
Replication in Early Evolution



Dissertation

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

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

vorgelegt von
Simon Alexander Lanzmich
aus Düsseldorf

München, September 2016

Erstgutachter:	Prof. Dr. Dieter Braun
Zweitgutachter:	Prof. Dr. Ulrich Gerland
eingereicht am:	29.09.2016
Datum der mündlichen Prüfung:	05.12.2016

Zusammenfassung

Unser Verständnis davon, was Leben ist, wurde wesentlich vom Konzept Darwinscher Evolution geprägt. Tatsächlich wird die Fähigkeit zu Darwinscher Evolution oft als eine definierende Eigenschaft von Leben herangezogen: Ein lebendes oder lebensähnliches System muss in der Lage sein sich zu replizieren und sein genetisches Material zu vererben. Die Replikation muss dabei in gewissem Maße Mutationen zulassen, welche über die Fitness der Nachkommen selektiert werden. Darwinsche Evolution ist nicht auf komplexe Systeme wie Organismen oder Zellen beschränkt, sondern findet auch in einfachen molekularen Systemen statt. Ein Studium auf dieser Ebene erlaubt es außerdem, Randbedingungen an die Entstehung lebender Systeme zu stellen.

Im ersten Teil dieser Arbeit wird eine einfache physikalische Umgebung untersucht, in der ein thermisches Nichtgleichgewicht einen Selektionsdruck zugunsten der Replikation von langen Nukleinsäuren vor kürzeren erzeugt. Diese Selektion überwindet den inhärenten Fitness-Vorteil der kürzeren, den diese durch ihre geringere Größe haben, und löst somit ein grundlegendes Problem früher Evolution. Die Umgebung umfasst einen submillimetergroßen, länglichen Hohlraum mit transversalem Temperaturgradienten. Zur Analyse des Selektionsverhaltens wurde die Replikation von DNA mittels Polymerase-Kettenreaktion (PCR) als Modell verwendet. PCR wird von Temperaturoszillationen angetrieben, die hier durch das Zusammenspiel von thermischer Konvektion, Thermophorese und Diffusion hervorgerufen werden. Die Selektion entsteht durch einen äußeren Fluss, der die Konvektion innerhalb des Hohlraums verändert. Eine theoretische Beschreibung der Experimente modelliert das Verhalten solcher thermogravitativen Fallen quantitativ. Der beschriebene Laboraufbau imitiert poröse Felsformationen in der Nähe hydrothermalen Quellen auf dem Meeresgrund, welche somit zu einem potentiellen Schauplatz früher Evolution werden.

Der zweite Teil behandelt ein DNA-Reaktionsnetzwerk, welches Abfolgen kurzer DNA-Sequenzen repliziert. Eine solche stückweise Replikation von Oligonukleotiden ist vom genetischen Code inspiriert, in dem Aminosäuren als Trinucleotid-Codons codiert werden. Die Struktur der einzelnen Moleküle des DNA-Netzwerks ist von Transfer-RNA abgeleitet. Die Replikation erfolgt kreuzkatalytisch, wird von Temperaturoszillationen angetrieben und verzichtet auf chemische Ligation. Sie benötigt lediglich die Hybridisierung komplementärer Nukleotidomänen und würde somit auch mit RNA oder potentiellen RNA-Vorläufern funktionieren. Bezüglich der Replikation einzelner Nukleotide wirkt das Replikationsschema als Korrekturmechanismus eines vorgeschalteten Polymerisationsprozesses.

Abstract

Our understanding of life is essentially shaped by the concept of Darwinian evolution. In fact, the ability to undergo Darwinian evolution is often regarded as a defining property of life. To this end, living or life-like systems must be able to replicate and pass on their genetic material. Replication itself needs to provide for some degree of variability, or mutations, which are selected via their effects on the fitnesses of the replicates. Darwinian evolution is not restricted to vastly complex systems such as organisms or single cells, but also acts on comparatively simple molecular systems. Studying the properties of evolution at this level allows to set boundary conditions on the origins of living systems.

In the first part of this thesis, a simple physical environment is studied, where thermal non-equilibrium exerts a selection pressure favouring the replication of longer nucleic acids over short ones. Selection is facilitated against the inherent fitness advantage of shorter molecules, which is due to their smaller size, and thereby overcomes a fundamental problem of early evolution. The environment consists of a submillimetre-sized, elongated cavity with a temperature gradient across. To probe its selection properties, the enzymatic replication of DNA in the polymerase chain reaction (PCR) was used as a model system. PCR is driven by temperature oscillations, which here are provided by the interplay of thermal convection, thermophoresis, and diffusion. Selection arises from an external flux, which alters the convection pattern inside the cavity. A theoretical treatment of the experiments quantitatively models selection and replication characteristics of such thermo-gravitational pores. The laboratory setup mimics porous rock formations in the vicinity of hydrothermal vents at the sea floor, which therefore qualify as a potential scene for early evolution and the origins of life on Earth.

The second part presents a reaction network of DNA strands, which replicates sequences of short DNA snippets. Replication of pieces of multiple nucleotides is inspired by the genetic code, where information about amino acids is encoded in trinucleotide codons. The structure of the individual molecules is derived from transfer RNA. Again, replication is driven by thermal oscillations, and proceeds cross-catalytically. It solely relies on base pairing of complementary nucleotide domains, and does not require any ligation chemistry. Therefore, it is detached from the details of the nucleic acids, and would also work with RNA or analogues discussed as potential prebiotic precursors. Considering the replication of individual nucleotides, the replication scheme effectively acts as a proofreading mechanism, improving the fidelity of an upstream polymerization process.

Contents

1. Introduction	9
1.1. Problems of early evolution	10
1.2. Potential answers	13
2. Replication and selection in thermo-gravitational reactors	15
2.1. Introduction	16
2.2. Results	18
2.2.1. Accumulation	18
2.2.2. Size selection	19
2.2.3. Selective survival of replicating populations	26
2.3. Discussion	31
2.4. Materials and methods	33
2.4.1. Experimental setup	33
2.4.2. Gel electrophoresis	34
2.4.3. Numeric calculations	34
2.4.4. Stochastic simulations	35
3. Transfer RNA as a codon replicator	37
3.1. Introduction	37
3.2. Results	38
3.2.1. Replication Mechanism	38
3.2.2. Strand design	40
3.2.3. Complex formation yields	42
3.2.4. Templating kinetics	44
3.2.5. Thermally driven amplification	45
3.2.6. Sequence replication	47
3.2.7. Replication fidelity	48
3.3. Discussion	49
3.4. Materials and methods	51
3.4.1. Strand design	51
3.4.2. Thermal cycling assays	53
3.4.3. Product analysis	54

Contents

3.4.4. Thermal melting curves	54
3.4.5. Sequences	55
4. Conclusions	57
A. References	59
B. Remarks	69
C. Danksagung	73
D. Publications	75

1. Introduction

Understanding the emergence of life in the universe is one of the major questions that has been driving human research throughout history. Nowadays, researchers studying this field come from all branches of the natural sciences: astronomy, geology, chemistry, biology, and physics. The point of this interdisciplinary endeavour is not necessarily to answer how life originated on Earth, as many traces of earlier life on Earth are irrevocably lost. However, since life on Earth is the only life known, it must serve as a starting point for investigation and careful deduction of general principles.

Here, the transition from non-living matter to a state one would consider living or at least life-like is of fundamental interest. To draw a line between living and non-living systems, a definition of life is required, but finding one has proven to be inherently difficult, if not impossible [8]. Especially when such a definition is to be used for identifying forms of life yet unknown. On the other hand, even a loosely defined understanding of life allows to study its properties and possible routes of how it can originate.

In the past decades, substantial progress has been made to pin down the chemical inventory of the early Earth and young planets in general. Organic compounds, the building blocks of life as we know it, are common in solar nebula and protoplanetary disks [15, 41, 68]. Similarly, a variety of amino acids, nucleobases, sugars, and activated phosphates has been detected in meteorites [13]. A fraction of these molecules has been (and still is) delivered to Earth, making it entirely plausible that the emergence of life was at least partially fuelled by organics of extraterrestrial origin. In the meantime, significant progress has been achieved in finding pathways for turning simple organic compounds into larger polymers under conditions argued to be prebiotically plausible. The non-enzymatic synthesis of nucleic acids, most prominently RNA [85], but also of peptides and lipids [80] has been reported recently.

While these findings allow to trace a possible history of life on Earth with an increasing level of detail, I will approach the topic on a more abstract level, addressing some fundamental challenges to the onset of life. The central driving force for the emergence of life is evolution, in particular the Darwinian type. To be capable of Darwinian evolution, a system must possess three properties. (1) It needs to be able to replicate itself, (2) its replication must allow for some rate of mutation, and (3), there must be a means of selection either intrinsic to the system itself or originating from interaction with the environment.

These requirements set a number of boundary conditions on the entities of replication.

1. Introduction

Self-replication implies a network of autocatalytic or cross-catalytic reactions. Mutation rates, which are nowadays restricted by the highly evolved replication machineries of cells, are estimated to be rather too high in non-enzymatic reactions. And finally, selection must not lead into dead ends but allow for a trajectory towards the evolution of complex functionality. Before turning to the presentation of the results of this thesis, I briefly recapitulate some problems accompanying the emergence of dynamic systems that eventually could develop into what one would consider alive.

1.1. Problems of early evolution

A driving environment. Living systems are out of equilibrium. They dissipate free energy from the environment to maintain their internal order and to replicate. In doing so, they produce entropy in their surroundings. Therefore, a continuous source of free energy is required to prevent degradation¹ [79, 103]. Different sources of energy are being studied: lightning discharge and ultraviolet radiation [58, 70, 109], gradients in pH or ion concentrations [62, 106, 120], or temperature gradients [10, 54].

Models of the early Earth and the evolution of our solar system predict that the UV radiation that reached the surface of the early Earth was significantly stronger than today, caused by a different composition of the atmosphere and a possibly higher solar activity [16]. Recent experimental findings suggest that UV irradiation was a major driver for chemical pathways producing precursors of RNA, amino acids and lipids [68, 75, 80]. In contrast, other authors stress the degrading effects of UV radiation [56]. For temperature gradients, volcanic submarine hot vents (black smokers) could provide a solution [4, 7], but temperatures of about 350 °C and short lifetimes on the order of decades are problematic with regard to the production of relevant organic molecules [47]. A particularly interesting environment are alkaline hydrothermal vents (white smokers), which are of non-volcanic origin and not only provide stable temperature gradients at more moderate temperatures, but also chemical gradients.

Focusing on a particular environment does not only specify a driving force, but also constrains available chemistries. When ultraviolet irradiation is assumed as a catalytic necessity, the setting must be close to the surface, such as areas around hot springs or geysers. The same holds when a rather direct coupling to the atmosphere is supposed. On the other hand, when the sea floor is deemed a more likely environment due to the shielding provided by the overlying ocean, reactions have to run on the reagents present in the ocean or emerging from vents in the crust.

¹In fact, the argument can also be approached from the other end, implying that life emerged as a energy-dissipating process in the first place [32, 102].

The concentration problem. Even though still widespread, the notion of the ocean of the early Earth as a primordial soup exhibits significant thermodynamic shortcomings [56]. The bulk ocean is homogeneous, and it lacks locally available energy sources that drive reactions out of equilibrium. Further, it is implausible that any combination of processes on the early Earth produced organic compounds in amounts to achieve significant concentrations across the whole ocean. However, these are required locally to drive chemical reactions at rates faster than decay. This is especially true for high-energy intermediates, which are formed during a reaction and required for subsequent steps. In polymerization reactions, this holds in a similar way². At concentrations below the dissociation constant of the individual polymer bonds, almost only monomers exist.

The requirement for sufficiently high concentrations is also reflected in laboratory studies of prebiotically relevant chemical reactions. These are commonly studied at mmol/L educt concentrations in order to provide sufficiently high yields [35, 80, 114]. In contrast, concentrations of organic compounds in the early Earth's oceans are estimated to be on the order of $\mu\text{mol/L}$ [109].

Several ways around this have been proposed: Cyclic evaporation and rehydration of tidal pools or hydrothermal fields [18, 21], adsorption to catalytic mineral surfaces [37], or freeze-thaw cycles where solute molecules are excluded from the forming ice and accumulated in the interstitial brine. As low temperatures reduce degradation rates, such a setting also increases the polymerization of RNA [73, 112], and the activity of ligation ribozymes [1, 76, 117]. A different solution is suggested by the exponential accumulation in hydrothermal vent systems [2], which effectuates even stronger enhancements of polymerization processes [64, 74]. Finally, one could also argue for the local emergence or generation of reagents at a sufficiently high rate. However, if the source is connected to an ocean, reaction products at some step in the pathway could diffuse away and be lost for any downstream reactions.

Encapsulation. Eventually, the solution to the concentration problem is encapsulation. All known life is cellular, and at some time in evolution, chemical reactions started to be encapsulated into proto-membranes [12, 110] or mineral pores [95]. The exact step of this to happen on the hypothetical ladder towards the first organism is debated, but being separated from the environment has substantial consequences for the reactions that occur inside the compartment.

On the one hand, reagents and intermediates cannot escape, but remain available for reactions. On the other hand, new reaction substrate needs to be imported into the com-

²In a general reversible polymerization model, as studied in [64], the equilibrium mean polymer length scales approximately as the square root of the total nucleotide concentration. A short derivation is presented in section B.4.

1. Introduction

partment and potentially inhibiting waste products accumulate. If the proto-membrane does not happen to be semipermeable in a very distinct and favourable way, this may be circumvented by repeated breakup and reconstruction of the compartments [18].

The tyranny of the shortest. In absence of outside selection pressures, it is the speed of replication itself that a population of replicators will be selected for. This raises a very fundamental question regarding the evolution of early replicators and was first demonstrated in the late 1960s by Spiegelman and co-workers [71], who studied the evolution of RNA-dependent RNA polymerases (replicases) *in vitro*. The Q β replicase was extracted from a virus, and then allowed to self-replicate. To sustain replication as well as select for the fastest replicators, parts of the reaction were successively transferred into fresh replication medium. At the end of the experiment, the replicase had eliminated 83 % of its genome, presumably everything not directly required for replication itself.

Similar behaviour can be expected for other sequential replication mechanisms, because replicating a larger molecule usually takes longer than replicating a smaller one. This strongly constrains the ability of a simple replicator to acquire new functionality. Once acquired, a particular feature might provide a significant boost in fitness. On the way, however, the not yet functional mutation would impede replication and provide a selective disadvantage, eventually slowing down its own emergence or preventing it entirely. Processes analogous to horizontal gene transfer or recombination could provide some answers here, but that would still require the pre-existence of the “genes” that were to be transferred. A different way to circumvent this tyranny of the shortest can be found not in the replicator itself, but in its environment. If an environment rewards some physical property such as size, charge, diffusivity, or surface to volume ratio, it imposes a selection pressure opposite to the evolutionary drift towards ever shorter genomes.

The error threshold. Besides the presence of an inherent advantage of small replicators, the previous example illustrates another point about evolution: the ability and imperative of a replicating system to mutate. On the one hand, mutation is a core ingredient of evolution, as it allows to adapt to the environment and compete with others. On the other hand, mutation must not be too strong, as a replicator is only a replicator if what it produces is actually a copy of itself.

Modern cells possess a network of highly evolved proteins and ribozymes to faithfully replicate their genome. Part of that is a polymerase with very high specificity regarding the discrimination of individual bases, but it also includes proofreading and error correction mechanisms. Cells must also make sure that DNA strands are copied in full and do not miss anything in between. In total, this results in an error rate lower than one in a billion bases [67]. In RNA viruses, error rates for genome replication are estimated to be two to five

orders of magnitude higher [86]. However, given that their genome is also a lot smaller, this still results in the average substitution of less than one nucleotide per replication cycle. There are of course cases where higher mutation rates are beneficial, such as for a parasite to evade a host's immune response (and the host's immune response itself), but this does not hold for those parts of the genome that define the very nature of the replicating entity.

The same would have been true for a significantly less sophisticated organism or replicating system on the early Earth. The error rate per full replication of the replication machinery must be low enough to ensure that the product still works. The argument can also be considered from the other side. Given a replicator with a certain copying fidelity, there is an upper limit to the size it can sustain [28, 29, 66]. Considering the example of a small nucleotide that replicates itself by stringing together individual nucleotides with a fidelity q , its length L is limited by [90, 110]

$$L < \frac{\ln(k/k_{mut})}{1 - q}. \quad (1.1)$$

Here, k/k_{mut} is the ratio of replication rates of the original sequence and the mutated copy. Supposing a rather poor copying fidelity, the maximum length of such a replicator is commonly assumed to be below 100 nucleotides [90, 110]. The same argument holds for other replicators which copy themselves from some kind of building blocks, where q then describes the fidelity of correctly adding another block to a partial copy.

While a number of RNA or RNA-like polymerization reactions are known [17, 36, 43, 88], the details of the particular mechanism may provide a more profound estimation of its copying fidelity and maximum sustainable length. Even when the polymerization itself is not strongly selective, kinetic effects such as stalling upon misincorporation of a nucleotide [87] or other proofreading mechanisms could improve the fidelity of the replication reaction.

Even after the onset of self-replication of oligonucleotides, the path towards replication of a larger genome is far from obvious, in particular with regard to the above-mentioned tyranny of the shortest. Individual replication of different parts of the genome could be an intermediate step [110], where replication would then operate on two different levels: replication of the individual parts and replication on how these parts are assembled or interact.

1.2. Potential answers

In the present work, I address two of the above problems. In the first part, I describe an environment that exerts a selection pressure on replicating systems which counteracts above-mentioned tyranny of the shortest: *submarine hydrothermal vents*. Located at the sea floor, these vents originate from exothermic serpentinization reactions in the Earth's

1. Introduction

mantle. Water reacts with the mineral olivine in the mantle rock, producing an alkaline fluid that emerges from the vents at temperatures of about 70–90 °C [48, 106]. At the encounter with the cold, low pH sea water, strong gradients in temperature, but also methane, hydrogen, and pH arise.

Here, only one of these aspects will be considered: temperature gradients across submillimetre-sized cavities in these vents. Convection caused by such gradients has been shown to accumulate molecules in a small region at the bottom of the the cavity [2], and facilitates thermally driven replication reactions [63]. In the present work, it will be shown that, when the cavity is not considered in isolation but connected to its surroundings, replication and accumulation are augmented by a remarkable feature: the generation of a selection pressure favouring the replication of larger molecules over short ones.

The second part studies a mechanism of *cooperative replication* of oligonucleotides. In contrast to other approaches of replicating nucleotide systems, it does not pursue polymerization of individual nucleotides. Instead, short sequence snippets are assembled in a cross-catalytic reaction. The reaction does not require chemical activation of the individual molecules, but is solely driven by thermal oscillations. Assembly requires only the hybridization of complementary nucleotide domains, a property not restricted to DNA or RNA, but common to a whole family of nucleic acid analogues [33, 59, 99, 100]. Accordingly, replication does not require chemical ligation activity, and explores a reversible replication mechanism, which also allows for the re-use and re-assembly of the individual sequence snippets.

Replication on the basis of more than one nucleotide at once is inspired by the genetic code and modern protein translation, where amino acid sequences of proteins are encoded in codons comprising multiple bases. Transfer RNAs (tRNAs), being central to the translation process, are arranged by the ribosome in temporal sequence in order to re-code a sequence of codons into amino acids. The mechanism here described modifies this process by arranging potential proto-tRNAs in spatial sequence. Importantly, this is not performed by an external molecule, but instead proceeds autonomously. By operating on codons instead of single nucleotides, the replication mechanism constitutes a means of proofreading for an underlying, possibly non-enzymatic polymerization process.

2. Replication and selection in thermo-gravitational reactors

This chapter addresses one particular issue regarding the evolution of replicating (and potentially life-like) systems: the *tyranny of the shortest*. As introduced above, replication is one core necessity of life-like systems, but replication processes usually are faster for short molecules than for longer ones. I will discuss how a temperature gradient across submerged porous rock creates an environment in which oligonucleotides can be replicated. That replication, however, favours molecules above a certain length. Short oligonucleotides cannot survive in the environment while longer ones do. Applied to a replicating system of mutating length, this imposes a lower bound on the length, counteracting above dilemma that inevitably leads to loss of genetic information.

Modern biology features many instances of the same matter at different scales. Bacteria quickly dispose of genes they no longer require and which do provide an advantage to their survival [52, 65]. Lost features may include the ability to synthesize a particular metabolite which has become available in the environment [27, 125], or the ability to actively move using a flagellum [52], which may be prodigal in a particular environment. The latter case looks like a significant regression, but the cost of building and running a flagellum can be substantial and surmount 10 % of the total energy budget of the cell [72]. Another example is the loss of plasmids containing genes coding for e.g. antibiotic resistance or the expression of a fluorescent protein [5, 98]. In laboratory experiments utilizing the expression of a fluorescent protein, it is therefore beneficial to couple fluorescence to antibiotic resistance. As long as the respective antibiotic is present in the environment, both are preserved.

In all cases, the loss or preservation of genetic information is not a quality of the replicating system alone, but originates in the interaction with the environment. If the loss of a particular feature (or part of the genome) directly or indirectly increases the rate of replication, it will eventually be lost. For the replication of a simple RNA or DNA strand, the removal of any nucleotide generally speeds up its replication. Therefore, the environment must provide some sort of selective pressure to counter that drift. At a very early stage and in the absence of specific interactions, this selective pressure must be linked to some physical properties of the molecules such a size or charge.

The experimental as well as many of the numeric results presented in this chapter have been published in [54]. Figures are reproduced where indicated.

2.1. Introduction

A proxy replicator. Studying the interaction of a simple replicating system with its environment requires, of course, the replicator. As a proxy for general exponential replication reactions that are driven by thermal oscillations, polymerase chain reaction (PCR) was used. It provides a well-established, robust, and readily available model system. Moreover, it does not rely on the exact shape of the temperature oscillations.

Replication proceeds in cycles, defined by oscillations in temperature. A double-stranded DNA template is first thermally denatured, for which it needs to be heated above its melting temperature T_m . After subsequent cooling, short DNA primers can bind to the template and be elongated by the polymerase enzyme. Elongation of the bound primers works over a broad range of temperatures [14], whereat the particular enzyme that was used, Taq polymerase, exhibits maximal activity at 75–80 °C. The robustness of the particular replication system does not impede the validity of the findings presented here, even though a “real” early replicator most probably was slower and less robust, and may have operated at a different temperature range. It was, however, not confined by the relatively strict time constraints of feasible laboratory experiments, and its particular requirements only needed to match some of the conditions that have been present on the early Earth.

As a commonly used technique in molecular biology, PCR is usually done by altering the temperature of the whole reaction vessel. However, exponential replication can also be observed in convection systems with a wide range of different geometries [10, 11, 55, 121], not limited to the regular shape of the pores here.

Thermo-gravitational pores. The environment that exerts a selection pressure favouring the replication of longer nucleic acids emulates submillimetre-sized pores in submarine hydrothermal vents. Today a hot spot of microbial life, these vents feature a number of interesting characteristics relevant for the evolution of early replicating systems. A main property is that the elongated pores are spanned by transverse temperature gradients, which arise from the temperature difference between the hot vent fluid and the colder ocean [2, 47].

It has been shown that temperature gradients across closed pores can exponentially accumulate nucleic acids [2], enforce the polymerization of nucleotides [64] or drive the formation of vesicles from dilute fatty acids [12]. Even more interestingly, the environment can also sustain PCR, while the DNA (and, to a lesser extent, the PCR mix) is accumulated inside a thermophoretic trap [63].

Based on these results on closed pore systems, I present an open pore system capable of *continuous replication* and, at the same time, *selection*. Figure 2.1 presents a schematic view on the system under consideration. Inside the pore, selective continuous replication is evoked from the interplay of four processes. (1) The interplay of convection and thermo-

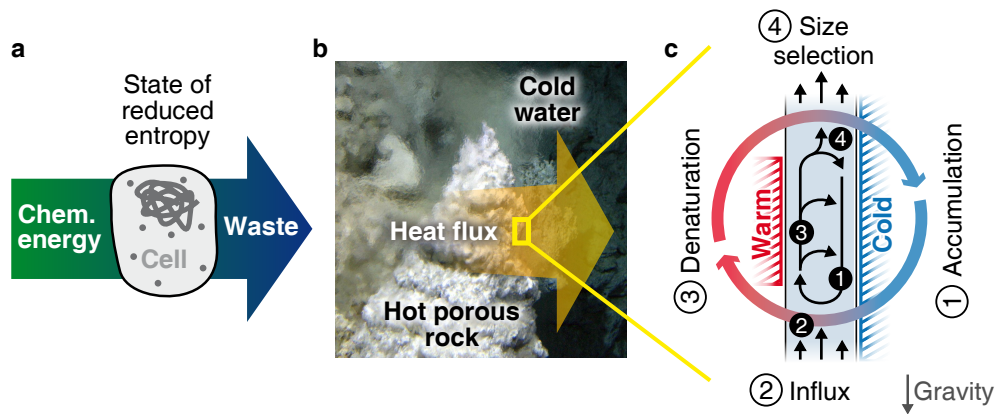


Figure 2.1. Replicating systems are driven by energy flux. This flux of chemical or thermal energy is required to sustain a state of locally reduced entropy. **a.** Modern cells feed on chemical energy to live and reproduce. **b.** Heat flux across pores in submarine hydrothermal vents, which provide an environment for replicating systems. **c.** Schematic representation of a thermo-gravitational reactor. (1) A thermal gradient across the pore effectuates convection and thermophoresis. Their interplay causes the accumulation of molecules. (2) Influx from below the pore feeds the replicator with nutrients. (3) Convection and diffusion move the molecules between cold and warm regions, allowing exponential replication by cyclic denaturation and elongation of nucleotides. (4) The interplay of thermophoresis, convection and external flux selects for replicating oligonucleotides above a minimal length. Shorter ones are flushed out of the pore. Figure reproduced from [54].

phoresis, both generated by the thermal gradient, accumulate molecules at the cold side of the pore. In vertical direction, strongest accumulation is located at the lower end of the heated region, where the temperature gradient is still mainly horizontal. (2) An external influx from below the cavity supplies the replication reaction with nutrients. (3) As accumulation is a dynamic state, DNA molecules diffuse in the thermal gradient and, from time to time, enter a region hot enough for their denaturation. One re-captured in the colder accumulation region, DNA primers can bind and template-based replication proceeds. (4) Besides supplying reagents for replication, the influx also deforms the otherwise almost symmetric convection pattern. This asymmetric velocity configuration causes short molecules to be washed out, while longer ones remain in the reaction cavity.

Experimental approach. For the experiments, hydrothermal pores were realized using borosilicate glass capillaries with an externally applied temperature gradient. Cooling on one side is performed using a water bath, heating on the other side is done either with a transparent indium tin oxide (ITO) electrode (if visual access is required) or an ohmic resistor. Inner cross sections of the capillaries were either $100 \times 2000 \mu\text{m}$ or $70 \times 1400 \mu\text{m}$.

Double-stranded DNA (dsDNA) samples recovered from the pores have been analysed using native polyacrylamide gel electrophoresis (PAGE) on 12.5 % gels. For the fractionation experiments using 20–200 bp dsDNA ladders, presented in figures 2.3 and 2.4, gels

2. Replication and selection in thermo-gravitational reactors

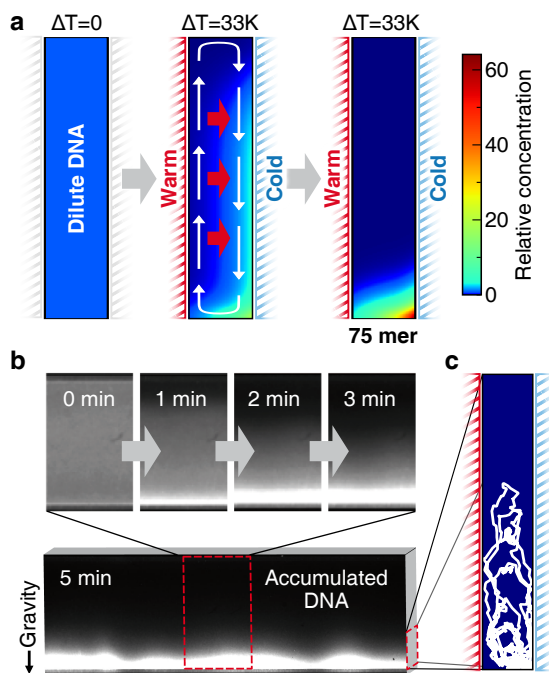


Figure 2.2. Dynamic accumulation of nucleic acids. **a.** A temperature gradient drives convection in a hydrothermal pore. Combined with thermophoresis pushing solute DNA towards the cold side, this results in exponential accumulation of oligonucleotides at the bottom. Colours give the relative concentration, normalized to the initial equal distribution. **b.** Accumulation kinetics are controlled by the convection velocity, which is on the order of $\mu\text{m/s}$ in a capillary of $100\ \mu\text{m} \times 2\ \text{mm}$. The measurement shows double-stranded DNA (100–1000 bp), visualized by SYBR Green I fluorescence, accumulating within minutes. **c.** A simulated trajectory of 500mer DNA reveals the dynamic nature of the accumulation inside the pore, as molecules travel between warm and cold areas. The length of the trajectory is about two minutes. Figure reproduced from [54].

were stained using SYBR Green I and recorded via fluorescent imaging. For the replication experiments, PCR products were generated from Cy5-labelled PCR primers, enabling a higher sensitivity during imaging. Details on the experimental methods are produced in section 2.4.

2.2. Results

Before considering the full-featured system, the individual aspects of the selective replication reaction are studied in more detail. On the one hand, this allowed to compare to earlier results on these reduced systems. However, it also provided fine-grained review of those simulation results which were directly observable.

2.2.1. Accumulation

The mechanism that drives accumulation and thereby balances mixing entropy is depicted in Figure 2.2a. A temperature gradient across a vertically oriented pore drives convection and thermophoresis. Convection is caused by the thermal expansion of water: water on the warm side rises, while that on the cold side descends. In the common approximation of constant viscosity η , i.e. $d\eta/dT = 0$, the resulting flux profile is given by a third order polynomial. Taking the temperature dependence of the viscosity into account, the velocity profile is slightly shifted to the hot side (cf. section 2.2.2 and Fig. B.1).

The thermophoretic and convective fluxes inside the pore occur on similar timescales, both on the order of minutes¹. Their combination results in a net downward transport that accumulates solute molecules at the bottom of the pore (Fig. 2.2b). While accumulation quickly reaches a steady state in such a closed pore system, the transient is significantly slower when the top of the pore is diffusively connected to a reservoir. In contrast, the open hydrothermal pores discussed below do not feature this downside.

The strength of the thermophoretic accumulation depends exponentially on the Soret coefficient S_T of the solute particles [2, 20] (cf. section B.1). For nucleic acids, it grows as the square root of strand length, because the higher charge contributes quadratically to the achievable accumulation [26, 82]. For DNA 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.

In contrast to the overall concentration distribution inside the chamber, which is experimentally accessible via fluorescence microscopy, the trajectories of individual solute molecules elude direct observation. However, knowledge about typical trajectories is central for a quantitative understanding of a replication process inside the pore. Figure 2.2c shows a trajectory of 500mer double-stranded DNA, simulated for two minutes. For most of the time, the particle is located at the bottom cold side. Nonetheless, it repeatedly escapes that region, travels some way with the upward flux along the warm edge, is pushed back to the cold side and returns to the region of strongest accumulation. This cyclic transit of regions of different temperatures is able to sustain replication reactions that require thermal separation of template and replicate, and will be further elucidated below. Details on the stochastic simulation are given in section 2.4.4.

2.2.2. Size selection

Closed systems cannot maintain continuous replication reactions, because at some point, all reagents will be used up and the reaction stalls. Hydrothermal pores with longitudinal openings pose a solution to that problem, as an external flux can provide new reagents and potentially also remove waste products. The interaction of such an external flow with thermophoresis and convection inside the pore is shown in Figure 2.3.

The external flow superimposes to the symmetric convection velocity, enlarging the region across the pore with upward flow velocity and increasing its magnitude (Fig. 2.3b, top). The Soret coefficient S_T of double-stranded DNA increases with length, such that longer DNA is pushed more strongly towards the cold edge of the pore than short oligomers. Due to the balancing of thermophoretic and diffusive fluxes, the concentration across the cavity follows an exponential distribution [25]. Its product with the flow velo-

¹An estimation is presented in section B.2.

2. Replication and selection in thermo-gravitational reactors

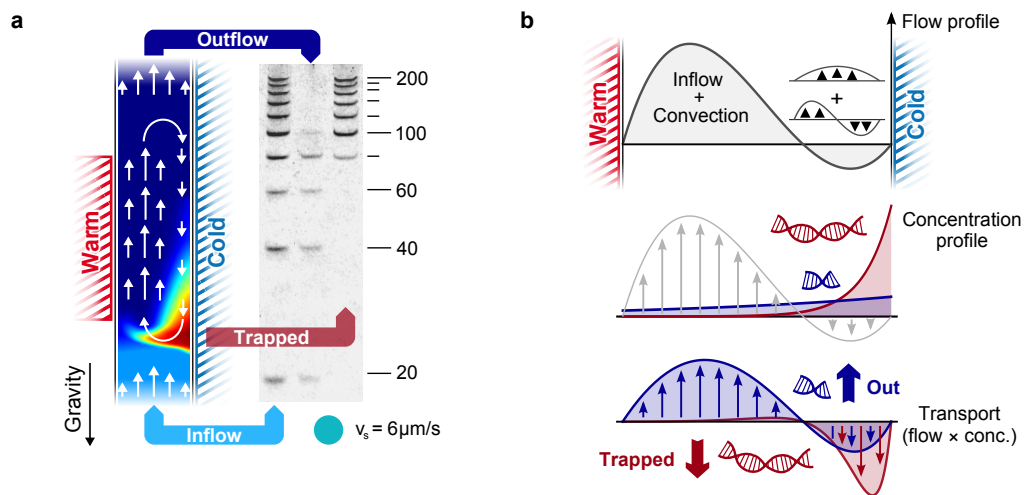


Figure 2.3. Heat-driven filter selecting for strand length. **a.** The pore, open at its top and bottom, was subjected to a continuous upward flow, creating a length-selective filter. It was loaded with a dsDNA ladder (20–200 bp, 20 bp steps) and subsequently flushed with buffer solution at $v_s = 6 \mu\text{m/s}$. Oligonucleotides with lengths ≤ 80 bp were washed out (central gel lane), while longer DNA stayed in the pore (right lane). The left lane shows the same sample as initially loaded. **b.** The parabolic inflow profile superimposes to the symmetric convection pattern, forming an asymmetric flow profile. Due to its higher Soret coefficient, long DNA (red) is pushed into the downward flow, while the effective transport of the short DNA is upward. The degree of asymmetry of the flow profile depends on the influx velocity v_s . Figure reproduced from [54].

city gives the net transport. Short molecules, having an upward net transport, are flushed out of the pore, while longer DNA has a downward net transport. The cut-off length can then be defined as the minimal length of DNA that is just not washed out of the pore.

Fig. 2.3a shows data measured in a capillary with an internal size of $70 \mu\text{m}$ along the temperature gradient. The heated region spans a length of about 3.5 mm. To study the size selection behaviour, the pore was loaded with a double-stranded DNA ladder containing oligonucleotides of 20–200 bp in equidistant steps of 20 bp. After loading, the capillary was rinsed with buffer solution at a mean velocity of $v_s = 6 \mu\text{m/s}$ and the outflow was collected. At the end of the experiment, the pore was washed at high fluid velocity, the outflux again collected for analysis on a denaturing polyacrylamide gel. The data show that DNA strands of lengths ≤ 80 bp were washed out, while those with lengths of 100–200 bp were retained inside the pore.

Depending on the magnitude of the external flux velocity v_s and thus the asymmetry of the velocity profile in the pore, the cut-off length changes. To probe the selection landscape as a function of v_s , the pore was again loaded with a dsDNA ladder of 20–200 bp. Starting at a velocity of $3.5 \mu\text{m/s}$, buffer was flushed through the capillary and the outflow was collected for PAGE analysis. After having collected about $2 \mu\text{l}$, the velocity was increased.

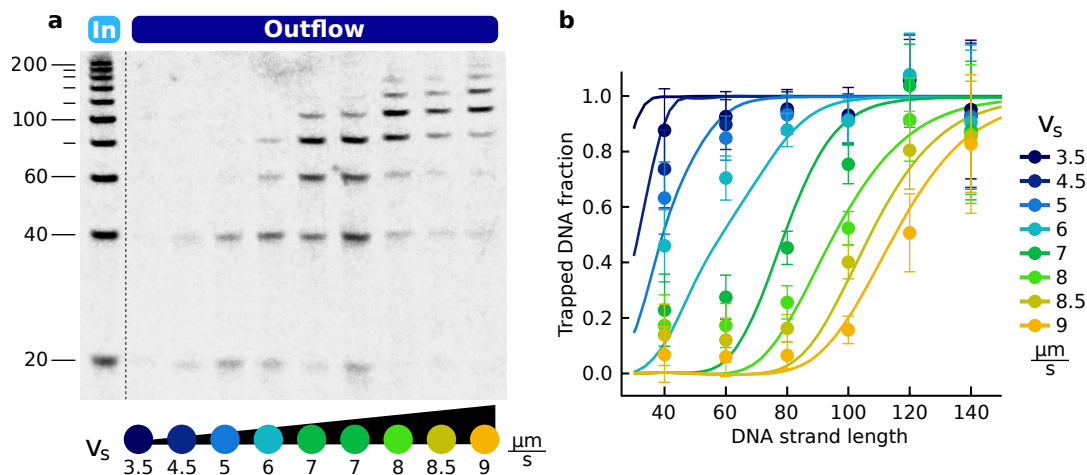


Figure 2.4. Selection Landscape. **a.** Increasing influx velocities shift the cut-off length to higher values. At the beginning of the experiment, the pore was loaded with the same dsDNA ladder as in Fig. 2.3A. The influx velocity was gradually increased from $3.5 \mu\text{m s}^{-1}$ to $9 \mu\text{m s}^{-1}$. **b.** The fraction of trapped DNA, measured from the electrophoresis gel in (A), reveals the selection landscape of this thermal habitat. Data points are in good agreement with the fluid dynamic model. Error bars depict the signal-to-noise ratio of the gel quantification (see section 2.4.2). Figure reproduced from [54].

This procedure was repeated up to a velocity of $9 \mu\text{l/s}$, the results are shown in Fig. 2.4. Details on the duration of each velocity step are given in [54], Supplementary Fig. S6.

Gel electrophoresis of the outflux at each velocity step v_s show the change in strand length composition (Fig. 2.4a). The actual selection landscape is obtained by integrating the concentrations of the different nucleic acids up to each value of v_s . It reveals a monotonous increase of the cut-off length as a function of v_s (Fig. 2.4b). Circles are calculated from the gel data, lines from the full numerical simulation of the pore.

Numeric analysis. In the experiment, the length distribution of DNA in the pore did not reach a steady state in each of the influx velocity steps. To estimate the difference between the apparent selection landscape from the experiment and a selection landscape at fixed time, the system was analysed numerically. Based on a two-dimensional finite element model of fluid flow, convection, diffusion and thermophoresis, the normalized DNA content of the pore was calculated as a function of time and strand length. For each influx velocity, a DNA pulse was flushed into the pore, and its propagation was observed for a duration of 10000 s. As only the total solute content was considered, this reduced the pore to an effectively zero-dimensional filter for all further steps. Details on the model are given in section 2.4.3.

To exhibit sharp length selection for strands that are also subjected to sufficient thermal cycling, the influx velocity must be similar to the mean convection velocity. For the applied

2. Replication and selection in thermo-gravitational reactors

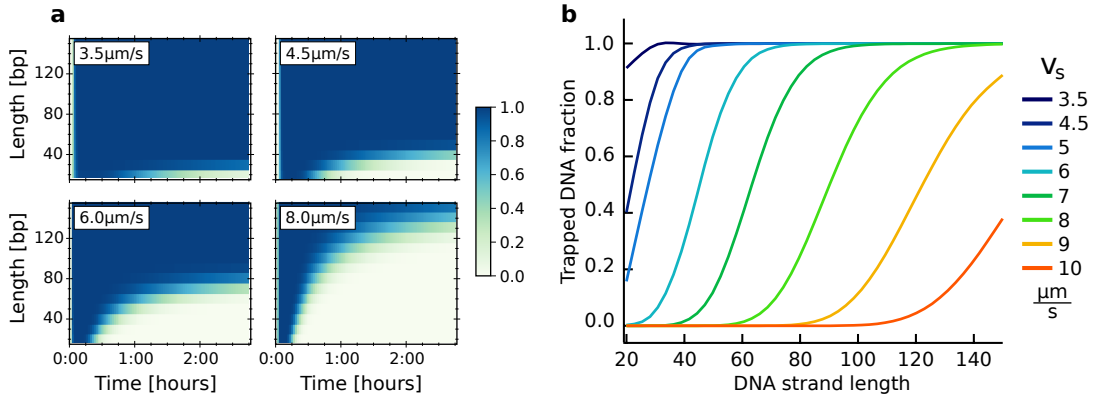


Figure 2.5. Time-dependent selection. **a.** Temporal evolution of strand length distribution inside the pore. Mean influx velocities v_s are listed in each plot, colours give the normalized total concentration of strands as a function of time and strand length. Close to the cut-off length, it takes a few hours before the length distribution reaches a steady state. **b.** Selection landscape at $t = 45$ minutes for a larger velocity range. The slope of the transition around the cut-off length decreases for increasing DNA strand lengths because the Soret coefficient $S_T(L)$ flattens out for larger lengths L .

temperature gradient from 61 °C to 94 °C across a pore of 70 μm, this is about 4.5 μm/s (cf. eq. B.5). Velocities half as high do not cause any significant outflow, except for very short oligonucleotides. Above 9 μm/s, only nucleotides above 150 bp can withstand the external flux and are retained inside the cavity. However, these accumulate so strongly at the cold side of the pore that they hardly undergo any thermal cycling.

The temporal evolution of the strand length distribution inside the pore is presented in Fig. 2.5. For all velocities, the distribution does not reach a steady state in the displayed time interval of about three hours. However, given replication rates on the order of 1–10 h⁻¹ (see below), the steady state length distribution is not necessarily of interest. Therefore, Fig. 2.5b presents the DNA length distribution inside the pore at $t = 45$ minutes. The selection landscape looks very similar to what was found experimentally (Fig. 2.4).

For longer retention times, the selection landscape shifts towards larger strand lengths. These are required by slower replicators, which is particularly relevant regarding non-enzymatic replication or other less evolved reactions in an early Earth context. However, this shift is compensated for by a slightly lower influx velocity, making the pore an equally feasible habitat for a wide range of time scales.

The selection properties of the thermo-gravitational pore can be summarized by a phase diagram in (influx velocity, strand length) space. Fig. 2.6 portrays the two possible behaviours of the pore: DNA is either retained (top left, blue) or washed out (bottom right, red). Lines separate parameter space regions where the pore contains more than 50 % of the initial content from regions with less content. The size of the transition region between the two behaviours is displayed for the last time step (darker blue area) indicating pore

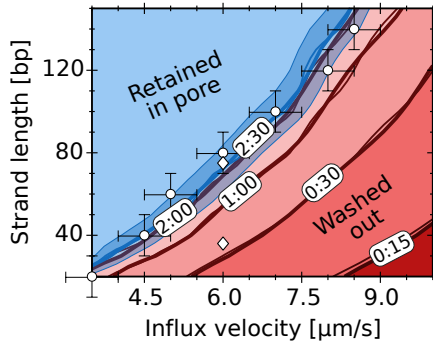


Figure 2.6. Selection phase space (i). Contour lines show the temporal evolution of the selection behaviour in (influx velocity, strand length) space, separating areas above and below 50 % pore content. In the blue region, DNA is retained inside the pore, red means it is washed out. Simulation times are denoted at each contour. For 2:30 h, the transition zone from 25 % to 75 % pore content is indicated by the darker blue area. Circles were measured using the 20–200 bp dsDNA ladder, diamonds represent the two DNA species used in the selective replication experiments.

contents from 0.25 to 0.75. Its width does not change significantly over time (omitted for clarity). Different lines represent the phase boundary at 0:15, 0:30, 1:00, 2:00 and 2:30 h, as indicated. Apparent selection behaviour changes most within the first hour, after which it seems to approach a steady state. Measured data points from the 20–200 bp dsDNA ladder shown good agreement with the fluid dynamic model.

Analytic modelling. Both experimental and numeric studies of the selection behaviour of the cavity are relatively time-consuming, the former especially when comparing different cavity widths. Therefore, an analytic model was developed. Its concept is explained in Fig. 2.3b. The selection behaviour concerning a particular species of molecules is characterized by the effective vertical transport. It is defined as the product of vertical fluid velocity and solute concentration. As the overall concentration of the solute should not have an impact, it is equally feasibly to define an effective velocity. This velocity v_{eff} is a function of the mean external velocity v_s (as introduced above) and solute properties, in particular the Soret coefficient S_T . Positive effective velocities mean that particles move upward and are eventually washed out of the pore. On the other hand, molecules are retained in the pore for $v_{\text{eff}} < 0$. The effective velocity is defined as the weighted average of vertical velocity:

$$v_{\text{eff}} = \frac{1}{c_{\text{tot}} w} \int_0^w c(x) v(x) dx. \quad (2.1)$$

x is the horizontal coordinate, $v(x)$ the vertical fluid velocity, and $c(x)$ the exponential concentration profile in the absence of any fluid velocity, i.e. subject to thermophoresis only:

$$c(x) = c_0 \exp\left(-\Delta T S_T \frac{x}{w}\right), \quad c_{\text{tot}} = \frac{1}{w} \int_0^w c(x) dx. \quad (2.2)$$

The steady state velocity \mathbf{v} inside the cavity is obtained by solving [45]:

$$(\text{div } \eta \text{ grad}) \mathbf{v} = -(\text{body force}) = \text{grad } P - \rho \mathbf{g}, \quad (2.3)$$

2. Replication and selection in thermo-gravitational reactors

where ρ and η are the solvent density and viscosity, g gravitational acceleration and P the pressure. The pressure gradient can be decomposed into a hydrostatic contribution $\bar{\rho}g$ and the spatially homogeneous external pressure gradient that drives the influx from below:

$$\text{grad } P = \bar{\rho}g - \text{grad } P_{ext}. \quad (2.4)$$

Here, $\bar{\rho}$ is the mean density of the fluid. Assuming that the vertical component v of the velocity $\mathbf{v} = (u, v)^T$ is constant along y , the second component of equation (2.3) reads:

$$\left(\left(\frac{\partial_x}{\partial_y} \right) \cdot \eta \left(\frac{\partial_x}{\partial_y} \right) \right) v = \frac{d}{dx} \left(\eta \frac{dv}{dx} \right) = (\rho - \bar{\rho})g - \frac{d}{dx} P_{ext}. \quad (2.5)$$

The temperature distribution inside the chamber is dominated by thermal conduction, being much faster than the fluid velocities which could create any significant convective contribution. Thus, the temperature is given by a constant gradient along the x dimension, parametrized as

$$T(x) = T_0 + (T_1 - T_0) \frac{x}{w} = \bar{T} + \Delta T \left(\frac{x}{w} - \frac{1}{2} \right). \quad (2.6)$$

Here, $\bar{T} \equiv \frac{1}{2}(T_0 + T_1)$ is the mean temperature inside the cavity and $\Delta T \equiv T_1 - T_0$. The density ρ is approximated linearly:

$$\rho = \rho(\bar{T}) - \alpha_T T(x) = \bar{\rho} \left(1 - \alpha_T \Delta T \left(\frac{x}{w} - \frac{1}{2} \right) \right), \quad (2.7)$$

with α_T being the thermal expansion coefficient. Inserting this into equation (2.5) and integrating once, we find:

$$\eta \frac{dv}{dx} = -\frac{g\bar{\rho}\alpha_T\Delta T}{2w} x(x-w) - \delta_{ext}x + C_1. \quad (2.8)$$

The integration constant C_1 will be determined to satisfy the boundary condition of vanishing velocity at $x = w$. $\delta_{ext} \equiv \frac{d}{dx} P_{ext}$ will be expressed as a function of the experimentally more meaningful mean velocity v_s , which is defined as

$$v_s = \frac{1}{w} \int_0^w v(x) dx. \quad (2.9)$$

In addition to the temperature dependence of the density, that of the viscosity η is also considered. In a temperature gradient from 60 to 92 °C, the viscosity varies between about 0.3 and 0.5 mPa s [50], making it a worthwhile correction. Beyond that, the consideration of locally varying viscosity is also relevant for mixed solvents, where thermophoresis affects the ratio of the two components along the temperature gradient. In first order approxi-

ation, η is given by:

$$\eta = \frac{\bar{\eta}}{1 + \beta \Delta T \left(\frac{x}{w} - \frac{1}{2} \right)}. \quad (2.10)$$

β is determined from the experimental data reported in [50], $\beta \Delta T$ amounts to about 0.4. The differences between the cases of constant and non-constant viscosity for the parameters of the experiments are displayed in Fig. B.1 in the appendix.

To simplify expressions before integration, let:

$$\xi = \frac{x}{w} - \frac{1}{2}, \quad A_0 = g\bar{\rho}w\alpha_T\Delta T, \quad A_1 = \beta \Delta T. \quad (2.11)$$

Inserting these and eq. (2.10) into eq. (2.8) yields:

$$\frac{1}{w} \frac{dv}{d\xi} = \left(-\frac{A_0}{2} \left(\xi^2 - \frac{1}{4} \right) - \delta_{ext} \left(\xi + \frac{1}{2} \right) w + C_1 \right) \frac{1 + A_1 \xi}{\bar{\eta}}. \quad (2.12)$$

Integration gives:

$$v(\xi) = -\frac{4\xi^2 - 1}{\bar{\eta}} \left(A_0 w \left(\frac{A_1}{32} \xi^2 + \frac{3A_1^2 - 20}{40(A_1^2 - 12)} \xi - \frac{A_1}{640} \right) - \frac{3}{2} \bar{\eta} v_s \left(\frac{8A_1}{A_1^2 - 12} \xi - 1 \right) \right) \quad (2.13)$$

where

$$C_1 = \frac{(A_1 + 6) \delta_{ext} w - A_0}{12} \quad \text{and} \quad \delta_{ext} = \frac{2 A_0 A_1 w - 360 \bar{\eta} v_s}{5 (A_1^2 - 12) w^2} \quad (2.14)$$

were substituted. Setting $\beta = 0$ and thus $A_1 = 0$ reproduces the familiar case of constant viscosity [20, 50].

In a similar way to the above numeric treatment, v_{eff} can be used to calculate a phase diagram in (influx velocity, strand length) space of the selection behaviour of the pore (Fig. 2.7). It shows excellent agreement with both experimental data and the computationally much more expensive numeric results. Data points and white lines are the same as in Fig. 2.6. Numbers in the white boxes give values of v_{eff} at the respective contour. All contours exhibit positive curvature, which corresponds to less abrupt selection behaviour for longer strands. This is because the Soret coefficient scales as $S_T \propto \sqrt{L}$ [26, 54], such that a difference in e.g. 40 bp does not produce an equally big difference as for shorter strands.

Fig. 2.7b shows phase space boundaries at zero effective velocity for an extended phase space section and pore widths ranging from 50 μm to 100 μm . All other parameters were left unchanged, in particular the boundary temperatures T_0 and T_1 . Smaller pores provide useful selection characteristics only at very low external velocities. Such a scaling is expected from the fact that the convection velocity scales in first order approximation with the square of the pore width. Likewise, wider cavities exhibit sharp selection behaviour for relatively short oligonucleotides, and retain longer molecules up to higher influx velocities

2. Replication and selection in thermo-gravitational reactors

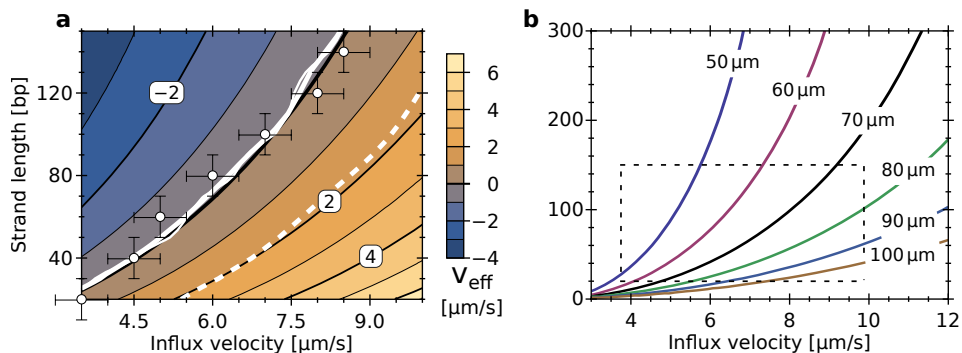


Figure 2.7. Selection phase space (ii): analytic treatment. **a.** The effective trapping velocity v_{eff} predicts a phase diagram that agrees with the findings from the fluid dynamic model. In regions corresponding to negative velocities (blue), DNA is retained in the pore. In yellow areas ($v_{\text{eff}} > 0$), particles are effectively washed out at the respective velocity. Circles show the measurement of the 20–200 bp dsDNA ladder, white lines represent the 2:30 h (solid) and 0:30 h (dashed) contours from Fig. 2.6. Time points 0:15 h, 1:00 h and 2:00 h were omitted for clarity. **b.** Contours at $v_{\text{eff}} = 0$ for pore widths from 50 μm to 100 μm and a larger section of phase space. The dashed rectangle refers to the phase space section shown in panel a. At all pore widths, the boundary temperatures T_0 and T_1 had the same values.

due to their higher convection velocities.

When positive, v_{eff} provides an estimate for how long solute DNA is delayed inside the pore before being washed out. At 2 $\mu\text{m/s}$, it takes 29 minutes to travel a distance of 3.5 mm, which is the size of the heated part of the cavity. This excellently agrees with the results from the numeric model at 0:30 h (dashed line in Fig. 2.7a). A good estimation of the relevant timescales is particularly important considering the replication of molecules inside the cavity or similar reactions with an intrinsic timescale.

2.2.3. Selective survival of replicating populations

In addition to length-selective accumulation, the thermo-gravitational pore also induces thermal cycling of the solute molecules. To study the properties of the thermal oscillations, stochastic simulations were conducted. They comprised a random walk-type model of diffusion, which was based on the temperature gradient and fluid velocity fields calculated using the fluid dynamical simulation presented in the previous section. Fig. 2.8a shows example temperature trajectories of DNA molecules of the two lengths that were studied in a competing replication scenario.

Exponential replication. Thermal cycling parameters were extracted from the trajectories using a simple algorithm. For each DNA species, a cycle was defined using two threshold temperatures, the melting temperature of the primer-template complex T_p and the melting temperature of the full length duplex T_m (dashed lines in Fig. 2.8a). A temperature cycle

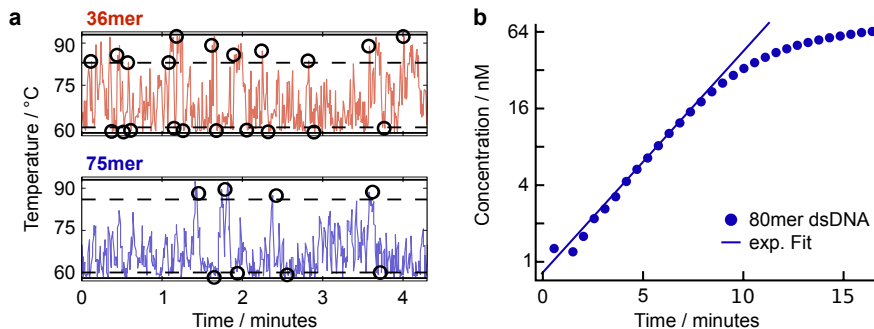


Figure 2.8. Exponential replication by diffusive thermal cycling. **a.** Stochastic single particle simulations reveal temperature trajectories of the two DNA species. Due to their weaker thermophoretic trapping and higher diffusion coefficient, 36mer DNA cycles more than two times faster than 75mer DNA. However, the short species is flushed out of the pore with a time constant of about 5 minutes, while the long DNA is contained for 19 minutes (see Fig. 2.9 for details). **b.** Exponential replication of 80mer DNA with a doubling time of 102 seconds without influx, recorded via SYBR Green I fluorescence. Figure reproduced from [54].

is then defined as the time it takes for a particle to move from the region with $T < T_p$ to a region where $T > T_m$, and back. In Fig. 2.8a, transition events between the different temperature regions are depicted as black circles. A thorough study of the thermal cycling inside the pore is produced in [57].

As both species used the same primer sequences for replication, they share the same value of T_p . For the duplex melting temperature, $T_m^{(S)} = 83^\circ\text{C}$ for the 36mer and $T_m^{(L)} = 86^\circ\text{C}$ for the 75mer DNA were assumed. (The superscripts (S) and (L) identify the short and long species, respectively.) The trajectories show that DNA of 36 bp cycles more than two times faster than 75mer DNA, giving it an intrinsic advantage in replication velocity. This is due to the smaller Soret coefficient $S_T \propto \sqrt{L}$ and higher diffusion coefficient, which approximately scales as $D \propto L^{-0.75}$ [26], yielding a scaling of the thermal diffusion coefficient as $D_T \propto L^{-0.25}$. Given these scaling relations, the reason for the slower cycling of longer DNA is apparent. Diffusive motion of 75mer DNA is restricted more strongly to the cold side of the pore, and detours to the denaturing region of $T > T_m$ are less frequent than for the short DNA.

Despite their erratic shape, the temperature trajectories can support exponential replication reactions. Fig. 2.8b shows the kinetics of the PCR of 80mer DNA without external flow. DNA concentration grows exponentially with a doubling time of 102 seconds, as revealed by an exponential fit to the data in the non-saturating regime. Compared to the replication presented in Fig. 2.10, this was done at excess primer concentrations to speed up the reaction.

A fitness model. In a conventional PCR-like reaction in the non-saturating regime, the growth of the DNA concentration can be modelled using a simple exponential model. Let $c_n^{(i)}$ be the concentration of species i in the n^{th} cycle. The concentration increase compared to the previous cycle is given by $c_n^{(i)} - c_{n-1}^{(i)} = \varepsilon c_{n-1}^{(i)}$, where ε is the replication efficiency. The value of the concentration in cycle n is then given by $c_n^{(i)} = c_0^{(i)} (1 + \varepsilon)^n$. For a continuous replication setting as in the gravitational pore, a continuous time description is more appropriate. It can be obtained by substituting $n \rightarrow t/\tau_c$, where τ_c is the mean period of the temperature oscillations. For the time-dependent concentration, this yields:

$$c_t^{(i)} = c_0^{(i)} \exp\left(k_{rep}^{(i)} t\right) \quad \text{with} \quad k_{rep}^{(i)} = \log(1 + \varepsilon) / \tau_c \quad (2.15)$$

In an analogous way, the effect of the continuously diluting external flux can be modelled using an outflux or dilution rate $k_{dil}^{(i)}$. The combination of both gives the simple exponential expression

$$c_t^{(i)} = c_0^{(i)} \exp\left(\left(k_{rep}^{(i)} - k_{dil}^{(i)}\right) t\right), \quad (2.16)$$

which, however, describes the kinetics of the replication reaction sufficiently well. Due to fluctuations in the experimentally recovered amounts of DNA, which were as low as 0.1 fmol, the relative concentration of the 75mer DNA $\theta_t^{(L)}$ provided are more precise observable than the absolute concentration $c^{(L)}$. The model predicts $\theta_t^{(L)}$ as

$$\theta_t^{(L)} = \frac{c_t^{(L)}}{c_t^{(L)} + c_t^{(S)}} = \left(1 + \frac{c_t^{(S)}}{c_t^{(L)}}\right)^{-1} = \left(1 + \frac{c_0^{(S)}}{c_0^{(L)}} \exp(-\Delta k t)\right)^{-1}, \quad (2.17)$$

where

$$\Delta k = \left(k_{rep}^{(L)} - k_{rep}^{(S)}\right) - \left(k_{dil}^{(L)} - k_{dil}^{(S)}\right) \quad (2.18)$$

is the effective differential growth rate. When Δk is positive, the fraction $\theta_t^{(L)}$ grows with time towards one, and the population eventually consists only of species (L).

Dilution rates were determined from the stochastic simulations presented in Fig. 2.8a. To this end, particles were initially placed at the bottom end of the pore, about 20 pore widths below the heated region. In the lateral dimension, positions were initialized from an equal distribution. The simulation was run for 1000 particles and 10^5 seconds. Particles that moved more than about ten pore widths above the heated region were removed from the simulation, as the fluid velocity points exclusively upwards in that region and returning against the flow is very unlikely (cf. section 2.4.4 for details).

Fig. 2.9a presents the number of particles inside the simulated pore and its temporal evolution for influx velocities v_s around $6 \mu\text{m/s}$. After an initial delay and a slow initial decay phase, the number of particles in the pore can be approximated by a simple exponential decay. The rate constant of that decay is the dilution rate $k_{dil}^{(i)}$. For influx velocities up to

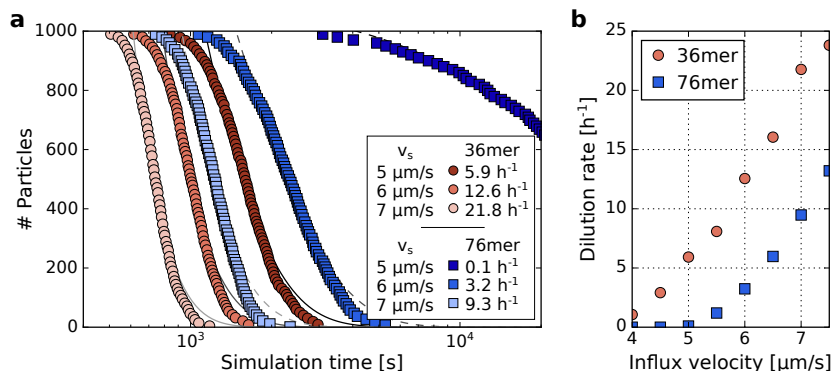


Figure 2.9. Determination of dilution rates. **a.** Simulated particle content for influx velocities of $v_s = 5\text{--}7 \mu\text{m/s}$. The data can be approximated by a simple exponential decay. Dilution rates $k_{dil}^{(i)}$ of the two DNA species are given in the legend. For $v_s = 6 \mu\text{m/s}$, the long DNA has a characteristic abundance time of $1/k_{dil}^{(L)} \approx 19 \text{ min}$, while the short DNA is flushed out after about 5 min. The simulation was run for 1000 particles and 10^5 seconds. **b.** Dilution rates for velocities of $v_s = 4\text{--}7.5 \mu\text{m/s}$. Up to $5 \mu\text{m/s}$, the 76mer DNA is hardly washed out of the pore. This matches the data presented in Fig. 2.4.

$5 \mu\text{m/s}$, the 76mer DNA is hardly affected by the selecting flux, while the shorter 36mer DNA already suffers some dilution (Fig. 2.9b). For velocities $\geq 5.5 \mu\text{m/s}$, the effective dilution rates of both DNA lengths increase with similar slopes.

Selection against the shortest. Using the combination of length-selective accumulation, replication, and an external feeding flow, the thermo-gravitational pore can overcome Spiegelman’s dilemma of the tyranny of the shortest. This was studied using a mixed DNA population of two lengths, 36 bp and 75 bp. A capillary of 2.5 mm length was loaded with unlabelled dsDNA containing 1 nM of each species. A temperature gradient from $61 \text{ }^\circ\text{C}$ to $94 \text{ }^\circ\text{C}$ was applied across the pore. After loading, an external flow with a mean velocity of $v_s = 6 \mu\text{m/s}$ was applied. It consisted of Taq reaction buffer containing polymerase, nucleotides, and 7 nM fluorescently labelled DNA primers.

The experiment was run for seven hours, during which the reaction volume was exchanged approximated 150 times. The outflux of the continuously running reaction was collected in aliquots for PAGE analysis. Fig. 2.10a presents Cy5 fluorescence intensities from the gel electrophoresis of samples drawn at four different reaction times. Coloured areas indicate the local baseline correction that was applied to each peak.

Initially, the population of the 36 mer DNA is bigger than that of the 75 mer. This is because both species initially are at the same concentration, and the shorter strands replicate faster than the long ones. However, accumulation of the short DNA is strongly affected by the external flux, such that the strands remain inside the pore for only about five minutes on average (cf. Fig. 2.9). The 36mer DNA is incapable of replicating sufficiently fast as to maintain its population size and dies out. On the other hand, the concentration of the

2. Replication and selection in thermo-gravitational reactors

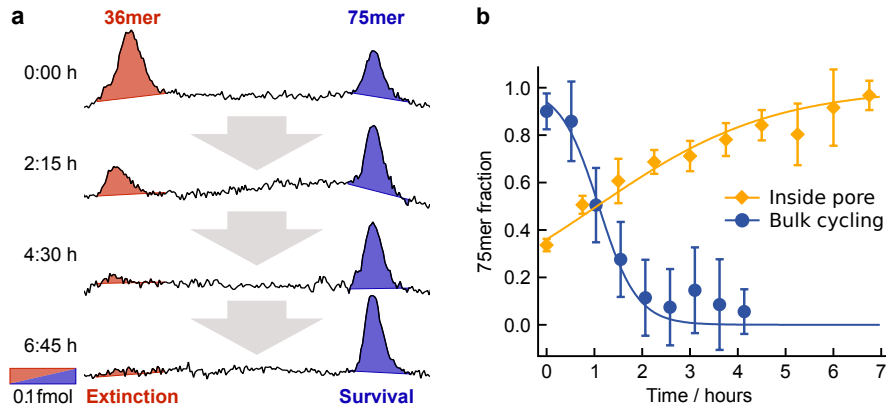


Figure 2.10. Selection of a binary population of replicating DNAs. **a.** An initially loaded binary DNA population replicated inside the pore. Quantitative gel electrophoresis shows that the long DNA species survived, while the short one goes extinct. The coloured rectangle in the bottom left gives the area corresponding to 0.1 fmol DNA. Details on the quantitative analysis are given in section 2.4.2. **b.** Temporal evolution of the fraction of 75mer DNA inside the pore (yellow diamonds) and, as a reference, in a well-mixed situation (blue circles). The selection pressure created by the temperature gradient favours the longer DNA species over the short one. Absolute fitness values were 1.03 for the 75mer and 0.87 for the 36mer. In the absence of the selection pressure, the short DNA quickly dominates the population, reproducing Spiegelman’s tyranny of the shortest. Error bars show the signal-to-noise ratio of the gel images. Figure reproduced from [54].

75mer DNA rises slowly over the course of the experiment. The fraction of the long DNA of the total concentration $\theta_t^{(L)}$ as a function of time is presented in Fig. 2.10b.

The stochastic model yields dilution rates of $k_{dil}^{(L)} = 3.2 \text{ h}^{-1}$ and $k_{dil}^{(S)} = 12.5 \text{ h}^{-1}$. Replication rates were experimentally determined in a PCR cycler using a protocol that matches the temperature cycling times inside the pore (Fig. 2.8a). They were measured as $k_{rep}^{(L)} = (3.3 \pm 0.4) \text{ h}^{-1}$ for the long strands and $k_{rep}^{(S)} = (12.05 \pm 0.06) \text{ h}^{-1}$ for the short strands. In the absence of dilution, the short strands replicate 3.7 times faster inside the pore than the longer ones, partly reflecting their a priori evolutionary advantage of a higher replication efficiency ε . In addition, they experience faster temperature cycles (19 s) than the long strands (44 s).

The absolute fitness for a specific genotype is defined by the ratio of individual (strands) before and after selection. During the lifetime $1/k_{dil}^{(L)} = 18.8 \text{ min}$ of long strands in the pore, their population grows by around 3 %, while the population of short strands shrinks by 13 %. This translates to 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 read 1.01 and 0.96, respectively.

In Fig. 2.10b, the experimental values (yellow diamonds) are presented alongside the model prediction (yellow line). In the length-selective environment produced by the pore, the relative concentration of the 75mer DNA $\theta_t^{(L)}$ slowly rises to a value of 1, meaning

that the long DNA species survives while the 36 bp species goes extinct. In contrast, in the well-mixed case lacking the pore's selection pressure, short DNA quickly takes over the population (blue line and circles).

2.3. Discussion

The experimental and theoretical 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 was likely abundant because of porous, partially metallic volcanic rock, both near the surface and at submarine sites [49, 96]. Due to their more than 100-fold larger thermal conductivity than water [113], metallic inhomogeneities next to a pore can focus the thermal gradient from centimetres down to a micrometre-sized cleft. 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 (cf. Supplementary Fig. S1 and S2 from [54]), instead of using the simple geometric setting of a single rectangular pore. At the bottom, a wide pore could provide the necessary temperature difference for replication [11]. On its top, the outflow would be constricted through one or more thin, but longer selecting pores. Their increased length of several centimetres instead of few millimetres compensates for the reduced temperature difference linearly [2]. Moreover, accumulation works equally well with thermal gradients 1000 fold less steep [46]. While timescales are too slow for laboratory experiments, this would not interfere with the overall ability to drive selective replication.

Selectivity requires that replication and selection occur on similar timescales. In particular, replication must not be too fast to excel selection. This is expected to be readily fulfilled in an early Earth scenario, as one would assume early replication mechanisms to have had lower per-cycle efficiencies than the highly optimized Taq polymerase. Lower efficiencies would be caused by a lesser intrinsic performance or lower concentrations of primers or polymerase, both being plausible in the context of early evolution. Thereby, the advantage of shorter abundance in the colder regions of the pore is reduced, as replication velocity is limited by the time required for the actual replication process.

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

2. Replication and selection in thermo-gravitational reactors

and selection studied here will also work for replication systems that require colder temperatures, including ribozymes [42, 104] or Q β replicase [40]. However, early replication systems are likely to rely on high temperatures for thermally induced strand separation.

For the PCR used in the experiment, strand lengths are highly controlled by the primers. In comparison, reactions involving ligations have a tendency to extend the strands with partial templating [35] and initiate the length extension of the genetic polymers. To extend this work and achieve Darwinian evolution, the replication process would require a substantial mutation rate that also affects 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.

However, even for replicating systems of variable length, replication in the described environment will not produce arbitrary large polymers, a problem analogous to the *tar problem* encountered in the original Miller-Urey experiment. The Soret coefficient's scaling as the square root of nucleotide length holds for lengths well beyond several kilobases [26]. When molecules become too large, accumulation to the cold side of the pore gets too strong, and they do no longer undergo thermal oscillations. The latter, however, is required for their replication, which means that they are excluded from any further replication cycles. Thereby, the pore prevents mutating replicating systems from diverting most of their resources into (large) dead ends.

By providing an upper and lower limit to the size of a replicating system, hydrothermal pores also constrain sequence space to some extent. While there were most probably several mechanisms which constrained the part of sequence space that was sampled by evolution², an upper limit to the length of self-replicating oligonucleotides may have contributed as one of the constraints.

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 favour 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 binding of aptamers [3].

Conclusion. The experiments reveal how temperature gradients, the most simple out-of-equilibrium setting, can give rise to local environments that stabilise molecular replication against entropic tendencies of dilution, degradation and negative length selection. A

²As sequence space growth exponentially with nucleotide length, its size quickly exceeds the estimated number of atoms in the observable universe, which is about 10^{80} [116]. This corresponds to a length of $N = 80 / \log_{10}(4) \approx 133$ nucleotides.

thermal gradient drives replication of 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 neighbouring 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.

2.4. Materials and methods

2.4.1. Experimental setup

Microfluidics. Borosilicate glass capillaries were connected to a feedback controlled syringe pump (neMESYS, Cetoni, Germany) via high purity PFA tubing (HPFA+, Upchurch Scientific, USA). 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 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 thoroughly prior to loading into the system in order to avoid formation of air bubbles 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, and at 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 ambient air, tubes were maintained at 90 °C during injection of the assay into the system.

Real-time fluorescent imaging. Fluorescent imaging of DNA was realised with a 90°-tilted upright microscope (Axioscope A1, Zeiss), using a 2.5x objective (Plan-Neofluar 2.5x 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).

2.4.2. Gel electrophoresis

Native polyacrylamide gel electrophoresis (PAGE) was performed in gels containing 12.5 % acrylamide. Gels were run in 1x TBE buffer at electric field strengths of 60 V/cm at 30 °C for 13 minutes. Afterwards, gels were stained in 1x TBE buffer containing SYBR Green I (Invitrogen, Germany) at 1x concentration for four minutes, followed by a one minute washing step using 1x TBE buffer. *Denaturing* gel electrophoresis was conducted adhering to a standard protocol [61] in 12.5 % gels containing 50 % urea. Before loading to the gel, DNA samples were denatured in a formamide glycerol buffer at 95 °C for 2 minutes, followed by shock cooling on ice. Electrophoresis was performed in two steps, first for 5 minutes at 7.5 V/cm, then at 60 V/cm in TBE buffer at 45–50 °C for 13 minutes.

Imaging. Gels stained with SYBR Green I were imaged 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) LED excitation (LED 470 nm, Thorlabs, Germany). Cy5 labelled reaction products recovered from the extra-cellular selection and replication experiment were illuminated with two spectrally filtered LEDs (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, OD2 blocking, Newport, USA, resulting in an excitation rejection of OD8+) by an actively cooled CCD camera (Orca 03-G, Hamamatsu, Japan).

Quantitative analysis. Gel image quantification was done with a self-written LabVIEW program, after point-like outliers have been removed using NIH ImageJ [101]. Before integrating intensities, images were corrected for the inhomogeneous illumination produced by the LED illumination of the gel. If necessary, they were rotated to ensure that the lanes are aligned vertically. To improve the signal to noise ratio, the intensity of each gel lane was then integrated along the horizontal axis, i.e. perpendicular to the direction of migration. For each band, a local linear background was subtracted, leaving only the actual DNA content (shaded regions Fig. 2.10a). The uncertainty of this integral was estimated using the standard deviation of the values around the base points of the linear background.

2.4.3. Numeric calculations

Finite element simulations of nucleic acid accumulation in a thermal gradient were performed in COMSOL Multiphysics, similar to simulations published before [2]. The simulation included temperature (conductive heat transfer module) and velocity fields (incompressible Navier-Stokes module), as well as concentration (convection, diffusion, and thermophoresis modelled using the general form PDE module).

Simulation of the fractionation of a DNA ladder, as presented in Fig. 2.4 (section 2.2.2), required multiple steps. First, the length-dependent propagation and trapping of a mixed-length DNA pulse through the trap was simulated for different influx velocities. From this data, the length distribution of strands that have been flushed were calculated as a function of time. Time steps were chosen to match the experimental steps in inflow velocities. For each velocity step, the concentration of the DNA leaving the trap was normalised to the step duration.

2.4.4. Stochastic simulations

Stochastic information of single particle motion was generated using a self-written computer program. This information was required to calculate single particle cycling rates, and was also used for calculating particle dilution rates. Individual particles were traced on a biased random walk trajectory inside the combined temperature and velocity fields that were computed using the finite element method. The bias accounted for convective and thermophoretic drift.

For the selective replication experiment, a cavity with a thickness of 70 μm and a heated region of 3.5 mm was considered. The simulated area extended 1.5 mm below and 1.0 mm above the heated region. Taking into account that the accumulation pattern extends somewhat into the area below the heated region (cf. Fig. 2.3a), this was sufficiently large as to exclude effects of the top and bottom boundaries of the simulation.

3. Transfer RNA as a codon replicator

3.1. Introduction

The question of simple self-replicating systems is one of the hardest problems in the quest for understanding the origins of life. Replication is the central feature of living systems, required to evade degradation and permit evolution. In biology, replication happens at many different levels from the overall structure of multicellular organisms or cells down to the replication of the genome or individual proteins. Translation of proteins also replicates information, which is first encoded in a sequence of codons in a messenger RNA strand, and then converted into a sequence of amino acids (disregarding degeneracies in the genetic code). Despite the significance of replication in general and translation in particular, their origins remain largely subject to speculation [22, 93, 122].

The RNA world hypothesis argues for RNA as candidate for a first replicator [9, 38], but RNA may also be just the last step that preceded modern DNA/RNA/protein life [90]. As alternatives to RNA, numerous polymers have been studied, substituting sugars, nucleobases or backbones for smaller or simpler moieties [33, 59, 99, 100]. These do not only form duplexes similar to DNA or RNA, but can also fold into well-defined structures [83]. Depending on the particular differences between any of these XNAs and DNA or RNA, some have compatible helix geometries and can form heteroduplexes with DNA, RNA, or both. On the other hand, it is not yet clear if any of these analogues could have been abiotically produced before RNA. Given that, the question arises whether information replication can be achieved disregarding the particular chemical nature of a polymer and rely only on its very basic property: the ability to form double helical structures or similar duplexes.

Another issue is the fact that simple non-enzymatic replication chemistries were probably rather restricted in their fidelity, which limits the length of polymers they can support [90, 110]. One way to circumvent this is the replication of sequence snippets. Distinguishing between sequences can be achieved more easily compared to single bases, and larger polymers can be produced from less pieces. For the individual oligomers, a somewhat lower copying fidelity of synthesis can then be tolerated, because the number of bases that needs to be faithfully replicated is lower.

In addition, reversibility may play a major role for an early replication mechanism, as it allows to re-use parts instead of requiring their re-synthesis. This is similar to present-day translation, where tRNA molecules are re-used in temporal sequence instead of the spatial

3. Transfer RNA as a codon replicator

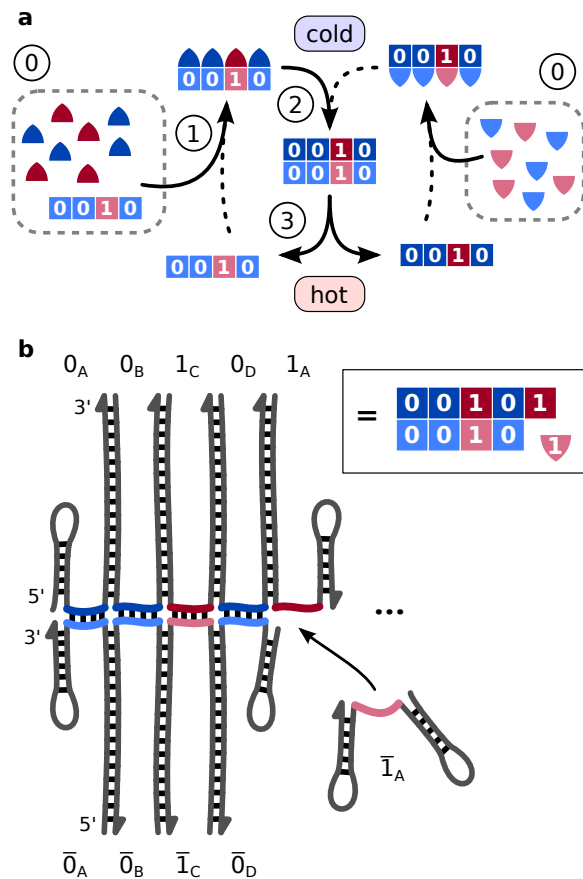


Figure 3.1. Heat-driven replication mechanism. **a.** Schematic representation of the thermally driven replication mechanism. (1) Strands with matching codon domains bind to the template. (2) Fluctuations in the bound strands' hairpins facilitate the hybridization to neighbouring strands. (3) Subsequent heating splits replicate from template, freeing both for the next cycle. **b.** Secondary structure of a complex of several DNA strands during template replication (steps 1+2). Each unit is an 82–84 nt DNA strand comprising two hairpin loops with an interjacent unpaired domain of 15 nt length (here: strand $\bar{1}_A$). Backbone domains are drawn in grey, codon domains are coloured. Information coding (red/blue) is achieved by different sequences. In the notation of this chapter, the displayed complex will be labelled 00101:0010 when sequence information is included, or 5:4 if only the size of the complex is of interest.

ordering presented here.

In this chapter, I show that replication of biological information can be realized via fully reversible base pairing and stacking interactions. Instead of any chemical ligation mechanism, the replication mechanism is driven by thermal oscillations. This makes the mechanism independent of any particular non-enzymatic ligation chemistry [24, 31, 81, 84, 94, 119]. The replication scheme is a generalization of a simpler system of only four interacting nucleotides [53], and is able to encode information without an a priori imposed limit.

3.2. Results

3.2.1. Replication Mechanism

The replication mechanism is a template-based replication, where, instead of single nucleotides, information is encoded in a succession of oligomers. Their structure is derived from that of transfer RNA. With just a few point mutations, tRNA secondary structure can be changed from the usual cloverleaf structure to one comprising two hairpin loops that surround the anticodon and a few neighbouring bases [53]. Using a number of these double-hairpin strands, a cooperatively replicating set of sequences was constructed (Fig. 3.1b).

Instead of manually mutating tRNA sequences to achieve the desired foldings, sequences were generated de novo using a computer algorithm (see section 3.4.1). To keep the size in the range of an average tRNA molecule [105], stem-loops consisted of 30–33 nucleotides, and the information-coding interjacent domains had lengths of 15 nt. As the mechanism is expected to work equally well in DNA as in RNA, it was implemented in DNA.

The strands' hairpin domains are sequentially complementary, such that the 3' hairpin of one strand is complementary to the 5' hairpin of the next. In order to not conceptually limit the length of the resulting meta-sequences, these hairpins, serving as backbone, have to be repetitive. Here, four different backbone domains were used, which makes the minimal cyclic complex large enough to keep the coding domains accessible even in cyclic configuration. For the information encoding (“codon”) domains, a binary code was used (drawn in red/blue). For replication, two sets of strands replicate sequences in a cross-catalytic manner (Fig. 3.1a). The two sets feature complementary coding domains and orthogonal backbone domains.

Replication is driven by thermal oscillations, similar to a two-temperature polymerase chain reaction (PCR). (1) At the annealing temperature, strands with matching codon domains bind to a template of already assembled strands. (2) Thermal fluctuations cause repeated unfolding of the hairpins. In free solution, the hairpin usually just re-folds into its initial state. However, when a strand is bound to the template, next to a strand with a complementary hairpin, the hairpins bind to each other and thereby connect adjacent strands. The formed duplex is generally more stable than the two single hairpins, as it features more paired bases. (3) Subsequent heating splits the newly formed replicate from the template, such that both can serve as template in the next cycle.

A simple model. The presented replication mechanism can be described by a very simple cross-catalytic model. It considers only the concentrations of complexes 0000 (c) and $\overline{0000}$ (\bar{c}) and is defined by

$$\frac{d}{dt}c(t) = k\bar{c}(t) + k_0, \quad \frac{d}{dt}\bar{c}(t) = \bar{k}c(t) + \bar{k}_0. \quad (3.1)$$

Here, k and \bar{k} are the effective per-cycle mutual catalysis rates, and k_0 and \bar{k}_0 are rates of spontaneous product formation, which all depend on the shape of the particular temperature oscillations. For $c \approx \bar{c}$, the two equations collapse into simple exponential growth. This description abstracts the whole network of differently sized complexes, which cross-catalyse their mutual formation, onto two concentrations. Further, it provides temporal resolution only on a per-cycle basis, which, however, is what was analysed experimentally.

The model can be solved in closed form, but does not account for saturation effects due to depletion of monomers. Therefore, it is not valid for concentrations similar to the total

3. Transfer RNA as a codon replicator

concentration of each strand. The solution is

$$\begin{pmatrix} c \\ \bar{c} \end{pmatrix} (t) = \frac{1}{\sqrt{k\bar{k}}} \begin{pmatrix} k\bar{c}_0 + k_0 \\ \bar{k}c_0 + \bar{k}_0 \end{pmatrix} \sinh(\sqrt{k\bar{k}}t) + \begin{pmatrix} c_0 + \bar{k}_0/\bar{k} \\ \bar{c}_0 + k_0/k \end{pmatrix} \cosh(\sqrt{k\bar{k}}t) - \begin{pmatrix} \bar{k}_0/\bar{k} \\ k_0/k \end{pmatrix}. \quad (3.2)$$

For the DNA sequences that were used in the experiment, symmetric rates $k = \bar{k}$ and $k_0 = \bar{k}_0$ were assumed. This is justified by the symmetry of binding energies and melting temperatures (Fig. 3.2 and 3.3 below). Using that assumption, equation (3.2) is simplified to

$$\begin{pmatrix} c \\ \bar{c} \end{pmatrix} (t) = \begin{bmatrix} \bar{c}_0 \\ c_0 \end{bmatrix} + \frac{k_0}{k} \sinh(kt) + \begin{bmatrix} c_0 \\ \bar{c}_0 \end{bmatrix} + \frac{k_0}{k} \cosh(kt) - \frac{k_0}{k}. \quad (3.3)$$

Most of the experiments below are started with zero initial concentration of c , i.e. $c_0 = 0$. Inserting this simplifies equation (3.3) to

$$c(t) = \bar{c}_0 \sinh(kt) + \frac{k_0}{k} (\exp(kt) - 1), \quad (3.4)$$

$$\bar{c}(t) = \bar{c}_0 \cosh(kt) + \frac{k_0}{k} (\exp(kt) - 1). \quad (3.5)$$

3.2.2. Strand design

To allow the thermally driven replication mechanism to work, binding energies of the different duplexes in the system must fulfil certain criteria. Duplexes of complementary hairpins must be the strongest in the system. These backbone bonds connect consecutive strands (e.g. $0_A0_B0_C$), and must in particular be more stable than binding between a template and a newly formed product complex (e.g. between $0_A0_B0_C$ and $\bar{0}_A\bar{0}_B\bar{0}_C$). Otherwise, any newly formed product would be destroyed on the way of being released from its template. Finally, hairpin stems must not be too stable in order to not freeze out the thermal fluctuations required during step (2).

DNA sequences were generated using the NUPACK software [124] and manually adapted to achieve a better symmetry in binding energies and melting temperatures. A detailed description of the design process is given in section 3.4.1, sequences are listed in Table 3.1. The software was also used to compute free energies ΔG and melting curves. For free energies, the reference level $\Delta G = 0$ corresponds to configurations with only unpaired bases. Melting curves were generated from equilibrium concentrations over a range of different temperatures.

Computed free energies of all two-molecule complexes at 35 °C show that the set fulfils above requirements (Fig. 3.2a). ΔG values of backbone and codon duplexes, and all undesired bonds split into three distinct groups. Backbone duplexes are significantly more stable than codon duplexes, and undesired bonds are suppressed.

As the relation of free energies at a particular temperature does not necessarily coincide

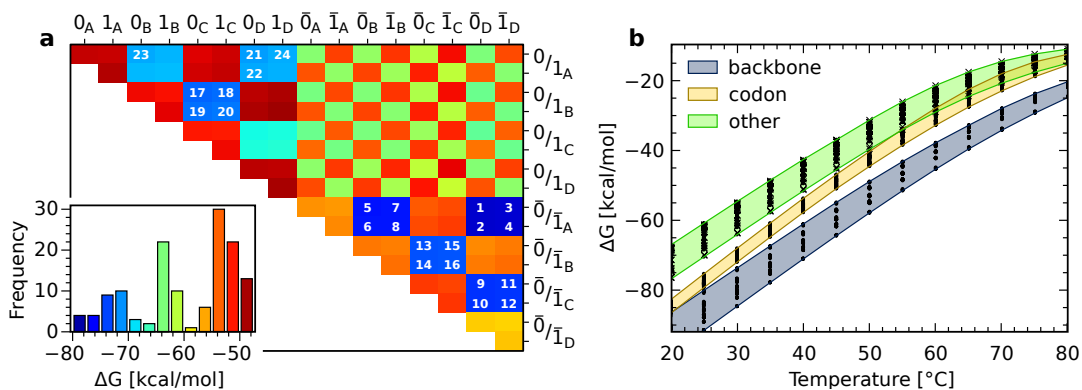


Figure 3.2. Calculated free energies of all DNA dimers. **a.** Free energy matrix at 35 °C. Backbone bonds are strongest (blue), followed by coding dimer bonds (yellow, green). Undesired bonds (e.g. $1:\bar{0}$ or $X_A Y_C$ with $X, Y \in \{0, 1\}$) are weaker than any target bonds. The histogram in the inset shows that binding energies are split into three groups. **b.** Free energies ΔG as a function of temperature. Coloured areas show the full range of data points in each category. Black circles represent heterodimers, homodimers are indicated by crosses. Backbone dimer binding (blue) is strongest over the whole temperature range. Below 50 °C, codon bonds (yellow) are more stable than all undesired dimers (green).

with how melting temperature relate [69], such an analysis was repeated for temperatures from 20 °C to 80 °C. The data is summarized in Fig. 3.2b. Backbone duplexes are the strongest over the whole temperature range. Codon bonds have a lower ΔG than all undesired dimers up to 50 °C. Above, neither of these two groups form at all, because binding is too weak.

Oscillation temperatures. To determine a suitable range for thermal cycling, melting curves of some representative dimers were measured and compared to the simulation results (Fig. 3.3a). Single codon duplexes (e.g. $0_A \bar{0}_A$, red lines) have a melting temperature of about 48 °C, as determined using UV absorption measurements and from simulation. Backbone bonds, as for example $0_A 0_B$, have melting temperatures between 76 °C and 79 °C. Simulation results deviated from measurements by at most +3 °C.

Computational analysis of a mixture of 200 nM of each of the six strands 0_A-0_C , $\bar{0}_A-\bar{0}_C$ at temperatures between 35 °C and 92 °C supports the feasibility of the intended replication mechanism. Figure 3.3b shows equilibrium concentrations of all complexes comprising up to six molecules. The only complexes at non-negligible concentrations are the desired configurations $0_A 0_B 0_C : \bar{0}_A \bar{0}_B \bar{0}_C$, $0_A 0_B 0_C$, $\bar{0}_A \bar{0}_B \bar{0}_C$, $0_A 0_B$, $\bar{0}_A \bar{0}_B$, and monomers. Starting from e.g. 45 °C, the first melting transition that is encountered when increasing the temperature is that of complex $000:\bar{0}\bar{0}\bar{0}$, which dissociates into $0_A 0_B 0_C$ and $\bar{0}_A \bar{0}_B \bar{0}_C$. These trimers remain stable up to about 75 °C, before they disintegrate.

The exact locations of the thermal oscillation temperatures ($T_{\text{base}} = 45$ °C, $T_{\text{peak}} = 67$ °C) were reassessed in actual cycling experiments (Fig. 3.6b and 3.8a). There, each of the

3. Transfer RNA as a codon replicator

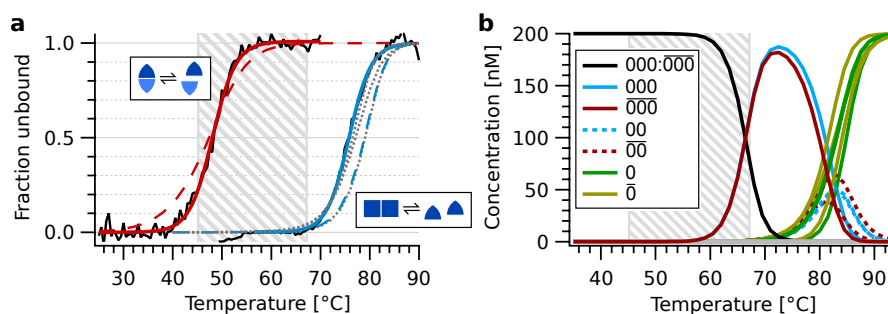


Figure 3.3. Determination of thermal oscillation temperatures. **a.** Melting temperatures of complementary codon domains (red, strands $0_A+\bar{0}_A$, determined via quenching of the Cy5 dye attached to strand 0_A) and backbone domains (blue, strands 0_A+0_B , determined via UV absorption). Dashed lines show simulation data. Dotted grey lines depict simulated melting curves for 0_B+0_C , 0_C+0_D , 0_D+0_A . The hatched area indicates the thermal oscillation range of 45 °C to 67 °C. **b.** Simulated equilibrium concentrations in a reaction mixture containing 200 nM of each of 0_A-0_C , $\bar{0}_A-\bar{0}_C$. The peak temperature of the thermal oscillations mostly melts the bond between the trimers $0_A0_B0_C$ and $\bar{0}_A\bar{0}_B\bar{0}_C$, but is below the melting transition of $0_A0_B0_C$ or $\bar{0}_A\bar{0}_B\bar{0}_C$ (yellow, green).

two temperatures was varied around the values derived from the simulations, and the choice of temperatures was made to optimize reaction yields while maintaining background rates as low as possible. Moreover, equilibrium data does not provide information about (un-)binding kinetics.

3.2.3. Complex formation yields

Strand assembly was analysed using native polyacrylamide gel electrophoresis (PAGE). To this end, different combinations of strands were mixed at concentrations of 200 nM per strand and annealed from 95 °C to 25 °C (cf. section 3.4.2). All strands assembled as targeted and illustrated in Fig. 3.1b. Comparing the bands produced by different subsets of strands allowed to identify all gel bands (Fig. 3.4). Despite their branched tertiary structure, all complexes could be resolved.

Di- and trimers appeared as double bands (lanes 4, 5, 10), which correspond to different spatial configurations of hybridized backbone domains and the neighbouring hairpins. Similar migration patterns of branched DNA complexes have been reported elsewhere [78]. For quantitative analysis, the intensity of these bands was combined.

Partially assembled complexes of two or three strands bound to a four-strand template could be resolved (lanes 13–15, 17, 19). Complexes containing single codon bonds are not stable during electrophoresis, and break at that bond (lanes 2, 3, 16, 18). This allows to differentiate fully assembled complexes from those where monomers are bound to a template but have not formed backbone duplexes.

The assembled complexes showed significantly higher friction coefficients during gel elec-

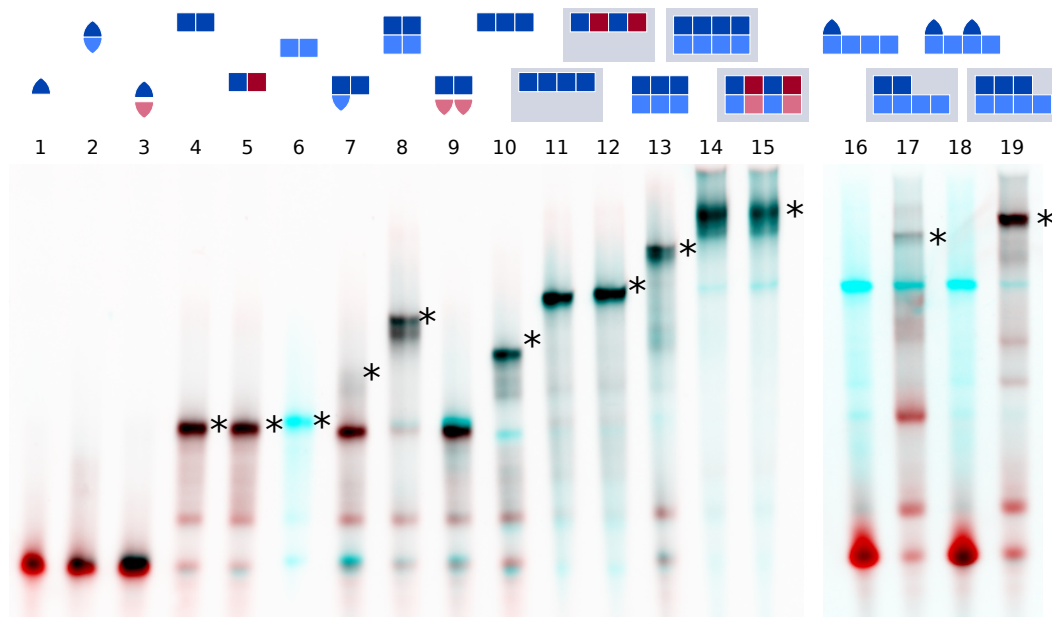


Figure 3.4. Assembly yields. Assembly yields of different subsets of the cross-replicating system of strands. Samples were annealed from 95 °C to 25 °C as described in section 3.4.2 and contained strands at 200 nM concentration each. Lane contents are indicated at the top of each lane. Comparison of different lanes allowed for the attribution of bands to complexes. Complexes incorporating all loaded strands are marked with an asterisk (*). The red channel shows the intensity of 0_A -Cy5, the cyan channel shows SYBR Green I fluorescence. Single codon bonds were mostly not stable during electrophoresis (lanes 2, 7, 16, 18). Complexes of 2–4 strands bound to a three- or four-strand template are resolved (lanes 13–15, 17, 19).

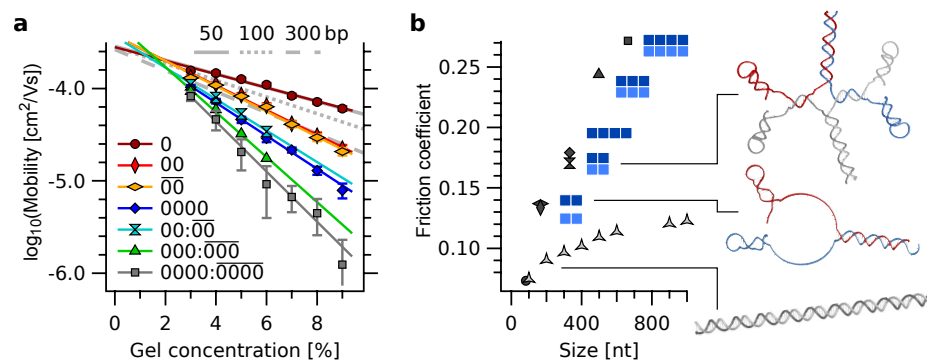


Figure 3.5. Gel mobilities of different complexes compared to linear dsDNA. **a.** Ferguson plot of differently sized complexes compared to linear dsDNA (grey lines) of comparable mass. The slopes of $\log(\text{mobility})$ vs. gel concentration are proportional to the friction constants of the molecules [92]. **b.** Linear dsDNA shows significantly lower friction constants than any of the complexes of at least two molecules. This is due to the branched structure of the complexes and conforms with the suggested assembly geometry. Symbolic complexes are indicated next to the data points. Idealized tertiary structures of complexes 00 and 00:00, and 100 bp dsDNA are given as size reference.

3. Transfer RNA as a codon replicator

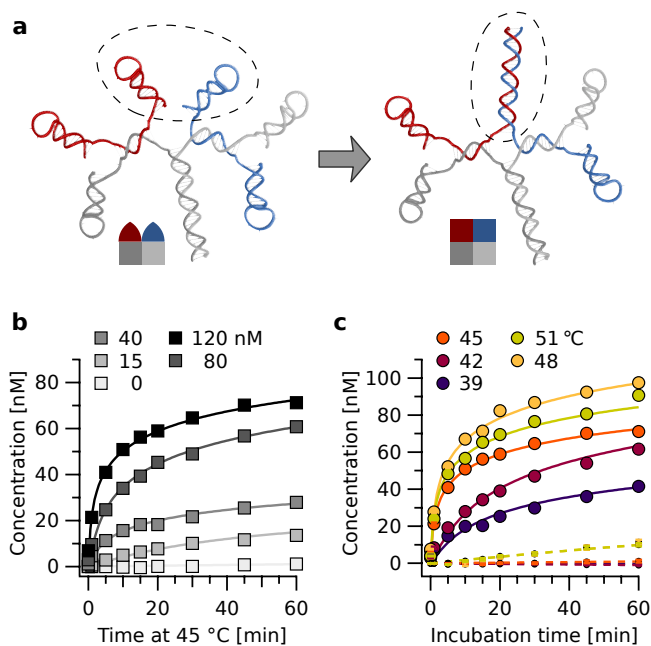


Figure 3.6. Template assisted product formation. **a.** Tertiary structures for the formation of a single backbone duplex. **b.** Kinetics of tetramer formation at 45 °C with different concentrations of template 0000. Data include concentrations of all complexes containing strands of length 4 (grey boxes in Fig. 3.4). **c.** Product formation proceeds over a broad temperature range. Large symbols show data for reactions at 120 nM template, small symbols show the spontaneous contribution only. The latter increases at $T > 45$ °C. Above 48 °C, binding of monomers to the template gets weaker, slowing down template assisted formation. This is consistent with the melting temperature of the codon domains, as determined experimentally and from simulations (Fig. 3.3).

trophoresis than linear double-stranded DNA of similar mass (Fig. 3.5). Complexes of two to four strands show a 1.6–1.8 fold higher friction coefficient than dsDNA. For larger complexes (4:4, ca. 660 nt), this ratio is about 2.4. The increased friction is due to the branched structure of the complexes, and conforms with the suggested strand assembly geometry.

3.2.4. Templating kinetics

Formation of backbone bonds (illustrated in Fig. 3.6a) is catalysed by the presence of already assembled complexes, which act as a template. Assembly kinetics at 45 °C were recorded in reactions containing 200 nM of each strand for a range of template concentrations. Templated reactions with 15–120 nM template surpassed 40 % yield within 10 to 15 minutes (Fig. 3.6b). The untemplated, spontaneous reaction proceeded much slower, and only produced 2 nM within 120 minutes.

Assembly rates showed a strong dependence on incubation temperature (Fig. 3.6c). At 39 °C, the reaction progressed a lot slower than at 42 °C or 45 °C. At this relatively low temperature, hairpin fluctuations are strongly damped and start being frozen out. Binding between complementary codon domains still occurs, but the formation of bonds between neighbouring strands is the limiting step.

On the other hand, template-directed assembly is again slowed down at temperatures above 48 °C. At these temperatures, codon duplexes are destabilized too strongly, and strands do not stay attached to the template for a significantly long time. Codon melting temperatures were determined as 48 °C (Fig. 3.3a).

The slower kinetics of template-directed product formation are partially overlaid by the

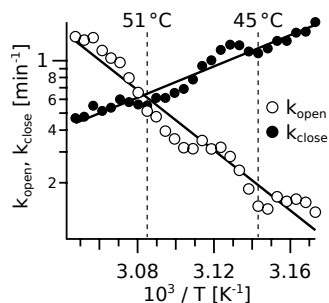


Figure 3.7. Hairpin fluctuation kinetics. Kinetics were extracted from hysteresis in thermal melting curves, as described in section 3.4.4. Data were measured on strand 0_D using SYBR Green I fluorescence. Slopes of $\ln(k_{\text{open}})$ and $\ln(k_{\text{close}})$ on the $1/T$ axis correspond to the energies of hairpin opening and closing (E_{open}/R and E_{close}/R). The fit revealed $E_{\text{open}} = 39 \text{ kcal mol}^{-1}$ and $E_{\text{close}} = -20 \text{ kcal mol}^{-1}$. Extrapolated to $39 \text{ }^\circ\text{C}$, the opening rate drops to 0.06 min^{-1} .

spontaneous formation of bonds, which was measured in reactions lacking an initial template of strands $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ (small circles in Fig. 3.6c). This spontaneous rate is the source of sequence mutations and will be further discussed in section 3.2.7.

Fluctuation kinetics. Kinetics of template formation were compared to hairpin fluctuation rates, which were measured from hysteresis in thermal melting curves as described in [69] and summarized in section 3.4.4. As the measurement included both hairpins of the strand, the determined value corresponds to the average value of the two hairpins. At $45 \text{ }^\circ\text{C}$, an opening rate of 0.2 min^{-1} was found (Fig. 3.7). Extrapolated to $39 \text{ }^\circ\text{C}$, the opening rate decreases to 0.06 min^{-1} , which is more than threefold lower.

What was considered as open state in above analysis does not necessarily correspond to zero unpaired bases. Melting curves were measured using the intercalating dye SYBR Green I, which binds only to double-stranded DNA regions with a minimal number of paired bases. However, the process of backbone duplex formation does not need the hairpins to be completely open. Instead, partially open hairpins are sufficient. Due to the bulged nucleotides in the hairpins' stems, simulation data suggest a two-step melting, where the base pairs next to the codon domain melt first. Opening of the remaining base pairs then proceeds via strand displacement. Therefore, the determined values reflect the relevant kinetics relatively well.

The spontaneous formation rate is caused by strands which are not bound to a template, but bind to each other in free solution. For this to happen, hairpins must be open with a sufficiently high probability. This corresponds to temperatures not much lower the hairpin melting temperature, which was measured as $51 \text{ }^\circ\text{C}$ for strand 0_D .

For both aspects of the temperature dependent templating kinetics, the measured fluctuation rates agree with the observed template formation rates (Fig. 3.6c).

3.2.5. Thermally driven amplification

Preliminary to actual replication, amplification reactions containing only strands encoding for 0, i.e. 0_A-0_D , $\bar{0}_A-\bar{0}_D$, were studied. Strands were subjected to thermal oscillations between $45 \text{ }^\circ\text{C}$ and upper temperatures between $67 \text{ }^\circ\text{C}$ and $78 \text{ }^\circ\text{C}$. The lower temperature

3. Transfer RNA as a codon replicator

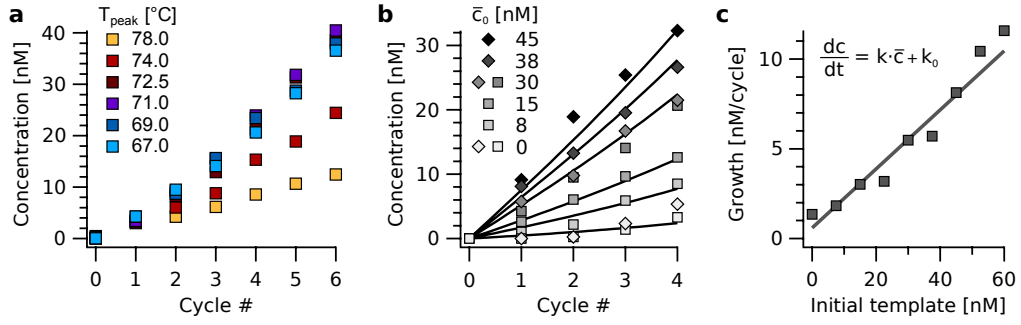


Figure 3.8. Exponential amplification of a restricted sequence subset. **a.** Amplification time traces for sequence subset 0000. Oscillation peak temperatures ranged from 67 °C to 78 °C, the base temperature was 45 °C. Reactions initially contained 30 nM of complex $\overline{0000}$ as template. Strands 0_A-0_D , $\overline{0}_A-\overline{0}_D$ were at 200 nM concentration each. Data points show concentrations of complexes 4:4. **b.** Reaction kinetics during the first four cycles ($T_{\text{peak}} = 67$ °C) for template concentrations from 0 to 45 nM. Data were fitted using the cross-catalytic model from equation (3.4). **c.** Initial reaction velocity as a function of initial template concentration \overline{c}_0 . Data points show good agreement with the line calculated from the fits in panel b. Fit coefficients were $k = 0.16 \text{ cycle}^{-1}$ and $k_0 = 0.4 \text{ nM cycle}^{-1}$.

was held for 20 minutes, the upper for about one second. Transients added up to 20 ± 1 seconds. This asymmetric shape accords with differences in kinetics of the elongation step and the melting of codon bonds, and is typical for temperature trajectories in thermal convection settings [11].

Amplification proceeded robustly for peak temperatures below 74 °C (Fig. 3.8a), which is about the melting temperature of the weakest backbone duplex (Fig. 3.3b). At higher temperatures, a large fraction of backbone duplexes breaks down in each high-temperature spike, resulting in a diminished yield.

Experiments with different initial concentrations of the template $\overline{0000}$ revealed an almost linear dependence of the initial reaction velocity on the initial amount of template (Fig. 3.8b). Data for initial concentrations up to 45 nM were fit to the simplistic cross-catalytic model described in section 3.2.1. Fitting revealed kinetic parameters of $k = 0.16 \text{ cycle}^{-1}$ and $k_0 = 0.4 \text{ nM cycle}^{-1}$. The relation between the initial growth rate $\Gamma_{t_2 \leftarrow t_1}$ and the initial template concentration \overline{c}_0 is a line given by equation (3.5)

$$\Gamma_{t_2 \leftarrow t_1} = \frac{c(t_2) - c(t_1)}{t_2 - t_1} = \Gamma_1 \overline{c}_0 + \Gamma_0, \quad (3.6)$$

with

$$\Gamma_1 = \frac{\sinh(k t_2) - \sinh(k t_1)}{t_2 - t_1} \quad \text{and} \quad \Gamma_0 = \frac{k_0 (\exp(k t_2) - \exp(k t_1))}{k (t_2 - t_1)}. \quad (3.7)$$

It shows good agreement with measurement data (Fig. 3.8c), implying the cross-catalytic nature of the amplification reaction.

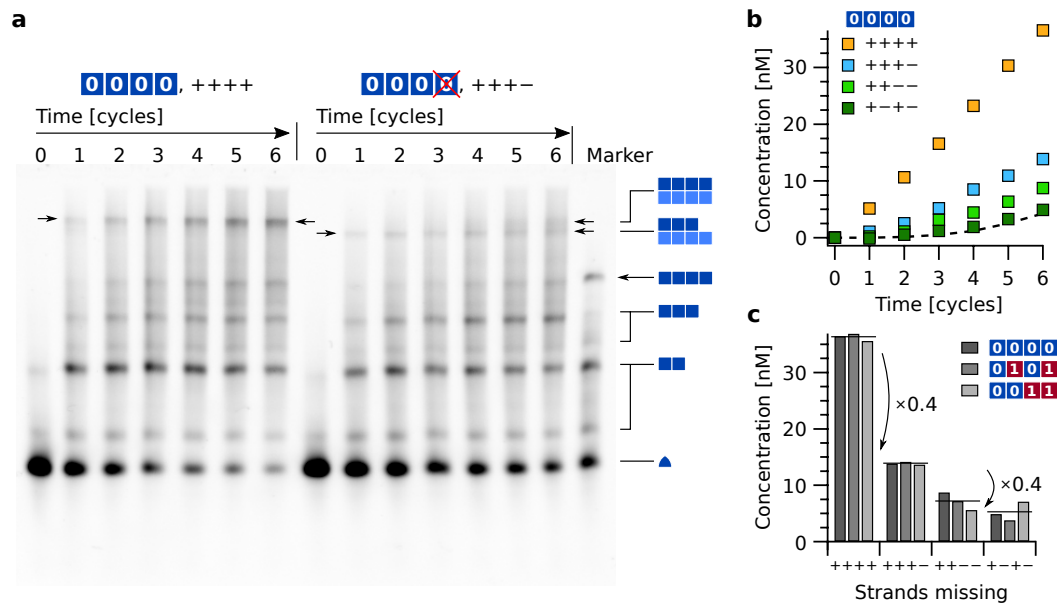


Figure 3.9. Sequence replication. **a.** Replication of sequence 0000. PAGE results comparing the reaction of all 16 strands (“++++”) with the reaction lacking strand 0_D (“+++−”). Reactions were started with 15 nM initial template $\overline{0000}$. Strands were at 100 nM each. The defective set “+++−” mostly produces complexes 3:4 instead of 4:4 (black arrows). The overall yield of tetramer-containing complexes is greatly reduced. The marker lane contained complexes 0000, 000, 00, and monomers as a size reference. **b.** Reaction time traces of the whole sequence pool (yellow) and three defective sets. Data were integrated from electrophoresis gels, and included all complexes containing tetramers. Reactions lacked strands 0_D (“+++−”), $0_C/0_D$ (“++−−”), and $0_B/0_D$ (“+−−−”). Reactions were initiated with 15 nM of $\overline{0000}$. The dashed line shows data from reaction “++++” without template. **c.** End point comparison of reactions with templates $\overline{0000}$, $\overline{0101}$, and $\overline{0011}$ after 6 cycles. Horizontal lines indicate averages of the three sequences. A single missing strand reduced product yield by 60 %, two missing strands by 80–85 %.

3.2.6. Sequence replication

To study actual replication of information, reactions containing all 16 strands (i.e. 0_A-0_D , 1_A-1_D , $\overline{0}_A-\overline{0}_D$, $\overline{1}_A-\overline{1}_D$), and one out of three different templates $\overline{0000}$, $\overline{0101}$, and $\overline{0011}$, were performed. The amount of product during six thermal oscillations in the full reaction was compared to the output of three defective reactions, i.e. such lacking one or more of the monomers required to replicate the template.

For example, a one-strand defect in the reaction with template $\overline{0000}$ was achieved by leaving out strand 0_D (labelled “+++−”). PAGE analysis showed that the formed product was mostly complex $0_A0_B0_C:\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ (3:4), which was expected given the lack of strand 0_D . If the complex contained a significant fraction of strands 1 or $\overline{1}$, the difference between reactions “++++” and “+++−” would not be as big. Moreover, the full reaction produced almost no complexes 3:4 or 4:3. This means that complexes $0001:\overline{000}$ formed in the

3. Transfer RNA as a codon replicator

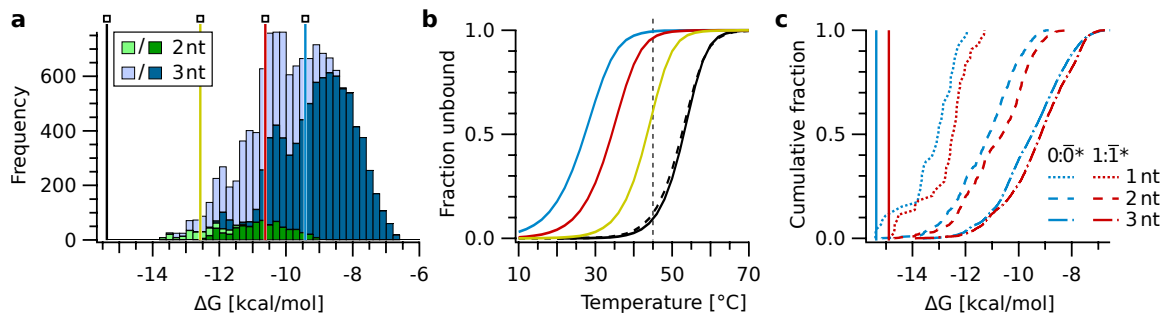


Figure 3.10. Effects of anticodon mutations. **a.** Simulated free energies of the codon domain duplex $0:\bar{0}$ (solid black line) and duplexes $0:\bar{0}^*$. Codons $\bar{0}^*$ can be derived from $\bar{0}$ by two (green) or three (blue) point mutations. Data were split into sequences with only internal mutations (dark colours) and others (light colours). 99 % of the duplexes $0:\bar{0}^*$ with any three mutations have free energies of $\Delta G \geq -12.5$ kcal/mol (yellow line). This also holds for 99.9 % of duplexes with 2–3 internal mutations only. **b.** Melting curves of codon duplexes $0:\bar{0}$ (solid black), $1:\bar{1}$ (dashed black), and three duplexes $0:\bar{0}^*$ with three mutations. Colour coding indicates free energies from panel a. Even the $0:\bar{0}^*$ duplexes at the low end of the ΔG distribution (yellow) have melting temperatures of about 10 °C below that of $0:\bar{0}$. **c.** Cumulative free energy distributions of all codon dimers with up to 3 point mutations. Contributions at the low end result from mutations at terminal bases. For single mutations, this corresponds to the little affected fraction of $2/15 \approx 0.13$.

defective reaction would quickly gain a strand $\bar{1}$ and end up as complex $0001:\bar{0001}$ on the gel. However, the production of sequence 0001 as well as complex $0001:\bar{0001}$ was strongly inhibited.

Removal of a further strand either directly next to the first one (“+ + – –”, missing strands $0_C/0_D$) or at a different location (“+ – + –”, $0_B/0_D$ missing) reduced the yield of tetramers even further. Reaction time traces were extracted by integrating the intensities of all gel bands including tetramers (Fig. 3.9b). The traces quantify the reduction in yield of a single missing strand to about 60 %. Reaction curves of the other two sequences 0101 and 0011 were very similar to the data for 0000, all showed almost equal yields. End points after 6 cycles are given in Fig. 3.9c for each of the three sequences as well as a sequence average (black lines). With a single defect, yield of tetramer complexes dropped by about 60 %. In reactions with two strands missing, production was reduced by 80–85 % compared to the full reaction.

3.2.7. Replication fidelity

The observed rate of erroneous product formation can be attributed to the spontaneous background rate (Fig. 3.9b, dashed line). Reaction “+ – + –” proceeded the same as the untemplated control, as it did not contain any strands which could bind next to each other to the template and form a backbone duplex. For reactions “+ + + –” and “+ + – –”, templating worked for partial sequences, producing yields between the two reactions.

The fact that the reaction with a single defect (i.e. missing strand) had 40 % of the yield of the full reaction (and ca. 16 % for two defects) translates into a per-codon replication fidelity of $1/1.4 = 71$ %. To derive a per-base fidelity, the properties of codon duplex $0:\bar{0}$ were compared to duplexes $0:\bar{0}^*$, where $\bar{0}^*$ differs from $\bar{0}$ by a fixed number of point mutations. 99 % of all duplexes with $\bar{0}^*$ containing three point mutations have a $\Delta G \geq -12.5$ kcal/mol compared to $\Delta G(0:\bar{0}) = -15.4$ kcal/mol at 45 °C. In terms of melting temperatures, this translates into values at least 10 °C lower than the original codon duplex $0:\bar{0}$ (Fig. 3.10).

Assuming that the replication does not differentiate between codon $\bar{0}$ and any codon $\bar{0}^*$ with up to K point mutations, the per-codon fidelity $q_K(N)$ is given by a cumulative binomial distribution

$$q_K(N) = \sum_{k=0}^K \binom{N}{k} q^{N-k} (1-q)^k. \quad (3.8)$$

Here, N is the codon length, and q the per-base fidelity. Assuming $K = 2$ and using the measured value of $q_2(15) = 0.71$, one finds a per-base fidelity of $q = 87.5$ %. This is when one only considers the 15 bases of the codon. Including the whole length of the proto-tRNAs of about 83 bases, the per-base fidelity would read 97.8 %.

In fact, above assumption that the reaction cannot distinguish between codon $\bar{0}$ and $\bar{0}^*$ with up to two mutations (i.e. $K = 2$) is essentially owing to mutations at the terminal bases. Codons $\bar{0}^*$ with mutations at two internal bases all show similar properties as codons with a total of three mutations, and 99 % have melting temperatures more than 10 °C lower than dimer $0:\bar{0}$ (Fig. 3.10). Including this refinement, the per-base fidelity reads 92 %.

3.3. Discussion

I present a replication mechanism that is capable of cross-catalytically replicating a succession of short nucleic acid stretches without the need for any ligation chemistry. Instead, nucleic acids are connected via hybridization of complementary domains. Replication is driven by thermal oscillations, does not require other fuel, and does not generate waste products which could interfere with the reaction later on. The reaction is relatively fast, and proceeds within a few thermal oscillations of 20 minutes each. This is comparable to other replicators [51], cross-ligating ribozymes [91], or autocatalytic DNA networks [123].

Codon sequences are replicated with a per-codon fidelity of about 70 %. Replication on a codon basis effectively constitutes a proofreading mechanism for a putative upstream polymerization process [17, 36, 43, 60] that would generate the proto-tRNAs. It rejects sequence snippets above a certain error ratio and thereby increases the effective fidelity of that replication process. The per-codon fidelity can be translated to a per-nucleotide basis, which is estimated to 88 to 92 %. Therefore, the underlying polymerization process could

3. *Transfer RNA as a codon replicator*

feature a relatively low fidelity¹, as that would only affect the concentration of “correct” molecules, and thus the velocity of replication, but not its fidelity.

Overall replication fidelity is limited by a spontaneous formation rate, which originates from the interaction of strands not bound to a template but in free solution. At lower concentrations, as one would imagine in an prebiotic setting, this rate would decrease at the expense of an overall slower reaction. To some degree, such a background rate is inherent to hairpin-fuelled DNA or RNA reactions [39, 123].

A similar selection mechanism for nucleic acids is constituted by the highly sequence-specific gelation of DNA [74]. Here, DNA at very high local concentrations forms hydrogels of up to 100 μm in size. The required concentrations are generated by hyperexponential accumulation of the DNA in thermo-gravitational pores [64]. The selectivity of the process is caused by the structure of the hydrogel, which consists of a network of short oligomers, connected by base pairing of complementary domains. Already the mutation of single nucleotides is sufficient to prevent gelation. In fact, such a phenomenon could serve as a pre-selection mechanism for hairpin-driven replication mechanisms, as it promotes self-complementarity in nucleotide sequences and thereby selects for hairpins. A process with similar selection properties is the biased hydrolysis of nucleotide backbones [77], where double-stranded regions less likely to be cleaved than single-stranded domains.

Thermal oscillations like those discussed here are typical for laminar convection in thermal gradients [11]. Depending on the envisioned environment, the mechanism could also be driven by thermochemical oscillations [6] or oscillations in pH. In the latter case, denaturation of codon duplexes would be due to alkaline pH instead of high temperature.

The presented nucleic acids may appear to be rather large, as only 18 % of the nucleotides actually encode for information. However, modern ribosomes can make up more than 15 % of the mass of a cell [34], such that the total required mass need not be a major concern. Moreover, the reversibility of the bonds between the proto-tRNAs makes the strands reusable, which further reduces the cost entailed by the non-coding parts.

Nevertheless, the replication mechanism would also work with shorter strands. For this study, the length of the strands was inspired by the size of modern tRNAs. Individual hairpins as well as codon and backbone duplexes had relatively high melting temperatures and slow kinetics, which eased experimental handling. Smaller strands would be equally feasible, as long as the order of the melting temperatures of codon and backbone duplexes is preserved. For smaller strands, the requirements on the underlying polymerization process are also somewhat lower, as sequence space shrinks exponentially with decreasing molecule size. Moreover, binding of shorter codon duplexes would discriminate even single mismatches, resulting in an increased selectivity of the proofreading mechanism. For strands

¹The polymerization process could not be completely random, as the size of sequence space of all 83 nt strands is 10^{50} .

of length 30, sampling the whole sequence space requires two micromoles of molecules.

The constraints regarding the stability of the backbone duplexes would be lifted by the combination with a proposed non-enzymatic ligation at short overhangs of RNA duplexes [107]. Such overhangs at each strand were present in the sequences used in this study and did not interfere with recognition or binding at the codon domains. In the bound configuration, such a ligation would proceed at the backbone duplexes and join successive strands. Another compatible mechanism would consist of a cleavage reaction at the codon domains [19], which would cut out the backbone duplexes and be followed by a ligation of the codon domains.

Considering the origins of translation, the double-hairpin configuration of the strands could suggest a link towards a simple translation system. Codon domains (containing what is the anticodon in modern tRNAs) are close to the 3' termini of unbound strands. The formation of short peptide-RNA hybrids [44], combined with specific interactions between amino acids and the codons, could have given rise to a primitive genetic code. The spatial arrangement of strands that is replicated by the presented mechanism would then translate into a spatial arrangement of the amino acid or short peptide tails attached to the strands. The next stage would then be the detachment and linking of the tails to form longer peptides and eventually proteins.

Outside the context of translation, mechanisms similar to the one described here could also be relevant as a mutable assembly strategy for larger functional RNA molecules. Hairpin loops are a ubiquitous secondary structure motif, commonly separated by stretches of unpaired bases. Unlike RNA systems with actual ribozyme activity [76, 115], the presented system is more symmetric. However, as catalytic functionality in RNA can emerge from something as simple as a one-nucleotide bulge in a short duplex [118], the structural regularity is no major roadblock.

On the question about the nature of the first functional RNA, the replication mechanism is compatible with the idea that their functionality need not be related to replication itself [111]. By relying on non-enzymatic polymerization of the RNA sequences instead of replicase activity, the assembled RNA complexes could rather have served structural or metabolic purposes. This also holds for potential chemical ligation activity, which is not required for replication, but could prove beneficial and evolve at a later stage.

3.4. Materials and methods

3.4.1. Strand design

DNA double-hairpin sequences were designed using the NUPACK software package [124]. The algorithm calculates free energies, melting temperatures and probabilities of different secondary structure configurations using the nearest neighbour model [97]. The model

3. Transfer RNA as a codon replicator

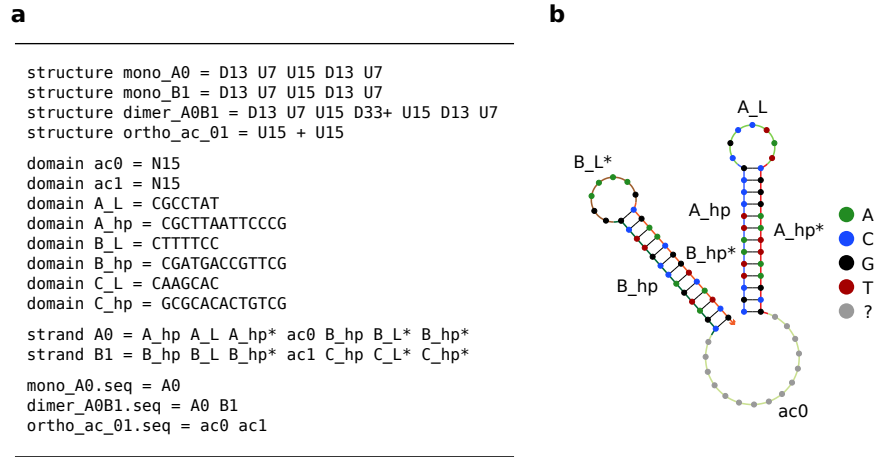


Figure 3.11. Example NUPACK input. **a.** Excerpt from the NUPACK input used to generate the sequences of the replicating set of DNA strands, listed in table 3.1. The example defines three “real” secondary structures (monomers 0_A , 0_B and dimer 0_A0_B), eight domains, and two strands. The structure *ortho_ac_01* is defined to ensure the orthogonality of the coding sequences. **b.** Visualization of secondary structure *mono_A0*, the 0_A double hairpin molecule. Labels of the seven domains are indicated next to the domains.

accounts for Watson-Crick and wobble base pairing energies as well as stacking interactions between neighbouring base pairs. It contains contributions from mismatched bases, dangling ends, internal and external loops, corrections for salt concentrations (NaCl and MgCl₂) and temperature, and other subtleties.

For calculations, a RNA or DNA sequence is represented as a *polymer graph* [23], where the strand is laid out in a circle and base pairs are represented by lines connecting nodes (bases) on that circle. When several strands are included in the calculation, they are concatenated in each permutation and analysed separately. For most cases, base pairs must be strictly nested, i.e. the structure must be free of pseudoknots². The free energy ΔG_s of a secondary structure is given by the sum of the free energies of its constituents:

$$\Delta G_s = \sum_k \Delta G_k. \quad (3.9)$$

The probability of finding a particular secondary structure is given by

$$p_s = \frac{1}{Z} e^{-\Delta G_s/k_B T}, \quad Z = \sum_{s \in \Omega} e^{-\Delta G_s/k_B T}. \quad (3.10)$$

Z is the partition function, Ω the state space.

The NUPACK design algorithm is controlled by an input script containing parameters

²A limited set of pseudoknots is allowed when analysing single RNA strands.

such as temperature and salt conditions as well as strand and target secondary structure definitions. The latter are given using the so-called DU+ notation, specifying stretches of paired (D*n*) and unpaired (U*n*) bases and backbone nicks (+). The substructure of a duplex immediately follows its definition. Figure 3.11a shows an excerpt of the input used to generate initial sets of sequences.

Target secondary structures were all individual double hairpin monomers and all pairs of consecutive dimer complexes (e.g. $0_A0_B, \bar{0}_C\bar{0}_B$). The sequence of each strand was decomposed into seven domains, as illustrated in Fig. 3.11b. In doing so, the complementarity of the 3' hairpin of one strand (e.g. 0_A) to the 5' hairpin of the next (i.e. 0_B) was easily specified. In addition, the orthogonality of the two anticodon sequences was specified. The strand length of 82–84 nt was chosen to lie in the range usual tRNA strand lengths [105].

From ten generated candidate sequence sets, the most suitable was chosen with regard to optimal homogeneity in the binding energies and the ordering of melting temperatures. Minor outliers in the binding energies were eliminated by manual mutation of some of the bases. More importantly, the predicted hairpin melting temperatures of the initial sequences were too high. While this would not fundamentally interfere with the replication mechanism, it would unnecessarily slow down templating kinetics. To facilitate sufficient degree of fluctuations in the hairpins at temperatures below the melting temperature of coding bonds, mismatches were introduced into the stem sequences. Finally, I added short 5' overhangs of four nucleotides in length to each strand. These overhangs can be used to covalently ligate short adapter hairpins to the backbone duplexes. Once ligated, the combined strands could be analysed using DNA sequencing or subjected to other denaturing downstream processing.

3.4.2. Thermal cycling assays

All reactions were performed in buffer containing 20 mM Tris, 150 mM NaCl, and 20 mM MgCl₂. DNA oligonucleotides (Biomers, Germany) were used at 200 nM concentration per strand in reactions containing a fixed-sequence subset of eight strands (e.g. $0, \bar{0}$ only) and 100 nM per strand in reactions containing all 16 different strands.

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

Template sequences were prepared using a two-step protocol. An annealing step from 95 °C to 70 °C within 1 hour was followed by incubation at 70 °C for 30 minutes. Afterwards, samples were cooled to 2 °C and stored on ice. When assembling complexes containing paired codon domains (e.g. $0_A0_B:\bar{0}_A\bar{0}_B$, Fig. 3.4), samples were slowly cooled down from 70 °C to 25 °C within 90 minutes before being transferred onto ice. DNA double hairpins

3. Transfer RNA as a codon replicator

were quenched into monomolecular state by heating to 95 °C and subsequent fast transfer into ice water.

3.4.3. Product analysis

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

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

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

For the determination of reaction yields, the intensities of all gel bands containing strands of the sequence length of interest were added up. For successions of four strands, these were the single tetramer as well as its complex with di- and tri- and tetramers. Single strands separate from their complements during electrophoresis (Fig. 3.4).

3.4.4. Thermal melting curves

Thermal melting curves (Fig. 3.3) were measured using either UV absorbance at 260 nm wavelength in a UV/Vis spectrometer (JASCO V-650, 1 cm optical path length), via fluorescence quenching of the Cy5 label at the 5'-end of strand 0_A (excitation: 620–650 nm, detection: 675–690 nm; only for $0_A:\bar{0}_A$), or using fluorescence of the intercalating dye SYBR Green I (excitation: 450–490 nm, detection: 510–530 nm). Fluorescence measurements were performed in a PCR cycler (Bio-Rad C1000). Samples measured via fluorescence were at 200 nM of each strand, those measured via UV absorption contained 1 μ M total DNA concentration to improve the signal-to-noise ratio. All data were corrected for baseline signals from reference samples containing buffer (and intercalating dye, if applicable) and analysed as described in [69].

Hairpin fluctuation kinetics (Fig. 3.7) were determined from hysteresis in thermal melting curves [69]. Melting curves were measured in a PCR cycler from 70 °C to 25 °C and reverse. Heating/cooling rates were 3 K/min.

3.4.5. Sequences

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

Table 3.1. Sequences of all DNA strands used in chapter 3. Strand 0_A is 5'-labelled with Cy5, all other strands have a 5'-terminal phosphate. Single underlines highlight hairpin loops, codon domains are indicated by wavy underlines.

4. Conclusions

In this thesis, two defining features of living or life-like systems were studied: replication and selection. First, it was shown that the selection pressure that is required to prevent the kinetically favoured loss of genetic material can be generated by a simple physical process: heat dissipation across submerged porous rock. Comprehensive numeric modelling of accumulation and selection inside such pores narrowed down parameter space and identified regions for experimental exploration. Moreover, an analytic description of the interplay of convection, external flux, diffusion, and thermophoresis provided an understanding of the underlying physics. On the other hand, thermal cycling of the replicating molecules in the narrow geometry of the pore is mainly determined by diffusion, and required a stochastic treatment. Simulated single-particle trajectories enabled a quantitative analysis of the cycling behaviour, which provided details about the kinetics of replication and dilution. Using these rates, the selection process was modelled as the competition of exponential replication and dilution.

The choice of DNA replication via PCR as a well-characterized model system allowed to study the properties of thermo-gravitational pores as a selective environment, while reducing the experimental challenges originating from handling the replicator itself. Constituting a proof of principle, the findings open the door to further-reaching experiments comprising less restricted replicators, ultimately allowing for the evolution of populations beyond the selection of the simple binary population studied here. Apart from its relevance for understanding early evolution, such experiments would also represent a continuous variant of conventional directed evolution methods [30, 89, 108], which however does not require sequential rounds of selection.

Considering simple replicating systems, the question about the nature of a simple yet generally functional replicase remains unanswered [111]. Following the idea that the first functional molecules may have possessed functions different from replication itself, a replication scheme was studied that operates on short nucleotide segments instead of single nucleotides. The mechanism then serves as a means of proofreading for a potentially non-enzymatic upstream polymerization process. As a novel approach to biological information encoding, it only relies on the formation of double-stranded duplexes, stabilized by base pairing and stacking. Despite being realized in DNA, this makes it compatible with DNA, RNA, and a variety of nucleic acids analogues discussed as putative proto-RNAs.

Replication proceeds by means of thermal oscillations, the energy to drive the replication

4. Conclusions

is stored in hairpin loops. Importantly, chemical activation is not required, which allows for the reactivation of the individual strands of the replicating system by a simple high-temperature spike. Due to the structural similarity of the individual strands to tRNA, such a codon-based replication scheme could even offer hints towards the origin of primitive but autonomous translation systems.

A. References

- [1] J. Attwater, A. Wochner, and P. Holliger, *In-ice evolution of RNA polymerase ribozyme activity*, *Nature Chemistry* **5** (2013), 1011–1018.
- [2] P. Baaske, F. M. Weinert, S. Duhr, K. H. Lemke, M. J. Russell, and D. Braun, *Extreme accumulation of nucleotides in simulated hydrothermal pore systems*, *Proceedings of the National Academy of Sciences of the United States of America* **104** (2007), 9346–9351.
- [3] P. Baaske, C. J. Wienken, P. Reineck, S. Duhr, and D. Braun, *Optical thermophoresis for quantifying the buffer dependence of aptamer binding*, *Angewandte Chemie International Edition* **49** (2010), 2238–2241.
- [4] J. L. Bada, *How life began on Earth: a status report*, *Earth and Planetary Science Letters* **226** (2004), 1–15.
- [5] M. I. Bahl, S. J. Sørensen, and L. H. Hansen, *Quantification of plasmid loss in Escherichia coli cells by use of flow cytometry*, *FEMS Microbiology Letters* **232** (2004), 45–49.
- [6] R. Ball and J. Brindley, *Hydrogen peroxide thermochemical oscillator as driver for primordial RNA replication*, *Journal of The Royal Society Interface* **11** (2014), 20131052–20131052.
- [7] J. A. Baross and S. E. Hoffman, *Submarine hydrothermal vents and associated gradient environments as sites for the origin and evolution of life*, *Origins of Life and Evolution of the Biosphere* **15** (1985), 327–345.
- [8] S. A. Benner, *Defining life*, *Astrobiology* **10** (2010), 1021–1030.
- [9] H. S. Bernhardt, *The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)*, *Biology Direct* **7** (2012), 23.
- [10] D. Braun, *PCR by thermal convection*, *Modern Physics Letters B* **18** (2004), 775–784.
- [11] D. Braun, N. L. Goddard, and A. Libchaber, *Exponential DNA replication by laminar convection*, *Physical Review Letters* **91** (2003), 158103.
- [12] I. Budin, R. J. Bruckner, and J. W. Szostak, *Formation of protocell-like vesicles in a thermal diffusion column*, *Journal of the American Chemical Society* **131** (2009), 9628–9629.

A. References

- [13] A. S. Burton, J. C. Stern, J. E. Elsila, D. P. Glavin, and J. P. Dworkin, *Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites*, *Chemical Society Reviews* **41** (2012), 5459.
- [14] A. Chien, D. B. Edgar, and J. M. Trela, *Deoxyribonucleic acid and polymerase from the extreme and thermophile thermus and aquaticus*, *Journal of Bacteriology* **127** (1976), 1550–1557.
- [15] F. J. Ciesla and S. A. Sandford, *Organic synthesis via irradiation and warming of ice grains in the solar nebula*, *Science* **336** (2012), 452–454.
- [16] I. Cnossen, J. Sanz-Forcada, F. Favata, O. Witasse, T. Zegers, and N. F. Arnold, *Habitat of early life: Solar x-ray and UV radiation at earth's surface 4–3.5 billion years ago*, *Journal of Geophysical Research* **112** (2007), E02008.
- [17] G. Costanzo, R. Saladino, G. Botta, A. Giorgi, A. Scipioni, S. Pino, and E. Di Mauro, *Generation of RNA Molecules by a Base-Catalysed Click-Like Reaction*, *ChemBioChem* **13** (2012), 999–1008.
- [18] B. Damer and D. Deamer, *Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life*, *Life* **5** (2015), 872–887.
- [19] V. Dange, R. B. Van Atta, and S. M. Hecht, *A Mn²⁺-dependent ribozyme*, *Science* **248** (1990), 585–588.
- [20] P. Debye, *Zur Theorie des Clusiussschen Trennungsverfahrens*, *Annalen der Physik* **428** (1939), 284–294.
- [21] V. DeGuzman, W. Vercoutere, H. Shenasa, and D. Deamer, *Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization*, *Journal of Molecular Evolution* **78** (2014), 251–262.
- [22] M. Di Giulio, *The origin of the tRNA molecule: implications for the origin of protein synthesis*, *Journal of Theoretical Biology* **226** (2004), 89–93.
- [23] R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. A. Pierce, *Thermodynamic analysis of interacting nucleic acid strands*, *SIAM Review* **49** (2007), 65–88.
- [24] N. Dolinnaya, N. Sokolova, O. Gryaznova, and Z. Shabarova, *Site-directed modification of DNA duplexes by chemical ligation*, *Nucl Acids Res* **16** (1988), 3721–3738.
- [25] S. Duhr and D. Braun, *Thermophoretic depletion follows Boltzmann distribution*, *Physical Review Letters* **96** (2006), 168301.
- [26] ———, *Why molecules move along a temperature gradient*, *Proceedings of the National Academy of Sciences of the United States of America* **103** (2006), 19678–19682.

- [27] D. Dykhuizen, *Selection for tryptophan auxotrophs of escherichia coli in glucose-limited chemostats as a test of the energy conservation hypothesis of evolution*, *Evolution* **32** (1978), 125.
- [28] M. Eigen, *Selforganization of matter and the evolution of biological macromolecules*, *Naturwissenschaften* **58** (1971), 465–523.
- [29] M. Eigen, J. McCaskill, and P. Schuster, *Molecular quasi-species*, *The Journal of Physical Chemistry* **92** (1988), 6881–6891.
- [30] A. D. Ellington and J. W. Szostak, *In vitro selection of RNA molecules that bind specific ligands*, *Nature* **346** (1990), 818–822.
- [31] A. E. Engelhart, B. J. Cafferty, C. D. Okafor, M. C. Chen, L. D. Williams, D. G. Lynn, and N. V. Hud, *Nonenzymatic ligation of DNA with a reversible step and a final linkage that can be used in PCR*, *ChemBioChem* **13** (2012), 1121–1124.
- [32] J. L. England, *Dissipative adaptation in driven self-assembly*, *Nature Nanotech* **10** (2015), 919–923.
- [33] A. Eschenmoser, *Chemical etiology of nucleic acid structure*, *Science* **284** (1999), 2118–2124.
- [34] F. Fegatella, F. Lim, S. Kjelleberg, and R. Cavicchioli, *Implications of rna operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium sphingomonas sp. strain rb2256.*, *Applied and Environmental Microbiology* **64** (1998), 4433–4438.
- [35] C. Fernando, G. von Kiedrowski, and E. Szathmary, *A stochastic model of nonenzymatic nucleic acid replication: “elongators” sequester replicators*, *Journal of Molecular Evolution* **64** (2007), 572–585.
- [36] J. P. Ferris, *Montmorillonite catalysis of 30–50 mer oligonucleotides: Laboratory demonstration of potential steps in the origin of the rna world*, *Origins of Life and Evolution of the Biosphere* **32** (2002), 311–332.
- [37] J. P. Ferris, A. R. Hill, R. Liu, and L. E. Orgel, *Synthesis of long prebiotic oligomers on mineral surfaces*, *Nature* **381** (1996), 59–61.
- [38] W. Gilbert, *Origin of life: The RNA world*, *Nature* **319** (1986), 618–618.
- [39] S. J. Green, D. Lubrich, and A. J. Turberfield, *DNA hairpins: Fuel for autonomous DNA devices*, *Biophysical Journal* **91** (2006), 2966–75.
- [40] I. Haruna and S. Spiegelman, *Specific template requirements of RNA replicases*, *Proceedings of the National Academy of Sciences of the United States of America* **54** (1965), 579–587.
- [41] T. Henning and D. Semenov, *Chemistry in protoplanetary disks*, *Chemical Reviews* **113** (2013), 9016–9042.

A. References

- [42] D. P. Horning and G. F. Joyce, *Amplification of RNA by an RNA polymerase ribozyme*, Proceedings of the National Academy of Sciences of the United States of America (2016), 201610103.
- [43] E. C. Izgu, S. S. Oh, and J. W. Szostak, *Synthesis of activated 3'-amino-3'-deoxy-2-thiothymidine, a superior substrate for the nonenzymatic copying of nucleic acid templates*, Chemical Communications **52** (2016), 3684–3686.
- [44] M. Jauker, H. Griesser, and C. Richert, *Spontaneous formation of RNA strands, peptidyl RNA, and cofactors*, Angewandte Chemie International Edition **54** (2015), 14564–14569.
- [45] R. C. Jones and W. H. Furry, *The separation of isotopes by thermal diffusion*, Reviews of Modern Physics **18** (1946), 151–224.
- [46] L. Keil, M. Hartmann, S. Lanzmich, and D. Braun, *Probing of molecular replication and accumulation in shallow heat gradients through numerical simulations*, Physical Chemistry Chemical Physics **18** (2016), 20153–20159.
- [47] D. S. Kelley, J. A. Baross, and J. R. Delaney, *Volcanoes, fluids, and life at mid-ocean ridge spreading centers.*, Annual Review of Earth and Planetary Sciences **30** (2002), 385–491.
- [48] D. S. Kelley, J. A. Karson, D. K. Blackman, G. L. Früh-Green, D. A. Butterfield, M. D. Lilley, E. J. Olson, M. O. Schrenk, K. K. Roe, G. T. Lebon, P. Rivizzigno, and the AT3-60 Shipboard Party, *An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30 degrees N*, Nature **412** (2001), 145–149.
- [49] 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, *A serpentinite-hosted ecosystem: The lost city hydrothermal field*, Science **307** (2005), 1428–1434.
- [50] J. Kestin, M. Sokolov, and W. A. Wakeham, *Viscosity of liquid water in the range -8°C to 150°C*, Journal of Physical and Chemical Reference Data **7** (1978), 941.
- [51] M. Kindermann, I. Stahl, M. Reimold, W. M. Pankau, and G. von Kiedrowski, *Systems chemistry: Kinetic and computational analysis of a nearly exponential organic replicator*, Angewandte Chemie International Edition **44** (2005), 6750–6755.
- [52] S. Koskiniemi, S. Sun, O. G. Berg, and D. I. Andersson, *Selection-driven gene loss in bacteria*, PLoS Genetics **8** (2012), e1002787.
- [53] H. Krammer, F. M. Möller, and D. Braun, *Thermal, Autonomous Replicator Made from Transfer RNA*, Physical Review Letters **108** (2012), 238104.

- [54] M. Kreysing, L. M. R. Keil, S. A. Lanzmich, and D. Braun, *Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length*, *Nature Chemistry* **7** (2015), 203–208.
- [55] M. Krishnan, V. M. Ugaz, and M. A. Burns, *PCR in a Rayleigh-Bénard Convection Cell*, *Science* **298** (2002), 793.
- [56] N. Lane, J. F. Allen, and W. Martin, *How did LUCA make a living? Chemiosmosis in the origin of life*, *BioEssays* **32** (2010), 271–280.
- [57] S. A. Lanzmich, *Accumulation and replication in prebiotic environments*, Master’s thesis, 2012.
- [58] A. C. Lasaga, H. D. Holland, and M. J. Dwyer, *Primordial oil slick*, *Science* **174** (1971), 53–55.
- [59] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, and F. E. Romesberg, *Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet*, *Journal of the American Chemical Society* **130** (2008), 2336–2343.
- [60] K. Leu, B. Obermayer, S. Rajamani, U. Gerland, and I. A. Chen, *The prebiotic evolutionary advantage of transferring genetic information from RNA to DNA*, *Nucleic Acids Research* **39** (2011), 8135–8147.
- [61] T. Maniatis, A. Jeffrey, and H. Van deSande, *Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis*, *Biochemistry* **14** (1975), 3787–3794.
- [62] W. F. Martin, F. L. Sousa, and N. Lane, *Energy at life’s origin*, *Science* **344** (2014), 1092–1093.
- [63] C. B. Mast and D. Braun, *Thermal Trap for DNA Replication*, *Physical Review Letters* **104** (2010), 188102.
- [64] C. B. Mast, S. Schink, U. Gerland, and D. Braun, *Escalation of polymerization in a thermal gradient*, *Proceedings of the National Academy of Sciences of the United States of America* **110** (2013), 8030–8035.
- [65] A. T. Maurelli, R. E. Fernández, C. A. Bloch, C. K. Rode, and A. Fasano, *“black holes” and bacterial pathogenicity: A large genomic deletion that enhances the virulence of shigella spp. and enteroinvasive escherichia coli*, *Proceedings of the National Academy of Sciences of the United States of America* **95** (1998), 3943–3948.
- [66] J. S. McCaskill, *A stochastic theory of macromolecular evolution*, *Biological Cybernetics* **50** (1984), 63–73.
- [67] S. D. McCulloch and T. A. Kunkel, *The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases*, *Cell Research* **18** (2008), 148–161.

A. References

- [68] C. Meinert, I. Myrgorodska, P. de Marcellus, T. Buhse, L. Nahon, S. V. Hoffmann, L. Le Sergeant d'Hendecourt, and U. J. Meierhenrich, *Ribose and related sugars from ultraviolet irradiation of interstellar ice analogs*, *Science* **352** (2016), 208–212.
- [69] J.-L. Mergny and L. Lacroix, *Analysis of Thermal Melting Curves*, *Oligonucleotides* **13** (2003), 515–537.
- [70] S. L. Miller, *A Production of Amino Acids Under Possible Primitive Earth Conditions*, *Science* **117** (1953), 528–529.
- [71] D. R. Mills, R. L. Peterson, and S. Spiegelman, *An extracellular darwinian experiment with a self-duplicating nucleic acid molecule.*, *Proceedings of the National Academy of Sciences of the United States of America* **58** (1967), 217–224 (eng).
- [72] J. G. Mitchell, *The influence of cell size on marine bacterial motility and energetics*, *Microbial Ecology* **22** (1991), 227–238.
- [73] P.-A. Monnard, A. Kanavarioti, and D. W. Deamer, *Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases*, *Journal of the American Chemical Society* **125** (2003), 13734–13740.
- [74] M. Morasch, D. Braun, and C. B. Mast, *Heat-flow-driven oligonucleotide gelation separates single-base differences*, *Angewandte Chemie International Edition* **55** (2016), 6676–6679.
- [75] G. M. Muñoz Caro, U. J. Meierhenrich, W. A. Schutte, B. Barbier, A. Arcones Segovia, H. Rosenbauer, W. H.-P. Thiemann, A. Brack, and J. M. Greenberg, *Amino acids from ultraviolet irradiation of interstellar ice analogues*, *Nature* **416** (2002), 403–406.
- [76] H. Mutschler, A. Wochner, and P. Holliger, *Freeze-thaw cycles as drivers of complex ribozyme assembly*, *Nature Chemistry* **7** (2015), 502–508.
- [77] B. Obermayer, H. Krammer, D. Braun, and U. Gerland, *Emergence of information transmission in a prebiotic RNA reactor*, *Physical Review Letters* **107** (2011), 018101.
- [78] E. A. Oussatcheva, L. S. Shlyakhtenko, R. Glass, R. R. Sinden, Y. L. Lyubchenko, and V. N. Potaman, *Structure of branched DNA molecules: gel retardation and atomic force microscopy studies*, *Journal of Molecular Biology* **292** (1999), 75–86.
- [79] R. Pascal, A. Pross, and J. D. Sutherland, *Towards an evolutionary theory of the origin of life based on kinetics and thermodynamics*, *Open Biology* **3** (2013), 130156.
- [80] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, and J. D. Sutherland, *Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism*, *Nature Chemistry* **7** (2015), 301–307.

- [81] V. Patzke, J. S. McCaskill, and G. von Kiedrowski, *DNA with 3'-5'-disulfide links-rapid chemical ligation through isosteric replacement*, *Angewandte Chemie International Edition* **53** (2014), 4222–4226.
- [82] R. Piazza and A. Guarino, *Soret effect in interacting micellar solutions*, *Physical Review Letters* **88** (2002).
- [83] V. B. Pinheiro and P. Holliger, *The XNA world: progress towards replication and evolution of synthetic genetic polymers*, *Current Opinion in Chemical Biology* **16** (2012), 245–252.
- [84] S. Pino, G. Costanzo, A. Giorgi, and E. Di Mauro, *Sequence complementarity-driven nonenzymatic ligation of RNA*, *Biochemistry* **50** (2011), 2994–3003.
- [85] M. W. Powner, B. Gerland, and J. D. Sutherland, *Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions*, *Nature* **459** (2009), 239–242.
- [86] K. V. Pugachev, F. Guirakhoo, S. W. Ocran, F. Mitchell, M. Parsons, C. Penal, S. Girakhoo, S. O. Pougatcheva, J. Arroyo, D. W. Trent, and T. P. Monath, *High fidelity of yellow fever virus RNA polymerase*, *Journal of Virology* **78** (2003), 1032–1038.
- [87] S. Rajamani, J. K. Ichida, T. Antal, D. A. Treco, K. Leu, M. A. Nowak, J. W. Szostak, and I. A. Chen, *Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication*, *Journal of the American Chemical Society* **132** (2010), 5880–5885.
- [88] S. Rajamani, A. Vlassov, S. Benner, A. Coombs, F. Olasagasti, and D. Deamer, *Lipid-assisted synthesis of rna-like polymers from mononucleotides*, *Origins of Life and Evolution of Biospheres* **38** (2008), 57–74.
- [89] D. L. Robertson and G. F. Joyce, *Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA*, *Nature* **344** (1990), 467–468.
- [90] M. P. Robertson and G. F. Joyce, *The origins of the RNA world*, *Cold Spring Harbor Perspectives in Biology* **4** (2010), a003608.
- [91] ———, *Highly efficient self-replicating RNA enzymes*, *Chemistry & Biology* **21** (2014), 238–245.
- [92] D. Rodbard and A. Chrambach, *Unified theory for gel electrophoresis and gel filtration*, *Proceedings of the National academy of Sciences of the United States of America* **65** (1970), 970–977.
- [93] A. S. Rodin, E. Szathmáry, and S. N. Rodin, *On origin of genetic code and tRNA before translation*, *Biology Direct* **6** (2011), 14.
- [94] R. Rohatgi, D. P. Bartel, and J. W. Szostak, *Nonenzymatic, template-directed ligation of oligoribonucleotides is highly regioselective for the formation of 3'-5' phosphodiester bonds*, *Journal of the American Chemical Society* **118** (1996), 3340–3344.

A. References

- [95] M. J. Russell and A. J. Hall, *The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front*, *Journal of the Geological Society* **154** (1997), 377–402.
- [96] M. J. Russell, A. J. Hall, A. J. Boyce, and A. E. Fallick, *100th anniversary special paper: > on hydrothermal convection systems and the emergence of life*, *Economic Geology* **100** (2005), 419–438.
- [97] J. SantaLucia and D. Hicks, *The thermodynamics of DNA structural motifs*, *Annual Review of Biophysics and Biomolecular Structure* **33** (2004), 415–440.
- [98] S. J. Sørensen, M. Bailey, L. H. Hansen, N. Kroer, and S. Wuertz, *Studying plasmid horizontal transfer in situ: a critical review*, *Nature Reviews Microbiology* **3** (2005), 700–710.
- [99] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, and A. Eschenmoser, *Chemical etiology of nucleic acid structure: The α -threofuranosyl-(3'→2') oligonucleotide system*, *Science* **290** (2000), 1347–1351.
- [100] J. G. Schmidt, L. Christensen, P. E. Nielsen, and L. E. Orgel, *Information transfer from DNA to peptide nucleic acids by template-directed syntheses*, *Nucleic Acids Research* **25** (1997), 4792–4796.
- [101] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis*, *Nature Methods* **9** (2012), 671–675.
- [102] E. D. Schneider and J. J. Kay, *Life as a manifestation of the second law of thermodynamics*, *Mathematical and Computer Modelling* **19** (1994), 25–48.
- [103] E. Schrödinger, *What is life*, Cambridge University Press, 1944.
- [104] J. T. Szcepaniski and G. F. Joyce, *A cross-chiral RNA polymerase ribozyme*, *Nature* **515** (2014), 440–442.
- [105] S. J. Sharp, J. Schaack, L. Cooley, D. J. Burke, and D. Soil, *Structure and transcription of eukaryotic tRNA gene*, *Critical Reviews in Biochemistry* **19** (1985), 107–144.
- [106] F. L. Sousa, T. Thiergart, G. Landan, S. Nelson-Sathi, I. A. C. Pereira, J. F. Allen, N. Lane, and W. F. Martin, *Early bioenergetic evolution*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **368** (2013), 20130088.
- [107] P. Stadlbauer, J. Šponer, G. Costanzo, E. Di Mauro, S. Pino, and J. E. Šponer, *Tetra-loop-like geometries could form the basis of the catalytic activity of the most ancient ribooligonucleotides*, *Chemistry A European Journal* **21** (2015), 3596–3604.
- [108] R. Stoltenburg, C. Reinemann, and B. Strehlitz, *SELEX—a (r)evolutionary method to generate high-affinity nucleic acid ligands*, *Biomolecular Engineering* **24** (2007), 381–403.

- [109] R. Stribling and S. L. Miller, *Energy yields for hydrogen cyanide and formaldehyde syntheses: The HCN and amino acid concentrations in the primitive ocean*, *Origins of Life and Evolution of the Biosphere* **17** (1987), 261–273.
- [110] E. Szathmary, *The origin of replicators and reproducers*, *Philosophical Transactions of the Royal Society B: Biological Sciences* **361** (2006), 1761–1776.
- [111] J. W. Szostak, *The eightfold path to non-enzymatic RNA replication*, *Journal of Systems Chemistry* **3** (2012), 2.
- [112] H. Trinks, W. Schröder, and C. K. Biebricher, *Ice and the origin of life*, *Origins of Life and Evolution of Biospheres* **35** (2005), 429–445.
- [113] T. M. Tritt, *Thermal conductivity: Theory, properties, and applications*, *Physics of Solids and Liquids*, Kluwer Academic/Plenum Publishers, 2004.
- [114] R. M. Turk, M. Illangasekare, and M. Yarus, *Catalyzed and spontaneous reactions on ribozyme ribose*, *Journal of the American Chemical Society* **133** (2011), 6044–6050.
- [115] N. Vaidya, M. L. Manapat, I. A. Chen, R. Xulvi-Brunet, E. J. Hayden, and N. Lehman, *Spontaneous network formation among cooperative RNA replicators*, *Nature* **491** (2012), 72–77.
- [116] D. Valev, *Estimations of total mass and energy of the universe*, *Physics International* **5** (2014), 15–20.
- [117] A. V. Vlassov, B. H. Johnston, L. F. Landweber, and S. A. Kazakov, *Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world*, *Nucleic Acids Research* **32** (2004), 2966–2974.
- [118] A. V. Vlassov, S. A. Kazakov, B. H. Johnston, and L. F. Landweber, *The RNA world on ice: A new scenario for the emergence of RNA information*, *Journal of Molecular Evolution* **61** (2005), 264–273.
- [119] G. von Kiedrowski, *A self-replicating hexadeoxynucleotide*, *Angewandte Chemie International Edition* **25** (1986), 932–935.
- [120] G. Wächtershauser, *Groundworks for an evolutionary biochemistry: The iron-sulphur world*, *Progress in Biophysics and Molecular biology* **58** (1992), 85–201.
- [121] E. K. Wheeler, W. Benett, P. Stratton, J. Richards, A. Chen, A. Christian, K. D. Ness, J. Ortega, L. G. Li, T. H. Weisgraber, K. Goodson, and F. Milanovich, *Convectively Driven Polymerase Chain Reaction Thermal Cycler*, *Analytical Chemistry* **76** (2004), 4011–4016.
- [122] Y. I. Wolf and E. V. Koonin, *On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization*, *Biology Direct* **2** (2007), 14.

A. References

- [123] P. Yin, H. M. T. Choi, C. R. Calvert, and N. A. Pierce, *Programming biomolecular self-assembly pathways*, *Nature* **451** (2008), 318–322.
- [124] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce, *NUPACK: Analysis and design of nucleic acid systems*, *Journal of Computational Chemistry* **32** (2010), 170–173.
- [125] S. Zamenhof and H. H. Eichhorn, *Study of microbial evolution through loss of biosynthetic functions: Establishment of “defective” mutants*, *Nature* **216** (1967), 456–458.

B. Remarks

B.1. Exponential accumulation

The strength of accumulation in a thermo-gravitational pore can be described as the ratio of concentrations at the bottom of the pore and outside. It depends exponentially on the length of the pore, the temperature difference ΔT across the pore and particles' Soret coefficient S_T [2, 20]:

$$c_{bottom}/c_{top} = \exp(Q S_T \Delta T r). \quad (\text{B.1})$$

Here, r is the aspect ratio of the compartment, and $S_T = D_T/D$ gives the ratio of the thermal and mass diffusion coefficients D_T and D . The function

$$Q \equiv \frac{q/120}{1 + q^2/10080} \quad \text{with} \quad q \equiv \frac{\alpha \rho}{6\eta} \cdot \frac{g \Delta T}{D} w^3 \quad (\text{B.2})$$

depends on the properties of solvent and solute, and the pore geometry. g is the value of gravitational acceleration, and α , ρ and η are the thermal expansion coefficient, mass density and viscosity of water. w is the width of the pore. Under optimal conditions, $Q \approx 0.42$.

B.2. Characteristic time scales

The timescale of the thermophoretic flux τ_T in a thermo-gravitational pore can be estimated using the width of the pore w and the thermophoretic drift velocity v_T as

$$\tau_T = \frac{w}{v_T} = \frac{w^2}{S_T D \Delta T} \approx \frac{(100 \mu\text{m})^2}{0.1 \text{ K}^{-1} 100 \mu\text{m}^2 \text{ s}^{-1} 30 \text{ K}} \approx 30 \text{ s}. \quad (\text{B.3})$$

As above, ΔT is the temperature difference across the pore, D and S_T are the diffusion and Soret coefficients. Similarly, the timescale of the convective flux is given by $\tau_c = h/v_c^{mean}$. Here, h is the vertical dimension of the pore and v_c^{mean} the mean convection velocity. In the approximation of constant viscosity, the velocity profile across the pore is given by

$$v_c(x) = -\frac{\alpha g \rho \Delta T}{6\eta} w^2 \left(\frac{x}{w} - \frac{1}{2} \right) \left[\frac{1}{4} - \left(\frac{x}{w} - \frac{1}{2} \right)^2 \right]. \quad (\text{B.4})$$

B. Remarks

g is the value of gravitational acceleration. α , ρ and η are the thermal expansion coefficient, mass density and viscosity of water. Averaging over the pore width yields

$$v_c^{mean} = \frac{1}{w/2} \int_0^{w/2} v_c(x) dx = \frac{w^2}{32} \frac{\alpha g \rho \Delta T}{6\eta} \approx 8 \mu\text{m/s}. \quad (\text{B.5})$$

For the timescale τ_c , one gets

$$\tau_c = h/v_c^{mean} \approx \frac{1 \text{ mm}}{8 \mu\text{m/s}} \approx 2 \text{ minutes}. \quad (\text{B.6})$$

B.3. Effects of local viscosity changes

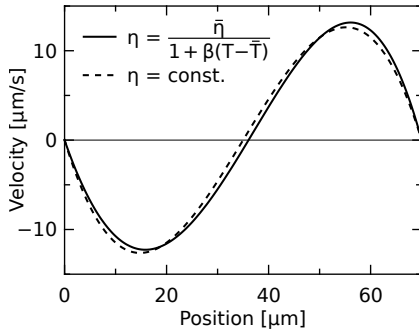


Figure B.1. Effects of temperature-dependent viscosity on the velocity profile across a thermo-gravitational pore. Data were calculated for the parameters of the experiments presented in section 2.2.2: a temperature gradient of 60 to 92 °C across a pore of 70 μm . Compared to the case of constant viscosity $\eta = \bar{\eta}$ (dashed), a first order approximation (solid) of $\eta(T)$ shifts the velocity profile a little to the hot side (right).

The convection velocity across a thermo-gravitational pore is commonly described under the assumption of a constant viscosity. However, as the temperature gradients discussed above are on the order of 30 K, a first order approximation of the temperature dependence of the viscosity was included in the calculations (“Analytic modelling” on page 23). While the shape of the profile is not strongly altered by that correction, the profile is shifted somewhat to the hot side (Fig. B.1). The fact that the overall impact is rather low is because the profile is determined by the boundary condition of zero total velocity.

B.4. Mean equilibrium polymer length

The dynamics of a general polymerization process with reversible bonds, as studied in [64], is described by

$$\frac{d}{dt}c_n = \frac{1}{2} \sum_{i+j=n} (k_{on}c_i c_j - k_{off}c_n) + \sum_{j>0} (k_{off}c_{j+n} - k_{on}c_j c_n). \quad (\text{B.7})$$

c_n denotes the concentration of strands of length $n > 0$. The model includes ligation of two polymers forming a larger one (terms containing k_{on}) and cleavage of one polymer into two smaller ones (terms containing k_{off}). Further, the system is assumed to be limited by the bond formation and breakup chemistry, such that the rates of association k_{on} and dissociation k_{off} are independent of the length of the molecules.

The mean equilibrium length $\langle L \rangle$ is given by

$$\begin{aligned} \langle L \rangle &= \frac{\sum_n n c_n}{\sum_n c_n} = c_0 \left(\sum_{n=1}^{\infty} c_1^n \left(\frac{k_{on}}{k_{off}} \right)^{n-1} \right)^{-1} = \frac{c_0}{c_1} \left(\sum_{n=0}^{\infty} \left(\frac{c_1}{K_D} \right)^n \right)^{-1} \\ &= \frac{c_0}{c_1} \left(1 - \frac{c_1}{K_D} \right) = \frac{c_0}{K_D} \left(\frac{K_D}{c_1} - 1 \right). \end{aligned} \quad (\text{B.8})$$

$c_0 = \sum_{n=1}^{\infty} n c_n$ is the total monomer concentration. the Using equation [S2] from [64]

$$\frac{c_1}{K_D} = 1 + \frac{K_D}{2c_0} - \frac{\sqrt{(K_D(K_D + 4c_0))}}{2c_0} = \frac{2c_0 + K_D - K_D \sqrt{1 + \frac{4c_0}{K_D}}}{2c_0}, \quad (\text{B.9})$$

the mean polymer lengths simplifies to¹

$$\begin{aligned} \langle L \rangle &= \frac{c_0}{K_D} \left(\frac{2 \frac{c_0}{K_D}}{2 \frac{c_0}{K_D} + 1 - \sqrt{1 + 4 \frac{c_0}{K_D}}} - 1 \right) \\ &= \left(\frac{2 \frac{c_0}{K_D}}{1 - \sqrt{1 + 4 \frac{c_0}{K_D}}} \right)^2 - \frac{c_0}{K_D} = \frac{1}{2} \left(1 + \sqrt{1 + 4 \frac{c_0}{K_D}} \right) \\ \langle L \rangle &= \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{c_0}{K_D}}. \end{aligned} \quad (\text{B.10})$$

¹The last simplification step is done using $2x = -\frac{1}{2}(1 + \sqrt{1 + 4x})(1 - \sqrt{1 + 4x})$
 $\Rightarrow (2x/(1 - \sqrt{1 + 4x}))^2 - x = \frac{1}{4}(1 + \sqrt{1 + 4x})^2 - x = \frac{1}{2}(1 + \sqrt{1 + 4x})$.

C. Danksagung

Das Entstehen dieser Arbeit ist der Mitwirkung vieler Menschen zu verdanken. Dafür gilt Euch mein Dank. Besonders bedanken möchte ich mich bei:

Dieter für die Möglichkeit, in seiner Arbeitsgruppe spannende Wissenschaft zu betreiben, für seine Unterstützung, Kreativität, seinen unkomplizierten Umgang mit administrativen Fragen, und seine begeisternde Art, neue Ideen zu verfolgen; außerdem dafür, dass ich an zahlreichen Konferenzen teilnehmen konnte, was kaum zu überschätzen ist.

Der gesamten AG Braun und ihren ehemaligen Mitgliedern für ein erfrischendes Arbeitsklima, das gegenseitige Hilfe, wissenschaftlichen Austausch und viele Diskussionen erst ermöglicht hat; für das Wissen, nie allein Espresso trinken zu müssen sowie die vielen Unternehmungen abseits des Laborbetriebs.

Christof für die Gewährleistung des Betriebs sämtlicher Geräte, die durch einen Computer gesteuert werden; für seine Antworten auf mindestens 98 % aller wissenschaftlichen Fragen, anregende Gespräche auch weniger wissenschaftlicher Natur und ein angenehmes Nerd-Niveau.

Lorenz für die gute Zusammenarbeit an einem herausfordernden Projekt.

Thomas für seinen Einsatz und seine Ausdauer auch bei ausgedehnten Messungen, die die Entstehung dieser Arbeit erheblich beschleunigt haben.

Susi für eine effiziente Kooperation, zahlreiche Abende mit oder ohne Getränk und noch mehr gute Gespräche.

Peter für die aufopferungsvolle Pflege der Espressomaschine, ohne die vieles mit Sicherheit langsamer und eine Spur bitterer gelaufen wäre.

Zhenya für ihre Feststellung, ich spräche als einziges verständliches Deutsch.

Georg für seine Anmerkungen zum Manuskript.

Meiner Familie und meinen Freunden, auch unter den oben genannten, ganz besonders meinen Eltern, für Anerkennung und Motivation in den vergangenen Jahren.

C. Danksagung

Nicht zuletzt danke ich **Katrin** für ihre Unterstützung besonders in frustrierenden Augenblicken, für ihre Nachsicht, wenn die Tage im Labor länger wurden und so vieles mehr.

D. Publications

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

Moritz Kreysing^{*}, Lorenz M. R. Keil^{*}, Simon A. Lanzmich^{*} & Dieter Braun

Nature Chemistry **7**, 203–208 (2015)

doi: 10.1038/nchem.2155

Figures 1 to 4 adapted with permission. Full article reprinted with permission.

© 2015 Macmillan Publishers Limited

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

Lorenz Keil, Michael Hartmann, Simon Lanzmich & Dieter Braun

Physical Chemistry Chemical Physics **18**, 20153–20159 (2016)

doi: 10.1039/c6cp00577b

Full article reprinted with permission.

© 2016 Royal Society of Chemistry

3. Thermophoresis in Nanoliter Droplets to Quantify Aptamer Binding.

Susanne A. I. Seidel, Niklas A. Markwardt, Simon A. Lanzmich & Dieter Braun

Angewandte Chemie International Edition **53**, 7948–7951 (2014)

doi: 10.1002/anie.201402514

Full article reprinted with permission.

© 2014 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

^{*}These authors contributed equally to this work.

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 1960s¹¹ 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 Earth¹² (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

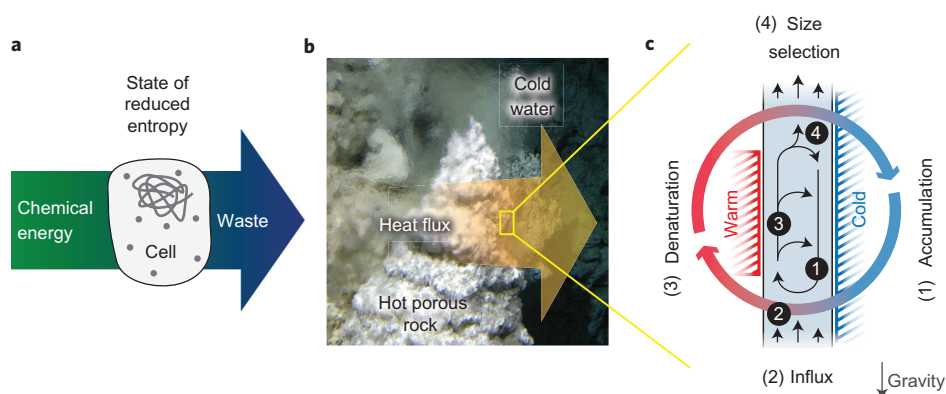


Figure 1 | Reduction of local entropy is key for living systems and can be caused by the flux of thermal energy. **a**, Modern cells feed on chemical energy, which enables them to host, maintain and replicate information-coding polymers, processes necessary for Darwinian evolution. **b**, The flux of thermal energy across geological cracks near a heat source (the white smoker²⁸ 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 $6\ \mu\text{m s}^{-1}$. 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).

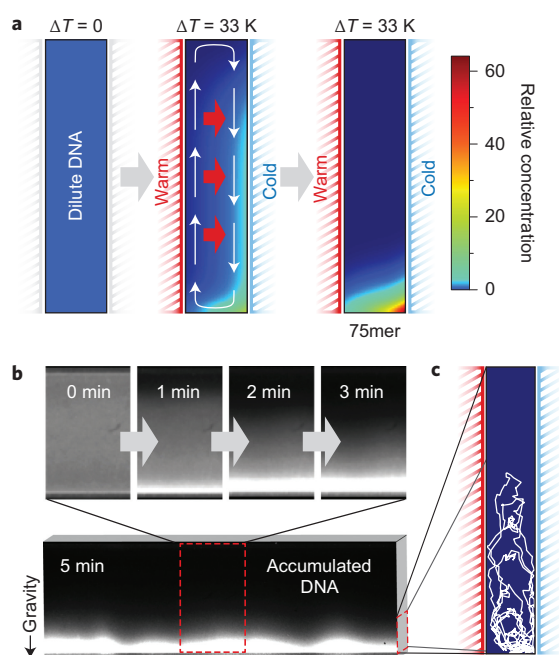


Figure 2 | Accumulation of oligonucleotides. **a**, The temperature gradient drives oligonucleotides horizontally from warm to cold by thermophoresis and simultaneously triggers the vertical thermal convection of water. Its combination results in a length-dependent accumulation at the bottom of an elongated 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 oligonucleotides²². 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}.

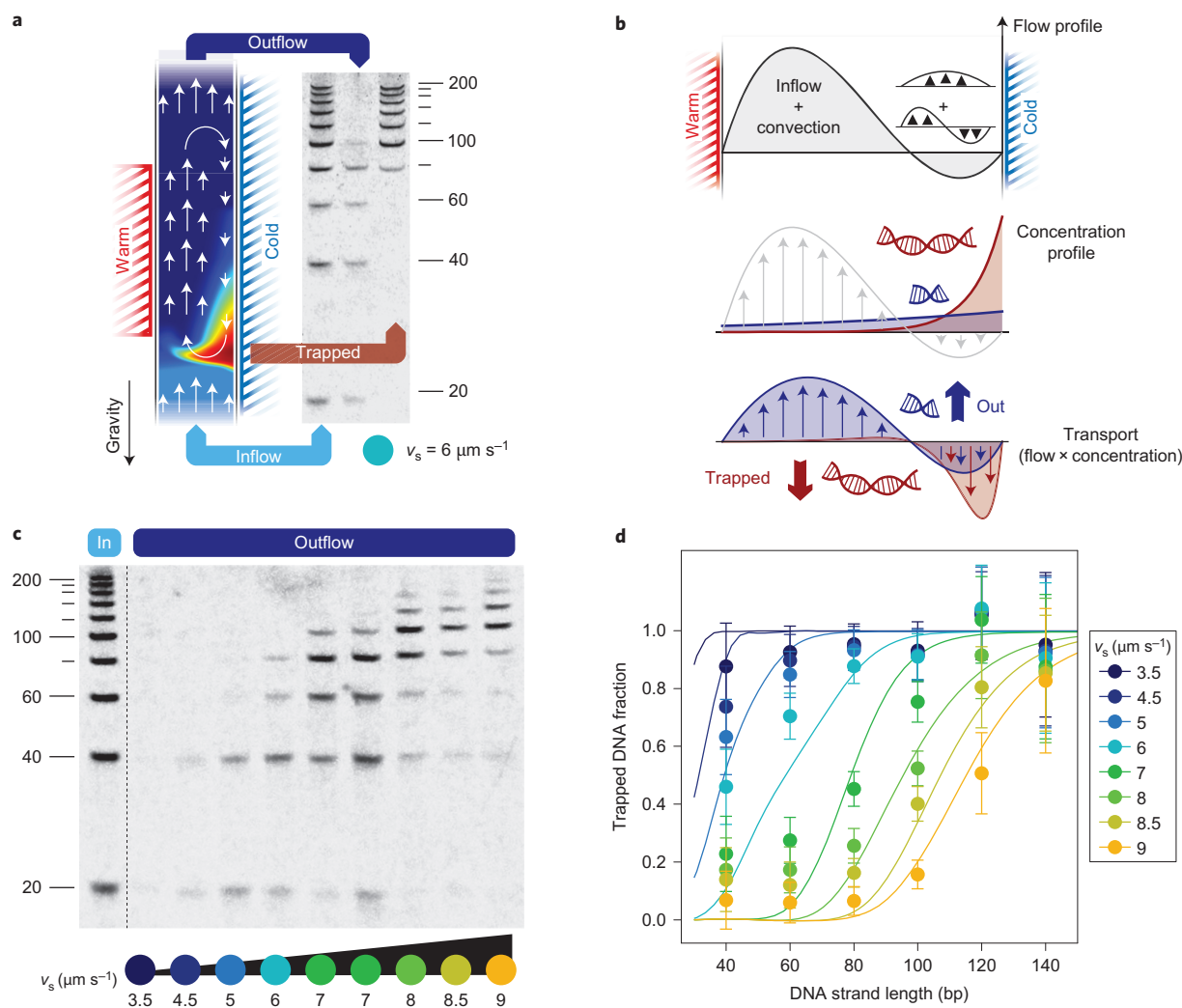


Figure 3 | Heat-driven filter selecting for strand length. **a**, A steady upwards feeding flow is triggered by opening the asymmetrically heated pore. A ladder of dsDNA (20–200 bp, 20 bp steps) was injected into the trap. Subsequent flushing of the capillary with pure buffer at a single velocity ($v_s = 6 \mu\text{m s}^{-1}$) 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 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 information¹¹. 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 $6 \mu\text{m s}^{-1}$. 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 $dc_i/dt = (\text{rep}_i - \text{dil}_i)c_i$ 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

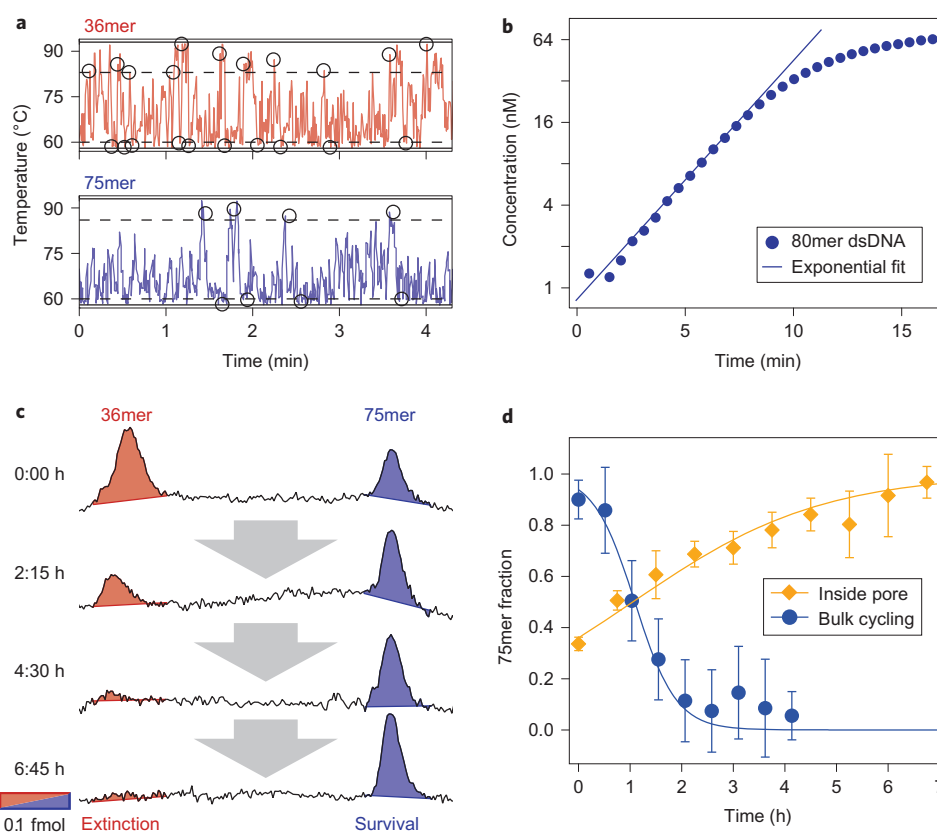


Figure 4 | Selection of a replicating DNA population that occupies the thermal habitat. **a**, Strands are subjected to temperature oscillations by the combination of thermophoresis, convection, feeding flow and diffusion. Simulations of stochastic molecule traces show that strands of 75 bp cycle inside 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 $\Delta k = (\text{rep}_L - \text{rep}_S) - (\text{dil}_L - \text{dil}_S)$ is the differential growth rate. We experimentally found that, inside the pore, long strands (L) outcompete shorter ones (S) with $\Delta k = 0.55 \text{ h}^{-1}$ (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 $\Delta k = -2.5 \text{ h}^{-1}$ (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 water²³, 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 replication²⁴. At its top, the outflow would be constricted through one or more thin, but longer, selecting

pores. Their increased length of several centimetres instead of 3.5 mm compensates linearly for the reduced temperature difference¹³.

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 templating²⁵ 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 aptamers²⁶.

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 $1.2 \text{ W m}^{-1} \text{ K}^{-1}$. 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 $1 \times$ 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 ($10 \mu\text{g ml}^{-1}$, 100–1000 bp, ten equidistant bands, weight equalized) was used in combination with $0.5 \times$ SYBR Green I²⁷. The applied temperature gradient from 22 °C

to 88 °C resulted in temperatures from 38 °C to 71 °C inside the capillary (inner dimensions, $100 \mu\text{m} \times 2,000 \mu\text{m}$ and $70 \mu\text{m} \times 1,400 \mu\text{m}$, as specified).

Fractionation experiments. A DNA ladder (20–200 bp, ten equidistant bands) was suspended in a $1 \times$ PCR buffer that included 0.1% Tween20. Fractionation was carried out in a vertically oriented capillary (inner dimensions, $70 \mu\text{m} \times 1400 \mu\text{m}$) 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, $70 \mu\text{m} \times 1,400 \mu\text{m}$, heated along 2.5 mm) at a mean solvent velocity of $6 \mu\text{m s}^{-1}$. 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 $\text{dil}_s = \text{dil}_t = 5.8 \text{ h}^{-1}$ 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 $\text{rep}_s = (5.8 \pm 0.6) \text{ h}^{-1}$ (36mer) and $\text{rep}_t = (3.3 \pm 0.4) \text{ h}^{-1}$ (75mer).

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

References

- Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
- Sievers, D. & Von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224 (1994).
- Mansy, S. S. *et al.* Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125 (2008).
- Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
- Paul, N. & Joyce, G. F. A self-replicating ligase ribozyme. *Proc. Natl Acad. Sci. USA* **99**, 12733–12740 (2002).
- 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).
- Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **3**, 1–14 (2012).
- 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).
- 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).
- Schrödinger, E. *What is Life?* (Cambridge Univ. Press, 1944).
- 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).
- Lay, T., Hernlund, J. & Buffett, B. A. Core–mantle boundary heat flow. *Nature Geosci.* **1**, 25–32 (2008).
- Baaske, P. *et al.* Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proc. Natl Acad. Sci. USA* **104**, 9346–9351 (2007).
- 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).
- 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 Thermoeffusion und Thermosiphonwirkung. *Naturwissenschaften* **26**, 546 (1938).
17. Debye, P. Zur Theorie des Clusiuschen 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.

HEAT FLUX ACROSS AN OPEN PORE ENABLES THE CONTINUOUS REPLICATION AND
SELECTION OF OLIGONUCLEOTIDES TOWARDS INCREASING LENGTH

Moritz Kreysing^{1,2,*}, Lorenz Keil^{1,*}, Simon Lanzmich^{1,*} and Dieter Braun¹

¹ Systems Biophysics, Physics Department, Center for Nanoscience,

Ludwig-Maximilians-Universität München, 80799 Munich, Germany

² now: Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

*: contributed equally

Table of Contents

Supplementary Methods	3
Temperature gradients i) Ohmic heating with a transparent electrode.....	3
Temperature gradients ii) Ohmic heating at monitored temperatures.....	3
Microfluidics.....	4
Imaging of fluorescently labelled oligonucleotides.....	4
Gel electrophoresis and documentation.....	4
Fractionation experiments.....	5
Quantitative gel analysis.....	5
Replication model.....	6
Diffusion and screening length parameters of DNA.....	7
Computer simulations and analytical trapping model.....	7
Calculation of temperature cycling times.....	8
Modelling of fractionation experiments.....	8

Supplementary Figures	9
Supplementary Fig. S1 Focussing of a temperature gradient in a millimetre-sized pore.....	9
Supplementary Fig. S2 Separation of accumulation and thermal cycling in multipore system..	10
Supplementary Fig. S3 Experimental set-up to generate temperature gradients across rectangular borosilicate capillaries by electrically heating a transparent electrode.....	11
Supplementary Fig. S4 Cross-sectional drawings of flow-through set-up employed for fractionation and selection & replication experiments (Figures 3, 4), side and top view.....	12
Supplementary Fig. S5 Injection of a DNA pulse for the fractionation experiments shown in Figure 3.....	13
Supplementary Fig. S6 Determination of critical trapping velocities for each 20 bp ladder fragment.....	14
Supplementary Fig. S7 Analytic fit functions for the Soret and diffusion coefficients of DNA and RNA based on published measurements of short DNA.....	15
Supplementary Fig. S8 Visualisation of random walk simulations of 36 mer (left group) and 75 mer (right group) DNA inside a 70 μm wide, asymmetrically heated pore.....	16
Supplementary Fig. S9 Thermal cycle times of individual nucleotide particles inferred from trajectories.....	17
Supplementary Fig. S10 Temperature cycle statistics obtained from random walk simulations.	18
Supplementary Fig. S11 Quantification of native polyacrylamide gel data from the selection and replication experiment presented in Fig. 4 b, d.....	19
Supplementary Fig. S12 Quantification of native polyacrylamide gel data from the serial dilution experiment presented in Fig. 4d.....	20
Supplementary Videos	21
Supplementary References	23

Supplementary Methods

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, Supplementary Fig. S3) in a radio frequency sputtering chamber (LS320, Von Ardenne, Germany)¹⁻³, 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.

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. S4). 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 stabilised by a PID-controlled feedback loop acting on the cooling side (± 50 mK)⁴, 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.

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.

Imaging of fluorescently labelled oligonucleotides

Fluorescent imaging of DNA was realised with a 90°-tilted upright microscope (Axioscope A1, Zeiss), using a 2.5 \times objective (Plan-Neofluar 2.5 \times 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 \times 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 \times 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 protocol⁵. 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 labelled 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).

Fractionation experiments

A weight equalised double stranded DNA ladder (10 equidistant bands, 20 bp–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 MgCl₂, 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. S5). 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 in Figure 3c is shown in Supplementary Fig. S6c.

Quantitative gel analysis

Gel image quantification (as shown in Supplementary Fig. S11, S12) was done with a custom LabVIEW program, after point-like outliers have been removed using NIH ImageJ⁶. 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. S11d, S12c). The uncertainty of this integral was estimated using the standard deviation of the values around the base points of the linear background.

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 our experiment, molecules are also subjected to a continuous outflow. The latter is modelled 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 $dc_i/dt = c_i(rep_i - dil_i)$. Its solution is $c(t) = c_i^0 \cdot e^{(rep_i - dil_i) \cdot t}$, the relative concentration of 75mer DNA is given by $c_L/(c_S + c_L) = c_L^0 \cdot e^{k_L \cdot t} / (c_L^0 \cdot e^{k_L \cdot t} + c_S^0 \cdot e^{k_S \cdot t})$, 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 $c_L/(c_S + c_L) = (1 + A \cdot e^{\Delta k \cdot t})^{-1}$.

Taking into account separately determined parameters for temperature gradients, Soret coefficients, diffusion coefficients, and inflow velocities, the fluid-dynamic model yields dilution rates of $dil_L = 3.2 \text{ h}^{-1}$ and $dil_S = 12.5 \text{ h}^{-1}$ 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. S9 for their determination). We found replication rates of $rep_L = (3.3 \pm 0.4) \text{ h}^{-1}$ for the long strands and $rep_S = (12.05 \pm 0.06) \text{ h}^{-1}$ for the short strands. 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, cf. also Supplementary Fig. S12). 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. During the lifetime $\tau = 18.8 \text{ min}$ of long strands in the pore defined by on/off-rate decay $c/c_0 = 1/e$, the population of long strand grows 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 our exponential growth model yields relative growths of $c_L(7h)/c_L(0) \approx 2.0$ and $c_S(7h)/c_S(0) \approx 0.04$ for short and long strands, respectively. This is in good agreement with the experimental results of the selection and replication experiment of $c_L(7h)/c_L(0) = 1.7 \pm 0.3$ and $c_S(7h)/c_S(0) = 0.1 \pm 0.1$.

Diffusion and screening length parameters of DNA

The Debye length is a major determining parameter for the strength of thermophoresis. It was estimated as $\lambda_{DH} = 1.30$ nm for the used PCR buffer (10mM TRIS, 50mM KCl, 1.5mM MgCl₂) at 75°C. Measurements of DNA at this Debye length were interpolated from the previously measured two-dimensional data set⁷. As seen in Supplementary Fig. S7a, the Soret coefficients of single and double stranded DNA do not significantly differ. Interestingly, RNA shows very similar thermophoretic properties⁷. We therefore used all measured values and interpolate them with a square root function, resulting in $S_T = -0.0063 + 0.0115 \cdot (\#bases)^{0.5}$. Here, $\#bases$ is the number of bases on a single strand. A very similar approach of fitting was used previously⁸.

To infer the diffusion coefficient from the same measurements⁹, the radius of dsDNA was fitted from the same data set as plotted in Supplementary Fig. S7b⁷. The radius shows good agreement with a line fit according to $R = (0.8 + \#bases \times 0.059)$ 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 $D = 6.69 \times 10^{-19} / (8 \times 10^{-10} + (\#bases) \times 5.9 \times 10^{-11})$ m²/s.

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 before⁷. 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).

Calculation of temperature cycling times

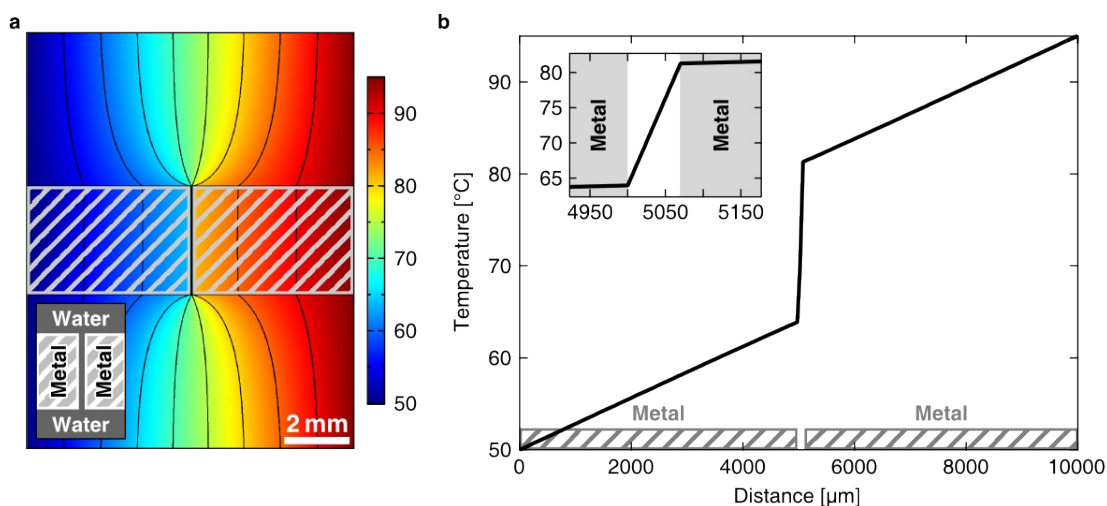
Typical results of these stochastic particle tracing simulations are visualised in Supplementary Fig. S8, showing individual trajectories for 36mer and 75mer DNA in the trapping geometry relevant to the selection and replication experiment of Fig. 4d. Using these trajectories, statistical data on thermal 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. S9 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. S10. 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.

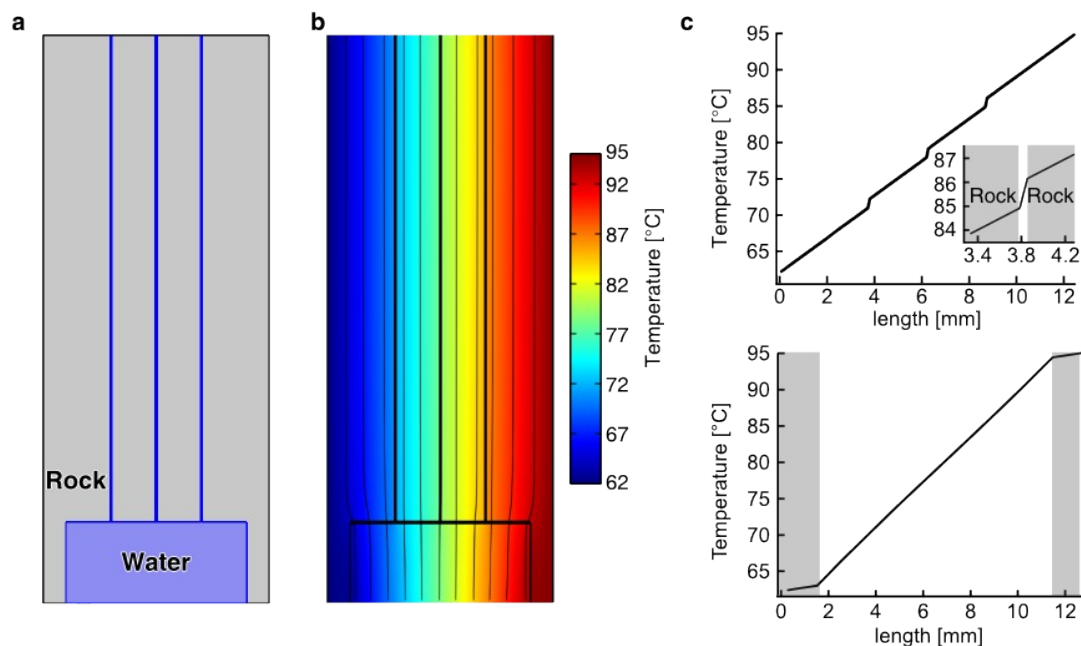
Modelling of fractionation experiments

Simulations of the fractionation of a DNA ladder, as presented in Fig. 3, required multiple steps. First, we simulated the length-dependent propagation and trapping of a mixed-length DNA pulse through the trap for different influx velocities. From this data, we 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. S6a,b. For each velocity step, the concentration of the DNA leaving the trap was normalised to the step duration.

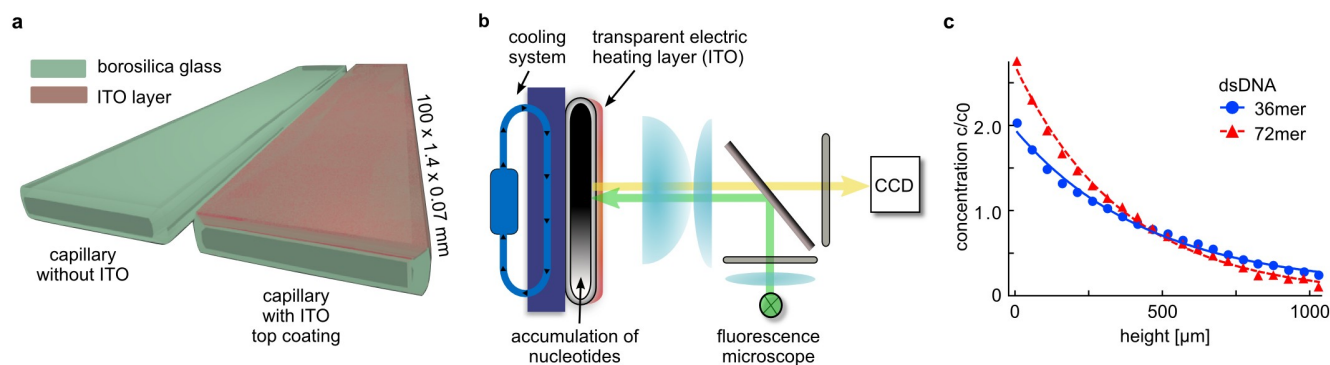
Supplementary Figures



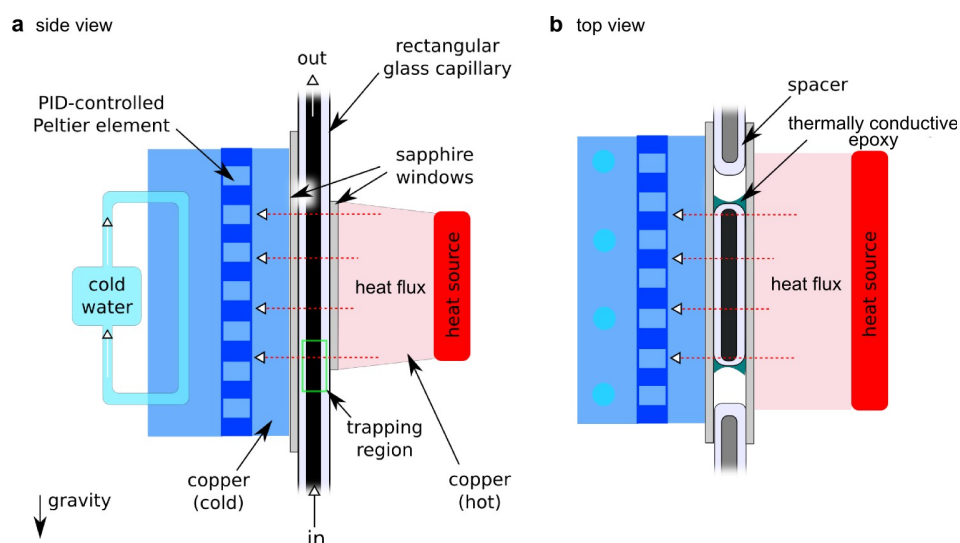
Supplementary Fig. S1 | Focussing of a temperature gradient in a millimetre-sized pore. **a**, Millimetre-sized metal inclusions (grey hatched) focus a temperature gradient across a millimetre-sized pore. As metals show about 100-fold higher thermal conductivities than water, the temperature gradient is strongly focussed 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).



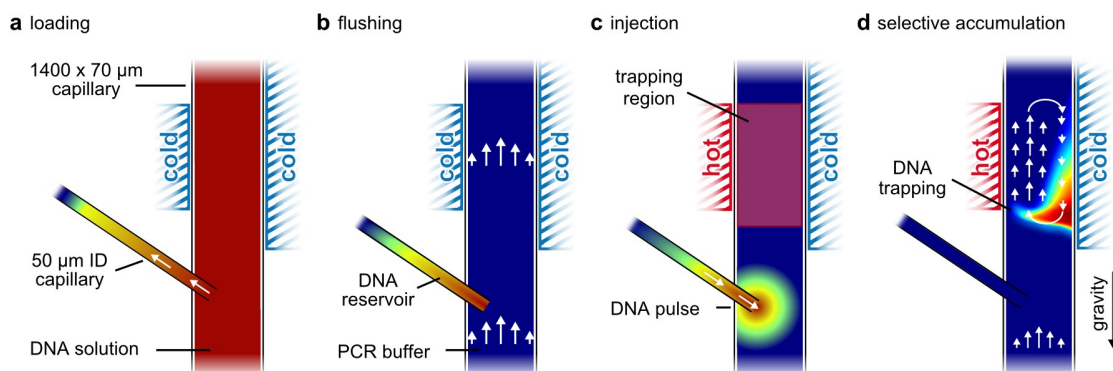
Supplementary Fig. S2 | Separation of accumulation and thermal cycling in multipore system. **a**, A multipore system combining a large 1 cm sized pore with multiple 70 μm small sized pores. The small pores with a length of 3 cm are able to accumulate DNA 10^3 -fold and the accumulation grows exponentially 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.



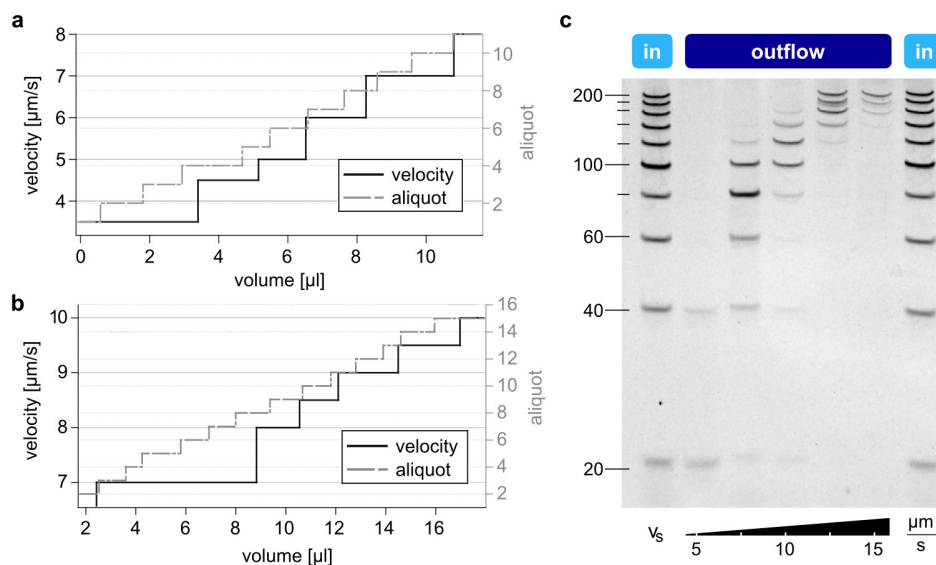
Supplementary Fig. S3 | Experimental set-up to generate temperature gradients across rectangular borosilicate capillaries by electrically heating a transparent electrode. **a**, Schematics of the capillary before and after coating with a transparent layer of indium tin oxide (ITO, here red). **b**, Capillary 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-labelled dsDNA templates 36 mer and 72 mer reveal exponential accumulation characteristics with a stronger spatial confinement of longer oligonucleotides. The experiment was carried out in a capillary with internal dimensions of $70\ \mu\text{m} \times 1400\ \mu\text{m}$ 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.



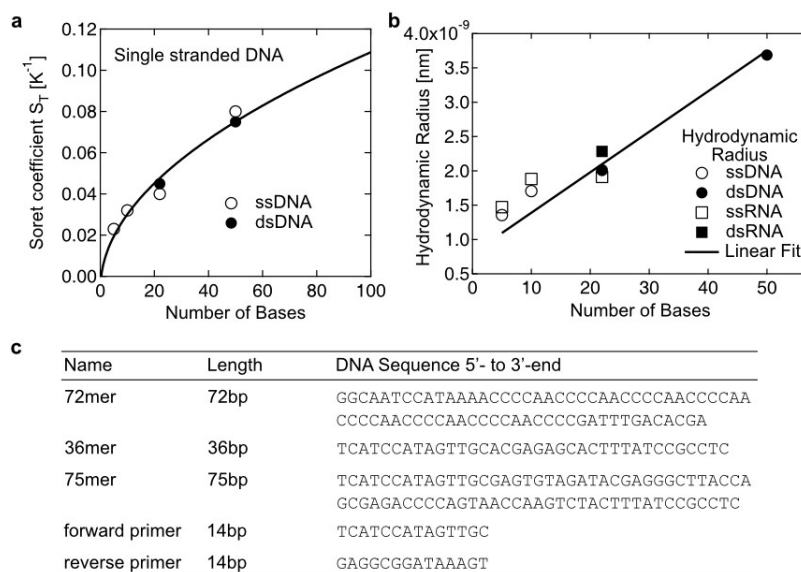
Supplementary Fig. S4 | Cross-sectional drawings of flow-through set-up employed for fractionation and selection & replication experiments (Figures 3, 4), side and top view. A rectangular glass capillary is sandwiched between two sapphire windows. Thermal coupling is achieved by thermally conductive 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.



Supplementary Fig. S5 | Injection of a DNA pulse for the fractionation experiments shown in Figure 3. **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.

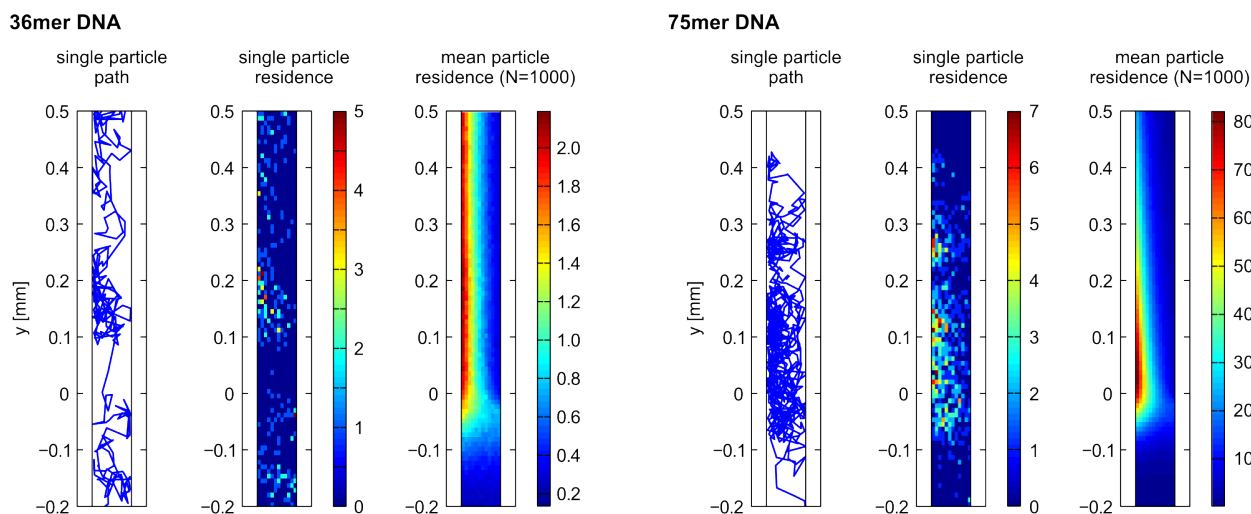


Supplementary Fig. S6 | Determination of critical trapping velocities for each 20 bp ladder fragment. Fractionated DNA was obtained from the output of the microfluidic system in volumes between 0.9-1.5 μl and assigned to the step-wise increased velocities. **a**, The influx was gradually increased from 3.5 nl/s to 10 nl/s. At an inflow of 3.5 $\mu\text{m/s}$, 4.5 $\mu\text{m/s}$, 5 $\mu\text{m/s}$, 6 $\mu\text{m/s}$ and 7 $\mu\text{m/s}$, DNA fragments with a length of 20bp, 40bp, 60bp, 80bp and 100bp started to be flushed out of the trapping region, respectively. **b**, At an inflow of 7 $\mu\text{m/s}$, 8 $\mu\text{m/s}$, 8.5 $\mu\text{m/s}$ and 9 $\mu\text{m/s}$, DNA fragments with a length of 100bp, 120bp, 140bp and 160bp were flushed out, respectively. **c**, Gel electrophoresis of thermally fractionated double stranded DNA ladder (20-200bp) under same conditions as data presented in main text (Fig. 3c), but at coarser velocity steps and with a wider range of inflow velocities v_s .

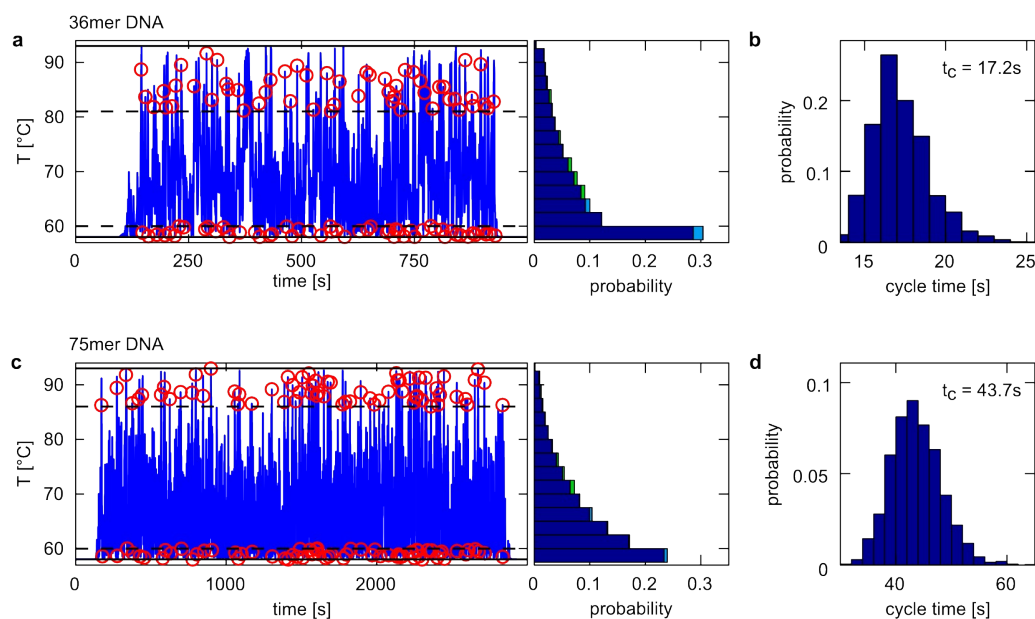


Supplementary Fig. S7 | Analytic fit functions for the Soret and diffusion coefficients of DNA and RNA based on published measurements of short DNA.

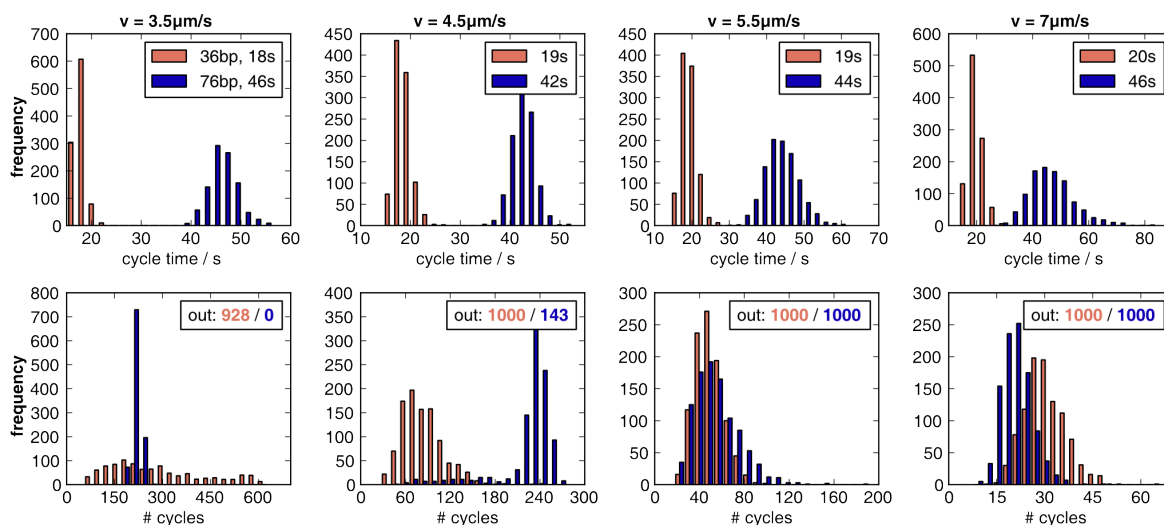
a, The fit of the Soret coefficient is used as input parameter for the fluid dynamic computer simulation of the thermal trap. **b**, Hydrodynamic radius of single and double-stranded DNA/RNA. **c**, Sequences of 72mer, 36mer, 75mer, and forward and reverse primers.



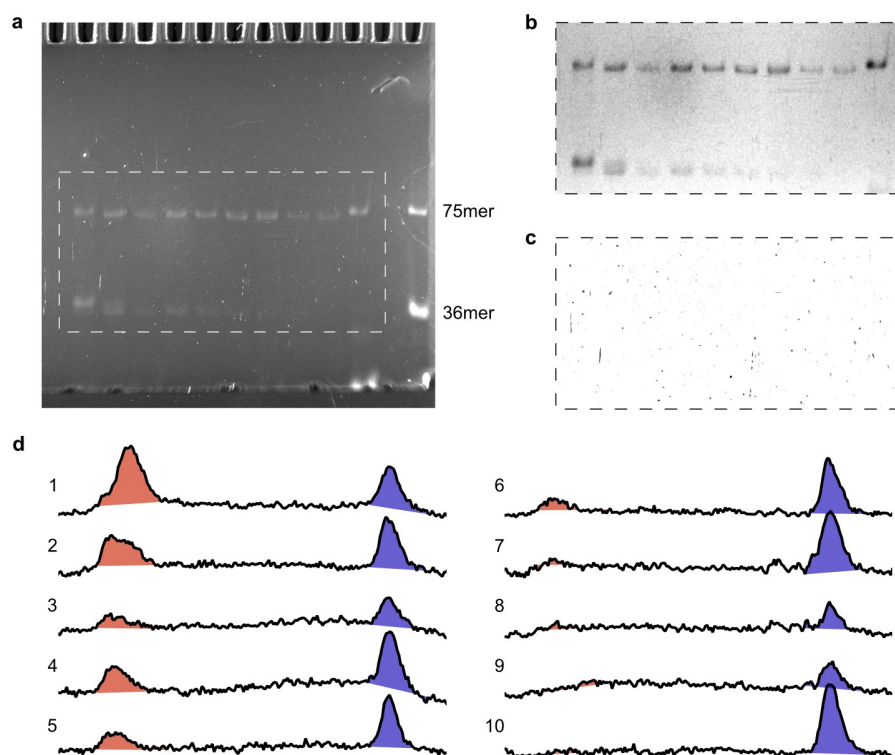
Supplementary Fig. S8 | Visualisation of random walk simulations of 36 mer (left group) and 75 mer (right group) DNA inside a 70 μm wide, asymmetrically heated pore. A mean influx velocity of 4 $\mu\text{m/s}$ is applied from below. Individual panels show a single particle trajectory (left), the corresponding 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 colour 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.



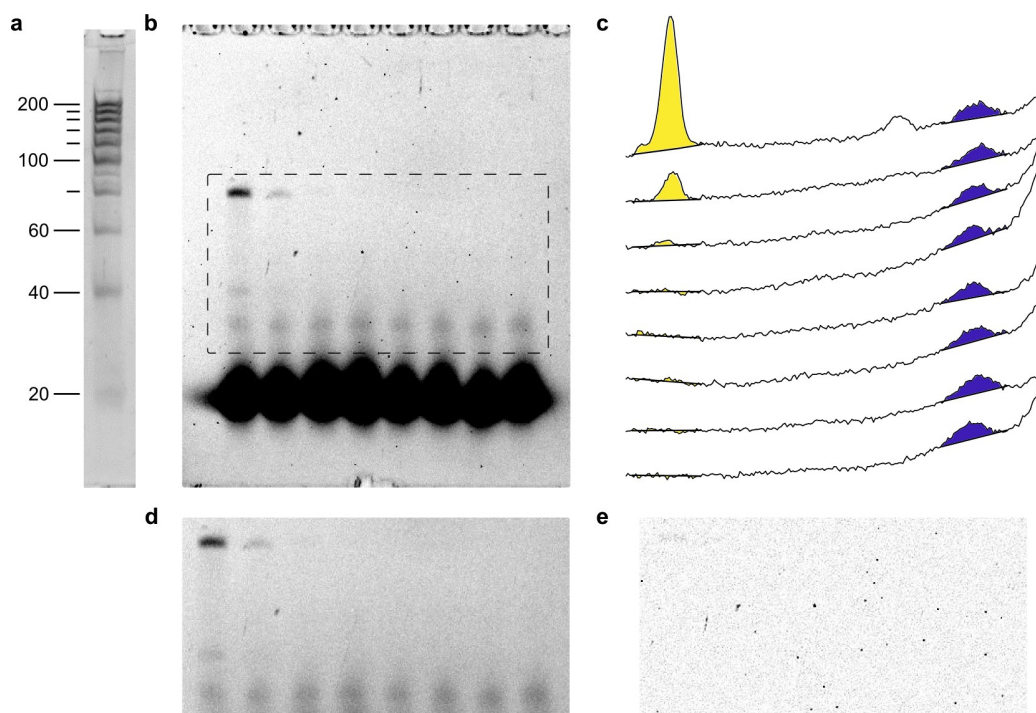
Supplementary Fig. S9 | Thermal cycle times of individual nucleotide particles inferred from trajectories. The underlying simulation has been performed for $N=1000$ particles at an influx velocity of $5.5 \mu\text{m/s}$. **a, c**, Temperature experienced during the lifetime of an individual particle. 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 summarise 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.



Supplementary Fig. S10 | Temperature cycle statistics obtained from random walk simulations. N=1000 particles have been simulated for 10.000 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, colour 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 abundance time, allowing them to cycle more often than the short ones.



Supplementary Fig. S11 | Quantification of native polyacrylamide gel data from the selection and replication experiment presented in Fig. 4 b, d. **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. 3d and 4d were calculated from the contribution of noise at the end points of the local baselines to the integrated intensities.



Supplementary Fig. S12 | Quantification of native polyacrylamide gel data from the serial dilution experiment presented in Fig. 4d. **a**, Size reference: equidistant 200 bp dsDNA ladder. **b**, Raw gel image and selection of the region of interest. **c**, Integrated intensities from panel **d**, of all 8 gel lanes. **d**, Region of interest for quantitative analysis. Outliers were removed before quantification. **e**, Outliers removed from panel **d**.

Supplementary Videos

1) nchem.2155-s2.mp4

Supplementary Video 1

Qualitative visualisation of cyclic trajectories inside a 100 μm thick and 2000 μm high rectangular capillary by fluorescent tracer particles, observed through the heated glass wall. A gentle flow perpendicular to the planes of cycling was applied to aid visualisation of the cycling trajectories, gravity pointing downwards. Total length of the video: 15 minutes.

2) nchem.2155-s3.mp4

Supplementary Video 2

Accumulation of DNA in an electrically heated capillary. The capillary was filled with 0.5 $\mu\text{g}/\text{ml}$ weight-equalised double stranded DNA ladder with 10 equidistant bands from 100 bp to 1000 bp in combination with 0.5 \times SYBR Green I. The capillary was heated through a transparent semiconductor layer. A temperature gradient ranging from 22 $^{\circ}\text{C}$ to 88 $^{\circ}\text{C}$ was applied at the outer walls of the capillary, resulting in a temperature difference from 38 $^{\circ}\text{C}$ to 71 $^{\circ}\text{C}$ inside the capillary. Inner dimensions of the capillary were 100 $\mu\text{m}\times$ 2000 μm . The video was recorded over a time frame of 300 seconds.

3) nchem.2155-s4.mp4

Supplementary Video 3

Length selective trapping. A capillary with dimensions of 70 $\mu\text{m}\times$ 1400 μm was filled with FAM labelled 75mer and a Cy5 labelled 25mer DNA. An alternating excitation of both dyes was used to separately record changes in molecular concentration. In order to perform selective pressure on trapped molecules, an external inflow of diluted DNA is applied and gradually increased from 2 $\mu\text{m}/\text{s}$ to 6 $\mu\text{m}/\text{s}$. Below 4 $\mu\text{m}/\text{s}$, both strands accumulate. In the region between 4 $\mu\text{m}/\text{s}$ and 5 $\mu\text{m}/\text{s}$, the 75mer DNA withstands the inflow whereas the 36mer DNA is diluted by the external inflow. Total recording time: 5.2 hours.

4) nchem.2155-s5.mp4**Supplementary Video 4**

Exponential replication of 80mer double-stranded DNA inside a rectangular glass capillary of 100 μm \times 2000 μm cross-sectional dimensions visualised by SYBR Green I fluorescence. View through heated side of the capillary, gravity pointing downwards. Total length of the video: 18 minutes.

Supplementary References

1. Minami, T., Sonohara, H., Kakumu, T., & Takata, S. Physics of very thin ITO conducting films with high transparency prepared by DC magnetron sputtering. *Thin Solid Films* **270**, 37–42 (1995).
2. Gupta, V. & Mansingh, A. Influence of postdeposition annealing on the structural and optical properties of sputtered zinc oxide film. *J. Appl. Phys.* **80**, 1063–1073 (1996).
3. Davidse, P. D. Theory and practice of RF sputtering. *Vacuum* **17**, 139–145 (1967).
4. Ziegler, J. G., Nichols, N. B. Optimum settings for automatic controllers. *trans. ASME* **64** (1942).
5. Maniatis, T., Jeffrey, A. & Van deSande, H. Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis. *Biochemistry* **14**, 3787–3794 (1975).
6. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
7. Reichl, M., Herzog, M., Götz, A. & Braun, D. Why charged molecules move across a temperature gradient: The Role of Electric Fields. *Phys. Rev. Lett.* **112**, 198101 (2014).
8. Reineck, P., Wienken, C. J. & Braun, D. Thermophoresis of single stranded DNA. *Electrophoresis* **31**, 279–286 (2010).
9. Stellwagen, E., Lu, Y. & Stellwagen, N. C. Unified description of electrophoresis and diffusion for DNA and other polyions. *Biochemistry* **42**, 11745–11750 (2003).



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*

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.

Received 26th January 2016,
Accepted 3rd May 2016

DOI: 10.1039/c6cp00577b

www.rsc.org/pccp

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 protobiomolecules, 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.

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

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

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.

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.

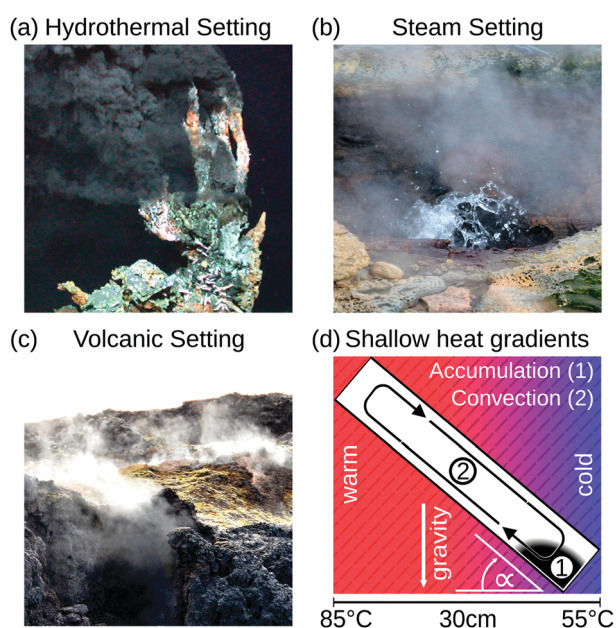


Fig. 1 Possible microthermal habitats for the origins of life. (a) Heat dissipation across submerged porous rocks and the cold ocean form steep temperature gradients which drive an efficient accumulation mechanism.^{22,23} (b and c) Shallow temperature gradients, approximately 100–2000-fold weaker than 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.

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 $\bar{v}_{TD} = -D \cdot S_T \cdot \nabla T$. 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 $T_{\text{low}} = 55$ °C and a high temperature of $T_{\text{high}} = T_{\text{low}} + \Delta T$ 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 convective 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 $\kappa = 3$ W mK⁻¹, a combination of quartz ($\kappa = 6.6$ W mK⁻¹) and olivine-melilitite ($\kappa = 1.7$ – 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:

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

with the molecule's diffusion coefficient D and Soret coefficient S_T , temperature T , the convective flow profile $\bar{v}(\alpha, w)$, 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 $\bar{v}(\alpha, w)$, 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 $c \cdot \nabla \cdot (\nabla T) = 0$; $c \cdot \nabla \cdot \bar{v}(\alpha, w) = 0$. 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:

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

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

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 $\Delta T = 30 \text{ K}$ over 30 cm and 3 cm , respectively. A temperature gradient in the range of $55 \text{ }^\circ\text{C}$ to $85 \text{ }^\circ\text{C}$, 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 $60 \text{ }^\circ\text{C}$ for elongation, followed by a high temperature region above $80 \text{ }^\circ\text{C}$ 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 $3 \times 10^5 \text{ h}$, corresponding to the steady-state reached after approximately $4.2 \times 10^5 \text{ h}$, 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 $D = 127 \text{ } \mu\text{m}^2 \text{ s}^{-1}$ and $S_T = 0.07 \text{ K}^{-1}$ for 100-mer RNA/DNA at a Debye length of $\lambda_D = 2.1 \text{ nm}$ at $70 \text{ }^\circ\text{C}$, both measured experimentally.^{40,47} Thermophoresis contributes with a net movement \bar{v}_{DT} along the gradient, resulting in a thermophoretic velocity of $|\bar{v}_{DT}| = 2.5 \text{ nm s}^{-1}$ for a 100-mer oligonucleotide. A convective flow profile with a maximum of $v_{\text{max}} \approx 8 \text{ } \mu\text{m s}^{-1}$ shuttles molecules between warm and cold areas. The displacement $\Delta s(x,y)$ of these particles is given by:

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

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

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 $10\text{--}400 \text{ K mm}^{-1}$, however, limit the scope of the approach to hydrothermal orifices and vapour heating with high flow rates. Here we show in simulations that $100\text{--}2000$ 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-melilitite) 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

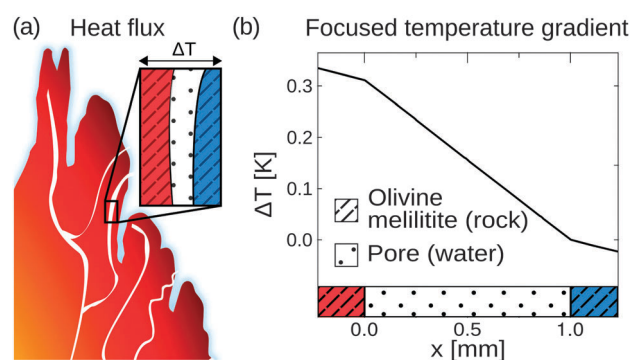


Fig. 2 Heat flux across porous rocks. (a) Formation of temperature gradients across porous rocks, heated by volcanic activity against the cold surrounding. Shallow temperature gradient settings can be found in various environments such as volcanic, water-vapour, or hydrothermal settings. (b) Simulations show that a difference in heat conductivity between the rock (olivine-melilitite $\kappa = 2 \text{ W mK}^{-1}$) and the water filled pore ($\kappa = 0.6 \text{ W mK}^{-1}$) 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.

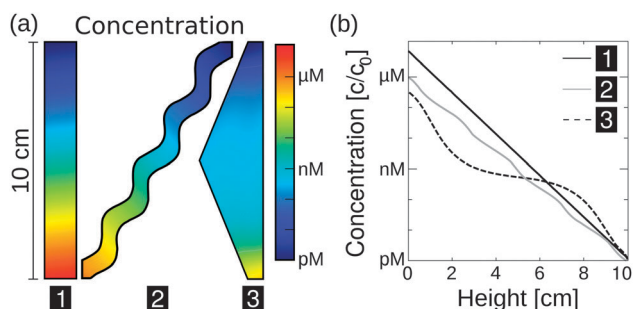


Fig. 3 Extreme accumulations of 100-mer oligonucleotides for various pore geometries via finite-element analysis, assuming shallow temperature gradients of 0.1 K mm^{-1} . Diffusion and Soret coefficients are based on experimental data.^{40,47} (a) The accumulation mechanism is found to be highly robust 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^+ ($D = 1029 \mu\text{m}^2 \text{ s}^{-1}$, $S_T = 0.0007 \text{ K}^{-1}$) and divalent ions such as Mg^{2+} ($D = 706 \mu\text{m}^2 \text{ s}^{-1}$, $S_T = 0.012 \text{ K}^{-1}$) and Ca^{2+} ($D = 792 \mu\text{m}^2 \text{ s}^{-1}$, $S_T = 0.013 \text{ K}^{-1}$).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 $\lambda_D = 0.79 \text{ nm}$) 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 ($\lambda_D = 2.1 \text{ nm}$) 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

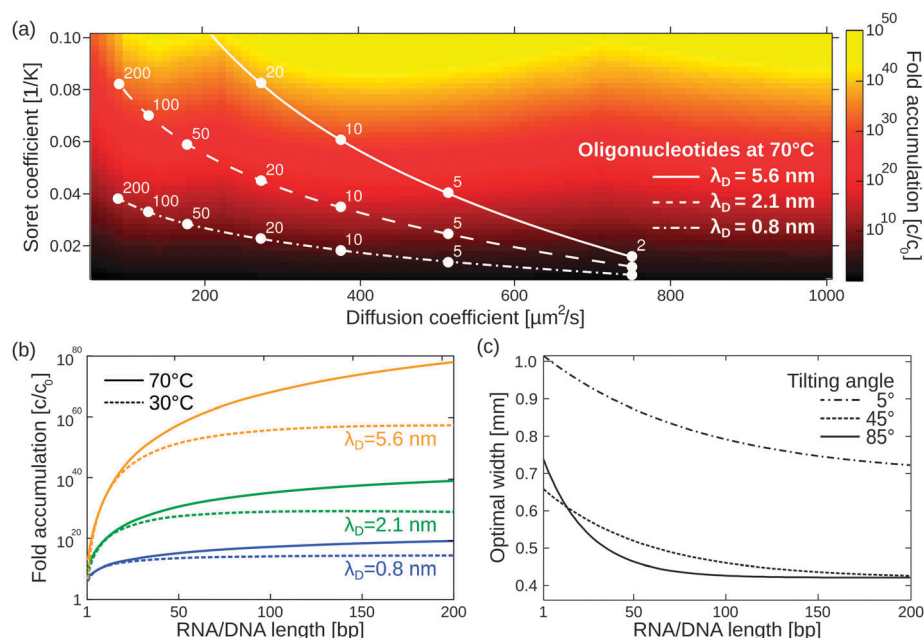


Fig. 4 Simulating exponential molecular accumulation. (a) Highly efficient accumulation of various molecules with a given diffusion coefficient and Soret coefficient. This allows the prediction of the accumulation for a very large range of molecules, including monovalent or divalent ions or single nucleotides. 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 $500 \mu\text{m}$, while longer oligonucleotides typically require pore widths below $500 \mu\text{m}$. Tilting the pore from the vertical case (90°) to almost horizontal case (5°) results in an increase in optimal pore width.

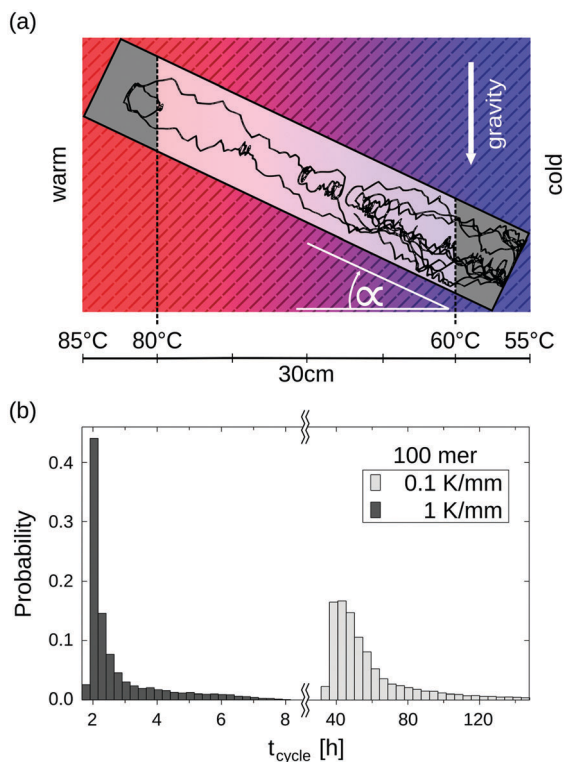


Fig. 5 Temperature cycle statistics of individual molecules derived by numerical simulations. (a) The trajectory of a 100-mer oligonucleotide inside an elongated, 45° tilted pore within a 0.1 K mm⁻¹ temperature gradient. A convective flow profile with a maximum velocity of $\sim 2 \mu\text{m s}^{-1}$ 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 $\lambda_D = 2.1 \text{ nm}$ are in the range of $\sim 80 \text{ }^\circ\text{C}$, 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 Nano-systems, 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. Kristjánsson and V. Marteinnsson, *Clays Clay Miner.*, 2002, **50**, 174–185.
- 45 P. M. Herzog 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.

Thermophoresis in Nanoliter Droplets to Quantify Aptamer Binding**

Susanne A. I. Seidel, Niklas A. Markwardt, Simon A. Lanzmich, and Dieter Braun*

Abstract: Biomolecule interactions are central to pharmacology and diagnostics. These interactions can be quantified by thermophoresis, the directed molecule movement along a temperature gradient. It is sensitive to binding induced changes in size, charge, or conformation. Established capillary measurements require at least 0.5 μL per sample. We cut down sample consumption by a factor of 50, using 10 nL droplets produced with acoustic droplet robotics (Labcyte). Droplets were stabilized in an oil–surfactant mix and locally heated with an IR laser. Temperature increase, Marangoni flow, and concentration distribution were analyzed by fluorescence microscopy and numerical simulation. In 10 nL droplets, we quantified AMP–aptamer affinity, cooperativity, and buffer dependence. Miniaturization and the 1536-well plate format make the method high-throughput and automation friendly. This promotes innovative applications for diagnostic assays in human serum or label-free drug discovery screening.

Molecular recognition is not only central to cell signaling, but it also represents the functional principle of pharmaceuticals and laboratory diagnostics. A variety of opportunities thus comes along with an in-depth understanding of biological binding events. From this perspective, it is not surprising to see an ever-growing interest in quantitative biomolecule interaction analysis. To this end, the directed movement of molecules along a temperature gradient, referred to as thermophoresis,^[1] has been successfully utilized in the last years.^[2,3] It is highly sensitive to molecular size, charge, and conformation. Based on binding induced changes in at least one of these parameters, affinity and concentration can be quantified, even in complex bioliquids.^[4]

In the well-established microscale thermophoresis (MST) approach, samples are measured in glass capillaries. Capillary MST has been applied for ions, small molecules, nucleic acids,

peptides, proteins, crude cell lysate, and untreated human blood serum.^[4–6] With circa 0.5 μL per capillary filling, the sample consumption is low compared to, for example, isothermal titration calorimetry.^[7] However, the actual measurement volume is significantly smaller: it lies in the range of 2 nL.^[2] The additionally consumed volume becomes essential when working with expensive or rare material such as patient samples. This is especially true if high-throughput analyses need to be performed, for instance, in diagnostics or drug discovery. Throughput and automation of conventional MST are further limited by the complicated handling of glass capillaries.

Therefore, we developed a capillary-free approach to measure thermophoresis in nL droplets under an oil–surfactant layer inside 1536-well plates (Figure 1). The water-in-oil system was experimentally characterized for temperature induced effects. The findings agreed with numerical simulations.

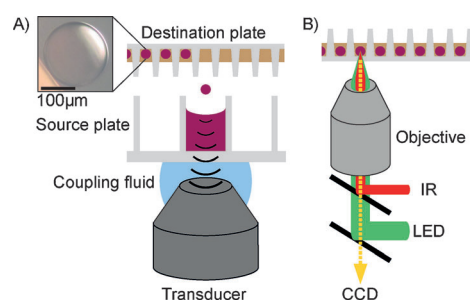


Figure 1. A) Droplet production. The liquid handler positions a destination plate above a source plate with a sample stock (purple). A transducer emits an acoustic pulse focused to the sample surface, whereby a 2.5 nL droplet travels into the destination well. To prevent evaporation, droplets are transferred into an oil–surfactant mix (brown). Inset: Samples were stable for several hours. 5 nL of 1:1 human serum/PBS. B) Inverted microscopic setup. The droplet center is heated with an IR laser. Thermophoresis is monitored by fluorescence (LED: light emitting diode; CCD: charge-coupled device camera).

The applicability of the system for biomolecule interaction studies was evaluated with a well-described nucleic acid aptamer. Aptamers were discovered more than 20 years ago.^[8] Owing to their three-dimensional conformation, these single-stranded oligonucleotides bind to various biomedically relevant targets, including proteins and small molecules.^[9,10] Just like antibodies, aptamers show high specificity and affinity. At the same time, these nucleic acid based ligands are superior to protein based ligands in production costs, storage conditions, and chemical modifiability.^[10] In vivo, their small size facilitates good delivery to the target tissue,

[*] S. A. I. Seidel, N. A. Markwardt, S. A. Lanzmich, Prof. D. Braun Systems Biophysics, Physics Department, NanoSystems Initiative Munich and Center for Nanoscience Ludwig-Maximilians-University Munich Amalienstrasse 54, 80799 Munich (Germany) E-mail: dieter.braun@lmu.de Homepage: <http://www.biosystems.physik.lmu.de>

[**] Financial support through a joint grant (BR2152/2-1) and project A4 within SFB 1032 from the Deutsche Forschungsgemeinschaft (DFG), by the Center for NanoScience (CeNS), and by the Nanosystems Initiative Munich (NIM) is gratefully acknowledged. The authors would like to thank Maximilian Weitz from the group of Friedrich C. Simmel for sharing his knowledge on microemulsions, Georg C. Urtel for support in building the setup and Christof B. Mast for programming support.

Supporting information for this article (including experimental details in Chapter 1) is available on the WWW under <http://dx.doi.org/10.1002/anie.201402514>.

whereas no immunogenicity and low toxicity have been reported.^[10,11] These benefits and the first marketed aptamer drug demonstrate the high potential of aptamers.^[12]

Aptamer binding studies were miniaturized employing a non-contact liquid handling system available commercially (Labcyte). The system delivers 2.5 nL portions from multi-well source plates into destination plates by acoustic droplet ejection (Figure 1 A).^[13] The deviation from the target volume is less than 2% (Supporting Information, Chapter S3a). To prevent evaporation, droplets were transferred into a protective layer of standard microbiology mineral oil supplemented with a surfactant mix according to Tawfik and Griffiths.^[14] For the presented experiments, we transferred four or eight 2.5 nL portions to yield 10 nL (270 μm) or 20 nL samples (340 μm). The positional accuracy of the transfer was reduced owing to deflection by the oil. To coalesce individual portions, destination plates with funnel-shaped wells were mildly centrifuged after transfer (≤ 500 g to avoid droplet damage). With our optimized procedure, we reproducibly obtained nL samples that were stable for several hours (Figure 1 A, inset). This allowed for multiple thermophoretic binding assays (10 min each).

Droplets were measured on a newly constructed microscopic setup (Figure 1 B). Similar to the previously described capillary instrument,^[2,5] thermophoresis was induced and analyzed all-optically. As an essential modification to the capillary setup, an inverted configuration was chosen so that the sample plate stayed upright to avoid oil dripping. While fixing the plate guaranteed that the droplets stayed in place, moving the optical parts allowed sequential measurements.

Before studying biomolecule affinity, we characterized the effects of local heating on aqueous nL droplets under oil. If asymmetrically applied, heating occasionally led to convective flows strong enough to move an entire droplet away from the laser spot. This was prevented by using plates with a small well floor area ($r=0.45$ mm). Utilizing the temperature dependence of the fluorescent dye Alexa 647, the radial temperature profile in the central horizontal plane of a 20 nL droplet was obtained 0.2 s after the IR laser had been turned on (Figure 2 A). For a temperature increase of $\Delta T_c = 11$ K in the heat spot center, the droplet periphery warmed up by $\Delta T_p = 4$ K. A Lorentz fit revealed an FWHM of 120 μm . In

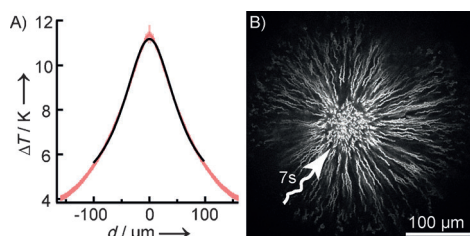


Figure 2. Local heating of 20 nL droplets. A) Radial temperature profile in the central horizontal plane (red). The temperature increased by $\Delta T_c = 11$ K in the center and by $\Delta T_p = 4$ K in the droplet periphery. A Lorentz fit (black) revealed FWHM = 120 μm . B) Flow profile of fluorescent polystyrene beads ($d = 1.0$ μm) integrated over 7 s during heating ($\Delta T = 15$ K). The beads moved toward the heat spot and out of focus with a peak velocity of 15 $\mu\text{m s}^{-1}$.

the following, ΔT denotes the average temperature increase of the central (30 \times 30) μm area.

Convective flows inside 20 nL samples were visualized with fluorescent polystyrene beads. Figure 2 B is integrated over 7 s of heating. The beads moved toward the central heat spot and out of focus, with peak velocities of 5–10 $\mu\text{m s}^{-1}$ for $\Delta T = 6$ K and 15 $\mu\text{m s}^{-1}$ for $\Delta T = 15$ K. To elucidate these flows, we performed full numerical simulations considering diffusion, convection, thermophoresis, and the temperature dependence of the dye. Simulations of 20 nL (Figure 3) and

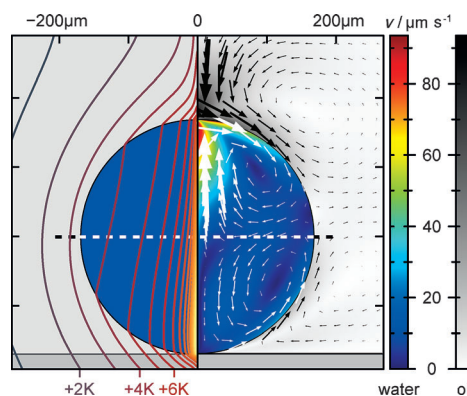


Figure 3. Numerical simulation of temperature and flow fields in a vertical cut through a 20 nL droplet after 0.2 s of heating. Left: Isotherms indicate the temperature increase. Right: The central horizontal plane (dashed) comprises the boundary of two toroidal flow vortices. The vortices are driven by Marangoni convection at the water–oil interface and have already reached the steady state.

10 nL droplets (Supporting Information, Figure S1) verified that the observed inward flow can be explained by Marangoni convection. This type of convection is caused by temperature-induced differences in interfacial tension. In our case, local heating decreased the interfacial tension between water and oil at the top and bottom of the droplet, triggering Marangoni fluid flow along the interface. Owing to the cylindrical symmetry, toroidal vortices arose in the upper and lower droplet hemisphere. Figure 3 shows the cross-sections of the tori in a vertical cut. The dashed line indicates the horizontal plane. Here, the flow is directed inward in the upper and lower vortex, which agrees with the experimental observation in this plane (Figure 2).

After flow field analysis, we recorded fluorescence time traces, the basis for our binding measurements, in 20 nL Alexa 647 samples (Figure 4 A). The experimental curves were highly reproducible and confirmed by simulation. A series of different events was identified in agreement with standard capillary measurements. When the heating laser was turned on, the fluorescence decreased owing to the temperature response (DTR) of the dye and thermophoretic molecule depletion. Thermophoresis and back-diffusion equilibrated within seconds. Subsequent slow warming of the entire sample slightly reduced the dye fluorescence intensity, but did not affect the measurement. When heating was turned off, fluorescence recovered owing to DTR and back-diffusion. A larger ΔT enhanced DTR and thermophoresis in experiment

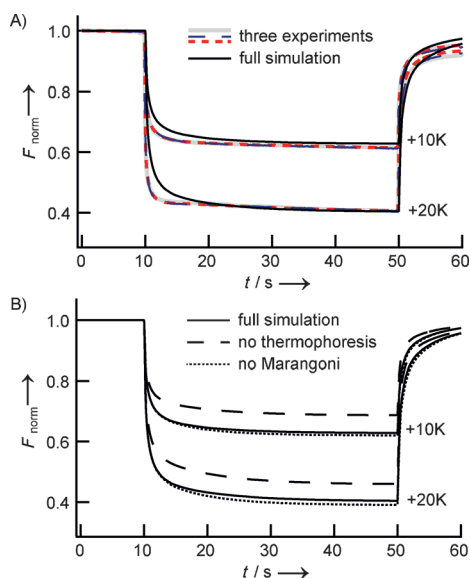


Figure 4. Fluorescence time traces from 20 nL droplets. A) Measurements of three Alexa 647 samples (gray, blue, red) overlap with minor deviations demonstrating the low batch-to-batch variation. Experiments and simulation (black) agree well. After the IR laser is turned on ($t=10$ s), the fluorescence decreases due to the temperature response (DTR) of the dye and thermophoresis. Thermophoresis and back-diffusion equilibrate within seconds. After heating ($t=50$ s), F_{norm} recovers due to DTR and isothermal back-diffusion. A larger ΔT enhances DTR and thermophoresis. B) Simulated contributions to the decrease in F_{norm} . Omitting Marangoni convection led to a negligible change of 0.008 (dotted); omitting thermophoresis changed the signal by 0.06 (dashed).

and simulation (Figure 4A). To assess the contribution of Marangoni convection and thermophoresis, simulations excluding either effect were performed (Figure 4B; implementation details are given in the Supporting Information, Chapter S2a). When neglecting Marangoni convection, the flow fields differed considerably, but the fluorescence signal was only slightly altered. Upon removal of thermophoresis from the simulation, the time traces changed significantly. This demonstrates that thermophoresis prevailed against the convective flows.

Having characterized thermophoresis in nL droplets under oil, we evaluated its applicability for biomolecule interaction studies. We analyzed a 25 mer DNA aptamer that binds adenosine and its phosphorylated analogues.^[15] This aptamer has previously been studied extensively.^[2,16] For nL-scale interaction studies, a constant concentration of fluorescently labeled aptamer ($c=2 \mu\text{M}$) was added to a serial dilution of adenosine-5'-monophosphate (AMP). As mentioned above, mild centrifugation in the funnel-shaped wells reliably coalesced individual AMP and aptamer portions. After coalescence, the concentration of AMP and aptamer equilibrated by diffusion. The short diffusion times through the small 10 nL or 20 nL samples guaranteed complete mixing within minutes. We found diffusive mixing to be as effective as manual premixing.

The mixed samples were locally heated by $\Delta T=6$ K. The resultant thermophoretic depletion of free aptamer significantly differed from its bound complex with AMP (Support-

ing Information, Figure S2). Furthermore, the temperature response of the aptamer dye (DTR) changed upon AMP binding.

The fluorescence after DTR and thermophoresis was divided by the fluorescence before heating as described in the Supporting Information, Figure S2 and previously.^[6] As this relative fluorescence can be approximated as linear to the bound aptamer fraction, it was directly fit to the Hill equation (Supporting Information, Chapter S3c).

Using the original selection buffer according to Huizenga and Szostak,^[15] we found $EC_{50}=(116 \pm 14) \mu\text{M}$ in 10 nL samples and $EC_{50}=(104 \pm 10) \mu\text{M}$ in 20 nL samples (Figure 5A). Both values agree with each other and the literature value of $(87 \pm 5) \mu\text{M}$ from capillary thermophoresis.^[2] The determined Hill coefficients of $n=1.2 \pm 0.1$ (10 nL) and $n=1.9 \pm 0.3$ (20 nL) indicate cooperative binding of more than one AMP, which is consistent with the previously reported tertiary structure of the complex (Figure 5, inset).^[17] Moreover, the Hill coefficients only slightly deviate from each other and confirm the literature value ($n=1.4$).^[2] As a control, we measured a DNA oligonucleotide with the same length as the aptamer but two point mutations. The dinucleotide

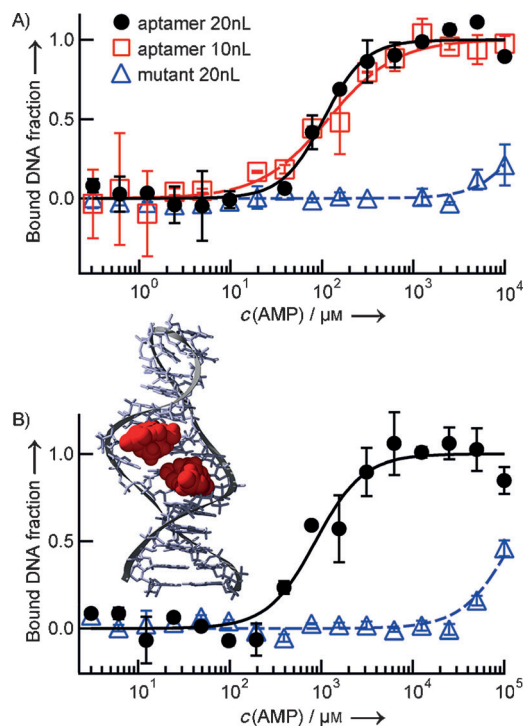


Figure 5. The specific signal change in DTR and thermophoresis upon AMP titration to labeled aptamer was fit to the Hill equation. Mean values of at least two individual nL samples; error bars: standard deviation. A) Selection buffer. The fit revealed $EC_{50}=(116 \pm 14) \mu\text{M}$ and $n=1.9 \pm 0.3$ in 10 nL (red squares) and $EC_{50}=(104 \pm 10) \mu\text{M}$ and $n=1.2 \pm 0.1$ in 20 nL (black circles). A dinucleotide mutant showed a 200-fold increased EC_{50} value of about 20 μM (blue triangles). B) PBS. $EC_{50}=(0.90 \pm 0.13) \text{mM}$ was found (black circles), confirming the buffer dependence of the aptamer ($n=1.6 \pm 0.4$). The mutant showed a 130-fold increased EC_{50} value of about 0.12 M (blue triangles). Inset: Determined Hill coefficients agree with the reported tertiary structure (NDB code 1AW4): an aptamer (gray) binds two AMP molecules (red).^[17]

mutant showed a 200-fold reduced AMP-affinity ($EC_{50} \approx 20$ mM). This demonstrates the specificity of the binding signal.

To quantify the reported buffer dependence of the AMP-aptamer,^[2] binding was measured in PBS (Figure 5B). An EC_{50} of (0.90 ± 0.13) mM was found, corresponding to a 10-fold affinity reduction compared to selection buffer. This reduction is not surprising, as the aptamer has originally been evolved in and thus optimized for its selection buffer.^[15,18] A dominant effect can most likely be ascribed to magnesium ions: while the selection buffer contained 5 mM $MgCl_2$, we used PBS without Mg^{2+} . Mg^{2+} do not only stabilize DNA, but can also neutralize AMP phosphate group and thus reduce repulsion to phosphates in the aptamer backbone.^[19] A reduction of the $MgCl_2$ concentration from 5 to 0 mM has been reported to significantly reduce AMP-aptamer retention in affinity chromatography.^[16] This is in accordance with the EC_{50} differences that we found in nL-thermophoresis. The Hill coefficient was not significantly affected by the buffer; it was $n = 1.6 \pm 0.4$ in PBS. The affinity of the mutant control was reduced 130-fold compared to the aptamer ($EC_{50} \approx 0.12$ M).

The successful quantification of affinity, cooperativity, and buffer dependence confirms the applicability of the presented method for aptamer analysis. This type of study is most likely to gain in importance now that the comprehensive aptamer patent portfolio, which presumably has suppressed many commercial applications, is starting to expire.^[20] Furthermore, nL-thermophoresis is a highly attractive analytical method for other biomolecules including peptides or proteins, and for complex bioliquids such as blood. The suitability for these studies remains to be tested, but can be expected judging from the application depth of capillary thermophoresis.^[4-6] Sample preparation is unlikely to be limiting, as the liquid handler can be deployed for various solution types. We, for example, produced stable nL droplets of 50% human blood serum (Figure 1 A, inset) as required for thermophoretic diagnostics. Diffusive mixing after nL transfer was successful. Therefore, an assay design in which a stock dilution series of a biomolecule target is tested against a high number of binding partners seems very practical, for example, for drug discovery. It could also be combined with our previously published diagnostic autocompetition approach.^[4] A stock dilution of an unlabeled tracer for the biomarker of interest would then be tested against multiple patient sera, supplemented with a constant amount of labeled tracer.

Compared to conventional capillary thermophoresis, the volume was reduced 50-fold. This leads to an enormous potential for high-throughput screens, even more so, as the easy-to-handle multi-well plates promote automation.

As a further advantage, the nL transfer is contact-free, which exempts from washing steps and minimizes cross-contaminations. After transfer, the sample is not in direct contact with the well surface, but forms a surfactant surrounded droplet inside the oil. This should significantly reduce nonspecific surface adhesion of biomolecules ("sticking"), an often encountered challenge in capillary thermo-

phoresis.^[6] The elimination of sticking represents a major benefit, even if surfactant and oil might have to be optimized for different sample types.

Considering these advantages, the miniaturization, and the extensive characterization in experiment and simulation, nL droplet thermophoresis promises diverse applications throughout the life sciences.

Received: February 17, 2014

Revised: April 11, 2014

Published online: June 4, 2014

Keywords: analytical methods · binding affinity · high-throughput screening · nanoliter thermophoresis · numerical simulation

- [1] a) C. Ludwig, *Sitzungsber. Akad. Wiss. Wien Math.-Naturwiss. Kl.* **1856**, 539; b) S. Duhr, D. Braun, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19678–19682.
- [2] P. Baaske, C. J. Wienken, P. Reineck, S. Duhr, D. Braun, *Angew. Chem.* **2010**, *122*, 2286–2290; *Angew. Chem. Int. Ed.* **2010**, *49*, 2238–2241.
- [3] a) C. J. Wienken, P. Baaske, U. Rothbauer, D. Braun, S. Duhr, *Nat. Commun.* **2010**, *1*, 100; b) L. C. Hinkofer, S. A. I. Seidel, B. Korkmaz, F. Silva, A. M. Hummel, D. Braun, D. E. Jenne, U. Specks, *J. Biol. Chem.* **2013**, *288*, 26635–26648.
- [4] S. Lippok, S. A. I. Seidel, S. Duhr, K. Uhland, H.-P. Holthoff, D. Jenne, D. Braun, *Anal. Chem.* **2012**, *84*, 3523–3530.
- [5] S. A. I. Seidel, C. J. Wienken, S. Geissler, M. Jerabek-Willemsen, S. Duhr, A. Reiter, D. Trauner, D. Braun, P. Baaske, *Angew. Chem.* **2012**, *124*, 10810–10814; *Angew. Chem. Int. Ed.* **2012**, *51*, 10656–10659.
- [6] S. A. I. Seidel, P. M. Dijkman, W. A. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, J. S. Joseph, P. Srinivasan, P. Baaske, A. Simeonov, I. Katritch, F. A. Melo, J. E. Ladbury, G. Schreiber, A. Watts, D. Braun, S. Duhr, *Methods* **2013**, *59*, 301–315.
- [7] T. Wiseman, S. Williston, J. F. Brandts, L.-N. Lin, *Anal. Biochem.* **1989**, *179*, 131–137.
- [8] a) C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–510; b) A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
- [9] A. D. Keefe, S. Pai, A. Ellington, *Nat. Rev. Drug Discovery* **2010**, *9*, 537–550.
- [10] A. Wochner, M. Menger, M. Rimmle, *Expert Opin. Drug Discovery* **2007**, *2*, 1205–1224.
- [11] R. S. Apte, M. Modi, H. Masonson, M. Patel, L. Whitfield, A. P. Adamis, *Ophthalmology* **2007**, *114*, 1702–1712.
- [12] E. S. Gragoudas, A. P. Adamis, E. T. Cunningham, M. Feinsod, D. R. Guyer, *N. Engl. J. Med.* **2004**, *351*, 2805–2816.
- [13] R. Ellson, M. Mutz, B. Browning, L. Lee, Jr., M. Miller, R. Papen, *JALA* **2003**, *8*, 29–34.
- [14] D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **1998**, *16*, 652–656.
- [15] D. E. Huizenga, J. W. Szostak, *Biochemistry* **1995**, *34*, 656–665.
- [16] Q. Deng, I. German, D. Buchanan, R. T. Kennedy, *Anal. Chem.* **2001**, *73*, 5415–5421.
- [17] C. H. Lin, D. J. Patel, *Chem. Biol.* **1997**, *4*, 817–832.
- [18] E. J. Cho, J.-W. Lee, A. D. Ellington, *Annu. Rev. Anal. Chem.* **2009**, *2*, 241–264.
- [19] H. Sigel, *Chem. Soc. Rev.* **1993**, *22*, 255–267.
- [20] P. Dua, S. Kim, D.-K. Lee, *Recent Pat. DNA Gene Sequences* **2008**, *2*, 172–186.