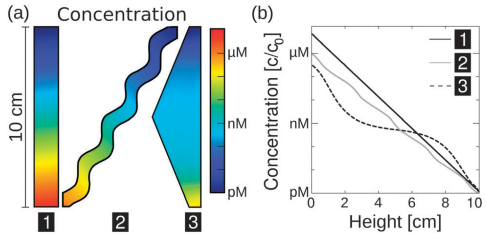


Fig. 3 Extreme accumulations of 100-mer oligonucleotides for various pore geometries *via* finite-element analysis, assuming shallow temperature gradients of 0.1 K mm^{-1}. Diffusion and Soret coefficients are based on experimental data.[40,47] (a) The accumulation mechanism is found to be highly robust with respect to the shape of the pore system. Pore systems of 10 cm in length achieve accumulations in concentration from pM to near μM. (b) The shape has only a minor impact on the magnitude of accumulation. The exact distribution of concentration is shown for the geometries in (a). A rectangular geometry shows the highest efficiency, accumulating molecules by a factor of 7 × 10^6. All pore geometries achieve at least a 10^5-fold accumulation.

molecules from pM to μM – is possible even for 10 cm pore systems (Fig. 3b). Typical pore lengths are significantly larger.[51] The accumulation efficiencies of longer pore systems can be determined since the concentrated material of a short pore, located at the bottom of the pore, serves as the starting concentration for an adjacent pore. For example, when stacking two triangularly shaped pores of 10 cm each, the concentration increases by a factor of 10^{12}, which is the product of both accumulations.

Besides its robustness with respect to the geometry of the pore, the mechanism also accumulates a large variety of molecules. Here, we extend previous simulations focused solely on oligonucleotides to a large pool of molecules (Fig. 4a), including monovalent ions like Li^+ (D = 1029 μm^2 s^{-1}, S_T = 0.0007 K^{-1}) and divalent ions such as Mg^{2+} (D = 706 μm^2 s^{-1}, S_T = 0.012 K^{-1}) and Ca^{2+} (D = 792 μm^2 s^{-1}, S_T = 0.013 K^{-1}).[40] To calculate the particle's accumulation efficiency, only its diffusion and Soret coefficients D and S_T, respectively, are necessary. Both parameters are strongly affected by salt concentration and ambient temperature, which is shown for 1- to 200-mer oligonucleotides (Fig. 4a and b). The diffusion and Soret coefficients were experimentally measured for Debye lengths of 0.79–5.6 nm, corresponding to salt concentrations in physiological solutions and more diluted solutions, respectively.[40] High salt concentrations (Debye length λ_D = 0.79 nm) and cold ambient temperatures (30 °C) result in relatively low accumulation efficiencies of $c/c_0 = 10^{14}$ for 200-mer oligonucleotides due to a drastic decrease in S_T. High ambient temperatures of 70 °C and medium salt conditions (λ_D = 2.1 nm) achieve a considerably higher accumulation of 10^{39} for a 200-mer. Still, even in the worst case scenario, the overall accumulation is high enough to provide a possible solution to the concentration problem of the origin of life, thereby affecting the outcome of chemical reactions considerably.[1,52] While the length of

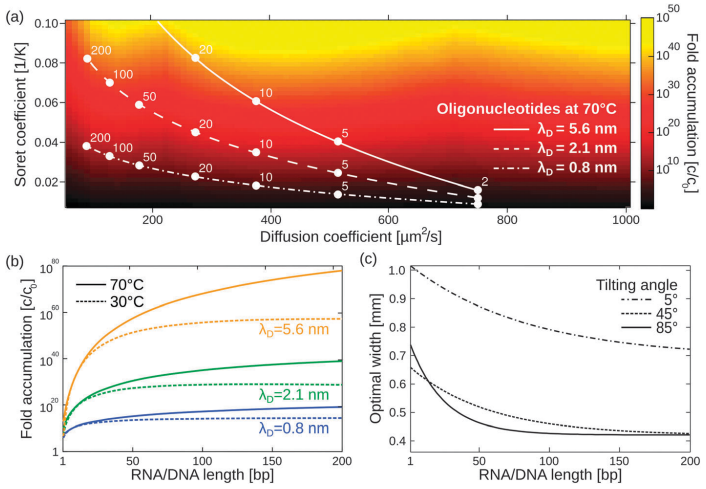


Fig. 4 Simulating exponential molecular accumulation. (a) Highly efficient accumulation of various molecules with a given diffusion coefficient and Soret coefficient. This allows the prediction of the accumulation for a very large range of molecules, including monovalent or divalent ions or single nucleotides. The curves show scenarios for the accumulation of 1- to 200-mer oligonucleotides at various salt concentrations. Diffusion and Soret coefficients were measured experimentally.[40,47] (b) The accumulation for RNA and DNA is shown as a line plot for better readability. DNA and RNA show very similar accumulation and are not distinguished. Debye lengths of 0.79 nm, 2.1 nm, and 5.6 nm denote high, medium, and low salt concentrations, respectively. (c) Optimal accumulation is achieved at a unique pore width. For example, short oligonucleotides accumulate best in a pore width above 500 μm, while longer oligonucleotides typically require pore widths below 500 μm. Tilting the pore from the vertical case (90°) to almost horizontal case (5°) results in an increase in optimal pore width.

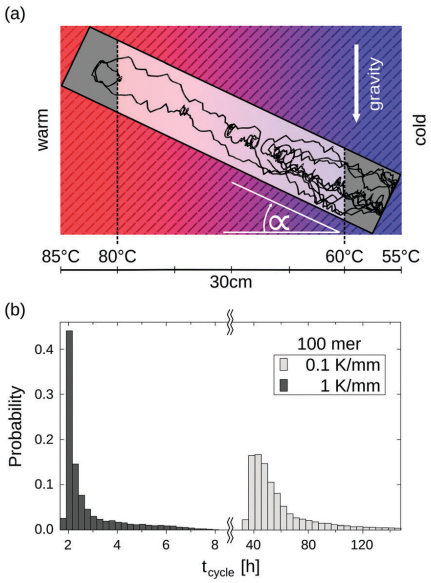


Fig. 5 Temperature cycle statistics of individual molecules derived by numerical simulations. (a) The trajectory of a 100-mer oligonucleotide inside an elongated, 45° tilted pore within a 0.1 K mm^{-1} temperature gradient. A convectional flow profile with a maximum velocity of ~2 μm s^{-1} shuttles molecules between warm and cold areas, enabling replication reactions by cyclic temperature changes. (b) Cycle time histograms of a 100-mer oligonucleotide. A temperature cycle is defined using two threshold temperatures for elongation (60 °C) and denaturation (80 °C). The oligonucleotides experience temperature cycles of 60 h and 3 h in temperature gradients of 0.1 K mm^{-1} and 1 K mm^{-1}, respectively. Thermal cycling statistics include long elongation- and short denaturation times.

the pore exponentially increases the accumulation efficiency, the optimal width of the pore varies for each molecule (Fig. 4c).

The largest accumulation for short oligonucleotides can be found for pores wider than 0.5 mm, while longer oligonucleotides typically require pore widths below 0.5 mm. Artificial molecular traps with a predefined pore width can therefore be used to accumulate a specific length regime of oligonucleotides, which has previously been shown for oligonucleotides of 20–200 bp.[28] This concept can be further extended to small molecules, such as mono- and divalent ions. The simulation also covers the case of geologically realistic, randomly aligned pores by adding a tilting angle for rectangular pores. Previous simulations focused on vertically aligned pores. Tilting the pore from a vertical case (90°) to an almost horizontal case (5°) results in a decrease in optimal pore width. Therefore, a certain pore width is able to accumulate various molecules depending on the arrangement of the pore.

Long-term storage and propagation of information, which is encoded in precursors of DNA – quite possibly RNA – at the origins of life, require environments that feature reliable replication mechanisms. The laminar convection within microthermal pores offers cyclic temperature changes and enables Watson–Crick-type replication mechanisms.[49,53,54] Random walk simulations derive temperature cycle statistics of 100-mer nucleotides. The oligonucleotides are placed inside a 0.75 mm thin, water-filled pore, which is embedded in a square of volcanic rock with a side length of 30 cm (Fig. 5a). The pore is tilted at $\alpha = 45^\circ$ and exposed to a shallow temperature gradient of 0.1 K mm^{-1}, which results in a slightly steeper gradient of 0.3 K mm^{-1} within the pore due to thermal conductivity differences of olivine-melilitite and water (see Fig. 2).

The particles perform temperature cycles within the pore by shuttling between the different temperature regions. 100-mer oligonucleotides take on average 60 h to complete a temperature cycle (Fig. 5b). For comparison, we evaluated the cycle statistics of 100-mer nucleotides assuming a ten fold steeper gradient (1 K mm^{-1}). Particles subjected to steeper gradients achieve considerably faster temperature cycles, requiring 3 h on average for completion. The molecules undergo thermal cycling comparable to regular polymerase chain reaction (PCR) protocols, comprising of long elongation and short denaturation times. Given that melting temperatures for random 100 mer oligonucleotides at a Debye length of $\lambda_D = 2.1$ nm are in the range of ~80 °C, the induced temperature cycles enable the denaturation of oligonucleotides and thus replication reactions.

Conclusion

The new theoretical findings suggest an abundance of potential sites on prebiotic Earth that would have been capable of solving the concentration problem of the origin of life. While previous studies were limited to large temperature gradients across hydrothermal porous rocks, we showed that shallow temperature gradients are sufficient to drive highly efficient molecular accumulation processes. As a result, various environments, such as steam- and volcanic settings, become available as potential sites for the origins of life. These sites, if located near the surface, also enable a supply to the pores of feedstock molecules that are synthesized on the surface. The results may prompt experiments with actual rock samples in more shallow gradients and longer time frames than used before, *e.g.* to probe the formation of long polynucleotides from low starting concentrations inside porous rocks.

Acknowledgements

Financial support from the Simons Collaboration on the Origin of Life, the NanoSystems Initiative Munich, the Ludwig-Maximilians-Universität Munich Initiative Functional Nanosystems, and the SFB 1032 Project A4 is acknowledged.

References

1 K. Dose, *BioSystems*, 1975, **6**, 224–228.
2 S. J. Mojzsis, T. M. Harrison and R. T. Pidgeon, *Nature*, 2001, **409**, 178–181.
3 N. Lane, *Life ascending: the ten great inventions of evolution*, Profile books, London, 2010.
4 S. L. Miller, *Science*, 1953, **117**, 528–529.

5 J. Oró, *Biochem. Biophys. Res. Commun.*, 1960, **2**, 407–412.
6 J. Oró and A. P. Kimball, *Arch. Biochem. Biophys.*, 1961, **94**, 217–227.
7 J. Oró and A. P. Kimball, *Arch. Biochem. Biophys.*, 1962, **96**, 293–313.
8 B. T. Burcar, L. M. Barge, D. Trail, E. B. Watson, M. J. Russell and L. B. McGown, *Astrobiology*, 2015, **15**, 509–522.
9 M. W. Powner, B. Gerland and J. D. Sutherland, *Nature*, 2009, **459**, 239–242.
10 H. Kuhn and J. Waser, *Nature*, 1982, **298**, 585–586.
11 M. Eigen, *Naturwissenschaften*, 1971, **58**, 465–523.
12 G. F. Joyce, *Nature*, 1989, **338**, 217–224.
13 D. Sievers and G. von Kiedrowski, *Nature*, 1994, **369**, 221–224.
14 R. Pascal, A. Pross and J. D. Sutherland, *Open Biol.*, 2013, **3**, 130156.
15 E. Schrödinger, *What is life?*, Cambridge University Press, Cambridge, 1944.
16 K. E. Nelson, M. P. Robertson, M. Levy and S. L. Miller, *Origins Life Evol. Biospheres*, 2001, **31**, 221–229.
17 J. Oro and A. Lazcano, *Adv. Space Res.*, 1984, **4**, 167–176.
18 T. Oberholzer and P. L. Luisi, *J. Biol. Phys.*, 2002, **28**, 733–744.
19 S. Koga, D. S. Williams, A. W. Perriman and S. Mann, *Nat. Chem.*, 2011, **3**, 720–724.
20 J. P. Ferris, A. R. Hill, Jr, R. Liu and L. E. Orgel, *Nature*, 1996, **381**, 59–61.
21 C. Huber, *Science*, 1998, **281**, 670–672.
22 M. J. Russell, A. J. Hall, A. J. Boyce and A. E. Fallick, *Econ. Geol.*, 2005, **100**, 419–438.
23 D. S. Kelley, J. A. Karson, G. L. Früh-Green, D. R. Yoerger, T. M. Shank, D. A. Butterfield, J. M. Hayes, M. O. Schrenk, E. J. Olson, G. Proskurowski, M. Jakuba, A. Bradley, B. Larson, K. Ludwig, D. Glickson, K. Buckman, A. S. Bradley, W. J. Brazelton, K. Roe, M. J. Elend, A. Delacour, S. M. Bernasconi, M. D. Lilley, J. A. Baross, R. E. Summons and S. P. Sylva, *Science*, 2005, **307**, 1428–1434.
24 I. Budin, R. J. Bruckner and J. W. Szostak, *J. Am. Chem. Soc.*, 2009, **131**, 9628–9629.
25 D. Braun and A. Libchaber, *Phys. Rev. Lett.*, 2002, **89**, 188103.
26 P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell and D. Braun, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9346–9351.
27 C. B. Mast, S. Schink, U. Gerland and D. Braun, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 8030–8035.
28 M. Kreysing, L. Keil, S. Lanzmich and D. Braun, *Nat. Chem.*, 2015, **7**, 203–208.
29 B. Herschy, A. Whicher, E. Camprubi, C. Watson, L. Dartnell, J. Ward, J. R. G. Evans and N. Lane, *J. Mol. Evol.*, 2014, **79**, 213–227.
30 S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobe, D. A. Treco and J. W. Szostak, *Nature*, 2008, **454**, 122–125.
31 A. Wochner, J. Attwater, A. Coulson and P. Holliger, *Science*, 2011, **332**, 209–212.
32 N. Paul and G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12733–12740.
33 Z. Yang, F. Chen, J. B. Alvarado and S. A. Benner, *J. Am. Chem. Soc.*, 2011, **133**, 15105–15112.
34 M. Morasch, D. Braun and C. B. Mast, *Angew. Chem.*, 2016, DOI: 10.1002/ange.201601886.
35 T. M. McCollom, G. Ritter and B. R. Simoneit, *Origins Life Evol. Biospheres*, 1999, **29**, 153–166.
36 B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy and J. D. Sutherland, *Nat. Chem.*, 2015, **7**, 301–307.
37 K. Clusius and G. Dickel, *Naturwissenschaften*, 1938, **26**, 546.
38 A. Würger, *Phys. Rev. Lett.*, 2016, **116**, 138302.
39 J. K. G. Dhont, S. Wiegand, S. Duhr and D. Braun, *Langmuir*, 2007, **23**, 1674–1683.
40 M. Reichl, M. Herzog, A. Götz and D. Braun, *Phys. Rev. Lett.*, 2014, **112**, 198101.
41 R. Büttner, B. Zimanowski, J. Blumm and L. Hagemann, *J. Volcanol. Geotherm. Res.*, 1998, **80**, 293–302.
42 P. B. Kelemen, E. Kikawa and D. J. Miller, *Proc. Ocean Drill. Program, Initial Rep.*, Ocean Drilling Program, 2004, vol. 209.
43 K. Midttømme, E. Roaldset and P. Aagaard, *Clay Miner.*, 1998, **33**, 131–145.
44 A. Geptner, H. Kristmannsdóttir, J. Kristjansson and V. Marteinsson, *Clays Clay Miner.*, 2002, **50**, 174–185.
45 P. M. Herzig and M. D. Hannington, *Ore Geol. Rev.*, 1995, **10**, 95–115.
46 P. A. Rona, E. E. Davis and R. J. Ludwig, *Proc. Ocean Drill. Program: Sci. Results*, 1998, **158**, 329–336.
47 P. Reineck, C. J. Wienken and D. Braun, *Electrophoresis*, 2010, **31**, 279–286.
48 S. Duhr and D. Braun, *Phys. Rev. Lett.*, 2006, **96**, 168301.
49 C. B. Mast and D. Braun, *Phys. Rev. Lett.*, 2010, **104**, 188102.
50 C. Clauser, *Landolt-Börnstein - Numerical Data and Functional Relationships in Science and Technologies, Group VIII: Advanced Materials and Technologies*, 2006, vol. 3, pp. 493–604.
51 M. J. Russell, A. J. Hall, A. J. Boyce and A. E. Fallick, *Econ. Geol.*, 2005, **100**, 419–438.
52 C. de Duve, *Blueprint for a cell—the nature and origin of life*, N. Patterson, Burlington, N.C., 1991.
53 M. Krishnan, *Science*, 2002, **298**, 793.
54 S. Rajamani, J. K. Ichida, T. Antal, D. A. Treco, K. Leu, M. A. Nowak, J. W. Szostak and I. A. Chen, *J. Am. Chem. Soc.*, 2010, **132**, 5880–5885.

Article in Nature Chemistry

Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length

by

Moritz Kreysing, Lorenz Keil, Simon Lanzmich and Dieter Braun

published in

Nature Chemistry doi: 10.1038/nchem.2155 (2015)
Reprinted from Kreysing *et al.* [44] with permission from the Nature Publishing Group

The publication in this Appendix was used as a basis for Chapter 4

nature chemistry
ARTICLES
PUBLISHED ONLINE: 26 JANUARY 2015 | DOI: 10.1038/NCHEM.2155

Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length

Moritz Kreysing†‡, Lorenz Keil‡, Simon Lanzmich‡ and Dieter Braun*

The replication of nucleic acids is central to the origin of life. On the early Earth, suitable non-equilibrium boundary conditions would have been required to surmount the effects of thermodynamic equilibrium such as the dilution and degradation of oligonucleotides. One particularly intractable experimental finding is that short genetic polymers replicate faster and outcompete longer ones, which leads to ever shorter sequences and the loss of genetic information. Here we show that a heat flux across an open pore in submerged rock concentrates replicating oligonucleotides from a constant feeding flow and selects for longer strands. Our experiments utilize the interplay of molecular thermophoresis and laminar convection, the latter driving strand separation and exponential replication. Strands of 75 nucleotides survive whereas strands half as long die out, which inverts the above dilemma of the survival of the shortest. The combined feeding, thermal cycling and positive length selection opens the door for a stable molecular evolution in the long-term microhabitat of heated porous rock.

From a wide range of exploratory experiments much is known about the capabilities and limitations of chemical replication systems[1–6]. It has become increasingly clear that such replicators are delicate systems that require a suitable supportive microenvironment to host non-equilibrium conditions. These conditions permit the sustainment of molecular evolution and the synthesis of molecules against equilibrating forces[1,7–9]. To the same end, modern cells provide active compartments of reduced entropy that protect genetic information against its thermodynamically favoured decay[8,10]. This is facilitated by sophisticated membrane-trafficking machinery and a metabolism that feeds on chemical low-entropy sources or light energy (Fig. 1a).

It has been known since Spiegelman's experiments in the late 1960s[11] that, even if humans assist with the assembly of an extracellular evolution system, genetic information from long nucleic acids is quickly lost. This is because shorter nucleic acids are replicated with faster kinetics and outcompete longer sequences. If mutations in the replication process can change the sequence length, the result is an evolutionary race towards ever shorter sequences.

In the experiments described here we present a counterexample. We demonstrate that heat dissipation across an open rock pore, a common setting on the early Earth[12] (Fig. 1b), provides a promising non-equilibrium habitat for the autonomous feeding, replication and positive length selection of genetic polymers. Previously, it has been argued that a temperature gradient spanning a submillimetre wide, closed compartment is able to accumulate dilute nucleotides, to enforce their polymerization or to concentrate lipids to form vesicles[13–15].

Here we extend the concept to the geologically realistic case of an open pore with a slow flow passing through it. We find continuous, localized replication of DNA together with an inherent nonlinear selection for long strands. With an added mutation process, the shown system bodes well for an autonomous Darwinian evolution based on chemical replicators with a built-in selection for increasing the sequence length. The complex interplay of thermal and fluid dynamic effects, which leads to a length-selective replication (Fig. 1c, (1)–(4)), is introduced in a stepwise manner.

Results

Accumulation. The accumulation mechanism responsible for counterbalancing the mixing entropy relies on the interplay of thermophoresis and gravitationally driven convection (Fig. 2a and Supplementary Movie 1). In the presence of a temperature difference, thermophoresis drives the molecules horizontally from the warm left side to the cold right side. On a similar timescale, the fluid moves vertically by convection and carries the molecules with it. Convection deflects the horizontal thermophoretic depletion and amplifies it to give a strong vertical molecule accumulation[16,17] (see Methods). This interplay of molecular movement and fluid flow therefore results in an efficient net transport of oligonucleotides to the bottom of the compartment; the experiment is visualized in Fig. 2b (also see Supplementary Movie 2).

For oligonucleotides with a length of 75 bases, concentrations increase by a factor of ten per millimetre pore length, which results in a millionfold concentration increase for a 6 mm high pore. Larger nucleic acids are exponentially better trapped because their higher charge contributes quadratically to the achievable accumulation[18,19]. This length-selective accumulation bias can be directly detected experimentally (Supplementary Fig. 3). The accumulation counterbalances diffusional dilution and offers a solution to the concentration problem associated with the origin of life.

Size-selective trapping from feeding flow. To establish efficient feeding with replication-relevant monomers, we opened the pore at both ends. This permitted an upwards feeding flow through the pore that originated from the overall large-scale upwards flow in a hydrothermal situation. Interestingly, this led to all-or-nothing

Systems Biophysics, Physics Department, Center for Nanoscience, Ludwig-Maximilians-Universität München, 80799 Munich, Germany. †Present address: Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany. ‡These authors contributed equally to this work.
*e-mail: dieter.braun@physik.lmu.de

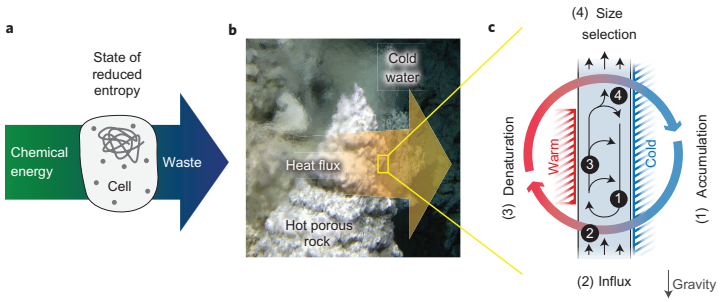


Figure 1 | Reduction of local entropy is key for living systems and can be caused by the flux of thermal energy. a, Modern cells feed on chemical energy, which enables them to host, maintain and replicate information-coding polymers, processes necessary for Darwinian evolution. **b**, The flux of thermal energy across geological cracks near a heat source (the white smoker[28] is adapted from an image courtesy of Deborah S. Kelley). **c**, (1) A thermal gradient across a millimetre-sized crack induces the accumulation of molecules by thermophoresis and convection. (2) A global throughflow imports nutrients into the open pore. (3) Exponential replication is facilitated by the local convection, which shuttles the molecules repetitively between warm and cold, and thus induces the cyclic denaturation of nucleotides. (4) The combination of influx, thermophoresis and convection selectively traps long molecules and flushes out short ones. The inflow speed determines the cut-off size of the resulting length selection. Mechanisms (1) to (4) are described in detail in this article.

trapping characteristics that depend on the strand length. We loaded an oligonucleotide ladder (20–200 base pairs (bp) dsDNA) in a 3.5 mm high and 70 μm wide pore and introduced an upwards flow with a velocity of 6 μm s^{-1}. Using gel electrophoresis, we observed that nucleic acids above a certain threshold length were trapped inside the pore, whereas shorter strands followed the upwards flow and were washed out of the pore (Fig. 3a and Supplementary Movie 3).

For a given velocity, this sharp length fractionation had a transition between 80 and 100 bp and can be understood by the interaction of the flow profile inside the trap with the thermophoretic concentration profile. The upwards feeding flow superimposes on the internal convection pattern, which generates an asymmetrical flow profile inside the trap (Fig. 3b). Long strands are pushed by thermophoresis into the descending flow at the cold side, transporting the molecules downwards. These are then localized against the upwards feeding flow at the bottom end of the heated section. Shorter strands experience weaker thermophoresis and the overall upwards flow drags them out of the trap.

The flow rate at which the solute nucleic acids start to move upwards and leave the pore depends monotonically on the strand length. Consequently, a gradual increase of the flow rate with time results in the sequential release of longer strands (Fig. 3c). The existence of the observed threshold length might come as a surprise, but a finite-element model that combines flow, diffusion and thermophoresis reproduces the behaviour of the trap in detail (Fig. 3d and Methods).

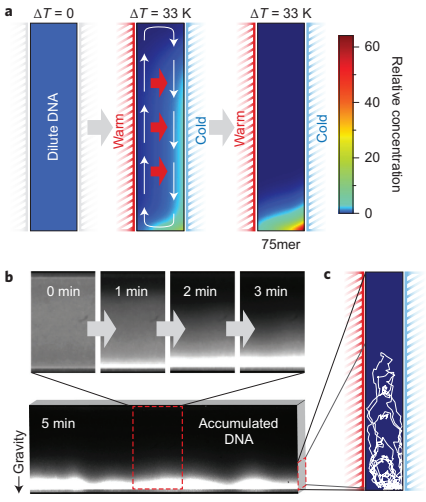


Figure 2 | Accumulation of oligonucleotides. a, The temperature gradient drives oligonucleotides horizontally from warm to cold by thermophoresis and simultaneously triggers the vertical thermal convection of water. Its combination results in a length-dependent accumulation at the bottom of an elongated pore within minutes (see Supplementary Movie 2). **b**, The accumulation of dilute double-stranded oligonucleotides (100–1,000mer) at the bottom is monitored within a 100 μm thin and 2 mm high capillary via SYBR Green I fluorescence. **c**, The accumulation is dynamic: the nucleotides cycle between the warm and cold sides, visualized in white for a single 500mer of DNA.

Exponential replication by convective thermal cycling. Besides continuous feeding and length-selective trapping, the asymmetrically heated pore offers another important feature relevant to the origin of life: laminar convective temperature cycling of the accumulated nucleic acids[20,21]. This opens the door to Watson–Crick-type replication mechanisms, which are otherwise hindered by the considerable energy costs required to separate double-stranded oligonucleotides[22]. The thermal cycling can be predicted from a fluid dynamics model that includes thermophoresis and diffusion (Fig. 4a). It is sufficient to separate cyclically double-stranded DNA (dsDNA) to drive exponential base-by-base replication with duplication times on the order of minutes, as documented by SYBR Green I fluorescence (Fig. 4b and Supplementary Movie 4). Our focus was to study the boundary conditions that enable early chemical systems for oligonucleotide replication. For this, we chose the polymerase chain reaction (PCR) as a fast and well-characterized placeholder for the large family of template-directed replication mechanisms that depend on temperature oscillations for long substrates[2–6].

NATURE CHEMISTRY DOI: 10.1038/NCHEM.2155 ARTICLES

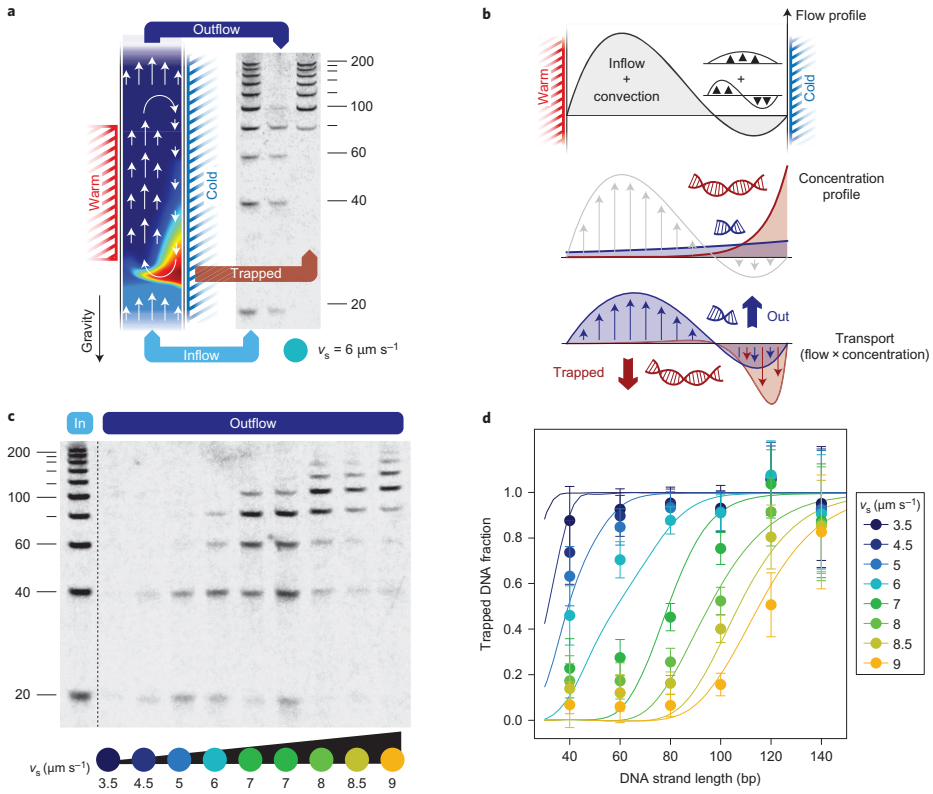


Figure 3 | Heat-driven filter selecting for strand length. a, A steady upwards feeding flow is triggered by opening the asymmetrically heated pore. A ladder of dsDNA (20–200 bp, 20 bp steps) was injected into the trap. Subsequent flushing of the capillary with pure buffer at a single velocity (v_s = 6 µm s^{-1}) revealed the filter's thresholding characteristics—lengths ≲80 bp flow through the pore whereas longer strands are trapped. **b**, An asymmetric flow pattern is generated by the superposition of the upwards flow and the convection. Thermophoresis pushes the long strands into the downwards flow and traps them. Short strands are subjected to the overall upwards flow and leave the pore. The trapping is a function of the feeding flow speed. **c**, The velocity of the external flow v_s tunes the fractionation of nucleic acids. As in the experiment before, a DNA ladder was initially introduced at a low flow velocity, which was then sequentially increased. The released DNA was measured using gel electrophoresis. **d**, The fraction of trapped DNA obtained from the electrophoresis gel constitutes a selection landscape of this thermal habitat in favour of long oligonucleotides. The velocity-dependent trapped fraction is described by a fluid dynamics model (see Methods). Error bars reflect the signal-to-noise ratio of the gel images (see Supplementary Fig. 11 for details).

Differential survival of replicating strands. Combining all of the above, we show how the joint thermally induced trapping and replication enables this arrangement to overcome Spiegelman's evolutionary dilemma of the degeneration of strand length and therefore loss of genomic information[11]. We followed the composition of a heterogeneous DNA population that replicates continuously inside the open pore. A 2.5 mm short capillary was seeded with a population of unlabelled template DNA strands with identical primer binding sites and a binary length distribution of 36 bp and 75 bp at a concentration of 1 nM each.

A temperature gradient from 61 °C to 94 °C was applied to a continuous upwards flux of template-free PCR buffer that contained nucleotides, polymerase and 7 nM fluorescently labelled primers and was run through the system at a speed of 6 µm s^{-1}. Over the course of the experiment (seven hours), the trapping volume was exchanged approximately 150 times with the template-free feeding buffer. Aliquots that contained the product of the continuously running reaction were taken from the outflow and analysed using gel electrophoresis. As the primers carried the labels, only replicated DNA strands were detected (Fig. 4c).

We observed that only the long strands were able to replicate sufficiently to withstand the diluting flow through the pore. This determined the increase of the relative concentration of the long, viable strands with respect to the total amount of DNA (Fig. 4d, yellow). The twofold shorter strands became diluted and then extinct.

This competitive replication and selection of two genetic polymers in favour of larger molecular lengths can be understood easily with a simple model. The determinants of the growth kinetics $dc_i/dt = (\mathrm{rep}_i - \mathrm{dil}_i)c_i$ for either the short or the long species $i = \{S, L\}$ are given by the replication rates rep_i and the dilution rates dil_i. Expressing the relative concentration of the long strands yields $c_L/(c_S + c_L) = (1 + Ae^{-\Delta kt})^{-1}$. $A = c_S^0/c_L^0$ is the initial

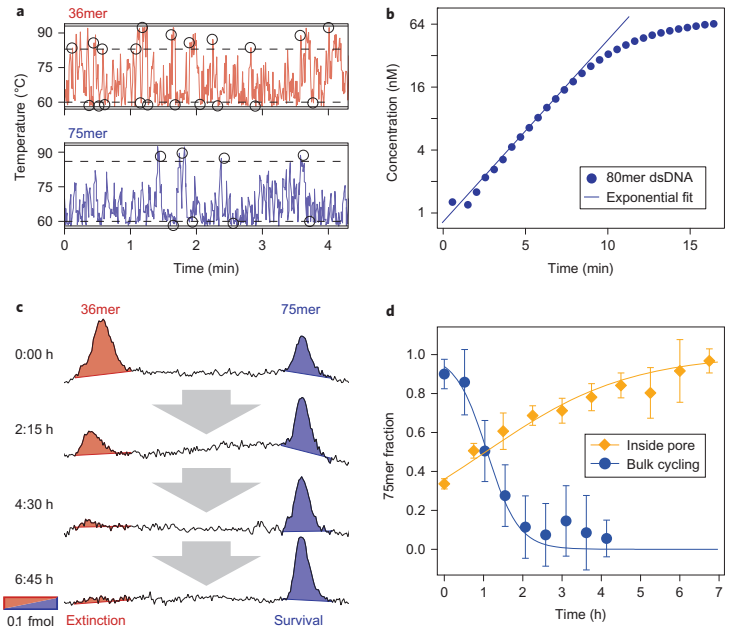


Figure 4 | Selection of a replicating DNA population that occupies the thermal habitat. a, Strands are subjected to temperature oscillations by the combination of thermophoresis, convection, feeding flow and diffusion. Simulations of stochastic molecule traces show that strands of 75 bp cycle inside the system for 18 minutes on average. In comparison, 36mers, owing to their enhanced diffusion, show faster temperature cycles, but are flushed out of the system after five minutes. **b**, Taq DNA polymerase-assisted replication of 80mer dsDNA by convective temperature cycling. Quantitative SYBR Green I fluorescence measurements show an exponential replication with a doubling time of 102 seconds (see Supplementary Movie 4). **c**, An open pore (see Fig. 1c) was seeded with a binary population of nucleic acids. Quantitative gel electrophoresis revealed sustainable replication for only the long strand. Short strands became diluted and then extinct despite their faster replication. **d**, Relative concentrations of the two competing species inside the thermal habitat. The selection pressure of the thermal gradient altered the composition of the binary population with time (yellow diamonds) in good agreement with an analytical replication model. The absolute fitness values were 1.03 and 0.87 for long and short strands, respectively. Without the thermal gradient, the short oligonucleotides won over the long strands (blue circles), analogous to the Spiegelman experiment. Error bars reflect the signal-to-noise ratio of the gel images (see Supplementary Fig. 11 for details).

concentration ratio and $\Delta k = (\mathrm{rep_L} - \mathrm{rep_S}) - (\mathrm{dil_L} - \mathrm{dil_S})$ is the differential growth rate. We experimentally found that, inside the pore, long strands (L) outcompete shorter ones (S) with $\Delta k = 0.55\,\mathrm{h}^{-1}$ (yellow curve). The length-selective fractionation model (Fig. 3c) confirmed that the shorter strands suffer from a fourfold higher dilution rate as compared to the trapped long strands. This selection of the longer replicating strand works best if the mechanism of replication is inefficient, such that the dilution of the short strand occurs before it can be replicated efficiently.

On the other hand, in a well-mixed situation, and hence in the absence of the selection pressure of the pore, we recovered Spiegelman's dilemma of the tyranny of the short. In a serial dilution experiment using a conventional thermal cycler with dilution rates that reproduce the pore conditions, the long strands died out rapidly with a differential growth rate of $\Delta k = -2.5\,\mathrm{h}^{-1}$ (Fig. 4d, blue curve).

Discussion

Our experimental findings conclusively show that, at the expense of dissipating free thermal energy, a habitat is created that drives and sustains the replication of long oligonucleotides by exploiting both convective temperature cycling and a selection pressure that supports the long over the short sequences. Therefore, heat dissipation enables the pore to overcome Spiegelman's classic problem for *in vitro* replication systems that create ever shorter genetic polymers, which results in the loss of genetic information.

On the hot early Earth, the pore system we describe was probably widespread because of porous, partially metallic volcanic rock, both near the surface and at submarine sites. As metals have a more than 100-fold larger thermal conductivity than water[23], metallic inhomogeneities near the pores can focus the thermal gradient from centimetres down to a micrometre-sized cleft (Supplementary Fig. 1). The kinetics of replication and selection were realized in the most simple geometrical setting of a single pore section with dimensions of 0.07 mm × 3.5 mm. Metallic inclusions do allow thermal gradients to be focused up to 100-fold to reach the thermal gradients of realistic geological settings (Supplementary Fig. 1). It is, however, important that the steepness of the thermal gradient can be further relaxed by at least one order of magnitude by separating replication and selection into two adjacent pores (Supplementary Fig. 2). At the bottom, a wide pore could provide the necessary temperature difference for replication[24]. At its top, the outflow would be constricted through one or more thin, but longer, selecting

NATURE CHEMISTRY DOI: 10.1038/NCHEM.2155 ARTICLES

pores. Their increased length of several centimetres instead of 3.5 mm compensates linearly for the reduced temperature difference[13].

Although the demonstrated length-selective trapping requires a temperature difference to work, the average temperature of the trap is not a critical parameter and can be tuned easily to fit the replication reaction. Therefore, the core mechanism of temperature cycling and selection studied here will also work for replication systems that require colder temperatures, including, for example, ribozymes or Q-beta replicase. However, many early replication systems are likely to rely on high temperatures for temperature-induced strand separation. For the PCR reaction used in the experiment, the strand lengths were highly controlled by the primers. In comparison, reactions that involve ligations have a tendency to extend the strands with partial templating[25] and initiate the length extension of the genetic polymers.

To extend this work to achieve Darwinian evolution in the demonstrated system, the replication process requires a significant mutation rate, including changes of the sequence length. The use of error-prone PCR with deep sequencing is therefore an interesting prospect for future experiments. At this point, the amount inside the pore is less than 1 pg, which prevents such an approach: the necessary strong preamplification would highly bias the obtained sequences and obscure their analysis.

Importantly, the thermophoretic selection pressure applies to each individual molecule of the population. As it is ultimately sensitive to the thermophoretic strength, the selection does not only favour the survival of long strands over short strands—it is possible that this mechanism could be tuned to select for the formation of macromolecular complexes or even for binding of aptamers[26].

Conclusion

Our experiments reveal how temperature gradients, the most simple out-of-equilibrium setting, can give rise to local environments that stabilize molecular replication against the entropic tendencies of dilution, degradation and negative length selection. A thermal gradient drives replication of oligonucleotides with an inherent directional selection of long over short sequence lengths. Interestingly, when replication and trapping inside the pore reach their steady state, the newly replicated molecules leave the trap with the feeding flow. This ensures an efficient transfer of the genetic polymers to neighbouring pore systems. Heat dissipation across porous rock was probably in close proximity to other non-equilibrium settings of pH, ultraviolet radiation and electrical potential gradients, all of which are able to drive upstream synthesis reactions that produce molecular building blocks. An exciting prospect of the presented experiments is the possible addition of mutation processes to achieve a sustained Darwinian evolution of the molecular population inside the thermal gradients of the early Earth. Accordingly, the onset of molecular evolution could have been facilitated by the natural thermal selection of rare, long nucleic acids in this geologically ubiquitous non-equilibrium environment.

Methods

Temperature gradients. Temperature gradients were generated across rectangular borosilicate glass capillaries (VitroTubes, VitroCom) with a cross-sectional aspect ratio of 1:20 and a thermal conductivity of 1.2 W m^{-1} K^{-1}. To this end, two different approaches were followed. (1) For the direct observation of the accumulation effect, glass capillaries were coated with a transparent conducting oxide layer that allowed for one-sided heating at a constant electric power with cooling from the other side. (2) Fractionation and replication experiments were performed in capillaries sandwiched between and thermally connected to temperature-controlled metal surfaces (compare the Supplementary Information and the figures therein for details of both approaches).

Accumulation-only experiments. dsDNA was diluted in 1 × Taq reaction buffer (New England Biolabs) that contained 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$ and 0.1% Tween20, with a pH of 8.3 at room temperature. A dsDNA ladder (10 μg ml^{-1}, 100–1000 bp, ten equidistant bands, weight equalized) was used in combination with 0.5 × SYBR Green I[27]. The applied temperature gradient from 22 °C to 88 °C resulted in temperatures from 38 °C to 71 °C inside the capillary (inner dimensions, 100 μm × 2,000 μm and 70 μm × 1,400 μm, as specified).

Fractionation experiments. A DNA ladder (20–200 bp, ten equidistant bands) was suspended in a 1 × PCR buffer that included 0.1% Tween20. Fractionation was carried out in a vertically oriented capillary (inner dimensions, 70 μm × 1400 μm) with an internal temperature gradient from 39 °C to 73 °C present over a capillary length of 3.5 mm (see the Supplementary Information for the details). The threshold trapping characterization was determined using a constant flow speed. Gradual fractionation was achieved by increasing the flow rate with time using a feedback-controlled syringe pump (neMESYS, Cetoni; see the Supplementary Information for a detailed protocol).

***In vitro* selection and replication.** Extracellular selection of replicating DNA strands was studied in a temperature gradient from 61 °C to 94 °C inside a thoroughly cleaned (DNA Away, Molecular BioProducts) capillary (inner dimensions, 70 μm × 1,400 μm, heated along 2.5 mm) at a mean solvent velocity of 6 μm s^{-1}. DNA replication was facilitated in a commercially available, glycerol-free master mix (fast cycling PCR Kit, Qiagen) that contained Taq polymerase, free nucleotides and standard concentrations of mono- and bivalent salts. The overall efficiency of DNA replication was reduced to less than 8% by means of a low concentration (7 nM) of each 14mer primer (forward (Cy5) and reverse primers; see Supplementary Fig. 7 for the sequences) in the feeding buffer. Unlabelled DNA templates (36mer, 75mer) were seeded into the region of replication through the system's output, leaving the feeding buffer template free. Reaction products that contained the incorporated Cy5 primer from the feeding buffer were extracted from the output of the artificial pore in 1.5 μl aliquots. Controls were performed in a conventional real-time PCR cycler (CFX96, Bio-Rad). A serial dilution experiment was performed to derive the replication efficiencies of the 36mer and 75mer DNA. Temperature cycles emulated the mean temperature cycle of 75mer DNA inside the pore, consisting of three seconds at 94 °C and 14 seconds at 60 °C (Supplementary Fig. 9). Including transition times, the total cycle time was 46.5 s. The initial concentrations were 2 pM (36mer) and 18 pM (75mer) for the PCR templates and 7 nM for the common primers. Every 40 cycles, the sample was diluted by a factor of 20 to yield a dilution rate of $dil_S = dil_L = 5.8\,h^{-1}$ that counterbalanced the concentration increase of the 36mer DNA within 40 cycles. This scheme prevented a depletion of the primer concentration and ensured that the efficiencies of the PCR reaction stayed constant over all 320 cycles. Replication rates were determined by comparison of the amount of DNA before each dilution using gel electrophoresis (Supplementary Fig. 12). The mean replication rates were determined to be $rep_S = (5.8 \pm 0.6)\,h^{-1}$ (36mer) and $rep_L = (3.3 \pm 0.4)\,h^{-1}$ (75mer).

Received 28 August 2014; accepted 2 December 2014; published online 26 January 2015

References

1. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459,** 239–242 (2009).
2. Sievers, D. & Von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **369,** 221–224 (1994).
3. Mansy, S. S. *et al.* Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454,** 122–125 (2008).
4. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332,** 209–212 (2011).
5. Paul, N. & Joyce, G. F. A self-replicating ligase ribozyme. *Proc. Natl Acad. Sci. USA* **99,** 12733–12740 (2002).
6. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc.* **133,** 15105–15112 (2011).
7. Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3,** 1–14 (2012).
8. Pascal, R., Pross, A. & Sutherland, J. D. Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics. *Open Biol.* **3,** 130156 (2013).
9. Powner, M. W., Sutherland, J. D. & Szostak, J. W. Chemoselective multicomponent one-pot assembly of purine precursors in water. *J. Am. Chem. Soc.* **132,** 16677–16688 (2010).
10. Schrödinger, E. *What is Life?* (Cambridge Univ. Press, 1944).
11. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl Acad. Sci. USA* **58,** 217–224 (1967).
12. Lay, T., Hernlund, J. & Buffett, B. A. Core–mantle boundary heat flow. *Nature Geosci.* **1,** 25–32 (2008).
13. Baaske, P. *et al.* Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl Acad. Sci. USA* **104,** 9346–9351 (2007).
14. Budin, I., Bruckner, R. J. & Szostak, J. W. Formation of protocell-like vesicles in a thermal diffusion column. *J. Am. Chem. Soc.* **131,** 9628–9629 (2009).
15. Mast, C. B., Schink, S., Gerland, U. & Braun, D. Escalation of polymerization in a thermal gradient. *Proc. Natl Acad. Sci. USA* **110,** 8030–8035 (2013).

16. Clusius, K. & Dickel, G. Trennung von Flüssigkeitsgemischen mittels kombinierter Thermodiffusion und Thermosiphonwirkung. *Naturwissenschaften* **26,** 546 (1938).
17. Debye, P. Zur Theorie des Clusiusschen Trennungsverfahrens. *Annal. Phys.* **428,** 284–294 (1939).
18. Piazza, R. & Guarino, A. Soret effect in interacting micellar solutions. *Phys. Rev. Lett.* **88,** 208302 (2002).
19. Duhr, S. & Braun, D. Why molecules move along a temperature gradient. *Proc. Natl Acad. Sci. USA* **103,** 19678–19682 (2006).
20. Krishnan, M., Ugaz, V. M. & Burns, M. A. PCR in a Rayleigh–Benard convection cell. *Science* **298,** 793 (2002).
21. Mast, C. B. & Braun, D. Thermal trap for DNA replication. *Phys. Rev. Lett.* **104,** 188102 (2010).
22. Rajamani, S. *et al.* Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.* **132,** 5880–5885 (2010).
23. Tritt, T. M. *Thermal Conductivity: Theory, Properties and Applications* (Kluwer Academic/Plenum, 2004).
24. Braun, D., Goddard, N. L. & Libchaber, A. Exponential DNA replication by laminar convection. *Phys. Rev. Lett.* **91,** 158103 (2003).
25. Fernando, C., Von Kiedrowski, G. & Szathmáry, E. A stochastic model of nonenzymatic nucleic acid replication: "Elongators" sequester replicators. *J. Mol. Evol.* **64,** 572–585 (2007).
26. Baaske, P., Wienken, C. J., Reineck, P., Duhr, S. & Braun, D. Optical thermophoresis for quantifying the buffer dependence of aptamer binding. *Angew. Chem. Int. Ed.* **49,** 2238–2241 (2010).
27. Wilhelm, J. & Pingoud, A. Real-time polymerase chain reaction. *ChemBioChem* **4,** 1120–1128 (2003).
28. Kelley, D. S. *et al.* An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. *Nature* **412,** 145–149 (2001).

Acknowledgements

We thank N. Osterman and C. Mast for the preliminary trapping experiments and discussions, M. Herzog and M. Reichl for thermophoresis measurements and S. Krampf for help with the gel electrophoresis. Financial support from the NanoSystems Initiative Munich, the Simons Collaboration on the Origin of Life, the Ludwig-Maximilians-Universität Munich Initiative Functional Nanosystems, the SFB 1032 Project A4 and the European Research Council (ERC) Starting Grant is acknowledged.

Author contributions

M.K., L.K. and S.L. contributed equally to this work and performed the experiments. M.K., L.K., S.L and D.B. conceived and designed the experiments, analysed the data and wrote the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.B.

Competing financial interests

The authors declare no competing financial interests.