A tRNA world

Hubert Krammer

München, Mai 2012

A tRNA world

Hubert Krammer

Dissertation
an der Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Hubert Krammer
aus Traunstein

München, den 23.03.2012

Erstgutachter: Prof. Dieter Braun
Zweitgutachter: Prof. Ulrich Gerland
Tag der mündlichen Prüfung: 22. Mai 2012

Zusammenfassung

Lebende Zellen enthalten ein komplexes Netzwerk aus biochemischen Reaktionen. Die Bedeutung einzelner Reaktionsschritte hängt von ihrer Kinetik ab. Mit Hilfe der neuartigen, molekularen Lock-in Mikroskopie vergleichen wir die Kinetik der DNA-Hybridisierung in Zellen mit in vitro Messungen. Dabei wird die Temperatur der Zelle mit Hilfe eines IR-Lasers periodisch modifiziert, und die Antwort der Hybridisierungsreaktion mit fluoreszenten Methoden beobachtet. Der Fit an eine komplexwertige Transferfunktion erlaubt die Messung charakteristischer Zeitkonstanten.
Die biochemisch komplexe Umgebung innerhalb der Zelle kann die Kinetik nach komplizierten Mustern beschleunigen oder verlangsamen. Der Effekt des "molecular crowding" spielt dabei nur eine geringe Rolle. Vielmehr sind es spezifische Proteine zur Stabilisierung von einzelsträngiger oder doppelsträngiger DNA wie *Rad52* oder*ProteinA*, die die Hybridisierungskinetik massiv beeinflussen. Die Messungen zeigen große Unterschiede zwischen der Kinetik in Zellen und in Pufferlösung und machen so eine neue Beurteilung der Relevanz zahlreicher in der Literatur vorhandener in vitro Kinetik-Messungen für biologische Prozesse notwendig.

Die Auswirkungen von Temperaturgradienten in porösen Gesteinen auf einen Pool von RNA Molekülen zeigt ein Gillespie-Modell im Zusammenhang mit dem Beginn der Evolution. Mit einem Zufluss von Monomeren und einem längenabhängigen Ausfluss, einer zufälligen Polymerisation und einer Zersetzung, die vom Hybridisierungszustand abhängt, beschreibt das Modell die Längen, Sequenzen und Sekundärstrukturen eines RNA-Reservoirs. Neben dem vermehrten Auftreten von hairpin-artigen Sekundärstrukturen und der Verschiebung der Längenverteilung zu langen Strängen ist vor allem ein Bruch im Sequenzraum überraschend. So bleiben Motive, die als Doppelstrang vorhanden sind, deutlich länger erhalten als rechnerisch erwartet, was den statistischen Fingerabdruck einer langsamen Replikation hinterlässt. Der entstehende RNA-Pool stellt eine viel versprechende Grundlage für proto-enzymatische Hybridisierungsreaktionen dar.

Die RNA-Welt Hypothese setzt die Replikation von genetischer Information ohne katalytische Proteine voraus, deren Spuren in der modernen Biologie noch zu finden sein sollten. Ausgehend von einer tRNA-Sequenz entwickeln wir ein exponentielles, autokatalytisches Replikationsschema, das genetische Information über die Abfolge von zwei Sequenzteilen kopiert. Die von Temperaturoszillationen getriebene Reaktion basiert auf Hybridisierung und ist damit von chemischen Energiequellen unabhängig. Sie zeigt in Experimenten eine gute Selektivität und Exponentialität, und ist mit einer Verdopplungszeit von etwa 30s die schnellste proteinfreie Replikation in der Literatur. Das Schema lässt sich auf natürliche Weise zu längeren Abfolgen von Sequenzteilen erweitern und eröffnet eine Brücke in Richtung moderner Proteinsynthese, die einen evolutionär wichtigen, aber bisher unbekannten Schritt in Richtung moderne Biologie dastellt.

Abstract

Knowledge about the kinetics of chemical reactions in cells is important for an understanding of signaling pathways and regulation. Even though there are many kinetic measurements of in vitro reactions in literature, methods for in vivo measurements are sparse. With help of Temperature Oscillation Optical Lock-in (TOOL) microscopy we measure the kinetics of DNA hybridization inside cells and detect significant acceleration or deceleration compared to in vitro measurements, dependent on the DNA sample. The differences can not be explained by molecular crowding effects. Only models that take the background interactions with genomic DNA and RNA as well as the activity of single stranded and double stranded binding proteins into account, can be fitted to data. The results imply that the biological relevance of kinetic rates measured in vitro has to be rejudged carefully.

The RNA world hypothesis predicts catalytic molecules based on RNA, as for example early replicators, as precursor of modern biology. But how can a pool of appropriate RNA molecules arise under early earth conditions? In a Gillespie-model, we observe the length distribution, secondary structure and sequences of a pool of RNA molecules in porous rocks like they appear near sites of volcanic activity. We assume a monomer influx, a length dependent outflux, a random, non-templated polymerisation and a degradation that is much stronger for single stranded than for double stranded RNA. After equilibrium is reached, the pool is populated with many hairpin-like structures due to the selection pressure for hybridized strands that can be bricks for RNA machines.
Once sequence motifs and their complements appear in the reactor, they protect each other and are present longer than statistically expected. This "protection by hybridization" has the same fingerprint as a weak replication. As a consequence, the pool does not cover the full sequence space but includes more similar sequences, which is an important condition for chemical reactions.

Replication of genetic information by RNA molecules is considered to be a key process in the beginning of evolution. It is so crucial that traces of this early replication are expected to be present in key processes of modern biology. We present a replication scheme based on hairpins derived from the sequence of tRNA that replicates the genetic information about a succession of sequence snippets. The replication is driven by temperature oscillations as they occur naturally inside of porous rocks in presence of temperature gradients, and independent on external chemical energy sources. It is selective for correct information and shows exponential growth rates with doubling times in the range of seconds to minutes and is thereby the fastest early replicator in the literature.
The replication scheme can naturally be expanded to longer successions by using double hairpins derived from full tRNA sequences by only few mutations. By charging double hairpins with amino acids or peptides, the proposed replication bridges the gap from the RNA world to modern biology by offering a rudimentary translation mechanism, that sorts amino acids to chains according to genetic information.

Für Maria und...

die hohe Kraft der Wissenschaft
der ganzen Welt verborgen.
Und wer nicht denkt,
dem wird sie geschenkt,
er hat Sie ohne Sorgen.
Goethe (Faust)

Contents

Part I.

Hybridization kinetics in vivo and in vitro

1. Hybridization kinetics inside cells

1.1. The importance of reaction kinetics for biology

Biochemical networks, thermodynamics and kinetics

Living cells contain a huge and complex network of coupled chemical reactions that are driven by chemical gradients and controlled by a variety of regulatory and sensing mechanisms[1]. Research progress in biochemistry was achieved on the identification of single components in this complex network and in the understanding of their interaction. However, in order to understand the cell regulation, knowledge not only about the chemical components involved and the reaction network, but also about the kinetics of these chemical reactions is needed.

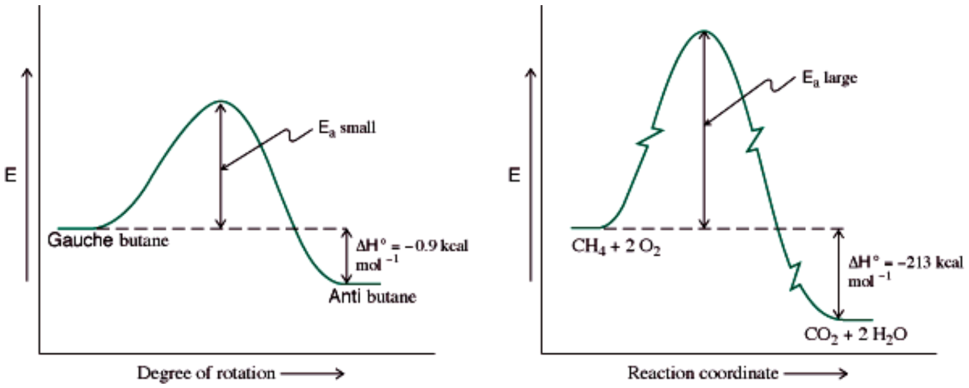

Figure 1.1.: Even though the combustion of Methane (right) is thermodynamically much more favorable, it is slower than the rotation in buthane (left). Kinetics is not so much dependent on ΔH but on the height of the energy barrier E_a. Figure taken from [42]

Thermodynamic features of chemical reactions do not by themselves determine the kinetic rates. The time needed to establish equilibrium is not so much defined by the difference in free energy and enthalpy, but rather by the activation energy barrier E_a. To illustrate this behavior, biochemical textbooks [42] compare for example the potential energy diagrams of a gauche-anti conversion (figure 1.1, left panel) in buthane with the combustion of methane (right panel). Even though the reaction in the right panel is thermodynamically more favorable, the reaction sketched on the left is much faster. The energy barrier E_a is lower and thus more molecules have enough energy to overcome it.
The hint towards the energy distribution already implies a dependence on temperature. The

[1] See for example the Roche Applied Science "Biochemical Pathways" wall charts

temperature dependence of the reaction rate is described by the Arrhenius equation:

$$K = A * e^{-E_a/RT} \tag{1.1}$$

Here, A stands for the maximal reaction rate that the reaction can have, in the case where all molecules have enough thermal energy to overcome the activation energy barrier E_a. With rising temperature, the exponent decreases and K approaches A. The reaction rate is defined in the law of mass action. For a reaction of 2nd order,

$$A + B \rightleftharpoons AB \tag{1.2}$$

with the rate k_{on} for association and the rate k_{off} for dissociation, the law of mass action defines the concentration dependence of the reaction constant K as:

$$K = \frac{[AB]}{[A] * [B]} \qquad and \qquad K = \frac{k_{on}}{k_{off}} \tag{1.3}$$

In vitro vs in vivo

Those formulas of classical chemistry are valid in 2-level systems in thermal and entropic equilibrium and in the absence of side reactions. Obviously, the dependencies in chemical reaction networks such as in a living cell are expected to be more complicated. Many partners participate in several reactions simultaneously, and side reactions do not allow a description with only two levels. The cell environment is inhomogeneous and non-equilibrium conditions drive all the crucial processes in a living cell. These arguments underline the importance of methods to measure kinetic rates in realistic conditions. Several studies of DNA-DNA and DNA-protein interactions have been done in vitro, in single molecule conditions, [39] in bulk solution and on surfaces,[61] with FCS[4] or AFM.[89] Most of these techniques were not used to measure reaction kinetics in a noisy and dirty surrounding like the intercellular space so far. Only few methods allow for a comparison of time constants in vitro and inside cells.[23] Therefore, the influence of the cell on kinetics is not determined and the validity of in vitro measurements for situations in living systems is questionable.

1.2. Tool Microscopy

Molecular Lockin

We use a fluorescence based molecular lock-in approach termed TOOL (Temperature Oscillation Optical Lock-in) microscopy[6] to determine the binding kinetics of short complementary DNA oligomers inside cells and compare it to kinetics in vitro.
Conventional Lock-in amplifiers multiply a periodically modulated measurement signal with a periodic reference signal of the same frequency (see figure 1.2).

The multiplied signal is integrated in a low pass integration. The result of the integration with different phase shifts allows for the analysis with help of a complex transfer function. It is thereby possible to siginificantly improve the signal-to-noise ratio.
In tool microscopy, the optical Lock-in approach is used.[6] The modulated measurement signal is caused by short double-stranded DNA that reports its binding state via the fluorescence of a fluorophore pair performing Förster-resonance energy transfer (FRET) in double strand

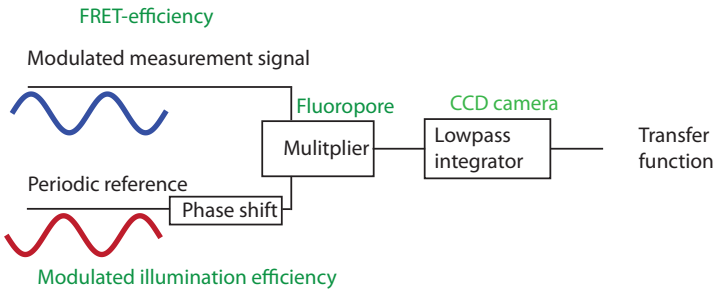

Figure 1.2.: A conventional lock-in amplifier multiplies a modulated measurement signal with a reference signal and integrates the result in a low-pass. In molecular Lock-In, a fluorophore multiplies the binding state dependent quantum efficiency with the illumination intensity. The emission intensity is integrated over the exposure time of a CCD-camera. In both cases, the result is analyzed by a complex transfer function.

conformation. Periodic temperature modulation caused by a defocussed infrared laser induces melting and annealing and therefore a modulation of the binding state and the quantum efficiency.
The modulated reference signal is the illumination intensity. The intensity is modulated so that it corresponds to an upper half of a sinus-function. The fluorophore multiplies the illumination intensity with the quantum efficiency which is determined by the binding state of the DNA strands, and reports the result by the intensity of fluorescent emission. A CCD camera with long exposure times serves as low pass integrator.

The setup is drafted in the left panel of Figure 1.3: An infrared laser heats the bottom of a cell culture chamber. The bottom of the chamber consists of a Si-chip, which is covered with a layer of Cr to absorb the laser light for thermal conduction. The metal is again covered with SiO_2 for compatibility with living cells and to provide an electrochemically inert surface for in vitro measurements. A $10\mu m$ spacer and a cover slip complete the chamber. Fluorescent readout is done through the cover slip.
The length and the binding energy of the dsDNA strand are chosen so that the melting temperatures are in a region where cells can survive many temperature cycles. (right panel of figure 5.7). The ends of the test strands are capped with L-enantiomeric bases to avoid degeneration by cellular DNA degradation mechanisms. To avoid strong binding to genomic DNA, a BLAST[2] search against the genome of the used HELA-cells was performed. For a FRET-pair, one strand is labeled with Rhodamine green while the complementary strand is labelled with 6-ROX (carboxy-X-rhodamine). Both labels are attached to the phosphate backbone to minimize degradation.

Fluorescence is recorded at different modulation frequencies. At slow frequencies, the reaction can follow the stimulus instantaneously. Amplitude is at its maximum and the phase is diminished. At higher frequencies, a phase delay between the stimulus and the reaction causes a diminished amplitude and a phase shift. The infrared laser intensity is modulated

[2]Basic Local Alignment Search Tool, available at `http://blast.ncbi.nlm.nih.gov/Blast.cgi`

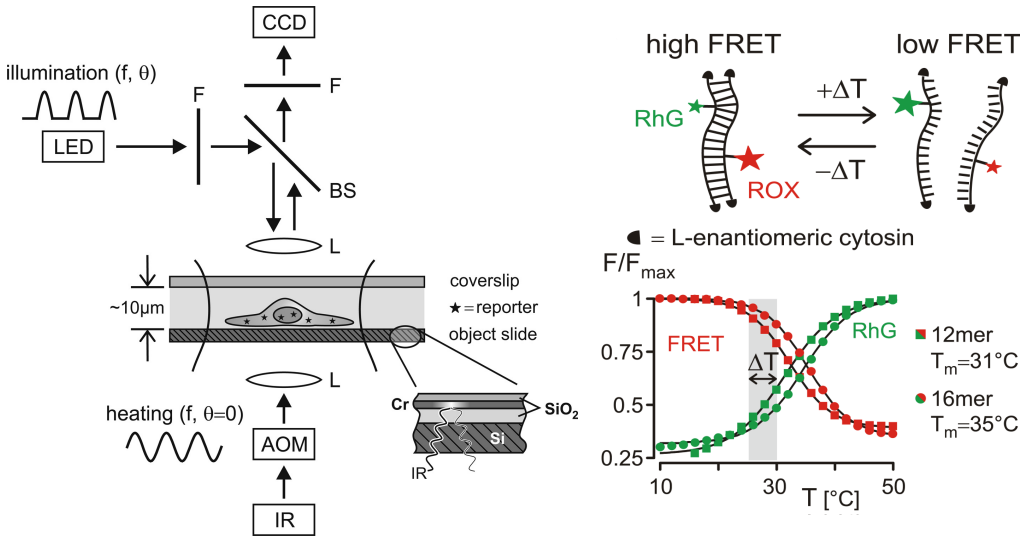

Figure 1.3.: Left: The setup used for TOOL microscopy. The temperature of a cell chamber is modulated by an IR-laser from the bottom, fluorescence is detected from the top. Right; A short DNA double strand is labeled with a FRET pair, strands end with L-enantiomeric bases to avoid degradation. Melting temperatures are in the physiologic range. Figure taken from [82].

with help of an acousto-optical modulator (AOM). Since the thermal relaxation of the cell chamber is fast compared to the measurement (see supplement of [82] in Appendix) the temperature follows the laser intensity. The modulation is synchronized with the illumination intensity that follows the shape of a half sine function. The synchronization includes also the recording with the CCD camera.
For a specific frequency, a series of pictures with different phase shifts between heating and illumination and one picture of the background illumination is recorded. The analytical complex valued transfer function is calculated from these pictures as given in [82]:

$$h(f) = \frac{4}{\pi} * \left(\frac{I_{0^\circ} - I_{180^\circ}}{I_{0^\circ} + I_{180^\circ} - 2I_{black}} + i \frac{I_{270^\circ} - I_{90^\circ}}{I_{270^\circ} + I_{90^\circ} - 2I_{black}} \right) \tag{1.4}$$

where index numbers represent the phase delay between temperature and illumination. Assuming small perturbations, the transfer function is fitted to:

$$h(f) = \frac{a}{1 + i * 2\pi f \tau} + b \tag{1.5}$$

where a represents the amplitude, τ the characteristic time constant, f the frequency and b an offset. The graphical representation as given in figure 1.4 allows a direct identification of these parameters. The transfer function is represented by a clockwise half circle in the complex plane. The amplitude is given by the diameter of the cycle, the offset b as an offset in the direction of $f \to \infty$ and the point of largest dissipation where the norm of the imaginary part has its maximum, is reached at a frequency of $f = (2\pi\tau)^{-1}$ and thereby yields the typical relaxation time of the reaction.

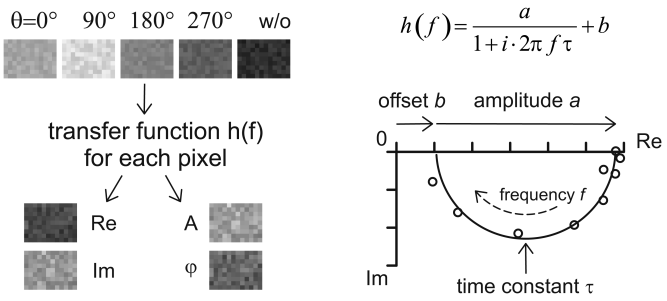

Figure 1.4.: Pictures with different phase delay are analyzed with help of a complex transfer function $h(f)$ reveal a characteristic time constant τ for the reaction. Figure taken from supplementary information of [82].

Spatially resolved data analysis

Since fluorescence is recorded with a CCD-camera providing spatial resolution and the measurement is sufficiently faster than cellular motion, a pixel-by-pixel analysis allows a comparison of kinetics in different compartments of the cell. In all graphs in [82] as well as in all figures of this chapter, we characterize the kinetic behavior by the characteristic time constant τ, which defined as

$$\tau^{-1} = \sqrt{k_{off} + k_{on} * [A] * [B]} \tag{1.6}$$

for a chemical reaction as described in equation 1.2, where k_{on} describes the association rate and k_{off} is the dissociation rate.

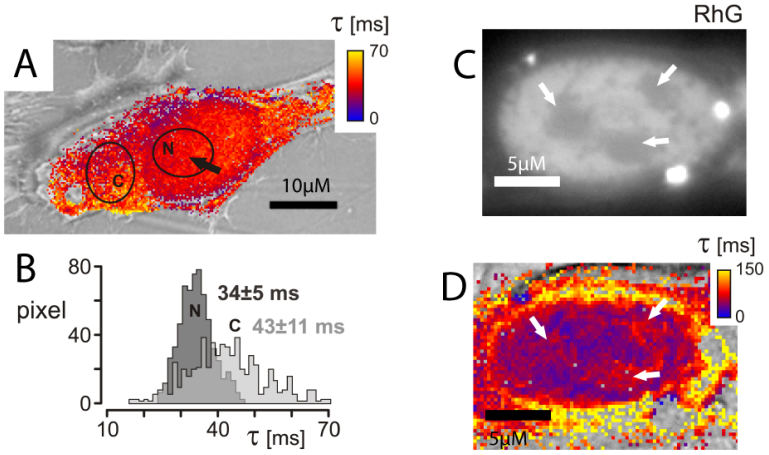

Figure 1.5.: Reaction Kinetics of dsDNA hybridization, spatially resolved in a living cell. A) Overlap of a brightfield-image of a cell and a pixel-by-pixel analysis of reaction Kinetics. B) a contrast between the cell nucleus is revealed by the histogram. C) Fluorescence of labeled DNA in the nucleus of a cell. Dark spots represent nucleoli (White arrows) D) At the nucleoli, hybridization speed is significantly slower. Figure taken from[82].

An overlay of a brightfield image and a kinetic map enables us to compare rates and concentrations between e.g. the nucleus of the cell and the cytosol. With additional help of an array scanner, that limits the focus depth of the light microscope, it is possible to image the nucleus by imaging the fluorescence of labelled dsDNA. As shown in the right side of 1.5, even the kinetic differences between nucleoli and the nucleus can be resolved.

1.3. Biologically relevant results

Modified Kinetics inside cells

The biologically most interesting result in [82] emerges from the comparison of in vitro vs in vivo data for the 12bp and 16bp dsDNA test samples. Not only that there is a significant difference between cells and bulk solution, but also a direction and size of the difference is probe-specific. The results are summed up in figure 1.6. Data from donor and acceptor signals as well as from the nucleus and the cytosole were fitted with equation 1.6 in order to obtain kinetic rates. While the 16bp DNA is about 7-fold faster in vivo ($k_{on} = 2,9 * 10^7 M^{-1}s^{-1}$) than in vitro ($k_{on} = 4,2 * 10^6 M^{-1}s^{-1}$), the 12bp DNA is about 5 fold slower in vivo ($k_{on} = 5,5 * 10^6 M^{-1}s^{-1}$) than in vitro ($k_{on} = 2,7 * 10^7 M^{-1}s^{-1}$). Interestingly, the in vivo data of the 12mer measurement could not be fitted sufficiently with equation 1.6.

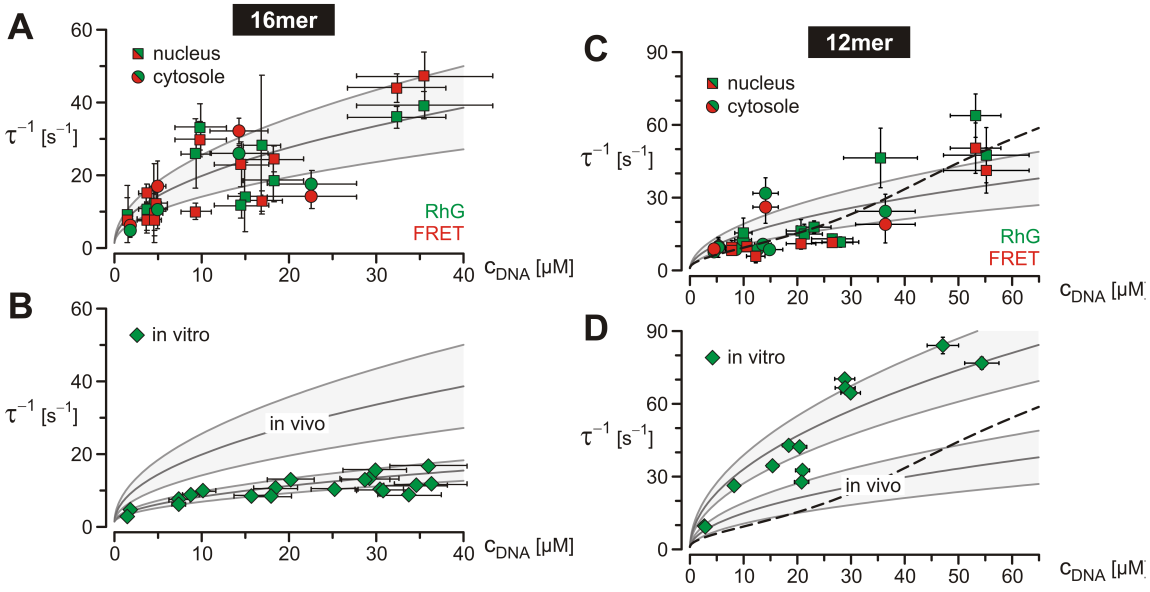

Figure 1.6.: Different Kinetics in vivo and in vitro for a 12bp and a 16bp DNA. Inverse time constants are plotted against concentration, solid fit lines represent reaction rates. The in vivo rate for the 16bp DNA is about 7-fold larger than in vitro, the in vivo rate for the 12bp DNA is about 5-fold lower. Figure taken from [82]

To explain this results, a much more detailed model accounting for the interactions between our sample DNA and genomic DNA and RNA as well as protein-DNA interactions, is needed.

For the interaction of the probe molecule with genomic nucleotides, a randomized genetic pool of the cells DNA and RNA would offer $4 \times 4^4 \approx 1000$ times more binding sites for a 12bp probe than for a 16bp probe. Assuming similar kinetic rates for labeled and unlabeled binding partners, these binding sites should ad to the probe concentration and speed up the reaction compared to the 16bp sample. However, in our measurements, we do not see this effect, but rather the opposite.
Proteins have highly specific tasks inside cells such as mediating hybridization or dehybridization of nucleotides as for example in DNA repair or transcription processes. Recombination mediator proteins as for example Rad52[65][90] catalyze the hybridization of DNA up to thousandfold reaction rates. Even though no oligos shorter than 15 bp were probed, the efficiency increases with length.[77][33] This effect may explain the acceleration of the 16bp probe. However it remains unclear whether such proteins are also present in the cytoplasm.
Other protein catalyzed processes can stabilize double stranded DNA or single stranded DNA and therefore slow down the reaction by blocking reaction partners. Examples are protein A for single stranded binding (ssb) proteins[104] or HMG-motif proteins for double strand binders (dsb).[8] A model that includes the protein activities was calculated and fitted to the measurements in dashed lines for the 12bp probe (see 1.6).

Molecular crowding

Living cells are densely packed with biomolecules. The effects on chemical reactions arising from the high concentrations of different molecule species are summed up under the loosely defined term "molecular crowding",[26] that is caused by the **excluded volume** that is occupied by macromolecules and the **limited diffusion** of reactants.

Excluded volume

To understand the effect of excluded volume, one can imagine a volume in a cell that is occupied by 30% with biomolecules. The center of mass of a newly added molecule can reach about 70% of the complete volume by diffusion. Nevertheless, a particle of a size comparable with the size of the molecules that are already there can only reach a small fraction of the volume that is left (see figure 1.7). With respect to chemical reactions, this leads to an effective concentration, that differs from the actual concentration of a molecule species. The ratio of effective concentration and actual concentration is termed activity coefficient.

The effect is strongly dependent on the size of the test particles in comparison to the biomolecules responsible for the crowding.[63][107]

Limited diffusion

To predict the influence of crowding on kinetic rates is hard. Reaction rates decrease with increasing crowding due to limited diffusion. However, for transition-state limited effects, reaction rate constants increase with increasing crowding.[108] The net result of those and other effects depends on the precise nature of the underlying process.
The effect of molecular crowding can experimentally be addressed with crowding agents. Those are organic macromolecules like branched dextranes or the spherical molecule Ficoll of various size and molecular mass.[26]
Interestingly, we found no significant influence of crowding agents on the hybridization kinetics of dsDNA in a series of measurements with dextranes of different molecular mass. Even

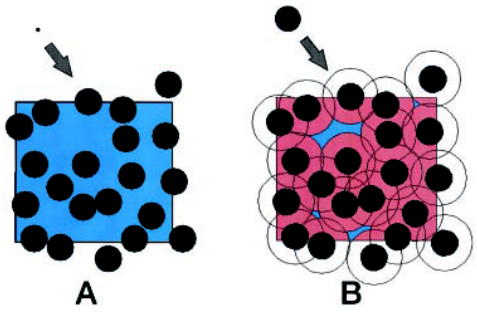

Figure 1.7.: Effect of excluded volume. A small particle added to a crowded volume can reach most of the unoccupied space, a particle with a size comparable to other biomolecules is excluded from most of the free space. Figure taken from[64].

though one can argue that the limited diffusion effect may be small due small size of the probe molecule compared to the crowding agents, this result is somehow surprising and is likely to give new input in theoretical considerations on molecular crowding.
The effect of divalent ions on hybridization stability and kinetics is well known[73] and of importance for any hybridization reaction such as DNA machines[106] or DNA Origami.[21] The increasing reaction speed with increasing concentration of $MgCl_2$ is demonstrated in figure1.8.

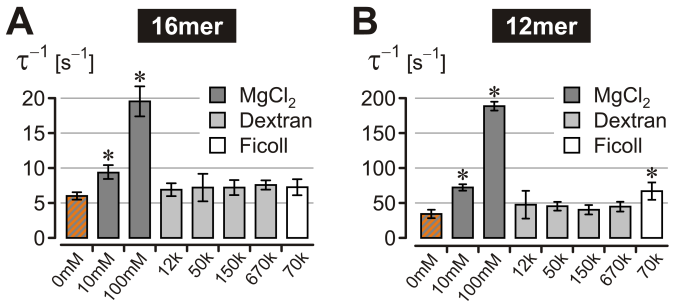

Figure 1.8.: Effect of divalent ions and crowding agents on hybridization kinetics for two different dsDNA samples. Bivalent ions speed up the reaction, whereas dextran does not change kinetics, Ficoll only shows a minor effect.

1.4. Outlook: Consequences for research in biochemistry

The results show that kinetic rates in living cells differ significantly from rates of the same sample in bulk solution. Furthermore, it is not possible to give a valid rule of thumb in which way both situations are connected, since both the kind and the strength of the influences that are responsible for the discrepancy are manifold, superimposed and not mutually independent. For research on kinetic rates of biomolecules this has two main consequences. Firstly, the biological relevance of kinetic rates measured in vitro, however under the usage of crowding agents, has to be judged very carefully from case to case. Secondly, there is an increasing

need for measurements of kinetic rates under in vivo conditions. Even though a couple of methods are available right now,[23] it would be desirable to increase these capabilities and to apply techniques to biologically and medically relevant questions. TOOL microscopy has the potential to address some of this questions, as for example the kinetics of DNA-protein-interactions and protein-protein interactions in the future.

A variety of reaction kinetics can be observed

TOOL microscopy allows the determination of kinetic rates for bimolecular biochemical reactions. In [82], hybridization and dehybridization of double stranded DNA was used as a typical representative reaction. However, the technique can approach a wide variety of reactions, which is contoured by the following conditions:

- A fluorescent readout with a quantum efficiency that depends on the binding state. In the experiment reported in [82], a FRET pair was used and the donor and acceptor fluorescence were recorded separately. This design allows several controls, that are necessary to prove reliability of the method. However in general, a single fluorescence signal is sufficient.

- Binding has to be temperature dependent. According to equation 1.1, any chemical reaction has a dependence on temperature. Nevertheless, for a detectable signal it is necessary to reversibly bind and release a significant fraction of molecules. In [82], a 16bp and 12bp double stranded DNA was used to reach a melting temperature that allows survival of cells. In general, other periodic stimuli than temperature are thinkable, e.g. light triggered reactions or oscillating chemical reactions. Stimuli have to cause reversible reactions of the tested system on a timescale that is similar to the kinetics in question.

- Reaction kinetics have to be in the kinetic range of the method. Since the dynamic range is not limited by the camera speed, typical timescales from $1\mu s$ to $1s$ are approachable. Compared with conventional techniques as FCS, Pump and Probe, temperature jump or time lapse methods, TOOL microscopy covers a dynamic range that is important for reactions of biologic interest.

Part II.

Chemical Evolution

2. The RNA world

The question about the origin of biology on earth is as old as human consciousness. Over centuries until today it is in the intersection of natural science, philosophy, religion and esoterica.[74] Even though the interest in this field is somehow academic, since it is impossible to finally prove the one or other scenario, the ambition to learn more about the mechanisms, that changed the prebiotic dead young earth to the green and blue planet it is today is so much part of human identity that an expanded motivation is almost unneeded. Notwithstanding many main questions in the field are of biological and biochemical nature, this thesis tries to ad new impulses from a physics point of view, arguing with terms like non-equilibrium, information, forces and fluxes, convective flows, thermophoresis and others.

Evolution

In order to simplify the following discussion, we define an evolutionary system as a system that is capable of **replication** with an appropriate **mutation** rate, and that is exposed to a **selection** pressure.[17] The evolutionary principle is nowadays widely accepted as the fundamental basis of biology. The approach to the origin of biology is therefore the search for a molecular system that is chemically plausible under early earth conditions and has the ability to replicate, mutate and withstand selection pressure. In modern biological systems, thereby the selection pressure approaches the phenotype, while the replication and mutation are at first hand limited to genomic information. This is important because the coupling between genotype and phenotype requires a highly evolved translation machinery, that is capable of replicating and reading genetic information and translate it into proteins with structure and function.
This replication and translation mechanism is based on a complex interplay between nucleotides, i.e. the carriers of information, and proteins, i.e. the carriers of catalytic functions. Proteins read and copy genetic information while the information about the assembly of this highly efficient catalysts is encoded in the sequence of nucleotides. In terms of evolution, this leads to the classical chicken-and-egg problem: Nucleotides without proteins to read them have no selective advantage, and proteins can only exist as soon as they are produced by a protein that can read nucleotide based information.

The RNA world Hypothesis

The solution of this chicken-and-egg-problem made a step forward with the discovery of ribozymes.[34][47] The groups of Altman and Cech independently found catalytic molecules or active subunits of molecules that consist of RNA and nevertheless not only carry sequential information, but also have complex biochemical catalytic function in different processes. Shortly thereafter, W. Gilbert came up with the idea of hypothetical ribozymes, that have the function of replicating genetic information, in the best case their own sequence.[31] Those molecular systems are called "early replicatiors". Even though also other researchers recognized the important role of RNA for biogenensis,[102][70][16] his classical paper formed

the term "RNA world" for a interdisciplinary field of research, in which scientists search for molecular candidates of early replicators as well as for models describing circumstances and properties of those molecules and of chemically plausible ways to gain such molecules under early earth conditions.

2.1. Early earth conditions

Geology of the archaeon

Only a minuscule part of the earth surface is composed of rocks old enough to give a geological or mineralogical record of the archaeon[1]. What we know about the geology of the Archaeon is that there was a hydrosphere and granitic an mafic components in the lithosphere.[67] Plate tectonics was present due to convective currents in the Astenosphere.[109] The presence of water tells us that the temperature of the earth surface has been relatively stable over at least 3.8 billions years, which is mainly an effect of green house effect, that arises from an CO_2 containing atmosphere.
The exact composition of the atmosphere in the archaeon is as well not confidently known,[30] but for sure has changed a lot during the same time scale. Oxygen was only a minor component before photosynthetic organisms started to produce it about 2.2 billion year ago. Until today it is assumed that the atmosphere was reducing and contained, besides nitrogen, small molecules such as CO, CO_2, CH_4, H_20, and NH_3. The absence of oxygen and an ozone-layer gives a hint towards high intensities of short wavelength-UV light from the sun on the surface, which is important in the context of nucleotide degradation.

Ingredients for the prebiotic soup

Most theories about the origin of biomolecules on early earth include a "prebiotic soup" that covered the earth.[69][35] This theory got experimental support from the Urey-Miller experiment[62] that produced simple amino acids from a mixture of CH_4, NH_3 and H_2 with heat and electric charges and follow-up experiments that produced further amino acids in different yields dependent on the assumptions about the composition of the atmosphere.[81] Protein synthesis in liquid phases has been also demonstrated.[28]
In [74], the discussion about the origin of nucleotides is summed up as *"The prebiotic chemistry of nucleotide bases is a contended question about experts."* Nevertheless, there exist different promising ideas. The synthesis of adenine as a crucial brick for many biochemical processes was already demonstrated in 1960,[71] and later also the synthesis of single nucleotide bases,[75] even though under improbably high concentrations of urea. Synthesis pathways based on formamide were presented from small molecules[80] to longer RNA polymers[14] in a promising chain of reactions with only a few gaps in it.
Whether the absorption on solid surfaces as it plays a role in volcanic rocks does accelerate reactions[13] by reduction of the concentration problem to two dimensions, or even decelerate due to increased degradation[66] is still under debate.

[1] $4400 - 3600Ma$ before today

Hot spots of early life: Hydrothermal vents

Which sites or places on early earth offer conditions suitable for chemical evolution? Besides others, hydrothermal vent systems show a broad variety of interesting chemical and physical properties in the context of the RNA world. They are found in all regions of the ocean[57] and arise in two different types:

Black and white

Black smokers arise in direct vicinity of magma chambers under the sea ground. They are driven by water that gets in close contact to the magma and reemerges at the vent sites. Water is typically acidic ($pH2-3$) and includes a significant amout of metallic ions such as $Fe(II)$, $Mn(II)$ as well as magmatic CO_2, H_2S and dissolved H_2 and CH_4. Temperatures appear in a wide range, from as hot as $400°C$ inside of the vent to as cold as the surrounding ocean $2°C$.[57] The appearance of black smokers is coupled to tectonic activity. Therefore their role for the origin of life is under debate, because they have comparably short lifetimes in the range of a couple of decades to centuries.
Most knowledge of **white smokers** is based on research on the Lost City Hydrothermal field (LCHF).[44][54] Those off-axis vents are located several kilometers off the spreading zone. Convection is driven by heat dissipation from the underlying mantle rocks and by exothermal chemical reactions. Chemistry is dominated by olivine, a mineral that is rich of iron and magnesium, and is the main product of serpentinization,[84] that forms chimneys of a porous consistence with a height of up to $60m$. The geological setting is highly alkaline, at $pH9-11$. White smokers are not directly coupled to tectonic activity and have much longer lifetimes. It is assumed that serpentinization was much more widespread in the Archaeon than nowadays, since its byproducts are found to an enormous extent on the sea ground.

Temperature gradients

Physically speaking, both types of vents consist of foamy, porous materials with a pore size in the μm to cm regime. Those pores are exposed to temperature gradients between the hot inside and the cold surrounding ocean and filled with water. This gradient is a non-equilibrium setting with physical implications such as convective flows and thermophoresis, that provide interesting clues to the riddles in the origin of life.
It is to mention that any pore, for example in basalt rocks as they are common near volcanic sites, empowers convective flows in combination with a temperature gradient. The assumptions about thermal convection and thermodiffusion made in [1], [46], [58] and others are not strictly limited to hydrothermal vents of any kind.

2.2. Riddles in the RNA world hypothesis

Within RNA world, there is a set of very general chemical and physical questions that have to be considered.

The Concentration Problem

As soon as there is a –probably inefficient– chemistry that produces RNA monomers or short polymers, the dilution and concentration problem comes into play. For entropic reasons, those molecules would get equally distributed in the premodial ocean over long timescales. To reach concentrations that enable relevant chemical reactions and the polymerization of longer sequences, effective concentration mechanisms are necessary. Entropy has to be reduced locally.
Besides the "warm little pond" proposed by Darwin[18], several theories describe ways to reduce entropy or to increase and maintain concentration, including evaporation of tidal pools,[18] adsorption to clays and surfaces,[29] or concentration in ice through melting and freezing processes.[94]
Hydrothermal vents provide spatially limited compartments in temperature gradients. The superposition of thermal convection and thermophoresis provides a highly efficient mechanism to concentrate molecules in a distinct corner of the pore. Simulations [1] and experiments [100]show that exponential trap efficiencies are dependent on aspect ratios. In analogy to similar processes in gaseous environment, this mechanism is referred to as "Clusius trap".

Encapsulation

Many definitions of life include encapsulation.[92] The definition of an inside and an outside gets crucial in the context of a metabolism that produces waste products. For molecular precursors of life, encapsulation is mainly necessary to avoid dilution. The appearance of a successful early replicator is therefore coupled to a meaningful compartmentalization, for example in lipid vesicles.[56] Experiments show that phospholipids in hydrothermal pores form vesicles mediated by the convective flows.[7] The pores of hydrothermal rocks supplies a compartmentalization at first hand. Furthermore, the high efficiency of the clusius trap especially for for longer or more complex molecules provides a "compartment without walls" that keeps molecules in the pores even though those are maybe leaky.
To escape the vent in the end, these borders have to be replaced by membranes similar to the phospholipid membranes of cells as we know them today, that has some kind of selective pores to enable a directed influx of nutrients and outflow of waste.

Nonequilibrium conditions

Life is a non-equilibrium system. In modern biology, gradients of protons or ions drive chemical reactions and signaling pathways. They are maintained by highly evolved proton pumps and membrane compartments in combination with pores that are selective for special ions. The chemical energy necessary is provided by metabolic processes.
Molecular systems as precursors of life lack semi-permeable membranes and an evolved metabolism, therefore some kind of gradient is necessary to keep reactions out of equilibrium. Hydrothermal activity provides gradients in ion concentration and in pH-value. Furthermore, hot vents in cold ocean water cause steep temperature gradients that can replace chemical gradients and are stable over the lifetime of a vent without any metabolism. The physical setting of porous compartments and temperature gradients represents non-equilibrium conditions that imply interesting perspectives on the concentration problem[1] and can thermally

drive replication mechanism similar to polymerase chain reaction. [58]

Chemical energy sources

Chemical energy sources such as ATP, activated nucleotide bases[97] or cofactors such as caffeine or theophylline [50] are only available in low concentrations if at all. So far, there is no consensus about a stable supply of early replication with chemical energy resources among experts.
Furthermore, a metabolism in the sense of modern biology produces side products that have to be selectively disposed without disposing the components of replication. Besides the "supply problem", chemical energy sources also rise the "waste problem". However, in replication reactions based on hybridization, the waste problem can be solved by reversible melting of side products at high temperature sites.

The Eigen paradox

In modern biology, genetic information is replicated by protein enzymes. The information about the amino-acid sequence of these proteins is encoded in genetic information. A classical chicken-and-egg problem arises when we ask the question how this interconnected system evolved, whichs often referred to as the Eigen-Paradox[25]: To fold a catalytically active protein, an amino-acid chain has to have a certain minimal length. However, without catalytically active proteins of that kind, it is not possible to read and replicate genetic information long enough to encode for such a protein with an error rate that is low enough to guarantee a high fidelity of the amino acid sequence. It is possible to derive an error-threshold from this paradox that defines the minimal fidelity of a replication and translation machinery that is not easily reached without protein driven proof-reading mechanisms.

Selective advantage of any intermediate step

As a fundamental consequence of evolution, any new step has to have a selective advantage to its precursor. The fitness with respect to the present selection pressure has to increase. In modern biology, mutation acts on the genotype level, while selection pressure attacks at the phenotype level, i.e. the translation of the genetic code into proteins. A protein with higher catalytic activity that arises due to a mutation in its genetic code can have a better fitness. For early replicators, the following question arises: what increases the fitness after a genetic mutation in the absence of a translation into proteins?
Early replicators have to withstand degradation by fast replication. Any mutation that increases the power or speed of replication increases fitness as well as anything that stabilizes against degradation, as for example the ability for Watson-Crick base pairing.

3. A model for an RNA reactor

How does an RNA pool in a hydrothermal pore system develop? What are the concentrations, and length distributions like? Furthermore, are secondary structures possible that have the potential to be bricks for ribozymes? And facing the huge sequence space of 4^n for an RNA molecule with n base pairs, is it probable to find enough molecules of one kind to perform any interesting reaction?
As soon as investigations concern the sequence space and bigger numbers of different sequences come into play, experiments get costly and time-consuming. For the analysis of inhomogeneous mixtures of short oligos with different sequences and a length distribution, we lack cost efficient methods so far.
Simulations can give a lot of insights and help to develop models with testable predictions. In cooperation with Ulrich Gerland and Benedikt Obermayer[1] a model for a pore was developed to have a closer look on the composition and structure of RNA molecules in a hydrothermal RNA reactor.

3.1. A hydrothermal pore model

The model considers an open volume in the part of a hydrothermal pore that has no borders and lies at the bottom, where the trapping mechanism according to [58] concentrates molecules. The pool of RNA molecules is modified by monomer influx, the outflow of molecules that is length dependent, random ligation and a cleavage that depends on the hybridization state. After all fluxes have reached equilibrium, length distribution, number of molecules, secondary structure and sequences are analyzed.

Monomer influx and ligation

The reactor is fueled with an influx of RNA monomers that are randomly chosen from the four nucleotides Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). Longer sequences are created by random ligation of monomers as well as ligation of oligos. It is crucial to mention that the ligation is neither template directed nor dependent on length or sequence of the ligating partners. Therefore it is completely unbiased in sequence and does not show replicator characteristics by itself.

Outflow and the Clusius trap

The outflow of molecules depends on the concentration of a species as well as on the length. The length dependence is a characteristic property of the clusius trap as described in [1] and has been experimentally tested in [100] and [58]. It arises in encapsulated pores inside a temperature gradient by an overlay of convective laminar flows and thermophoresis.

[1] Statistical and Biological Physics, Arnold Sommerfeld Center, LMU

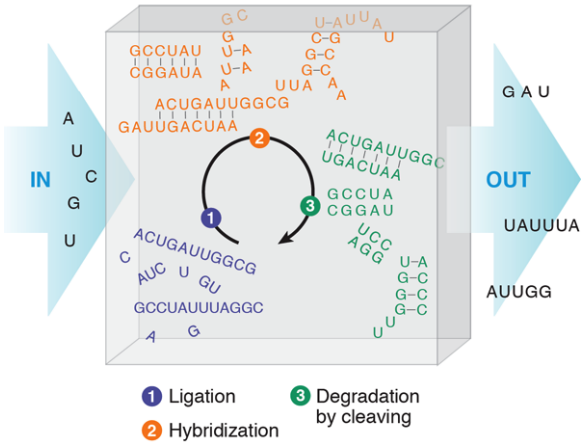

Figure 3.1.: Scheme of a hydrothermal RNA reactor: Influx of monomers, outflow of sequences according to characteristics of a hydrothermal trap, random polymerization and hybridization dependent binding modify pool of RNA molecules. Figure taken from [72].

It is expected that longer sequences are trapped more effectively in such pores. Nevertheless, the detailed description is dependent on temperature gradients and on the pore geometry as well as on Soret-coefficient of molecules and the ionic strength and composition of the solvent. Therefore in the model, length dependence is described in a term relative to a characteristic length L_C. The crossover scale L_C combines parameters such as the Soret coefficient, the trap geometry and the temperature difference across the trap. Because this complicated length dependence inhibits further analysis, it is set to $L_C \to \infty$ for the calculation of the steady state length distribution in the simulations.

Hybridization-dependent cleavage

RNA is exposed to different degradation processes in natural circumstances. For example, strong illumination with UV-light can connect single bases so that they are no longer able to form double strands, while the backbone stays intact.[83] RNA damage mechanisms like that are not considered in the model. It is reduced to damage processes that have an impact on the analysis, i.e. damages that break the backbone.
Furthermore, the cleavage rate is assumed to be dependent on the hybridization state of neighboring bases. It is known that double stranded RNA is more stable against degradation than single stranded RNA.[95] The model allows intermolecular hybridization as well as intramolecular hybridization. In terms of the evolutionary questions addressed with the model, the high cleavage rate of single stranded RNA presents a selection pressure for complementary sequences, including molecules that are able to form secondary structures. i.e. intramolecular hybridizations.

3.2. Simulation results

Four different scenarios

Four different scenarios werer simulated. Those scenarios were not chosen to be realistic under early earth conditions, but to illuminate the role of hybridization to the results. Low temperatures and high concentrations make hybridization favorable in case that complementary sequences are available. We compare four scenarios with increasing influence of hybridization:

- Without folding, (color code grey)
- At high temperatures ($60°C$) and low concentrations in the pM range (color code magenta)
- At low temperatures ($60°C$) and low concentrations in the pM range (color code light blue)
- At low temperatures ($60°C$) and high concentrations in the mM range (color code dark blue)

The color code is valid for all graphs that illustrate the simulation results.

Properties of the RNA pool

As a consequence of the selection pressure for hybridized bases, the length distribution shifts towards longer strands and bigger molecules. Since longer strand are created by ligation, the total number of molecules decreases.

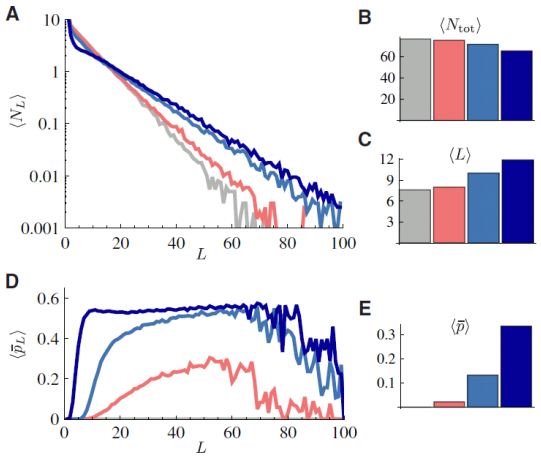

Figure 3.2.: Length distribution of molecules in the reactor. Scenarios with increased hybridization shift the length distribution towards higher values. Consequently, the total number of molecules is slightly lower. Figure taken from [68].

The four simulation scenarios are chosen to investigate the increasing selection pressure for hybridization and consequently for a lower degradation. In panel A of figure **??**, the length

distribution of the molecular pool is plotted. With increasing hybridization, the distribution is shifted to longer molecules. Panels B and C illustrate the consequent decrease of the total number of molecules and the increase of the mean length. It is to mention, that the number of molecules is small compared to the number of possible sequences so that the sequence space is not fully populated, even for short oligos. Panel D and E illuminate the binding probability dependent on length and the mean binding probability. With increasing length of a molecule, the probability to be part of a double helix increases. This behavior is even more pronounced for scenarios with higher selection pressure for hybridization.
In conclusion, the diminished cleavage rate for hybridized RNA leads to a selection of longer strand that are paired with their complements.

Metastable hairpins: fuel for hybridization driven reactions?

The analysis of secondary structures by means of the Vienna Package [37] reveals that secondary structures such as hairpins, double hairpins, hammerheads and others are very common in the reactor pool. In diluted solutions, RNA molecules with appropriate sequences tend to form hairpin loops. The persistence length of single stranded RNA as in the loop is as short as about $1-3nm$ or about $3-9bp$ dependent on the buffer conditions.[9] The hybridization dependent cleavage rate provides a strong selection pressure for sequences that are able to form such secondary structures.

Hairpin like RNA molecules can act as energy and entropy source for directed hybridiza-

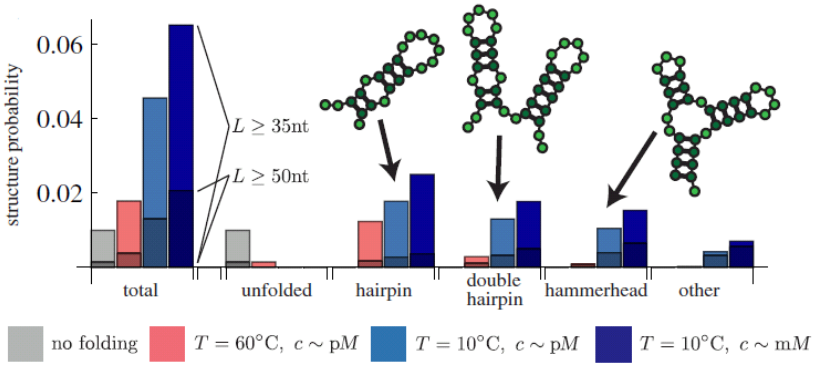

Figure 3.3.: Hybridization inside the reactor: Depending on temperature and concentration, a significant fraction of bases is in the hybridized state. RNA tends to fold secondary structures such as hairpins and double hairpins as consequence of a selection pressure for bound bases. Figures taken from [68].

tion reactions. Hairpin loops provide sterically trapped hybridization energy, since the single stranded bases in the loop can be transformed to double strands in structured processes.[32] In that sense they can act as source of chemical energy for RNA machines and early replicators. Since hairpins can open up and be closed by appropriate temperature cycles in case they are open or not properly hybridized, the general metabolic problem of waste production is avoided with hairpins as energy source. Entropy driven scenarios including hairpins similar to such known from the field of DNA machines are also thinkable.[106] Furthermore, RNA molecules playing key roles in modern biology include hairpin structures at crucial active

sites. Prominent examples are tRNA, the ribosomes catalytic site,[49] or splicosomes including RNAseP.[34] Most of these molecules are considered to be ancient and have ancestors in the RNA world scenario.

Broken Symmetry in the sequence space

The lifetime of sequence motifs in the reactor is extracted from simulations of all 4^l different motifs. Since the number of molecules is small compared to the combinatorially possible number of motifs, the sequence space is not fully covered, even for a short length of 3 base pairs. To gain suitably averaged observables, autocorrelations of sequence motifs as well as cross correlation functions of motifs and their complements are calculated and plotted in figure 3.2.
Taking a special motif out of the pool, the time development can be described by a death and

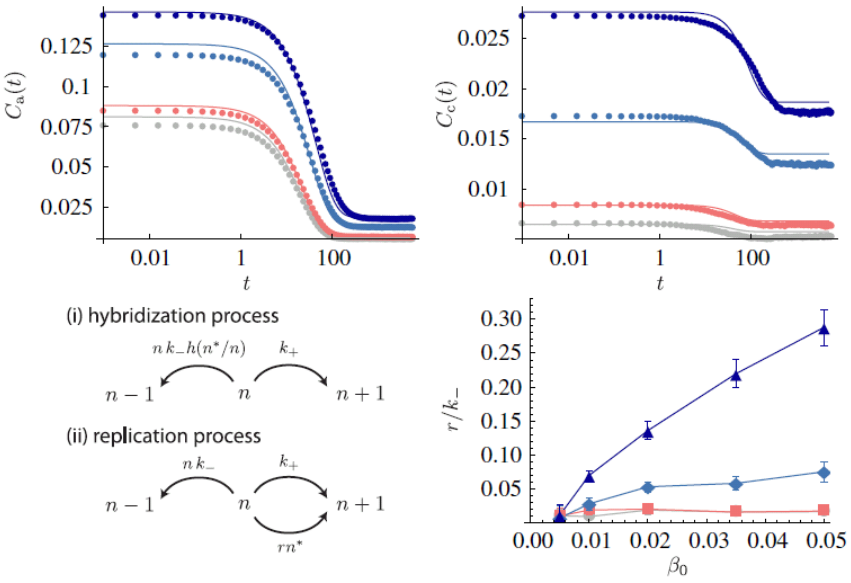

Figure 3.4.: Autocorrelation $C_a(t)$ and cross correlation $C_c(t)$ for simulation data with motif length of 6 base pairs together with analytical model. Figures taken from [68].

birth process with a birthrate k_+ and a death rate k_-. A usual process of this kind however does not take into account the protection of complementary sequences due to modified cleavage rates. Therefore the death rate is corrected by an additional factor $h(x) = 1 - (r/k_-) * x$ in which r accounts for the replication rate. The result of this modification is astonishingly equal to a replication scheme as shown in 3.2. Comparison of simulated cross catalysis results with analytical calculations allow to determine the replication rate.
A true replicator in the sense of template directed replication should show a $(r/k_-) << 1$, which is obviously not the case in our simulations. Nevertheless, the (r/k_-) we determine is higher than expected from a death birth process without protection by hybridization. The appearance of this *quasi replication* or *proto* replication has important implications:
The reactor is populated by related sequences. Sequence motifs that appeared once and found complementary partners stay longer in the reactor as statistically expected. Furthermore, if they are degraded in the end, the snippets still contain parts of the motif. New occurring mo-

tifs, generated from the old snippets, therefore tend to show similarities. By this mechanism, the sequence pool in the reactor will not cover the full sequence space, but rather stay in a corner of it and be populated with related sequences. Only the appearance of a hairpin pool with equal or related molecules in appreciable concentrations allows the discussion of RNA replication machines as described in [44][53] or[46](see Appendix and chapter 4 and 5).

3.3. Outlook: Experimental testability of the model

The predictions of the model about length distribution, secondary structures and sequence give new impulses for RNA world scenarios. Therefore it is of interest how far they can be tested by experimental means.

A hydrothermal pore in the lab

Hydrothermal traps can experimentally be addressed by means of glass capillaries in temperature gradients. Thereby convective flows can be driven by gravity [7] or mircrofluidic methods such as thermo viscous pumping.[58][100] To implement an influx and outflow of molecules with help of syringe pumps is possible and has already been realized in the lab. However, the volumes of such traps are small, so that it is a big effort to gain enough yield for analysis, for example with electrophoresis gels.
Polymerization of RNA to longer strands in the absence of ligation proteins has been reported under plausible conditions,[14] but at very slow rates and with so low yields that radioactive gel staining is necessary. One can think about protein mediated ligation processes with higher efficiencies as a model system.[45] Furthermore, other kinds of polymerization can be considered, for example short pieces of double stranded RNA molecules with sticky ends at both sides representing a monomer that is able to polymerize to longer chains.
Selective degradation under testable circumstances is hard to realize and has to be discussed and developed. Exposure to UV light or chemical hydrolysis might be one possibility under the condition that the mechanism of RNA degradation does lead to a backbone cleavage without destroying neighboring bases. Intense literature recherches and cooperations might be necessary to find a confident cleavage mechanism. Also, several kinds of RNase proteins could act as model system.[43]
The greatest experimental challenge will be to balance influx, outflow, ligation and degradation in a way that the steady state includes an ensemble of molecules that is good enough for detailed analysis.

Chemical analysis of the pool in the trap

The model described above makes predictions for length distribution, secondary structure and sequence space. To verify predictions on the length distribution, separation methods like high performance Liquid Chromatography (HPLC) may be an option. Due to the inhomogeneity and to the small peaces of RNA predicted, separation on high resolution electrophoresis gels will be hard to analyze with fluorescence methods due to small volumes and low concentrations. Even for radioactive staining methods as used in [14], high concentrations and yields are necessary.
Whenever predictions hit the sequence space, experiments get difficult and costly. Sequencing inhomogeneous mixtures of short oligos seem to be an experimental challenge. As long

as one manages to concentrate on short molecules and thereby on a small sequence space, mass spectrometry can possibly replace a real sequencing as long as the number of possible combinations can give a fingerprint. Secondary structures are more likely to be calculated by Vienna package from the result of sequencing than to be measured directly.

4. Early replicators

4.1. Starting point for evolution

Early replicators play a key role in the RNA world hypothesis. Molecular systems that are capable of replicating genetic information are considered as a starting point of evolution.[31] Evolutionary systems are defined by replication, mutation and selection pressure. It seems obvious that an early replicating system will most probably undergo mutations, since it lacks effective, protein catalyzed proofreading mechanisms as known from modern biology. Selection pressure is provided by a chemically active environment and the lack of a proper encapsulation as we know it from highly evolved semipermeable cell membranes and environmental influences such as intense UV-light.
Remarkably, in a first step, it is not urgently necessary that an early ribozymal replicator is capable of reproducing its own sequence base by base. Such a scenario with its high requirements of proofreading and fidelity is rather improbable. For an early replication of information, it is in a first step sufficient to replicate successions of RNA snippets according to a template sequence.[86] The information about a succession of snippets is the basic unit of information, and not principally different from the information about a succession of nucleotide bases. This simplification is used in many approaches to RNA replication. Thereby replication means the assembly of a template copy from two or only a few fractions of the copy.[53][97][87] This approach is also used for technological methods for sequence detection such as ligase chain reaction (LCR)[3] or hybridization chain reaction (HCR)[20] as well as in DNA machines to code and copy information.[99]

4.2. Physical and chemical boundary condtions

There is a row of requirements a molecular candidate for early replication has to fulfill. Those are given by the task of information transmission over several generations and by the geological and chemical boundary conditions in the Archaeon.

RNA or other nucleotides

Environmental conditions have to allow chemical pathways that produce bricks for replication, like monomers or short strands of nucleotides. Nucleotides such as RNA or DNA are predestined to serve as replicators because of their ability to form double strands. Complementary sequences contain equal information, even though the sequence is different. After strand separation both single strands carry full information about the sequence and can both act as a template for a new copy via the selective hybridization of bases with their complements. This construct points towards cross catalytic reactions. Even though there exist some ideas and experiments about protein made early replicators[40] and cross catalytic replication has been reported,[52] the RNA hypothesis is considered to be more plausible at the moment.

It is thinkable that not RNA as we know it today is the matter early replicators were made of, but rather some molecule with a different backbone such as DNA, PNA[93] or precursors of RNA.[85] However, all life as we know it, shares an universal code consisting on RNA and DNA, other coding polymers did not leave traces in biology. As proposed in [96], this is a consequence of horizontal gene transfer among replicators with the same or similar codes.

Replication rates faster than degradation

To allow for Darwinian evolution it seems desirable if not necessary to have exponential growth of information carriers.[98][91] To maintain information over significant time periods, replication has to be faster than all involved degradation processes. Furthermore, in a competition for resources such as monomers or short oligos, an exponential replicator has a sufficiently high chance to survive.
Exponential replication is a highly efficient screening method for correct template sequences. As soon as a good template is available, however in small concentrations, it can reach significant concentrations by replication. This property of exponential replication is technologically used in amplification methods such as PCR, LCR or HCR. [3][20][79] Cross catalytic reactions including nucleotide double strands and their separation often result in exponential growth rates. Experimentally, the kinetic characterization of early replicator systems is an important test for their biological relevance.

Spontaneous formation of information

Besides actual replication, non-templated polymerization reactions occur that generate RNA strands with random sequences.[14] Concerning experiments with replicators, this spontaneous formation of information carrying polymers is especially interesting. All known early replicators with an experimental realization assemble the replication product from sequence snippets that are present in abundance.[50][97][53] Therefore, side reaction that form the product in absence of a template are possible, due to high concentrations of educts in experiments. This type of reaction creates "false positives", since the product, although built without contact to a template, is indistinguishable from the product of template based replication. Even though these side reactions are usually slow and not exponential, they set a lower limit for the reaction speed of a true template based replication.
Products of false positive reactions nevertheless act as templates for replication. Exponential replicators are highly effective sequence detectors for correct sequences. As soon as a small amount of template arises from a side reaction, it is replicated exponentially.

Fuel and waste

Replicators by definition create information carrying molecules and thereby run uphill by means of entropy. Therefore, some source of energy is necessary, especially as soon as energy consuming ligation processes are involved. Chemical energy sources for ligation, such as activated single nucleotides, are hardly plausible candidates, since they have, if available at all, comparably short lifetimes and low concentrations.[97]
Reactions that create waste are problematic for early replicators. Waste products sum up over time, hinder reaction in the end and are costly in terms of the rare resources.
A reaction driven by hybridization instead of covalent chemistry can provide base pairing energy without external energy sources. For example, hairpin-like structures store the ability to

hybridize in bases that are trapped in the loop. If these loops are opened by some mechanism, loop sequences can hybridize and supply reactions with energy. In terms of waste, those structures can be recycled by heating and fast cooling in temperature oscillations. Heating opens all hybridization whereas a fast cooling rate closes hairpins again and makes them available for new reactions. At low concentrations and fast cooling rates, an intramolecular reaction such as hairpin formation is more probable than intermolecular reactions, i.e. hairpins are more likely to get closed than to bind to other binding partners.

Descendants in modern biology

It is a consequence of evolution that a molecule with a function so crucial as being an early replicator may leave its traces in modern biology. In many known examples, successful concepts are modified in evolution, they get new functions or positions, but keep their former structure or sequence. Thereby, comparison of genetic sequences allows an evolutionary classification of biology [103] and an estimation of the age of molecular history.[24]
Thinking the time line backwards, those facts give a road map for the search of early replicators. Possible descendants of early RNA replicators are RNA molecules that are present in all domains of modern biology, and therefore are older than the last universal common ancestor (LUCA),[51]. They are maintained in structure or sequence over several species. Examples for these class of molecules are the Ribosome[19], RNAseP.[10] RNA enzymes such as the R3C enzyme[53] or tRNA.

Long sequences and a bridge to translation

Even though ribozymes have the ability to catalyze chemical reactions, proteins are the powerful catalysts of modern biology. At some point, the early replicator has to offer the bridge to a replication of long sequences and towards a translation of genetic information into proteins.[11]
Most replication schemes in literature assemble the product from two parts. For a meaningful translation, a succession of only two code parts is probably not sufficient, an elongation to longer polymers with more information is desirable.
Intermediate steps of proto-translation do not have high requirements to reliable translation[96] since proteins with similar sequences can fulfill similar tasks. High specificities as known from modern translation are a product of evolution. Nevertheless, a link from RNA replication towards translation is missing.

5. The tRNA world

5.1. An early tRNA replicator

From modern tRNA to ancient replication or vice versa?

Here we propose a reaction scheme for exponential replication of information that is based on tRNA sequences and driven by temperature oscillations in convective flows. By only a few mutations, the characteristic clover leaf secondary structure of tRNA[15] changes to a double hairpin configuration. Figure 5.1 illustrates this transition after 3 mutations on the example of Methanobacterium Thermoautotrophicum Ala TGC. The sequence was taken from the Genomic tRNA database[12] and the secondary structure was calculated by means of the RNA vienna package.[37] Even though the transition ability was not systematically examined over the whole database with algorithms, we found it to be surprisingly robust. This is not completely intuitive, since nucleotides situated in the acceptor stem of the modern tRNA hybridize together with nucleotides in the anticodon loop with high complementarity. Some more examples are shown in figure 5.4.

Double hairpins consist of two hairpins that surround a $8 - 9bp$ long single stranded region.

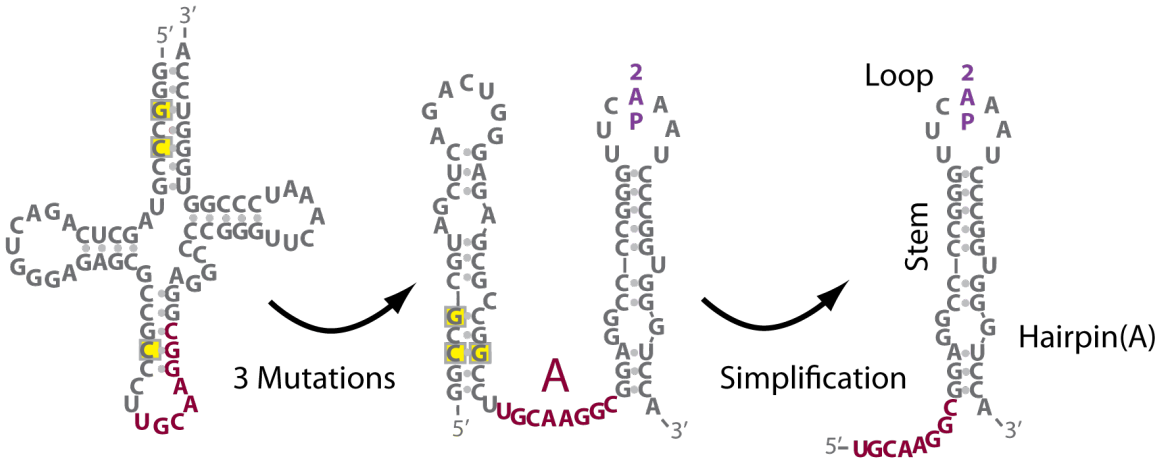

Figure 5.1.: Transition from modern tRNA to double hairpins. By 3 mutations only, the predicted secondary structure of tRNA changes from the characteristic cloverleaf structure to a double hairpin. A simplified reaction scheme with two letters requires hairpins, that arise by a cut next to the anticodon region.

Interestingly, this single stranded region includes the anticodon, which is the information defining sequence in modern tRNA, and some neighboring bases. This single stranded region, that is named **codon** in the following, acts as a monomer for the replication mechanism pre-

sented below.
To simplify the experiments and to explain the principle of replication, a further abstraction creates a single hairpin from the double hairpin by cutting the sequence next to the anticodon. Notably, a single cut of the tRNA yields a hairpin in absence of any mutations, as shown in figure 5.1

A replication scheme for information replication

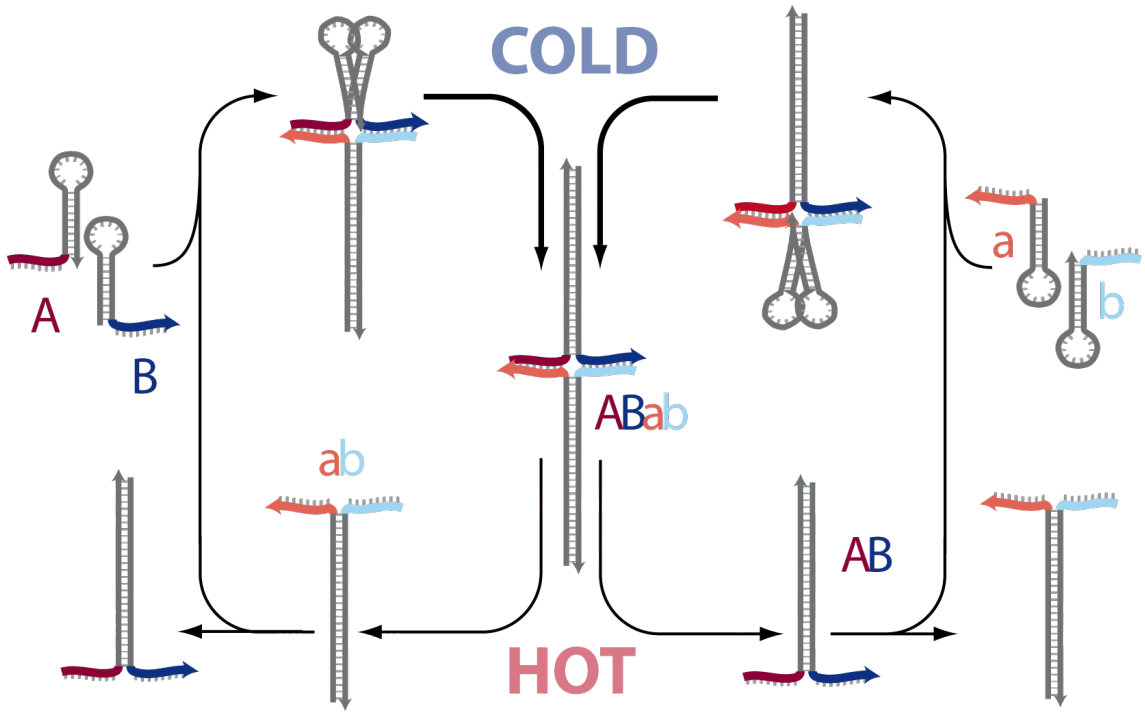

Figure 5.2.: Replication of a succession of two codons. Hairpins A and B bind to a template duplex ab. Local fixation catalyzes the formation of a quarduple $ABab$, that is separated into two duplexes AB and ab by a moderate temperature increase. The reaction is cross-catalytic, both duplexes act as template in a new reaction cycle.

The genetic information to be replicated is the information about a succession of two codons A and B in a duplex AB, or the information about the succession of codons a and b in duplex ab respectively, given that A and a are complementary as well as B and b. From a viewpoint of information, duplex AB and duplex ab are equivalent, even though they differ in sequence. The replication scheme is sketched in figure 5.2: Consider two hairpins A and B that consist of a hairpin structure and a codon. A template duplex ab that carries the information about the succession "a comes before b" acts as replication template at first hand. Due to complementarity, the codon of hairpin A hybridizes to the left half of the template, which is the codon of former hairpin a, as well as the codon of hairpin B hybridizes to the right side, which is the former hairpin of b. The kinetics of these reactions are in the second timescale for the binding as well as for the unbinding[39][32] and the open and close fluctuations of hairpins are

in the timescale of milliseconds.[4] The extremely high local concentration of hairpin A and hairpin B that arises due to the local fixation on the template causes a high probability to hybridize hairpins together to a duplex AB, which is by itself hybridized to a duplex ab. As soon as these transitions happen, the information about the succession of codons is replicated.

A small increase of temperature, as it occurs naturally in thermal convection, can separate both hairpins. That frees duplex ab for a new reaction cycle, meanwhile duplex AB acts as template for a symmetric replication reaction. Therefore the reaction acts cross-catalytic, since the left cycle creates duplex AB as template for the right cycle and vice versa. The number of duplexes should double in each cycle, and the information replication is expected to show an exponential growth of information carrier concentration. (section 5.2)
Expansion of the replication scheme to longer successions is straight forward. Besides the two pairs of hairpins with complementary codons, an additional pair of double hairpins is introduced, with complementary codons and the ability to hybridize with not only one, but two neighboring sequences. Apart from that, the replication scheme stays unchanged.
An at first hand present template trimer qur can catalyze the hybridization of a complemen-

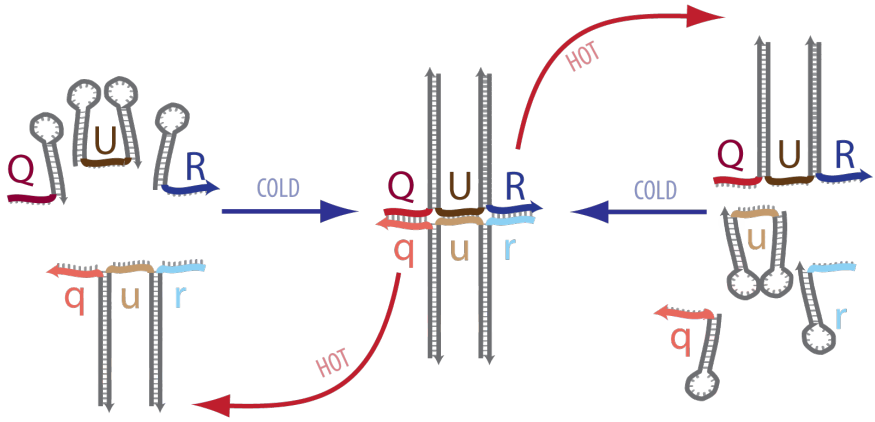

Figure 5.3.: Elongation of the replication scheme to a succession of 3 codons by replacement of a hairpinby a double hairpin. The elongation step can in principle be repeated several times.

tary duplex QUR, both duplexes can be separated by temperature jumps. (see figure 5.3)
This expansion of the replication is in principle not limited to a succession of three letters, but can be repeated several times. Aspects of this inductive expansion are discussed in chapter 5.4.

5.2. Experiments with the tRNA replicator

The replication scheme was realized experimentally and analyzed in detail. To this aim replicator sequences were designed from tRNA sequence and analyzed by fluorescence methods, gel electrophoresis and thermophoresis as elucidated in[46].

Sequence design

tRNA sequences are extracted from the Genomic tRNA database[1].[12] In addition to sequences, this database provides plots of the secondary structure of tRNAs. To test for the transition from tRNA to double hairpins, thermophilic species were chosen. This choice was made not so much to bring the experiments in the context of volcanic activities, but to yield replicator systems with melting temperatures higher than room temperature, which makes experiments easier to handle. The transition is not limited to hydrophilic species, as shown in figure 5.4. The tRNA was chosen to carry amino acids such as alanine or glycine that are considered to be old or arise from a chemistry that is thinkable on early earth.[62] Mutations

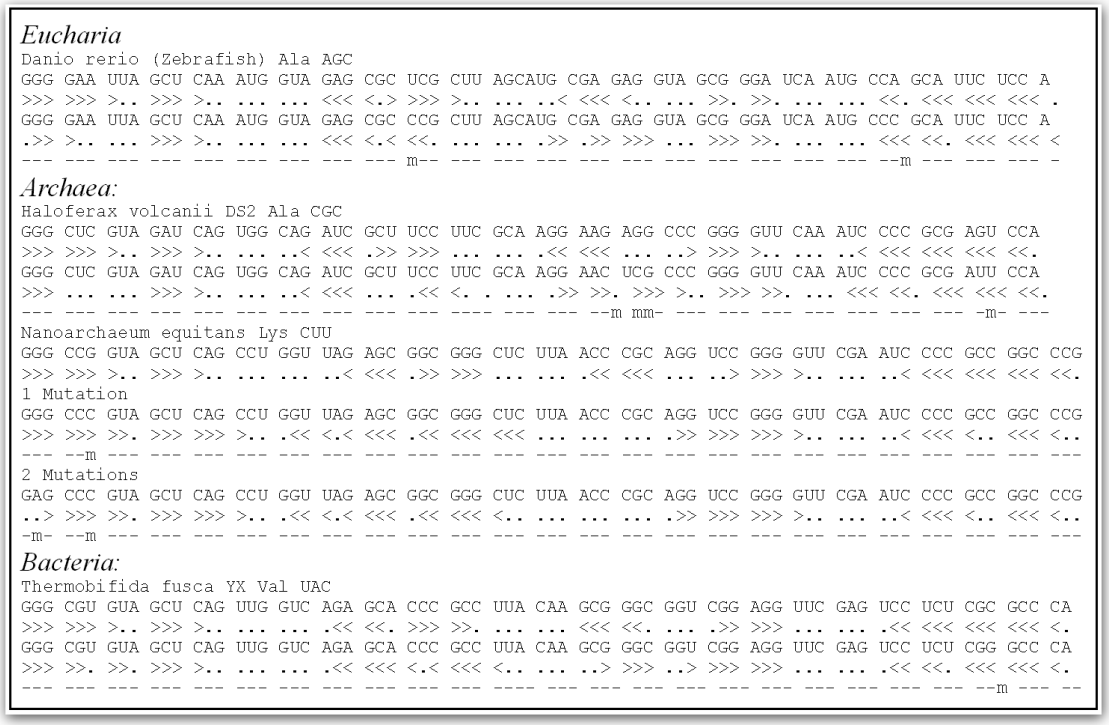

Figure 5.4.: Original tRNA sequences are plotted in first lines, modified double hairpins in second lines. Bases marked with ">" pair with bases further downstream marked with "<". Bases marked with "." are single stranded. In third lines, "m" labelles a mutation. Figure taken from supplementary material of [46] (see Appendix)

are introduced in regions that are modified in comparative sequence studies.[24] The resulting sequence is tested by the RNA vienna package[37] online tool [2] or other other software based on the Vienna algorithms such as the "Oligo Analyzer"[3]. The transition is amazingly robust. In many cases, parts that are originally separated in the anticodon loop and the amino acceptor stem hybridize with high complementarity, and only few mutations are necessary to destabilize the cloverleaf structure in favor of the double hairpin structure.

[1] URL:http://gtrnadb.ucsc.edu.
[2] http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi
[3] http://eu.idtdna.com/analyzer/applications/oligoanalyzer/default.aspx

As soon as the double hairpin conformation is stable, melting temperatures and hybridization energies are calculated. The experimental critera for a good model system are:

The melting transition of the codon region and its complement is well separated from the melting transition of the hairpin stem or the melting of duplexes and triplexes (see figure 5.6) Especially for longer successions this is necessary to gain clear signals in experiments. In principal, a partial opening of long successions is sufficient.

The loop size is long enough to store a sufficient amount of binding energy. To some extent this criteria is connected to the former. The number of monomers in the loop is the number of base pairs that a duplex gains compared to hairpin conformation. The energetic advantage increases with the loop length. A calculation of binding energies and entropic penalties is included in the supplement of [46]

The codon length is short enough to be separated, but long enough to allow for binding of a single codon at reachable temperatures and to provide selectivity.

Reaction partners are designed by means of complementarity

For the simplified replication reaction with four hairpins (see Figure 5.2), hairpin A was gained by cutting of tRNA of Methanobacterium Thermoautotrohpicum Ala TGC next to the anticodon (see figure 5.1) Reaction partners were generated by means of complementarity: With respect to hairpin A, hairpin a has a complementary codon, hairpin B has a complementary hairpin loop and a codon that is not binding significantly to hairpin A, hairpin b has a codon complementary to the codon of hairpin B and a loop that does not allow hybridization to any already given sequence. As a consequence, hairpin a is then completely determined. The sequences designed for the replication of three codons as sketched in figure 5.3 was designed accordingly.
It is crucial to mention, that in principle a one-to-one complementarity of neighboring loops is not necessary, since RNA helices tolerate mismatches and bubbles to high extent. However, for a lab system, we choose a perfect match to yield interpretable signals and to gain reaction times smaller than the lifetime of a student, which is not necessary for evolutionary processes. Furthermore it is to mention that all reaction partners except hairpin A do not directly come from an actual tRNA, but have a sequence with similar hybridization energies and secondary structures. It is a matter of former[76] and further investigations with bioinformatic methods to search for combinations of actual tRNAs that can be transformed to reaction partners for replication schemes. This does not influence the arguments about the similarities of tRNA and the replicator.

2-Aminopurine, a sensor for hybridization states

For fluorescence investigations of replication, labeling of hairpins b and r with 5-FAM (carboxyfluorescin) and hairpin Q with Cy5 at the 5'-end were chosen. Both spectra can easily be distinguished and have no significant FRET-transfer or channel crosstalk (see 5.5).
By means of electrophoresis, those labels enable us to analyze the distribution of species in the steady state after a reaction. To investigate the kinetic development, a fluorophore that reports the changes in hybridization state in real time is necessary. Considering the small size of the hairpin sequences in use, conventional FRET pairs or donor-quencher pairs may

Name	Sequence	Modification
hairpin (A)	UGC AAG GCG GAG GCC CCG GGU UCA AAU CCC GGU GGG UCC A	Mod. at Pos.24 with 2-AP
hairpin (B)	U GGA CCC ACC GGG AUU UGA ACC CGG GGC CUC CAC GUU CCG	not modified
hairpin (a)	G GGC CAC CCA GGU AAG UUU ACC UCC GGG GCC CGC CUU GCA	not modified
hairpin (b)	CGG AAC GUG GGC CCC GGA GGU AAA CUU ACC UGG GUG GCC C	Mod. 5'-end with Fam
hairpin(Q)	ACG UCG AAA GGC CGG CGC UCU CCC AGU CUG AGC UAC GCG GCC UU	Mod. at Pos 24 with 2-AP and Mod 5'-endCy5
doublehairpin(U)	AAG GCC GCG UAG CUC AGA CUG GGA GAG CGC CGG CCU UGC AAG GCG GAG GCC CCG GGU UCA AAU CCC GGU GGG UCC GA	Mod. at pos. 24 with 2-AP
hairpin(R)	UCG GAC CCA CCG GGA UUU GAA CCC GGG GCC UCC GAC GUU CCG	not modified
hairpin(q)	AAC CGG CGC AUC GAG UCU GAC CCU CUC GCG GCC GGU UUC GAC GU	not modified
doublehairpin(u)	CGG GCC ACC CAG GUA AGU UUA CCU CCG GGG CCC GCC UUG CAA CCG GCC GCG AGA GGG UCA GAC CG AUG CGC CGG UU	not modified
hairpin(r)	CGG AAC GUC GGG CCC CGG AGG UAA ACU UAC CUG GGU GGC CCG	Mod. 5'-end with Fam

Figure 5.5.: List of all sequences and modifications used.

disturb the system significantly. As an alternative, the base 2-Aminopurine (2-AP) is used. In terms of Watson-Crick base pairing, it behaves very much like its normal sister base adenine (6-Aminopurine).[27] The difference is a much longer fluorescence lifetime.[55] Interestingly, the fluorescence, which is in the UV region, is quenched by hybridization about a factor of $2-3$ [27] due to changed base stacking behavior.[41] Therefore, an adenine base in the loop of hairpin A and hairpin Q as well as double hairpin U is replaced by 2-AP. A transition from a single stranded hairpin loop towards a double stranded duplex lowers fluorescence due to quenching. We define quenching Q for all experiments as

$$Q_t = 1 - F_t/F_{t=0} \tag{5.1}$$

in which F_t is the time dependent fluorescence and $F_{t=0}$ is the fluorescence at the start of the reaction.

5.3. Results

Melting curves, prediction and thermophoresis measurements

Melting temperatures of duplex strands, hairpins and codons were predicted by means of the Vienna package[37] and verified by melting curve measurements. It is notoriously difficult to determine melting curves of RNA helices by means of UV-absorption, as widely used for DNA. Even though the absorption of light with a wavelength of $250nm$ is modified by hybridization in RNA melting transitions as well, there are many intermediate folding and hybridization states. The smear-out of the characteristic S-shaped temperature curves caused by these intermediate states anticipates a proper fit with a 2-level theory, even though the transition

from a double helix to single strands is clearly is of that kind. [60]
A more accessible way of mapping the transition is thermophoresis.[101] Thereby, it is not the UV absorption that maps the transition, but any change in fluorescence, temperature dependence of fluorescence or thermodiffusion. Since the Soret coefficient is dependent on surface charge density as well as on the organization of the hydration shell,[22] almost any change in hybridization or order causes a significant change in thermophoretic movement and can map a melting transition.

In figure 5.6, the important melting transitions for replication are mapped. To enable

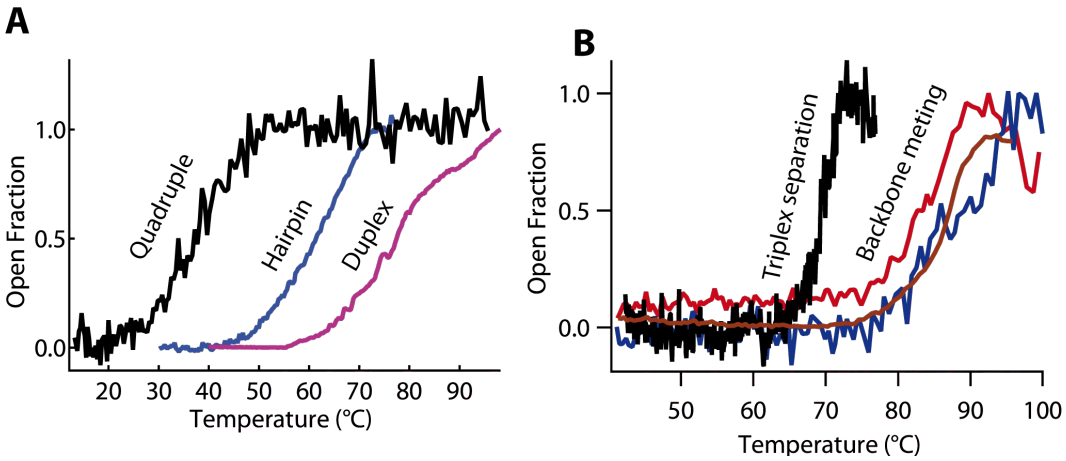

Figure 5.6.: A: Melting transitions of quadruples (black), hairpins (blue) and duplexes (purple) are well separated for a replicator with two letters. B) Melting curves for the replicator with 3 letters. Triplex separation(black) is well separated from hairpin opening(blue) and backbone hybridization(shades of red)

replication, the temperature profile has to fulfill the following conditions:

Low temperatures low enough to enable hybridization of a single codon to its complement. Vicinity of Mg^{2+} can help to stabilize this sensible binding.

High temperatures high enough to separate template and product but low enough not to destroy the product or open a significant portion of hairpins. Open hairpins are available partners for false positive reactions. Nevertheless, high cooling rates can close open hairpins again.

Cooling times long enough to allow a complete hybridization of hairpins to free template molecules, **hot times** short enough to keep side reactions slow.

The sequences in use as given in table 5.5 yield well separated melting transitions between product separation and degradation of hairpins or backbone hybridizations and allow replication with a wide variety of temperature profiles.

Chamber design and temperature control

For the kinetic measurements by means of fluorescence, a chamber was designed according to figure 5.7 that allows simultaneous temperature control and fluorescence readout. The

bottom is defined by a $0,6mm$ thin silicon slide. Silicon has a high heat conductivity[4] to guarantee fast thermal equilibration, is transparent for infrared light, easy to clean and shields autofluorescence originating from the ceramic surface of the peltier element below. The walls of the chamber are defined by a chemical inert and easy to handle rubber sheet with a hole. The top of the chamber is covered with a sapphire slide to allow the excitation of 2-AP with UV light and to fasten thermal equilibration. The dimension of the chamber is opitmized to a volume of about $10\mu l$ or more to allow electrophoresis gel analysis after the kinetic experiment.

The temperature control was realized in two different ways. In the left half of figure 5.7,

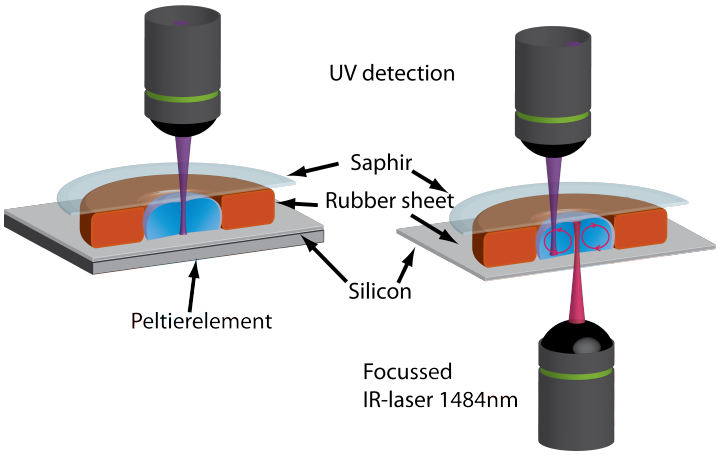

Figure 5.7.: Setup for Kinetic measurements. The liquid sample lies in a chamber with a Si-bottom and a top of sapphire. A chemical inert rubber sheet defines walls. Fluorescence excitation and detection comes from the top. Temperature control is provided in two different ways: Left: a Peltier element changes temperature homogeneously in the complete sample. The sample is in thermal equilibrium at any time. Right: Temperature oscillation by convection caused by punctual heating with an infrared laser.

temperature oscillations are introduced by a peltier element below the Si-layer. This high-end peltier-elements allow temperature changes as fast as $80K/s$. The chamber relaxation time is fast compared to the kinetics of the reaction, so that the sample can be considered to be in thermal equilibrium at any time. This simplifies comparison with theoretical models. Temperature measurements were performed with a BCECF-solution in $10mM$ Tris. This "molecular thermometer" shows a temperature dependent fluorescence and thereby allows temperature calibration. Temperature cycles were programmed at fix currents through the Peliter-element, calibrated before every measurement and then applied to the sample.
In the design sketched on the right hand in figure 5.7, the temperature is controlled by heating with an infrared laser ($1484nm$). Punctual heating of watery solution leads to convective flows. Molecules following flow lines in such a convection undergo temperature oscillations that are capable of driving replication.[5] Since the temperature distribution is inhomogeneous in such a convection cell, the laser power necessary to reach appropriate temperature oscillations is

[4] 150W/mK

estimated by finite element simulations of the reaction chamber as exemplary shown in figure 5.8. Thereby, the physical properties such as heat conductivity and heat capacity of bottom, top and spacer of the chamber, optical properties of IR laser, temperature dependent viscosity of water and so on are taken into account.
In the Peltier-design, reaction steps are time wise separated and happen one after another and homogeneously distributed in the complete chamber. In the laser heating design, reaction steps are spatially separated. Melting processes happen nearby the heating spot, annealing processes away from it. Molecules are transported between that areas by thermal convection.

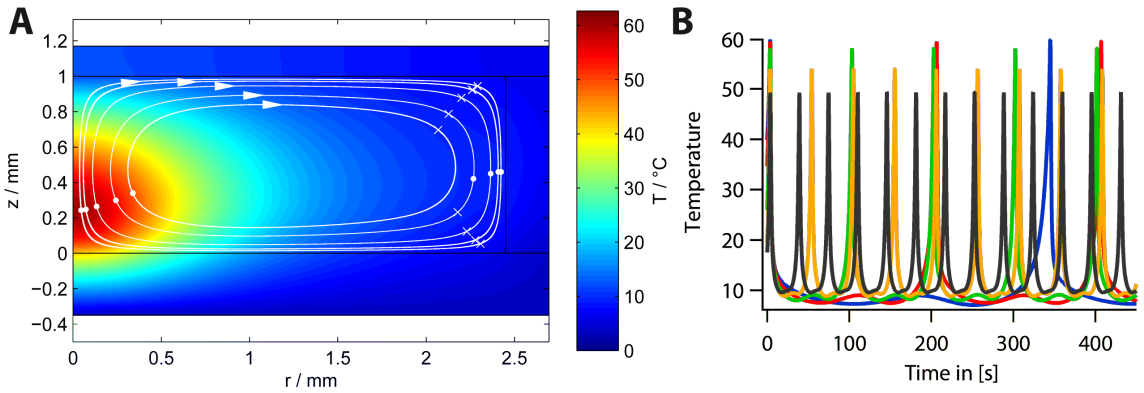

Figure 5.8.: Thermal convection as in a laser heated chamber. **A**:A focussed infrared laser heats the center of a circular chamber ($r = 0mm$). Between the cold surrounding and the hot spot, a convective flow is established. Flowlines are drafted in white. **B**: Temperature oscillations along flowlines. (Figures friendly provided by Simon Lanzmich)

Kinetic measurements with fluorescence

Fluorescence measurements are performed with an upright Zeiss light microscope. To enable the excitation of 2-AP, a quartz-objective was used and the excitation light was coupled into the optical pathway directly above the objective. Together with the sapphire cover glass, it is ensured that the UV-excitation light does not pass borosilicate glass components. Emission light was recorded with a photomultiplier tube.

Figure 5.9 shows recording tracks of a replication reaction as presented in figure 5.2 in black. The transition from hairpins to duplexes and quadruples causes a massive quenching increase. The quenching peaks in all curves measured with peltier element chambers are an overlay of the binding signal originating from codon hybridization and the temperature dependent quantum efficiency of 2-AP. In the convection chamber case, the detection site is placed away from the hot spot, an influence of heating on fluorescence is therefore not visible. Control experiments with only hairpins A and B show no significant quenching increase. Replication speed is determined directly via time characteristics in the first cycles, where the concentration of hairpins is sufficiently high.

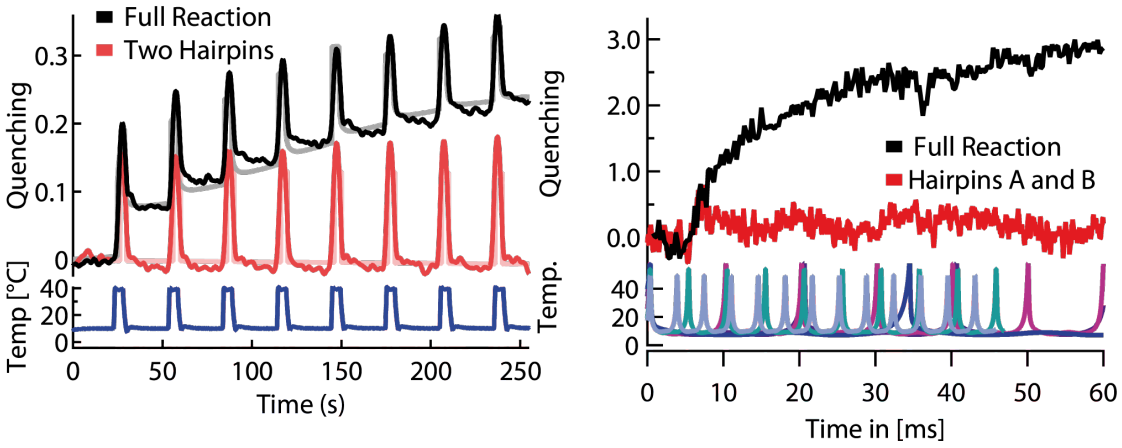

Figure 5.9.: Representative replication tracks. Left: Temperature oscillation with Peltier-element. A temperature profile t(blue) is applied to the chamber. Heating steps cause peaks due to unbinding events and temperature dependent fluorescence. Right: Thermal convection. A full reaction shows a fast and massive quenching compared to control experiments

The steady state product of the reaction is analyzed via high-resolution gel electrophoresis (see figure 5.10). This agarose is optimized to separate mixtures of small oligo-nucleotides up to a length of $1000bp$. Hairpins b, duplexes ab and quadruples in lanes $1-3$ are prepared by a slow annealing from $95°C$ to $5°C$ in two hours to act as a ladder. No significant fraction of side products is visible.

In lanes 4 and 5, control experiments show that the fraction of spontaneously formed

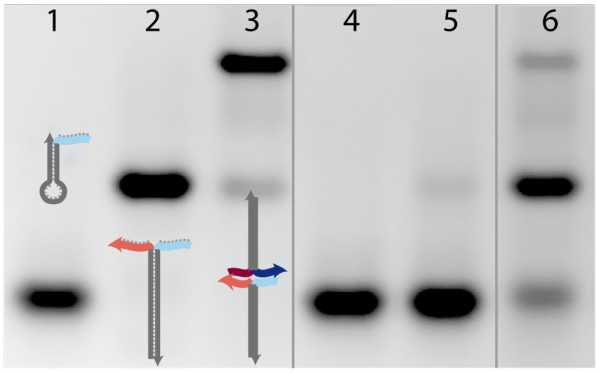

Figure 5.10.: High resolution agarose gels separate hairpins(1), duplexes(2) and quadruples(3). Temperature cycling keeps hairpins without reaction partners stable(4), false positive rate in absence of a catalyst is comparably low(5). A full reaction produces dupexes and quadruples(6).

duplexes or misfold hairpins is small in cycling experiments. Lane 6 analyzes the result of a full replication reaction as described in figure 5.2. Most hairpins are transformed to information carrying duplexes and quadruples.

Data analysis with a kinetic model

The quantity that is observed directly in all kinetic experiments is the quantum efficiency of 2AP. The fluorescent raw signal is a mixture of emission light originating from hairpins, duplexes, quadruples and other intermediate products that appear in the reaction. Those fluorescence contributions are added up by the photo detector. The sum is converted into a quenching rate according to formula 5.1 and corrected for the temperature dependence of 2AP by measurements of only hairpins A. How can this Q_t be interpreted as a time dependent concentration C_t ?
Figure 5.2 and the more simplified figure 5.3 present systems of chemical reactions that consist of various partial reactions that are coupled via the concentrations of products and reactants. In a detailed kinetic model, we formulate all relevant partial reactions in a chemical reactions of the form:

$$X + Y \rightleftharpoons XY \tag{5.2}$$

with k_{on} rates on the left and k_{off} rates on the right. Thereby we can describe the time development of all species concentrations by a system of coupled differential equations of the form:

$$d[X]/dt = -[X] * [Y] * k_{on} + [XY] * koff \tag{5.3}$$

Any reaction that includes the species $[X]$ adds two additional terms on the right side of equation 5.3. For any partial reaction, a k_{on} and a k_{off} rate are determined. Rates are fixed by own measurements and comparison with literature values.[106][39][89] All partial reactions that are included in the model and all rates in use for the two letter replicator are given in figure 5.11.
The model allows to calculate time developments for the concentration of any species. Any species that includes a hairpin A contributes to the quenching signal with a specific factor. Multiplication of that factor with the concentration and summation of all signals allows to predict a quenching curve that can be compared to measurements. The left panel in figure 5.9 plots measurements in strong, simulations in shade colors and demonstrates a good agreement. A detailed comparison is included in the supplement of [46](see Appendix).

Serial dilution experiments

In former reports on early replication, the power of replicators is typically tested by serial dilution experiments.[97][53] In those experiments, the product of a replication cycle is used as template in a new cycle with fresh educt material. Such a typical "serial dilution" experiment is represented in figure 5.12.
A replication reaction with four hairpins in $500nM$ concentration and duplexes ab as template was performed in a laser driven reaction chamber. The resulting solution was diluted $30x$ with a mixture of fresh hairpins and temperature cycling by IR laser was repeated. Successful replication can be repeated several times, demonstrating the ability of the replicator to maintain information over several generations. In green, the false positive reaction of hairpins A and B is tested. It represents the spontaneous generation of template duplexes. The serial dilution shows impressively the dominant role of the replicated succession against a succession spontaneously formed.

Reaction	koff T= 10°C	kon T=10°C	koff T=40°C	Kon T=40°C
AB+a ⇋ ABa	0.2	1.3	200	1.3
AB+b ⇋ ABb	0.2	1.3	200	1.3
A+ab ⇋ Aab	0.2	1.3	200	1.3
B+ab ⇋ Bab	0.2	1.3	200	1.3
ABa+b ⇋ ABab	1e-09	1.3	1e-06	1.3
ABb+a ⇋ ABab	1e-09	1.3	1e-06	1.3
A+Bab ⇋ ABab	1e-09	1.3	1e-06	1.3
B+Aab ⇋ ABab	1e-09	1.3	1e-06	1.3
AB+ab ⇋ ABab	0.0025	25	100	25
A+B ⇋ AB	1e-06	0.00015	0.00015	0.00015
a+b ⇋ ab	1e-06	0.00015	0.00015	0.00015
A+a ⇋ Aa	0.1	0.57	120	0.57
B+b ⇋ Bb	0.1	0.57	120	0.57
A+A ⇋ AA	0.00001	0.00015	0.001	0.00015
B+B ⇋ BB	0.00001	0.00015	0.001	0.00015
a+a ⇋ aa	0.00001	0.00015	0.001	0.00015
b+b ⇋ bb	0.00001	0.00015	0.001	0.00015
Aa+B ⇋ AaB	1e-06	0.0015	0.00015	0.0015
Aa+B ⇋ AaB	1e-06	0.0015	0.00015	0.0015
Bb+A ⇋ ABb	1e-06	0.0015	0.00015	0.0015
Bb+a ⇋ Bab	1e-06	0.0015	0.00015	0.0015
Aa+Bb ⇋ ABab	1e-09	0.00015	1e-06	0.00015
Aoh ⇋ A	1e-09	0.001	1e-09	0.5
Boh ⇋ B	1e-09	0.001	1e-09	0.5
aoh ⇋ a	1e-09	0.001	1e-09	0.5
boh ⇋ b	1e-09	0.001	1e-09	0.5

Figure 5.11.: Kinetic rates for all partial reactions as used in the simulation. k_{off} rates differ about one order of magnitude per $10K$.[89] Rates are compared with own measurements and literature values. Figure taken from supplementary information of [46].

Hybridization as energy source and thermal reset

One major advantage of a replication scheme that is purely based on hybridization is the independence of chemical energy sources and the avoidance of waste products. The energy necessary to overcome the entropic penalty is by far provided by hybridization energy of bases that are single stranded in the hairpin loop and double stranded in the duplex conformation. The transformation of two hairpins to a duplex results in $\Delta G_{duplex} = -45kcal/mol$. The binding of two codons to their complements delivers $\Delta G_{codon} = -13kcal/mol$. The complete gain in free energy in a transition of four hairpins to a quadruple is therefore given as $\Delta G_{replication} = 2 * \Delta G_{duplex} + 2 * \Delta G_{codon} = -116kcal/mol$. This is sufficient to compensate for the entropic penalty estimated as $\Delta G_{entropic} = RTln(Q)$ with $Q = ([hairpin]/c_0)^4/([quadruple]/c0)$. For a somewhat arbitrary reference point at which half of hairpins are used up, we get $[hairpin] = [quadruple] = 250nM$, $Q = 2,5 * 10^{-7}$ and estimate a $\Delta G_{entropic} = 26kcal/mol$, which is by far compensated by hybridization energy.

The reaction we present here is not based on the formation of covalent binding but on hybridization. Therefore, it is fully reversible. A heating step to temperatures higher than the melting temperature can separate any product of the reaction and side reaction. Oligos are requenched to hairpins in steep temperature flanks as they are typical for convective

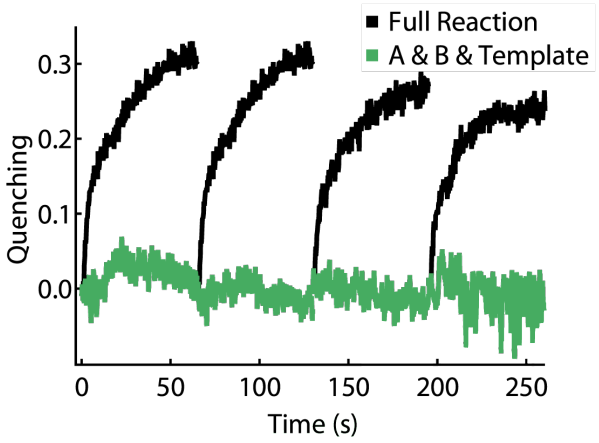

Figure 5.12.: Serial dilution of a full replication vs spontaneous generation of duplexes. Any step correlates with a $30x$ dilution. Information replication is maintained over several dilution steps, while spontaneously formed information can not withstand dilution.

flows.[5]
To demonstrate the reset, we performed a full replication with the same conditions as in 5.9. Afterwards, the temperature of the laser spot was increased in a way that a significant fraction of flow lines get warmer than $95°C$ and melt all double helices. The rapid cooling rate of the flow lines create hairpins from the part that can participate in a new replication reaction. The diminished amplitude is explained by the fact that only a fraction of the chamber volume gets hot enough to denature double helices. The fraction can be estimated by simulations.

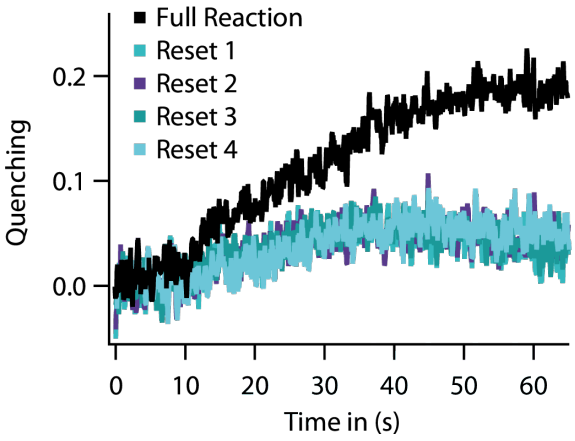

Figure 5.13.: Thermal reset of the reaction. The replication is reversible by thermal denaturation and re-quenching. The diminished amplitude origins from the fact that only a part of the chamber is hot enough to melt all helices.

Selectivity and isothermal reaction

The reliability of information transmission is an important quality feature of a replicator. Only if the information content does not decrease over several replication cycles, a replication can survive for long times.
The replication scheme presented here is as selective as the complementary hybridization of codon regions. Selectivity was tested by a hairpin that has the same hairpin sequence as hairpin a, but a codon region that is not complementary to the codon of hairpin A. The result is a breakdown of replication. (see figure 5.14)

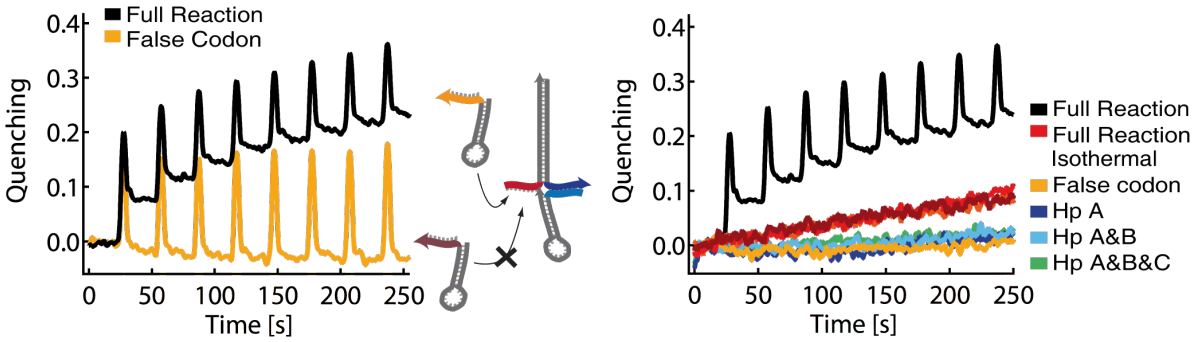

Figure 5.14.: Left: Replication is selective for correct codons. Exchange of a hairpin with same loop, but wrong codon suppresses replication. Right: In the absence of temperature oscillations, replication slows down dramatically (shades of red). With false codons and in the absence of a hairpin, background reaction is slow.

The codon region used here has a length of $9bp$ and the test was performed with a codon that has no matching parts. Sequences with one or more mismatches were not tested experimentally. According to calculations,[37] the stability of shorter strands even at temperatures as low as $10°C$ decreases rapidly with every mismatch. Stable hybridization in complementary sequences shorter than $3-4bp$ do only rarely form at all due to the decrease of the base stacking contribution. That means that codon sequences can tolerate only very few mismatches and thus selectivity for right sequences is high.
As a side remark, the modern codons of translation connecting tRNA with mRNA have a length of $3bp$. In cellular surrounding at physiological temperatures, a double strand does not form from such short sequences. It is one crucial task of the ribosome to catalyze this binding.
Left panel of figure 5.14 illustrates how temperature cycling overcomes template inhibition. At constant temperatures (shades of red) replication still takes place due to finite k_{off} rates, but is much slower and does not differ much from false positive formation (pale blue). Isothermal replication with slow rates has already been demonstrated.[53][97] It is to mention, that the isothermal reaction, as any chemical reactions, can be speed up by higher temperatures. In the case presented here, the k_{off} rates that separates template and product increase by an order of magnitude if temperature is rised about $10K$.[89] Nevertheless, higher temperatures lower the change to hybridize codons and start catalytic formation of product. Temperature cycles both open the catalyst for new reactions at high temperature and optimize codon binding at low temperatures.

Exponential growth rates

Exponential growth rates are an important criteria for the relevance of early replication.[91] The experimental prove, beside serial dilution experiments, are sigmoid curves in the concentration vs. time plot as well known from PCR experiments[79][58] and a plot of the concentration growth rate against the initial template concentration. Exponential growth of

information carrier C concentration shall follow an equation of the form:

$$dC/dt = ln(2)/\tau * C + k_0 \tag{5.4}$$

in which τ is the doubling time of information carrier concentration C and k_0 accounts for the spontaneous formation of information carriers (false positives).

Panel A of figure 5.15 shows several replication reactions with various starting concentra-

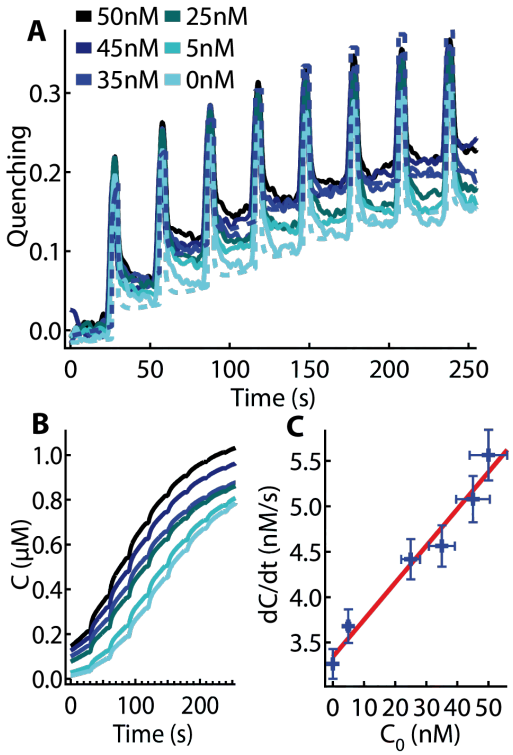

Figure 5.15.: Exponential growth rates in experimental test. (A) Replication reaction with various starting template concentrations. The good agreement between simulation and experiment is evident. (B) Simulated concentration yields sigmoidal curves (inset), logaritmic plots reveal exponential characteristics. (C) Fit to equation 5.4 shows a doubling time of $\tau = 28 \pm 5s$, a background rate of $k_0 = 3 \pm 0.3nM/s$ at a cycling time of $t_{cycle}30s$.

tions of duplex *ab* as template. With increasing concentration, the derivative of quenching over time, dQ/dt, increases. Every quenching curve was simulated with the kinetic model. Only the initial duplex concentration and the pipetting time that accounts for a certain amount of duplexes that arise from the isothermal replication reaction were changed between the simulation runs. In that sense, the model is parameter-free and describes the reactions in great detail.

The agreement between the model and the measurement is evident. A detailed plot for all single curves is plotted in the supplementary material of [46]. The concentration of information carriers is extracted from the model as:

$$[C]_{info} = [duplex\ AB] + [duplex\ ab] + 2 * [quadruple\ ABab] \tag{5.5}$$

and plotted in panel B of figure 5.15. In panel A, the quenching curves for several inital template concentrations are shown together with exemplary simulations. The good agreement of simulation and raw data allows the extraction of concentrations from the model. Concentration over time plots for different initial template concentrations reveal the classical sigmoid curves as comparable to PCR.
The range of initial template concentrations reachable with the measurement methods used is limited. The smallest concentration reachable is limited by the rate of spontaneous duplex formation and the pipetting time. The highest concentrations are determined by the initial hairpin concentration and the doubling time. In case all hairpins are used up in only very view cycles, interpretation of quenching curves is not confident any more.
A plot of dC/dt for the first three cycles against the start concentration allows a fit to equation 5.4 and reveals a doubling time of $\tau = 28 \pm 5s$, which is in the same range as the cycling time of $t_{cycle} = 30s$, and a comparably slow background rate of $k_0 = 3 \pm 0.3nM/s$. Compared to all known RNA replicators,[53][97] this the fastest replication mechanism by orders of magnitude.
The reactions speed expressed by the doubling time is dependent on the form of the reaction cycle and can vary dependent on the cold times, as long as those do not fall below the hybridization times of the codons, which are in the seconds timescale. In convective chambers as realized by laser heating, several flowlines with different temperature and time profiles occur simultaneously. As long as there is a fraction of flowlines with appropriate properties, replication will take place.

Longer sequences

The expansion of the replication scheme to a replication scheme for a succession of 3 codons by introduction of double hairpins according to figure 5.3 is straight forward even though the experiments are harder to tune and to interpret.

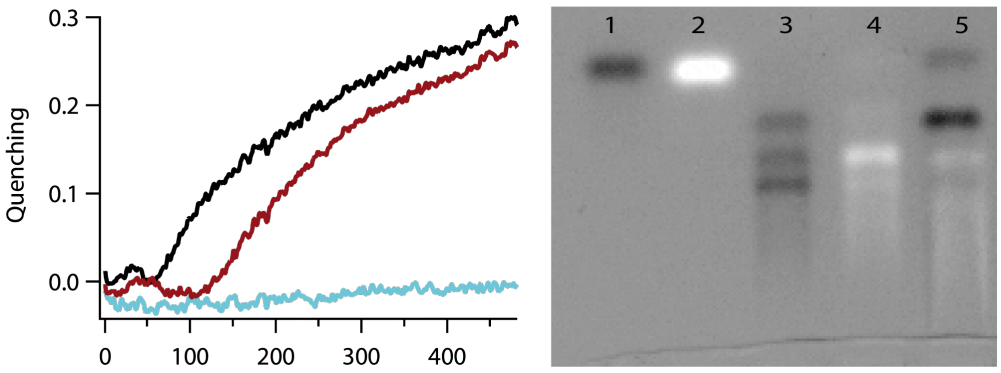

Figure 5.16.: Quenching of a replication reaction for a sequence of 3 codons. 1) hairpin r, dark label, 2) hairpin Q bright label, 3) trimer qur and side products, 4) trimer QUR and side products. 5) end products of a full cycling. Most material is in 2-mers, but still both trimers are available in a much higher amount as the starting concentration of the catalyst.

Results of a tree-codon replication are shown in figure 5.16. Hairpin Q is labelled with Cy 5 and plotted bright in the gel picture (lane 1), while hairpin r is labelled with FAM and plotted dark (lane 2). By a slow annealing of trimers qur in lane 3, besides trimers, there arise side products that consists of two hairpins or hairpins and double hairpins. The other half of the complex, trimer QUR forms properly, as shown in lane 4. The full replication reaction with all six educts and a small amount of template produces some trimers in line 5. Besides a bigger fraction of duplexes form hairpins and double hairpins, additional trimers can be found on the gel. As template the black colored half was used, the white band visible in the gel is the produced trimers. The yield can be increased by optimized parameters for the chamber geometry that influences the length and peak height in temperature oscillations. Quenching curves show the typical replication fingerprint of a replication. According to the gel, this is mainly the formation of dimers. For a proper replication of trimers, two steps are necessary: the formation of a trimer and a further addition of a hairpin. The interpretation of quenching curves with so many side reactions requires a detailed simulation that is a matter of future research.

5.4. A bridge to translation

The replication scheme as presented above implies a straight way from replication of genetic information to translation. In biology, translation is defined by the assembly of an amino acid chain according to a nucleotide sequence. This task is accomplished by the ribosome, a protein with catalytic subunits built from RNA.[2] In modern translation, different tRNAs specifically charged with amino acids at the 3'-end are sorted by complementary binding of the anticodon to single stranded messenger RNA (mRNA). This sorting of amino acids according to codons with a length of $3bp$ leads to formation of a protein with an amino acid sequence according to the codon sequence written on mRNA.

In a gedankenexperiment, we now expand the results presented in chapter 5.3 in two steps:

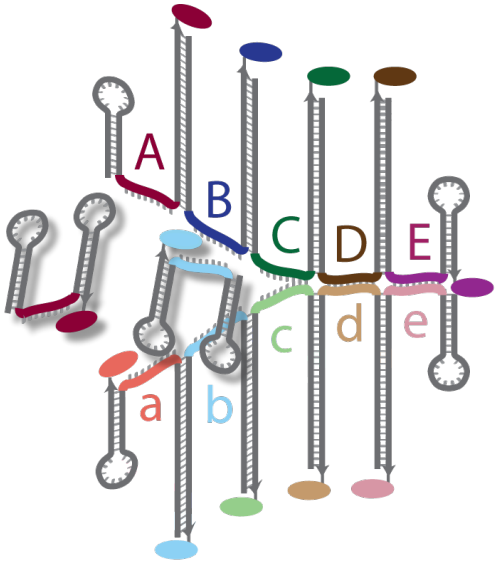

Figure 5.17.: A bridge to translation: Double hairpins are able to replicate longer codon successions. Specific charging of double hairpins with amino acids or peptides at the 3'-end of double hairpins allow a succession of peptides according to a succession of codons, very similar to modern translation.

Firstly, the replication is extended to longer successions by using several double hairpins

instead of single hairpins. Secondly, in analogy to the function of modern tRNA in translation we assume a more or less specific charging of double hairpins with amino acids or peptides on the 3'-end. Since this loading is realized by an esterification, both the carboxyl-group as well as the amino group are available for the formation of a peptide bond.
As illustrated in figure 5.17, the gedankenexperiment leads to a sorting of amino acids or peptides according to the succession of codons, very similar to modern translation. In that sense, it resembles older proposals by Kuhn,[48] but adds all the connections to modern biology that come with the relation between double hairpins and tRNA.

Implications of the Translation scheme

Replication of longer successions.

With an extended length of the codon sequence, the hybridization of template and product gets more stable. Nevertheless, a partial separation by a temperature increase is plausible. Bubbles larger than the length of a single codon are sufficient to allow partial replication, that can be enlarged by strand displacement. It has to be considered here that the hybridization of template and product is interrupted by holiday junctions[59] that destabilizes the hybridization by disturbing the base stacking after every codon.
The scheme as proposed in figure 5.17 makes predictions for the sequence of double hairpins and hence for tRNA. For example, the codon regions of different tRNAs should show an increased complementarity. In fact, such a complementarity of anticodons and their neighboring bases between different tRNAs of the same species was reported.[76]

Selective charging of double hairpins

The selectivity of this charging, nowadays guaranteed by a complex tRNA-synthetase protein,[78] is a hard problem for early translation. By means of diffusion time and degradation, Hopfield argued, that the spatial proximity of amino acid and codon in the double hairpin configuration can lead to selectivity, since some combinations of amino acid and codon sequence protect each other better from degradation than others.[38] Selectivity arises by assuming spatial separation of charging and translation and by a diffusion distance in between. It is under debate whether the selectivity that can be reached is sufficient for a faithful translation. Nevertheless, it has to be considered that an error prone translation at first hand has a selective advantage compared no translation at all.
The formation of selective binding pockets from RNA for special amino acids was also shown previously.[105]

Proteins or Peptides?

In addition to a charging of double hairpins with single amino acids that are in any case chemically plausible under early earth conditions,[62] a charging with short peptides is very well thinkable. This generalization could enhance the selectivity of charging and the catalytic activity of the end product.[88]
It is to mention here that already the binding of peptides or amino acids to the vertical strands in figure 5.17 can stabilize those connections for example by adding positive charges and thereby reducing backbone repulsion. By this stabilization, the replication of longer

strands is enabled which is a selective advantage against the degradation of single strands. Furthermore, in the absence of better alternatives, the demand on a reliable translation are not to high, since proteins with similar amino acid sequences have similar function and errors can be tolerated to a wide extent.[96]

5.5. Outlook: Proposal for further experiments

The implications of the reaction scheme propose a new sequence design. In this design, it is not only possible to replicate fixed successions of two or tree codons, but longer, random successions. Therefore two main generalizations are necessary:

A universal hairpin sequence that allows hybridization of different codons in random order. Thereby it is not sufficient to make the hairpin upstream the codon complementary to the hairpin downstream the condon. Double hairpins of that design would form big hairpins with the codon sequence as loop and the hairpin parts as each a single strand for the big stem. It is rather necessary to have two kinds of double hairpins that alternate, for example a spacer S and two others with different codons, as shown in figure 5.18

Different distinguisable codon sequences that can sequenced experimentally in a way. The sequence design has to offer a change to experimentally approach the succession of codons and to determine the distribution of successions, in order to find traces of replication.

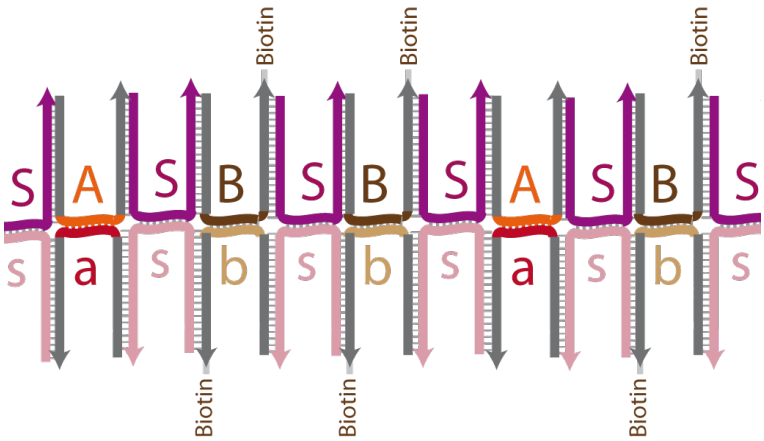

Figure 5.18.: Arbitrary succession of codons A and B with codon S as spacer. Free successions can be built from a 2-letter code. Double haripin B has a biotin linker on one end, that can be incubated with streptavidin to enable a sequencing by AFM imaging.

In figure 5.18 a sequence design enabling a arbitrary succession of codon A and codon B is proposed. A unfolded double hairpin S connects as spacer two coding double hairpins. Thereby, for A and B, hairpin sequences are possible that bind to S, but not to themselves. Of course it is thinkable to have additional double hairpins with the hairpin sequences of S and further codons. For experiments, this 3-partner solution is the simplest possible.

Double hairpin B carries a biotin linker on one end. After replication, incubation with streptavidin can mark the position of the codon B in a way that it can be imaged for example with an AFM.[36]
Somehow more tricky is the preparation of a template sequence as catalyst for replication. In principle, a single stranded DNA or even RNA can act as template to copy. Problems may occur, since it is not trivial to take into account the holiday junctions that appear between two codons. However, since the sequence design does not allow longer successions of polycodons, for example AAA, it may be possible to create template sequences similar to the protection chemistry that is used for DNA and RNA synthesis:
Let us assume we have a surface with an attached RNA sequence that is partially, but not fully complementary to the 5'-hairpin S. If we now add single stranded double hairpin sequences S at higher temperatures, S may bind to the surface. In a next step, we add open A, that binds to S, than again S. We can repeat this steps with A and B. A moderate heating step can release the finished template sequence from the surface. Before replication, the result can again be tested by AFM.

Conclusion

The tRNA world replicator presents a basic, protein-free replication of genetic information. Besides the remarkably fast time scales of replication, originating from temperature oscillations in convective flows, it can be extended to longer successions. By its nature originating from tRNA, it points towards an proto-replication, that can offer an escape from the RNA world towards a translation of proteins, the powerful catalysts of modern biology.

6. Acknowledgements

Good science is a task for a good team: All I learned I learned from people, and all I did I did with help of colleges and all new ideas I gained in discussions. Therefore it is my pleasure to acknowledge a great team:

Professor Dieter Braun, my supervisor, for well defined projects, for trust and support, for all the scientific enthusiasm and for the great atmosphere in the lab, that has it visible manifestation in the "golden espresso machine". And thousand thanks for all I learned besides science just by watching him being a great team leader, a great family person and an expert for parachutes, e-bikes and paradigm-shifting lab parties.

My scientific cooperators Ulrich Gerland and Benedikt Obermayer and Ingmar Schön for good team work. I learned far more from them as I could contribute.

Friederike Möller It is so easy to survive in the lab with people that share not only the project, but also friendship. As long as Bünting tea is in the lab, there is still hope.

Simon Lanzmich and Lorenz Keil for running experiments as long is I was absorbed in paper work, for good questions, and good simulations that gave good hints for experiments.

All my former and actual office and lab mates, Christof and Mario, Maren, Susi, Manuel, Soichi, Moritz, Georg, Ulrich, Andi, Franz W, Christoph, Svenja, Fabian, Mathias, the Nanotemper Crew and so on, it is the people that make it fun to go to work. A special thanks goes to all the prove-readers of this manuscript, first of all Carolin, Svenja und Friederike.

All neighbors form the Gaub Lab and the Rädler lab that provide all the help we need in discussions, with instruments, chemicals and diverse barbecue and snow events. Special thanks goes to Mathias, a couple of Stefans and Stephans, Ingo, Diana, Angelika and Tom as well as to both Susies. Nowadays they call it network, but is something special here.

Svenja Lippok for proofreading the manuscript of this thesis.

Carolin, Verena Felix, Peter, Maria and lots of other musicians from diverse bands, orchestras and Tanzlmusi. Its your music, that keeps me singing.

My family for unlimited trust, endless patience and all the support they give in desperate moments.

Maria for a very special time gone and a very special future coming up.

Bibliography

[1] 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*, 104(22):9346, 2007.

[2] N. Ban, P. Nissen, J. Hansen, P.B. Moore, and T.A. Steitz. The complete atomic structure of the large ribosomal subunit at 2.4 å resolution. *Science*, 289(5481):905, 2000.

[3] F. Barany. Genetic disease detection and dna amplification using cloned thermostable ligase. *Proceedings of the National Academy of Sciences*, 88(1):189, 1991.

[4] G. Bonnet, O. Krichevsky, and A. Libchaber. Kinetics of conformational fluctuations in dna hairpin-loops. *Proceedings of the National Academy of Sciences*, 95(15):8602, 1998.

[5] D. Braun, N.L. Goddard, and A. Libchaber. Exponential dna replication by laminar convection. *Physical review letters*, 91(15):158103, 2003.

[6] D. Braun and A. Libchaber. Lock-in by molecular multiplication. *Applied physics letters*, 83:5554, 2003.

[7] 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(28):9628–9629, 2009.

[8] M. Bustin. Regulation of dna-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Molecular and cellular biology*, 19(8):5237, 1999.

[9] G. Caliskan, C. Hyeon, U. Perez-Salas, RM Briber, SA Woodson, and D. Thirumalai. Persistence length changes dramatically as rna folds. *Physical review letters*, 95(26):268303, 2005.

[10] T.R. Cech. The chemistry of self-splicing rna and rna enzymes. *Science*, 236(4808):1532, 1987.

[11] T.R. Cech. Crawling out of the rna world. *Cell*, 136(4):599–602, 2009.

[12] P.P. Chan and T.M. Lowe. Gtrnadb: a database of transfer rna genes detected in genomic sequence. *Nucleic acids research*, 37(suppl 1):D93–D97, 2009.

[13] C.A. Cohn, T.K. Hansson, H.S. Larsson, S.J. Sowerby, and N.G. Holm. Fate of prebiotic adenine. *Astrobiology*, 1(4):477–480, 2001.

[14] G. Costanzo, S. Pino, F. Ciciriello, and E. Di Mauro. Generation of long rna chains in water. *Journal of Biological Chemistry*, 284(48):33206, 2009.

[15] F. Cramer, H. Doepner, et al. On the conformation of transfer rna. *Proceedings of the National Academy of Sciences of the United States of America*, 61(4):1384, 1968.

[16] F.H.C. Crick. The origin of the genetic code. *Journal of molecular biology*, 38(3):367–379, 1968.

[17] Charles Darwin. *The Origin of species*. Murray London, 1859.

[18] Francis Darwin. *The life and the letters of Charles Darwin volume 3*. John Murray, 1887.

[19] C. Davidovich, M. Belousoff, A. Bashan, and A. Yonath. The evolving ribosome: from non-coded peptide bond formation to sophisticated translation machinery. *Research in microbiology*, 160(7):487–492, 2009.

[20] R.M. Dirks and N.A. Pierce. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43):15275, 2004.

[21] S.M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W.M. Shih. Self-assembly of dna into nanoscale three-dimensional shapes. *Nature*, 459(7245):414–418, 2009.

[22] S. Duhr and D. Braun. Why molecules move along a temperature gradient. *Proceedings of the National Academy of Sciences*, 103(52):19678, 2006.

[23] S. Ebbinghaus, A. Dhar, J.D. McDonald, and M. Gruebele. Protein folding stability and dynamics imaged in a living cell. *Nature Methods*, 7(4):319–323, 2010.

[24] M. Eigen, B.F. Lindemann, M. Tietze, R. Winkler-Oswatitsch, A. Dress, and A. von Haeseler. How old is the genetic code? statistical geometry of trna provides an answer. *Science*, 244(4905):673, 1989.

[25] M. Eigen and P. Schuster. A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565, 1977.

[26] R.J. Ellis. Macromolecular crowding: obvious but underappreciated. *Trends in biochemical sciences*, 26(10):597–604, 2001.

[27] K. Evans, D. Xu, Y. Kim, and T.M. Nordlund. 2-aminopurine optical spectra: solvent, pentose ring, and dna helix melting dependence. *Journal of Fluorescence*, 2(4):209–216, 1992.

[28] JP Ferris, DB Donner, and AP Lobo. Possible role of hydrogen cyanide in chemical evolution: investigation of the proposed direct synthesis of peptides from hydrogen cyanide. *Journal of Molecular Biology*, 74(4):499–510, 1973.

[29] J.P. Ferris, A.R. Hill, R. Liu, and L.E. Orgel. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 1996.

[30] M.J. Gaffey. The early solar system. *Origins of Life and Evolution of Biospheres*, 27(1):185–203, 1997.

[31] W. Gilbert. Origin of life: The rna world. *Nature*, 319(6055), 1986.

[32] S.J. Green, D. Lubrich, and A.J. Turberfield. Dna hairpins: fuel for autonomous dna devices. *Biophysical journal*, 91(8):2966–2975, 2006.

[33] J.M. Grimme, M. Honda, R. Wright, Y. Okuno, E. Rothenberg, A.V. Mazin, T. Ha, and M. Spies. Human rad52 binds and wraps single-stranded dna and mediates annealing via two hrad52–ssdna complexes. *Nucleic acids research*, 38(9):2917–2930, 2010.

[34] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman. The rna moiety of ribonuclease p is the catalytic subunit of the enzyme. *Cell*, 35(3):849–857, 1983.

[35] J.B.S. Haldane. The origin of life. *Rationalist Annual*, 148:3–10, 1929.

[36] HG Hansma, J. Vesenka, C. Siegerist, G. Kelderman, H. Morrett, RL Sinsheimer, V. Elings, C. Bustamante, and PK Hansma. Reproducible imaging and dissection of plasmid dna under liquid with the atomic force microscope. *Science*, 256(5060):1180–1184, 1992.

[37] I.L. Hofacker. Vienna rna secondary structure server. *Nucleic acids research*, 31(13):3429–3431, 2003.

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

[39] S. Howorka, L. Movileanu, O. Braha, and H. Bayley. Kinetics of duplex formation for individual dna strands within a single protein nanopore. *Proceedings of the National Academy of Sciences*, 98(23):12996, 2001.

[40] K. Ikehara. Possible steps to the emergence of life: The [gadv]-protein world hypothesis. *The Chemical Record*, 5(2):107–118, 2005.

[41] J.M. Jean and K.B. Hall. 2-aminopurine fluorescence quenching and lifetimes: role of base stacking. *Proceedings of the National Academy of Sciences*, 98(1):37, 2001.

[42] N. E. Schore K. Peter, C. Vollhardt. *Organic chemistry*. Freeman and Co, New York, 1999.

[43] B.R. Kelemen, T.A. Klink, M.A. Behike, S.R. Eubanks, P.A. Leland, and R.T. Raines. Hypersensitive substrate for ribonucleases. *Nucleic acids research*, 27(18):3696–3701, 1999.

[44] 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, et al. A serpentinite-hosted ecosystem: the lost city hydrothermal field. *Science*, 307(5714):1428, 2005.

[45] M. Konarska, W. Filipowicz, and H.J. Gross. Rna ligation via 2'-phosphomonoester, 3'5'-phosphodiester linkage: requirement of 2', 3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase. *Proceedings of the National Academy of Sciences*, 79(5):1474, 1982.

[46] H. Krammer, FM. Möller, Lanzmich S., and Braun D. Thermal, autonomous replicator made from transfer rna. *PRL*, 2012.

[47] K. Kruger, P.J. Grabowski, A.J. Zaug, J. Sands, D.E. Gottschling, and T.R. Cech. Self-splicing rna: autoexcision and autocyclization of the ribosomal rna intervening sequence of tetrahymena. *Cell*, 31(1):147–157, 1982.

[48] H. Kuhn and J. Waser. Evolution of early mechanisms of translation of genetic information into polypeptides. 1982.

[49] J.A. Lake. Evolving ribosome structure: domains in archaebacteria, eubacteria, eocytes and eukaryotes. *Annual review of biochemistry*, 54(1):507–530, 1985.

[50] B.J. Lam and G.F. Joyce. Autocatalytic aptazymes enable ligand-dependent exponential amplification of rna. *Nature biotechnology*, 27(3):288–292, 2009.

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

[52] D.H. Lee, J.R. Granja, J.A. Martinez, K. Severin, and M.R. Ghadiri. A self-replicating peptide. *Nature*, 382(6591):525–528, 1996.

[53] T.A. Lincoln and G.F. Joyce. Self-sustained replication of an rna enzyme. *Science*, 323(5918):1229, 2009.

[54] K.A. Ludwig, D.S. Kelley, D.A. Butterfield, B.K. Nelson, and G. Früh-Green. Formation and evolution of carbonate chimneys at the lost city hydrothermal field. *Geochimica et cosmochimica acta*, 70(14):3625–3645, 2006.

[55] P.O. Lycksell, A. Gräslund, F. Claesens, LW McLaughlin, U. Larsson, and R. Rigler. Base pair opening dynamics of a 2-aminopurine substituted eco ri restricion sequence and its unsubstituted counterpart in oligonucleotides. *Nucleic acids research*, 15(21):9011–9025, 1987.

[56] S.S. Mansy, J.P. Schrum, M. Krishnamurthy, S. Tobé, D.A. Treco, and J.W. Szostak. Template-directed synthesis of a genetic polymer in a model protocell. *Nature*, 454(7200):122–125, 2008.

[57] W. Martin, J. Baross, D. Kelley, and M.J. Russell. Hydrothermal vents and the origin of life. *Nature Reviews Microbiology*, 6(11):805–814, 2008.

[58] C.B. Mast and D. Braun. Thermal trap for dna replication. *Physical review letters*, 104(18):188102, 2010.

[59] S.A. McKinney, A.C. Déclais, D.M.J. Lilley, and T. Ha. Structural dynamics of individual holliday junctions. *Nature structural biology*, 10(2):93–97, 2003.

[60] J.L. Mergny and L. Lacroix. Analysis of thermal melting curves. *Oligonucleotides*, 13(6):515–537, 2003.

[61] W. Michel, T. Mai, T. Naiser, and A. Ott. Optical study of dna surface hybridization reveals dna surface density as a key parameter for microarray hybridization kinetics. *Biophysical journal*, 92(3):999–1004, 2007.

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

[63] A.P. Minton. The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences. *Molecular and cellular biochemistry*, 55(2):119–140, 1983.

[64] A.P. Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *Journal of biological chemistry*, 276(14):10577, 2001.

[65] U.H. Mortensen, C. Bendixen, I. Sunjevaric, and R. Rothstein. Dna strand annealing is promoted by the yeast rad52 protein. *Proceedings of the National Academy of Sciences*, 93(20):10729, 1996.

[66] FG Mosqueira, G. Albarran, and A. Negron-Mendoza. A review of conditions affecting the radiolysis due to 40 k on nucleic acid bases and their derivatives adsorbed on clay minerals: Implications in prebiotic chemistry. *Origins of Life and Evolution of Biospheres*, 26(1):75–94, 1996.

[67] A.P. Nutman, C.R.L. Friend, and V.C. Bennett. Review of the oldest (4400-3600 ma) geological and mineralogical record: Glimpses of the beginning. *Episodes*, 24(2):93–101, 2001.

[68] B. Obermayer, H. Krammer, D. Braun, and U. Gerland. Emergence of information transmission in a prebiotic rna reactor. *Physical Review Letters*, 107(1):18101, 2011.

[69] A.I. Oparin. *Origin of life*. Dover Pubns, 2003.

[70] L.E. Orgel. Evolution of the genetic apparatus. *Journal of Molecular Biology*, 38(3):381–393, 1968.

[71] J. Oró. Comets and the formation of biochemical compounds on the primitive earth. 1961.

[72] A. Ott. Rna in cycles. *Physics*, 4:52, 2011.

[73] S.D. Patil and D.G. Rhodes. Influence of divalent cations on the conformation of phosphorothioate oligodeoxynucleotides: a circular dichroism study. *Nucleic acids research*, 28(12):2439–2445, 2000.

[74] Rauchfuss. *Chemische Evolution und der Ursprung des Lebens*. Springer Berlin Heidelberg, 2005.

[75] M.P. Robertson and S.L. Miller. An efficient prebiotic synthesis of cytosine and uracil. 1995.

[76] S. Rodin, S. Ohno, and A. Rodin. Transfer rnas with complementary anticodons: could they reflect early evolution of discriminative genetic code adaptors? *Proceedings of the National Academy of Sciences*, 90(10):4723, 1993.

[77] E. Rothenberg, J.M. Grimme, M. Spies, and T. Ha. Human rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proceedings of the National Academy of Sciences*, 105(51):20274, 2008.

[78] M.A. Rould, J.J. Perona, D. Soll, and T.A. Steitz. Structure of e. coli glutaminyl-trna synthetase complexed with trna (gln) and atp at 2.8 a resolution. *Science*, 246(4934):1135, 1989.

[79] R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullís, and H.A. Erlich. Primer-directed enzymatic amplification of dna with a thermostable dna polymerase. *Science*, 239(4839):487, 1988.

[80] R. Saladino, U. Ciambecchini, C. Crestini, G. Costanzo, R. Negri, and E. Di Mauro. One-pot tio2-catalyzed synthesis of nucleic bases and acyclonucleosides from formamide: Implications for the origin of life. *ChemBioChem*, 4(6):514–521, 2003.

[81] G. Schlesinger and S.L. Miller. Prebiotic synthesis in atmospheres containing ch 4, co, and co 2. *Journal of molecular evolution*, 19(5):376–382, 1983.

[82] I. Schoen, H. Krammer, and D. Braun. Hybridization kinetics is different inside cells. *Proceedings of the National Academy of Sciences*, 106(51):21649, 2009.

[83] W.J. Schreier, J. Kubon, N. Regner, K. Haiser, T.E. Schrader, W. Zinth, P. Clivio, and P. Gilch. Thymine dimerization in dna model systems: cyclobutane photolesion is predominantly formed via the singlet channel. *Journal of the American Chemical Society*, 131(14):5038–5039, 2009.

[84] M. Schulte, D. Blake, T. Hoehler, and T. McCOLLOM. Serpentinization and its implications for life on the early earth and mars. *Astrobiology*, 6(2):364–376, 2006.

[85] A.W. Schwartz. Speculation on the rna precursor problem. *Journal of theoretical biology*, 187(4):523–527, 1997.

[86] C.E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.

[87] D. Sievers and G. Von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369(6477):221–224, 1994.

[88] J. Söding and A.N. Lupas. More than the sum of their parts: on the evolution of proteins from peptides. *Bioessays*, 25(9):837–846, 2003.

[89] T. Strunz, K. Oroszlan, R. Schäfer, and H.J. Güntherodt. Dynamic force spectroscopy of single dna molecules. *Proceedings of the National Academy of Sciences*, 96(20):11277, 1999.

[90] T. Sugiyama, J.H. New, and S.C. Kowalczykowski. Dna annealing by rad52 protein is stimulated by specific interaction with the complex of replication protein a and single-stranded dna. *Proceedings of the National Academy of Sciences*, 95(11):6049, 1998.

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

[92] SP Thoms. *Ursprung des Lebens*. Fischer Kompakt, 2005.

[93] S. Thurley, L. Röglin, and O. Seitz. Hairpin peptide beacon: Dual-labeled pna-peptide-hybrids for protein detection. *Journal of the American Chemical Society*, 129(42):12693–12695, 2007.

[94] H. Trinks, W. Schröder, and C.K. Biebricher. Ice and the origin of life. *Origins of Life and Evolution of Biospheres*, 35(5):429–445, 2005.

[95] DA Usher and AH McHale. Hydrolytic stability of helical rna: a selective advantage for the natural 3', 5'-bond. *Proceedings of the National Academy of Sciences*, 73(4):1149, 1976.

[96] K. Vetsigian, C. Woese, and N. Goldenfeld. Collective evolution and the genetic code. *Proceedings of the National Academy of Sciences*, 103(28):10696–10701, 2006.

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

[98] G. Von Kiedrowski and E. Szathmáry. Selection versus coexistence of parabolic replicators spreading on surfaces. *Selection*, 1(1):173–180, 2001.

[99] T. Wang, R. Sha, R. Dreyfus, M.E. Leunissen, C. Maass, D.J. Pine, P.M. Chaikin, and N.C. Seeman. Self-replication of information-bearing nanoscale patterns. *Nature*, 478(7368):225–228, 2011.

[100] F.M. Weinert and D. Braun. An optical conveyor for molecules. *Nano letters*, 9(12):4264–4267, 2009.

[101] C.J. Wienken, P. Baaske, S. Duhr, and D. Braun. Thermophoretic melting curves quantify the conformation and stability of rna and dna. *Nucleic Acids Research*, 39(8):e52, 2011.

[102] C. Woese. The genetic code. 1967.

[103] C.R. Woese, O. Kandler, and M.L. Wheelis. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences*, 87(12):4576, 1990.

[104] M.S. Wold. Replication protein a: a heterotrimeric, single-stranded dna-binding protein required for eukaryotic dna metabolism. *Annual review of biochemistry*, 66(1):61–92, 1997.

[105] M. Yarus. A specific amino acid binding site composed of rna. *Science*, 240(4860):1751, 1988.

[106] D.Y. Zhang, A.J. Turberfield, B. Yurke, and E. Winfree. Engineering entropy-driven reactions and networks catalyzed by dna. *Science*, 318(5853):1121, 2007.

[107] S.B. Zimmerman et al. Macromolecular crowding effects on macromolecular interactions: some implications for genome structure and function. *Biochimica et biophysica acta*, 1216(2):175, 1993.

[108] S.B. Zimmerman and A.P. Minton. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annual review of biophysics and biomolecular structure*, 22(1):27–65, 1993.

[109] Geoffrey Zubay. *Origins of Life on the Earth and the Cosmos*. 2000.

Hybridization kinetics is different inside cells

Ingmar Schoen, Hubert Krammer, and Dieter Braun[1]

Systems Biophysics, Center for Nanoscience, Ludwig Maximilians Universität München, Amalienstraße 54, 80799 Munich, Germany

Edited by William A. Eaton, National Institutes of Health, Bethesda, MD, and approved October 21, 2009 (received for review February 10, 2009)

It is generally expected that the kinetics of reactions inside living cells differs from the situation in bulk solutions. Macromolecular crowding and specific binding interactions could change the diffusion properties and the availability of free molecules. Their impact on reaction kinetics in the relevant context of living cells is still elusive, mainly because the difficulty of capturing fast kinetics in vivo. This article shows spatially resolved measurements of DNA hybridization kinetics in single living cells. HeLa cells were transfected with a FRET-labeled dsDNA probe by lipofection. We characterized the hybridization reaction kinetics with a kinetic range of 10 μs to 1 s by a combination of laser-driven temperature oscillations and stroboscopic fluorescence imaging. The time constant of the hybridization depended on DNA concentration within individual cells and between cells. A quantitative analysis of the concentration dependence revealed several-fold accelerated kinetics as compared with free solution for a 16-bp probe and decelerated kinetics for a 12-bp probe. We did not find significant effects of crowding agents on the hybridization kinetics in vitro. Our results suggest that the reaction rates in vivo are specifically modulated by binding interactions for the two probes, possibly triggered by their different lengths. In general, the presented imaging modality of temperature oscillation optical lock-in microscopy allows to probe biomolecular interactions in different cell compartments in living cells for systems biology.

DNA | in vivo | molecular crowding | temperature oscillation | optical lock-in microscopy

How does a cell react to a certain stimulus? Numerous biochemical reactions are orchestrated in both space and time to transmit and process information between different locations inside a single cell. In the past, studies of signaling networks focused on the identification of molecules, their binding partners, and different mechanisms of action. Such typically static investigations in vitro are increasingly complemented by imaging techniques in vivo (1–3) and computational systems biology methods to model the dynamic response on several levels (4–7). However, the necessary reaction rates are generally taken from measurements in vitro because of the lack of quantitative in vivo data (8). To fill this information gap, methods are needed to measure the reaction kinetics in its natural context, namely inside living cells.

Experimental knowledge about how reactions are influenced by the intracellular environment is sparse. On one hand, molecules are densely packed in the cellular space and could potentially act as barriers for diffusion or restrict molecular motion. These effects are subsumed under the loosely defined term molecular crowding (9, 10). It is expected that molecular crowding affects equilibrium properties and reaction kinetics and could impair the quantitative significance of existing models (8). Often, anomalous diffusion is taken as an indicator of molecular crowding (11, 12). But how it affects the kinetics of cellular reactions has not been addressed experimentally. However, selective interactions with reaction partners might significantly modulate kinetics in vivo as compared with the situation in vitro.

Here, we describe a method termed temperature oscillation optical lock-in (TOOL) microscopy for imaging reaction kinetics in living cells with optical resolution. It allows us to compare reaction kinetics in vivo with that in vitro. We apply the method to measure hybridization of short dsDNA probes in HeLa cells and find clear indications for both accelerated and decelerated kinetics compared with identical solution measurements.

Principle of TOOL Microscopy

In general, ensemble measurements of reaction kinetics are performed by perturbing the system with an external stimulus and observing the relaxation behavior. The standard technique is to measure the response to a small temperature jump (13, 14). Here, we apply temperature oscillations instead of a jump (Fig. 1). At low oscillation frequencies, the concentrations of the reaction partners follow the stimulus instantaneously and oscillate with the temperature. At higher frequencies where the temperature oscillates faster than the reaction time constant, the concentrations oscillate with a phase delay and diminished amplitude. Both the delay time and the decreasing amplitude can be evaluated to obtain the reaction time constant. Mathematically, the exponential relaxation of the reaction translates to a transfer function in Fourier space that is used to fit the measurement data of a temperature oscillation protocol (see Eq. **A1**). The concept of oscillatory signals is widely used in electrical engineering and increasingly perceived as a useful tool to characterize biological networks and chemical reactions (15–20).

The TOOL approach requires a fluorescent readout of the reaction state (typically FRET) under external perturbation by small, laser-induced temperature oscillations (Fig. 1). The signal-to-noise ratio of fluorescence detection can be improved substantially by an optical lock-in scheme (21). Briefly, a frequency-locked stroboscopic illumination with four different delay times highlights certain phases of the reaction and allows retrieval of the amplitude and phase with a slow standard CCD camera (Fig. S1). This approach has been successfully used to image the conformational kinetics of a DNA hairpin (21). We adapted the method for the investigation of reaction kinetics in living cells (Fig. 1). Infrared laser light from below is absorbed by a chromium layer on a highly thermally conductive substrate, similar to previous approaches (22). HeLa cells can be seeded directly onto these object slides and kept in standard cell culture. Temperature oscillations with an amplitude of ≈2.5 K and a relaxation time constant of ≈200 μs are applied with a defocused infrared laser as measured with temperature-sensitive fluorescence (see *SI Text*).

Results

Monitoring DNA Hybridization with FRET. The hybridization of two complementary and antiparallel strands of nucleic acids is a specific example for a reversible bimolecular reaction. We monitored the opening and closing of 12- and 16-bp dsDNA probes with an internal FRET pair (23) of rhodamine green

Author contributions: I.S. and D.B. designed research; I.S. and H.K. performed research; I.S. and D.B. analyzed data; and I.S., H.K., and D.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: dieter.braun@physik.uni-muenchen.de.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901313106/DCSupplemental.

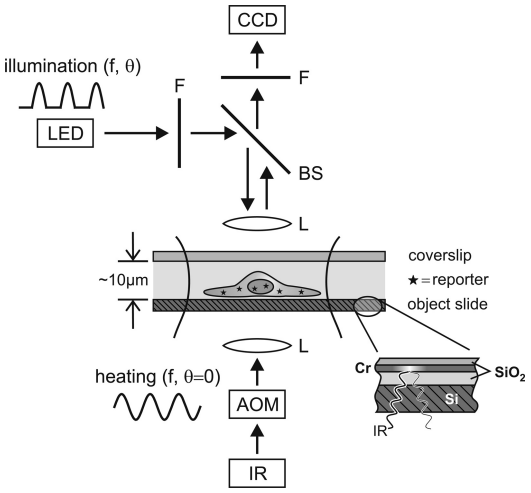

Fig. 1. TOOL microscopy. An IR laser (wavelength 1,455 nm) is heating the bottom of a cell culture chamber. Fast heat retraction is accomplished by a silicon substrate and thin chamber dimensions. Both heating and epi-illumination are modulated with a tunable phase shift θ and imaged with a standard CCD camera. AOM, acousto-optical modulator; L, lens; LED, light-emitting diode; F, filter; BS, beam splitter.

(RhG) and carboxyl-X-rhodamine (ROX). (Fig. 2*A*). Spacing at a distance of seven bases and short linkers avoided contact quenching (24). A BLAST search was performed with the sequence of the 16-mer to minimize specific interactions with genomic DNA or messenger RNA. Each strand was capped at both ends by an enantiomeric cytosine (L-nucleotide) (25) to suppress degradation by exonucleases (26). Lipofection was used for DNA delivery into the cells (26).

Excitation of the donor resulted in two emission peaks at 530 and 610 nm (Fig. S5*A*). With increasing temperature, donor fluorescence increased and FRET diminished; both yielding sigmoidal melting curves with a melting temperature of 31 °C (12-mer) and 35 °C (16-mer). As seen in Fig. 2*B*, temperature changes between 25 °C and 30 °C give rise to detectable fluorescence changes (gray underlay). The fact that the anticorrelated donor and FRET signals provide two separate measures for the same hybridization reaction is a good control for the origin of the signal.

The exogenous DNA probe was transferred into HeLa cells by lipofection and confocal images were taken to visualize its dissemination inside the cell (Fig. S5*B*). Both strands were evenly distributed over the cytoplasm and showed an enhanced concentration inside the nucleus (26). An overlay of donor and acceptor images showed that the two strands colocalized. The detection of FRET demonstrated the presence of duplexes throughout the cell and the stability of the labeled strands against degradation.

Imaging of the Reaction Kinetics. We applied periodic heating between 1 and 200 Hz to a cell transfected with the 16-mer and measured the donor and FRET signals by the described lock-in method. Both signals were then corrected against the temperature reference recorded in the vicinity of the cell (Fig. S4). The donor signal (Fig. 2*C*) described a lower half-circle in the complex plane that corresponds to a positive amplitude signal as expected from the positive slope of the RhG melting curve (see Fig. 2*B*). The transition occurred at a time constant of ≈35 ms. At high frequency, the signal settled around −0.8%/K that represents the intrinsic temperature sensitivity of the RhG dye. The FRET signal (Fig. 2*D*) shows negative amplitude in accordance with the negative slope of the FRET melting curve. The

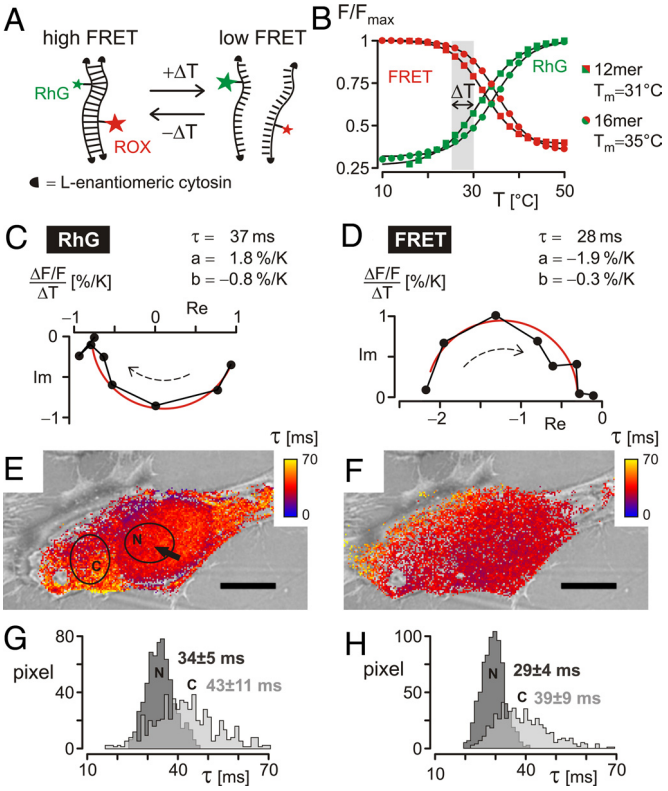

Fig. 2. Hybridization kinetics inside a single HeLa cell. (*A* and *B*) dsDNA probe design: (*A*) Complementary strands were labeled with the FRET pair RhG (donor) and ROX (acceptor) and left-handed chimeric cytosines at 3′ and 5′ ends to suppress degradation. Melting curves for the donor (green) and the FRET (red) signals were anticorrelated and fitted by melting temperatures of ≈31 °C for the 12-mer (squares) and ≈35 °C for the 16-mer (circles), respectively. (*C* and *D*) Transfer function and best fit (red line) at frequencies 1–200 Hz for a single pixel (arrow in *E*) for the 16-bp probe. The reaction amplitudes of the donor (*C*) and the FRET signal (*D*) show equal magnitude but opposite signs. (*E* and *F*) Cellular maps of the hybridization time constant show highly similar kinetics in both the donor and the FRET channel. (Scale bars: 10 μm.) (*G* and *H*) Histograms from nuclear (N) and cytoplasmic (C) regions (ellipses in *E*). Within the error bars, donor and FRET signals yielded identical results. Values are mean ± SD.

fitted time constant of 28 ms was similar to that derived from the RhG measurement. The residual temperature sensitivity of −0.3%/K is expected from the mixed sensitivities of ROX and RhG in the FRET channel. The good agreement of the fitted characteristic temperature sensitivities and the anticorrelation of the donor and FRET signals verify their origin in the hybridization reaction.

We reconstructed a cellular map of the reaction time constant for the donor and the FRET signal (Fig. 2 *E* and *F*). For both, the time constant varied between 20 and 70 ms depending on the location inside the cell. The kinetics were significantly faster inside the nucleus as compared with the rest of the cell as shown by histograms over representative regions (Fig. 2 *G* and *H*). Donor and FRET signals revealed an average time constant of 30 ms in the nucleus as compared with 40 ms in the cytoplasm.

We performed kinetic measurements with the 16-bp DNA probe for 16 individual cells and with the 12-bp DNA probe for 10 cells. Three examples of each are shown in Fig. 3 *A* and *B*, respectively. In all cases, the kinetics in the nucleus is distinctly faster than in the cytoplasm (see Fig. 3, histograms). In some cases, imaging resolved differences between nucleoli and the rest of the nucleus with slower kinetics in the nucleoli (Fig. S6), demonstrating kinetic imaging contrast of subcellular features.

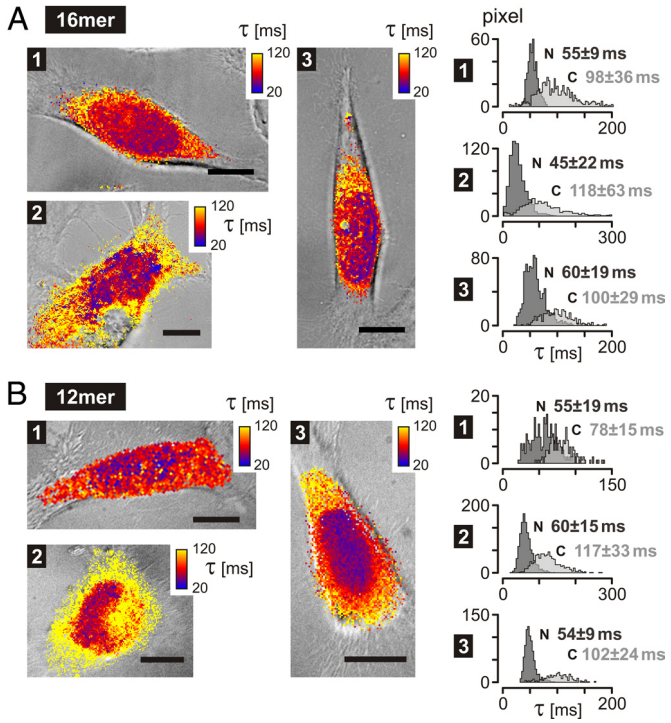

Fig. 3. Reaction speed in cellular compartments. (*A*) Shown is the 16-bp DNA probe. Maps of the reaction time constant (color-coded) for three individual cells and respective histograms from nuclear (N) and cytoplasmic (C) regions are shown. (*B*) As in *A* for a 12-bp DNA probe. Hybridization kinetics were always faster in the nucleus. (Scale bars: 10 μm.)

Concentration Dependence. Where do the above differences in the time constant arise from? One important parameter that influences the reaction speed of a second-order reaction is the concentration of reactants. If we neglect side reactions of the probe for now, we expect a time constant of the form $\tau^{-1} = k_{off} + k_{on}$ ([*D*]+[*A*]) (27) where the off rate k_{off} represents the dissociation of the duplex and its formation is determined by free donor and acceptor concentrations ([*D*], [*A*]) and an on rate k_{on}. For a 1:1 mixture ([*D*] [*A*]), and with the total DNA concentration c_{DNA} = [*A*] + [DA] we obtain

$$\tau^{-1} = \sqrt{k_{off}(k_{off} + 4k_{on}c_{DNA})} \qquad [1]$$

(see *Materials and Methods*). A plot of the inverse time constant versus the DNA concentration thus yields a horizontal half-parabola whose intercept with the τ^{-1} axis is given by the off rate and whose slope is determined by the on rate.

Confocal images of the acceptor brightness were used to measure the DNA concentration. As calibration, we determined the brightness of the DNA probe in solutions of known concentration (Fig. S7*A*). We correlated the hybridization kinetics in the nuclei or cytosolic regions of single cells with their respective DNA concentration. The data of ≈10 individual cells are plotted for the 16-mer in Fig. 4*A* and for the 12-mer in Fig. 4*C*. The reverse time constant increased with increasing concentration in agreement with Eq. **1**. On average, the data from the cytosole and nuclear regions followed the same trend, except for a few outliers for the cytosole.

For comparison, we measured the hybridization kinetics in free buffer solution. The in vitro data (Fig. 4 *B* and *D*) again followed a marked concentration dependence (Eq. **1**). Comparing the inverse time constant for the two probes, we found three times faster kinetics for the 12-mer compared with the 16-mer, as is expected from its shorter length. Surprisingly, a comparison

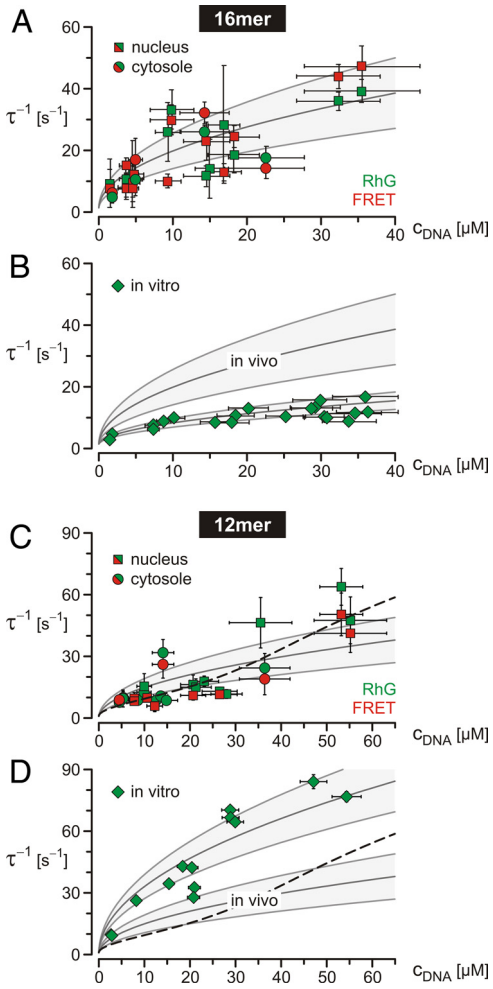

Fig. 4. Different kinetics in vivo compared with in vitro for the 16-bp DNA (*A* and *B*) and the 12-bp DNA (*C* and *D*). The inverse time constant τ^{-1} is plotted versus dsDNA concentration. Data are given as mean ± SD together with the best fit of Eq. **1** (solid line) and its 95% confidence interval (gray). In vivo experiments comprise data from nuclear and cytosolic regions of >10 individual cells. They overlapped and were treated as a single dataset during fitting. The in vivo data of the 16-mer (*A*) were well described by the parabolic fit, whereas the shape of the 12-mer data (*C*) was better described by a buffered kinetics (dashed line; see *SI Text*). As compared with in vivo, the experiments in vitro showed slower kinetics of the 16-bp DNA (*B*) whereas the kinetics of 12-bp DNA (*D*) was distinctly faster.

with the in vivo measurements revealed a completely different behavior for the two probes: whereas the kinetics of the 16-mer was distinctly faster within cells, the kinetics of the 12-mer was slowed down in vivo.

Reaction Rates. We fitted the data by Eq. **1** to elucidate the changes in the underlying reaction rates. For the 16-mer, we obtained an on rate of 2.9×10^7 $M^{-1}{\cdot}s^{-1}$ in vivo that was 7-fold larger than 4.2×10^6 $M^{-1}{\cdot}s^{-1}$ in vitro (Fig. 4 *A* and *B*, solid lines). In a distinct contrast, we obtain for the 12-mer an on rate of 5.5×10^6 $M^{-1}{\cdot}s^{-1}$ in vivo that was ≈5-fold smaller than the on rate of 2.7×10^7 $M^{-1}{\cdot}s^{-1}$ in vitro. In contrast to the in vitro data, the in vivo data of the 12-mer were not well described by the parabolic relation but rather tended to be smaller at small concentrations and larger at high concentrations. We will describe this deviation later. Because of the lack of data at very low concentrations, the off rate was ill-defined and thus held constant during fitting. The above results were independent of the

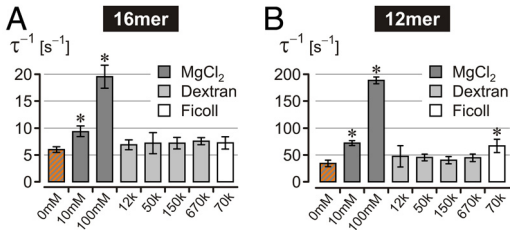

Fig. 5. Effect of divalent ions and crowding agents on the hybridization kinetics in vitro for the 16-bp probe (*A*) or the 12-bp probe (*B*). Magnesium chloride speeded up the reaction, whereas dextrans and Ficoll had only a minor impact on the kinetics. DNA concentration was 20 μM; crowding agents had a final concentration of 20% (wt/vol). Error bars represent the SEM of four independent experiments. Asterisks indicate a significant difference between the sample against the reference in pure PBS (orange) according to Student's *t* test to the level $P < 0.01$.

choice of off rates that could be equally well fitted within a range from 0.1 to 4 s^{-1}. In accordance with the melting curve in vitro and literature data (28) we fixed the off rate to 0.3 s^{-1} for the 16-mer and 1.0 s^{-1} for the 12-mer.

Kinetics with Crowding Agents. To probe the effects of molecular crowding and divalent ions, we measured both DNA probes in vitro under various buffer conditions (Fig. 5). We found a significant acceleration in the reaction kinetics for Mg^{2+}. Interestingly, crowding agents barely affected the hybridization kinetics. At a concentration of 20% (wt/vol) that approaches the situation in vivo (9), highly branched dextrans of various molecular mass or the globular molecule Ficoll-70 barely enhanced the kinetics, with the only significant difference in the case of the 12-mer. A deceleration of the hybridization kinetics was never observed for any of the tested conditions and probes.

Discussion

Unspecific Modulation of Hybridization Kinetics. The measured hybridization rates in vitro for the 16-mer and the 12-mer dsDNA (Fig. 5) agree well with literature data from other solution measurements (29–31). However, the in vivo findings for the two dsDNA probes are surprising because the kinetics of the 16-mer was substantially speeded up, whereas that of the 12-mer was considerably slowed down. As we will discuss below, this finding is hardly explained without the aid of introducing DNA binding partners.

Macromolecular crowding is considered to change reaction kinetics in two ways: on one hand, an excluded volume predicts accelerated kinetics as a result from an enhanced effective concentration (10). Estimates of the excluded volume in the cytoplasm range from 20% to 30% (9), thereby increasing the effective probe concentration and thus the on rate by a factor of 1.2–1.4. In the nucleus, crowding effects might be even more pronounced. For example, a complex nucleoprotein network confines DNA and accelerates DNA repair by homologous search (46). However, macromolecular crowding leads to hindered diffusion (9) and potentially slows down reaction kinetics. A power-law dependence of the diffusion coefficient inside the nucleus (11, 12) supports these ideas. However, our experiments neither showed significant kinetic differences between the cytosole and the nucleus (Fig. 4 *A* and *C*), nor between in vitro measurements with or without crowding agents (Fig. 5). Molecular crowding apparently has only a minor impact on the hybridization kinetics of short DNA strands.

We do not expect variations in the calibration of DNA concentration. An underestimation of the fluorescence-measured concentration of oligonucleotides in cells could be wrongly attributed to an enhanced on rate. For example, the concentration calibration can be misguided through intracellular fluorescence quenching by reducing agents (32) or bleaching. However, control measurements with glutathione and ascorbic acid at typical intracellular concentrations (33, 34) showed only an effect of a few percent on the determined concentration (Fig. S7*B*). The effect of bleaching was also small as evaluated from images taken at the beginning and the end of a cell measurement.

Hydrophobic or electrostatic interactions might differentially affect the stability of either the ssDNA, the duplex, or intermediate structures (10). For example, a differentially reduced electrostatic repulsion of ssDNA enhances the association rate of hybridization (35). Divalent ions show this effect as confirmed by in vitro measurements demonstrating a strong acceleration with increasing $MgCl_2$ concentration for both probes (Fig. 5). But even if cells do implement such considerably enhanced ionic shielding to explain the 7-fold enhanced on rate of the 16-mer, the conditions equally would affect the 12-mer.

Probe-Specific Modulation of Hybridization Kinetics. The mechanisms discussed so far cannot explain why the hybridization kinetics of the 12-bp probe was differentially modulated as compared with the 16-bp probe inside cells. In the following we discuss examples of binding partners that have the potential to distinguish between the two probes and eventually describe the differentially modified kinetics (Fig. 4 *A* and *C*).

Interference RNA, messenger RNA or freely available genomic sequences are probably able to interact with the probe dynamically and in a length-dependent manner. A randomized sequence space for these endogenous oligonucleotides would yield $4 \times 4^4 \approx 1{,}000$ times more abundant binding sites for the 12-mer as compared with the 16-mer. If this background binding shows similar on and off rates as the probe reaction, it enhances the reaction speed by adding with its concentration c_B to the probe concentration c_{DNA} in Eq. **1**. The plot of the reaction speed shifts to the left and primarily enhances the kinetics for small probe concentrations (see Fig. S8). The in vivo data did not show such an offset nor was the 12-mer stronger accelerated than the 16-mer; in fact, we observed quite the opposite. Both findings make the above scenario improbable.

Proteins in the cell can interact specific and unspecific with ssDNA and dsDNA and can thereby either speed up or slow down the reaction, depending on the type of interaction. We distinguish two cases:

Recombination mediator proteins, as for example Rad52, catalyze and thereby accelerate the hybridization of complementary ssDNA in the context of homologous recombination, DNA repair, and rescue of collapsed replication forks. Thousandfold accelerated annealing rates have been reported for Rad52 (36, 37). The annealing efficiency has been shown to be higher toward longer DNA strands, although no oligos shorter than 15 bp were investigated (38). This finding might hint toward a selective acceleration of the 16-mer in the nucleus; however, it remains unclear whether such a mechanism also exists in the cytosole as found in our measurements. In the simplest model, we account for these effects by enhancing the on rate in Eq. **1** as we did in the fit for the 16-mer (Fig. 4*A*, solid line).

Other DNA-binding proteins can stabilize ssDNA or dsDNA and thus slow down the annealing kinetics by reducing the concentration of free reactants. An example for a ssDNA-binding (SSB) protein is replication protein A (39), whereas HMG-motif proteins (40) are prototypical dsDNA-binding (DSB) proteins. Both usually tend to bind stronger to longer constructs (39, 40), and some evidence exists that they also reside in the cytosole (39). A stabilization of the DNA strands requires slower binding kinetics as compared with the probe reaction and renders the background binding quasi-stationary. We modeled the binding to either ssDNA or dsDNA with an equilibrium constant K_B and a binder concentration c_B as derived in *SI Text*.

The previously not described deceleration of the 12-mer compared with its in vitro kinetics can be convincingly fitted by either DSB interactions with $c_B = 36\ \mu M$ and $K_B = 5.4 \cdot 10^5$ (Fig. 4*C*, dashed line) or SSB interactions with $c_B = 135\ \mu M$ and $K_B = 6.1 \cdot 10^5$ (see Fig. S9*C*). Notably, both fits described the shape of the data better than the parabolic Eq. **1**.

Thus, the modulation of hybridization kinetics in vivo is likely the result of two opposite effects. On the one hand, recombination mediator proteins or divalent ions can describe the 7-fold increase of the association rate for the 16-mer in its magnitude. However, buffering of probe strands by e.g., DSB proteins can explain the slowed kinetics of the 12-mer. We implicitly assumed that the acceleration was specific for the 16-mer and the buffering only affected the 12-mer. However, both effects might act stronger, but in concert: our data are consistent with a 16-mer with strong enhancement of the association rate under minor buffering and a 12-mer with slight enhancement of the association rate and strong buffering (Fig. S9). This means that the 7-fold acceleration of the on rate and the 5-fold deceleration caused by buffering are lower bounds to the actual in vivo changes. Overall, in vivo imaging of DNA reaction kinetics in living cells indicates that hybridization kinetics is under significant differential control. In our case, factors of molecular crowding appear to be of limited importance.

TOOL Microscopy. The advantages of using the frequency space approach of TOOL compared with conventional temperature jump and time-lapse imaging are manifold: the kinetic range is not restricted by the camera speed, the method is compatible with standard fluorescence microscopy and thus applicable to cells, the signal-to-noise ratio is significantly enhanced by the lock-in approach, and neither the intrinsic temperature sensitivity of the dye nor bleaching hamper the relaxation analysis. As an established alternative, kinetics can be measured by fluorescence correlation spectroscopy (FCS) under true equilibrium conditions even in living cells (11, 12), yet with some restrictions. The short residence time of molecules in the focus restricts FCS to fast, typically intramolecular kinetics. Low concentrations are required to record an autocorrelation signal. Several extensions of the FCS principle have the potential to overcome these limitations in the future (41, 42). To date, TOOL microscopy is a useful complement of FCS-based techniques, because it works at high concentrations, has a kinetic range distinct from FCS, and provides fast and comparably simple direct imaging.

The requirements of TOOL microscopy with respect to the probe are very general: the readout can be any fluorescence property and the nature of the probe itself is of minor importance. For example, binding of small labeled DNA or RNA molecules to larger proteins could be visualized by changes in their fluorescence anisotropy. Protein–protein interactions could be monitored by FRET between two fluorescent proteins or two organic dyes fused to genetic tags (43). This will enable the comprehensive investigation of reactions in signaling networks. The kinetic range reaches from ≈10 μs to 1 s and is restricted on the one side by the retrieval of the phase at high frequencies and on the other side by long measurement times for very slow oscillations. TOOL will help to establish in vivo test systems for computational modeling and will be used to probe effects of molecular crowding in cellular compartments.

Outlook

To conclude, we developed TOOL microscopy to image kinetics at the tens of microsecond to second scale inside living cells. We measured the kinetics of DNA hybridization in the cytosole and the nucleus and found an unexpected strong, probe-dependent modification compared with in vitro measurements. The application of TOOL microscopy to protein–protein reactions is anticipated.

Materials and Methods

Lock-In Imaging. An upright microscope (Axiotech Vario; Zeiss) was equipped with a 100× oil-immersion objective (CFI Apochromat TIRF; Nikon). The beam of a fiber-coupled near-infrared laser (RLD-5–1455; IPG Laser) was modulated by an acousto-optical modulator (AA.DTS.XY.100; Pegasus Optik) and weakly focused (C240TM-C; Thorlabs) to a FWHM of 150 μm in the chamber. Illumination was provided by cyan or red light-emitting diodes (Luxeon III Star; Philips). A CCD camera (SensiCam QE; PCO) imaged with 1-s exposure time and 2 × 2 or 4 × 4 binning was used.

Signals were generated by two synchronized A/D cards (PCI-6229 and PCI-6221; National Instruments) under LabView control. The IR laser intensity followed $I_{IR} = 0.5\Delta I_{IR}[\sin(2\pi f \cdot t)+1]$ with a maximum power $\Delta I_{IR} = 2.5$ W. Illumination patterns were composed of upper halves of a sine wave $I_{LED} = \Delta I_{LED}\Theta\ [\sin(2\pi f \cdot t - \theta)]\sin(2\pi f \cdot t - \theta)$ with a phase lag θ (21). To correct for bleaching, images were taken in the order θ = 0°|180°|270°|90°|B|B|90°|270°|180°|0° (B = without illumination). Intracellular hybridization kinetics (Figs. 2–4) was investigated for f = 1…200 Hz. The temperature reference was recorded by lateral translation of the sample and recording Cy5 fluorescence next to the cell. Intracellular temperature kinetics (Fig. S2) was recorded for f = 20…1,000 Hz. Cell viability was routinely checked before and after the experiment (Fig. S3).

As a side note, the herein described optical lock-in (21) has to be distinguished from a more recent method with a similar name (44, 45) that uses filtering in Fourier space to remove nonperiodic contributions from a reversibly switched fluorescence signal.

Data Analysis. The complex transfer function $h(f)$ of the fluorescence response was reconstructed from the image series for each frequency by summation of image pairs at identical phases and applying

$$h(f) = \frac{4}{\pi}\left[\frac{I_{0^\circ} - I_{180^\circ}}{I_{0^\circ} + I_{180^\circ} - 2I_{back}} + i\frac{I_{270^\circ} - I_{90^\circ}}{I_{270^\circ} + I_{90^\circ} - 2I_{back}}\right] \qquad \textbf{[A1]}$$

(21). For temperature measurements, fluorescence amplitudes were translated to temperature changes via the Cy5 calibration. The RhG or FRET transfer function was divided by the respective temperature transfer function for each frequency (see Fig. S5). The resulting hybridization transfer function was fitted pixelwise by Eq. **1**. Artifacts from overexposed pixels were removed and poor fits were filtered out by a threshold to the mean square error. The amplitude was corrected against background fluorescence by taking the mean fluorescence intensity I_{back} in the vicinity of the cell and rescaling the fitted amplitude by $I/(I - I_{back})$, with I being the fluorescence intensity at the respective image pixel.

Transfer Function of Reaction. The temporal relaxation of the fluorescence signal after a small temperature jump adopts a single exponential time course $\Delta F(t) = a\cdot\exp(-t/\tau) + b\cdot\Theta(t)$ with a characteristic time constant τ, an amplitude a that depicts the concentration change for $t \to \infty$, and an instantaneous offset b caused by the intrinsic temperature sensitivity of the dye. The corresponding response in the frequency domain is derived from this formula by a Laplace transformation. It is given by the transfer function $h(f)$ that assigns a complex valued concentration response to a certain stimulation frequency f. The mono-exponential relaxation kinetics transforms to

$$h(f) = \frac{a}{1 + i\cdot 2\pi f\cdot\tau} + b \qquad \textbf{[A2]}$$

with a, b, and τ as above (21). A plot of $h(f)$ in the complex plane illustrates the characteristic transition (Fig. S1).

Determination of Local DNA Concentration. A multipoint confocal scanner (vtInfinity2; Visitron Systems) with an EMCCD camera (Cascade II; Photometrics) was mounted to a second output of the microscope. The illumination with either 488 or 561 nm was restricted to 5 mW to minimize bleaching. Laser intensity, scanning parameters, filters, and camera settings were identical for all confocal images, the only variable parameter was the exposure time. The fluorescence intensity of ROX was determined confocally, weighted with the exposure time, and calibrated against identically prepared chambers containing dsDNA concentrations of 3.6, 8, 16, 22, 28, and 36 μM in PBS buffer (see Fig. S7). Slides were pretreated by a plasma cleaning to prevent adsorption of DNA.

Concentration Dependence of the Time Constant. The relaxation time constant of the reaction $A + D \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AD$ is given by $\tau^{-1} = k_{off} + k_{on}\ ([D]+[A])$ (27).

Starting from the equilibrium condition $k_{off}[DA] = k_{on}[D][A]$ we set $[D] = [A]$ for the used 1:1 mixture, added $k_{off}[A]$ on both sides, and obtained $k_{off}([A] + [DA]) = k_{on}[A]^2 + k_{off}[A]$. We defined $c_{DNA} = [A] + [DA]$, replaced the bracketed term on the left side, solved for $[A]$, and found $[A] = (k_{off} + \sqrt{k_{off}^2 + 4k_{on}k_{off}c_{DNA}})/(2k_{on})$. Inserting this into the above relation for the time constant yielded Eq. **1**.

ACKNOWLEDGMENTS. We thank Hermann Gaub, Dean Astumian, Ulrich Gerland, Philipp Tinnefeld, Britta Person-Skegro, Hilmar Strickfaden, Franz Weinert, and Philipp Baaske for discussions; the referee for comments on an earlier version of the manuscript; Markku Kulomaa and Ann Fornof for critical reading and comments on early versions of the manuscript; and Philip Altpeter for technical assistance during chip fabrication. This work was supported by the Nanosystems Initiative Munich and the Center for Nanoscience.

1. Lippincott-Schwartz J, Snapp E, Kenworthy A (2001) Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* 2:444–456.
2. Miyawaki A (2003) Visualization of the spatial and temporal dynamics of intracellular signaling. *Dev Cell* 4:295–305.
3. Giepmans BNG, Adams SR, Ellisman MH, Tsien RY (2006) The fluorescent toolbox for assessing protein location and function. *Science* 312:217–224.
4. Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: Systems biology. *Annu Rev Genomics Hum Genet* 2:343–372.
5. Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L (2006) From in vivo to in silico biology and back. *Nature* 443:527–533.
6. Fisher J, Henzinger TA (2007) Executable cell biology. *Nat Biotechnol* 25:1239–1249.
7. Zhang S, Jin G, Zhang XS, Chen L (2007) Discovering functions and revealing mechanisms at molecular level from biological networks. *Proteomics* 7:2856–2869.
8. Takahashi K, Arjunan SNV, Tomita M (2005) Space in systems biology of signaling pathways: Toward intracellular molecular crowding in silico. *FEBS Lett* 579:1783–1788.
9. Ellis RJ (2001) Macromolecular crowding: Obvious but underappreciated. *Trends Biochem Sci* 26:597–604.
10. Minton AP (2006) How can biochemical reactions within cells differ from those in test tubes? *J Cell Sci* 119:2863–2869.
11. Weiss M, Elsner M, Kartberg F, Nilsson T (2004) Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophys J* 87:3518–3524.
12. Guigas G, Kalla C, Weiss M (2007) The degree of macromolecular crowding in the cytoplasm and nucleoplasm of mammalian cells is conserved. *FEBS Lett* 581:5094–5098.
13. Eigen M (1954) Methods for investigation of ionic reactions in aqueous solutions with half-times as short as 10^{-9} sec. *Faraday Discuss* 17:194–205.
14. Kirschner K, Eigen M, Bittman R, Voigt B (1966) The binding of nicotinamide-adenine dinucleotide to yeast D-glyceraldehyde-3-phosphate dehydrogenase: Temperature-jump relaxation studies on the mechanism of an allosteric enzyme. *Proc Natl Acad Sci USA* 56:1661–1667.
15. Pörschke D, Nolte G (1994) Deconvolution of electrooptical data in the frequency domain: Relaxation processes of DNA from rigid rods to coiled spheres. *Macromology* 27:590–595.
16. Lipan O, Wang WH (2005) The use of oscillatory signals in the study of genetic networks. *Proc Natl Acad Sci USA* 102:7063–7068.
17. Berthoumieux H, Jullien L, Lemarchand A (2007) Response to a temperature modulation as a signature of chemical mechanisms. *Phys Rev E* 76:056112.
18. Bennet MR, et al. (2008) Metabolic gene regulation in a dynamically changing environment. *Nature* 454:1119–1122.
19. Ingolia NT, Weissman JS (2008) Reverse engineering the cell. *Nature* 454:1059–1061.
20. Mettetal JT, Muzzey D, Gómez-Uribe C, van Ourdenaarden A (2008) The use of oscillatory signals in the study of genetic networks. *Science* 319:482–484.
21. Braun D, Libchaber A (2003) Lock-in by molecular multiplication. *Appl Phys Lett* 83:5554–5556.
22. Zondervan R, et al. (2006) Laser-driven microsecond temperature cycles analyzed by fluorescence polarization microscopy. *Biophys J* 90:2958–2969.
23. Cardullo RA, et al. (1988) Detection of nucleic acid hybridization by nonradiative fluorescence resonance energy transfer. *Proc Natl Acad Sci USA* 85:8790–8794.
24. Marras SAE, Kramer FR, Tyagi S (2002) Effciencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res* 30:e122.
25. Damha MJ, Giannaris PA, Marfey P (1994) Antisense L/D-oligodeoxynucleotide chimeras: Nuclease stability, base-pairing properties, and activity at directing ribonuclease H. *Biochemistry* 33:7877–7885.
26. Sixou S, et al. (1994) Intracellular oligonucleotide hybridization detected by fluorescence resonance energy transfer (FRET). *Nucleic Acids Res* 22:662–668.
27. Mortimer RG (2008) *Physical Chemistry* (Elsevier Academic, London).
28. Ikuta S, Takagi K, Wallace R, Itakura K (1987) Dissociation kinetics of 19 base-paired oligonucleotide-DNA duplexes containing different single mismatched base pairs. *Nucleic Acids Res* 15:797–811.
29. Parkhurst KM, Parkhurst LJ (1995) Kinetic studies by fluorescence resonance energy transfer employing a double-labeled oligonucleotide: Hybridization to the oligonucleotide complement and to single-stranded DNA. *Biochem* 34:285–292.
30. Sekar MMA, Bloch W, St John PM (2005) Comparative study of sequence-dependent hybridization kinetics in solution and on microspheres. *Nucleic Acids Res* 33:366–375.
31. Gao Y, Wolf LK, Geordiadis RM (2006) Secondary structure effects on DNA hybridization kinetics: A solution versus surface comparison. *Nucleic Acids Res* 34:3370–3377.
32. Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy* (Kluwer Academic, New York).
33. Winkler BS, Orselli SM, Rex TS (1994) The redox couple between glutathione and ascorbic acid: A chemical and physiological perspective. *Free Radical Biol Med* 17:333–349.
34. Valko M, et al. (2006) Free radicals, metals, and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact* 160:1–40.
35. Williams AP, Longfellow CE, Freier SM, Kierzek R, Turner DH (1989) Laser temperature jump, spectroscopic, and thermodynamic study of salt effects on duplex formation by dGCATGC. *Biochemistry* 28:4283–4291.
36. Mortensen UH, Bendixen C, Sunjevaric I, Rothstein R (1996) DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci USA* 93:10729–10734.
37. Sugiyama T, New JH, Kowalczykowski SC (1998) DNA annealing by Rad52 Protein is stimulated by specific interaction with the complex of replication protein A and single-stranded DNA. *Proc Natl Acad Sci USA* 95:6049–6054.
38. Rothenberg E, et al. (2008) Human Rad52-mediated homology search and annealing occurs by continuous interactions between overlapping nucleoprotein complexes. *Proc Natl Acad Sci USA* 105:20274–20279.
39. Wold MS (1997) Replication protein A: A heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* 66:61–92.
40. Bustin M (1999) Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol* 19:5237–5246.
41. Burkhardt M, Schwille P (2006) Electron multiplying CCD based detection for spatially resolved fluorescence correlation spectroscopy. *Opt Expr* 14:5013–5020.
42. Ries J, Schwille P (2006) Studying slow membrane dynamics with continuous wave scanning fluorescence correlation spectroscopy. *Biophys J* 91:1915–1924.
43. O'Hare HM, Johnsson K, Gautier A (2007) Chemical probes shed light on protein function. *Curr Opin Struct Biol* 17:488–494.
44. Mao S, et al. (2008) Optical lock-in detection of FRET using synthetic and genetically encoded optical switches. *Biophys J* 94:4515–4524.
45. Marriott G, et al. (2008) Optical lock-in detection imaging microscopy for contrast-enhanced imaging in living cells. *Proc Natl Acad Sci USA* 105:17789–17794.
46. Minsky A (2003) Strural aspects of DNA repair: The role of restricted diffusion. *Mol Microbiol* 50:367–376.

Supporting Information

Schoen et al. 10.1073/pnas.0901313106

SI Text

Materials and Methods

Substrate and Cell Culture. A silicon wafer (double side polished, 1 μm SiO_2; MEMC Electronic Materials) was coated with a 60-nm chromium and a 69-nm SiO_2 layer on top and cut into 10 mm × 10-mm slides. The slides were cleaned with detergent (5% Tickopur R36; Bandelin), rinsed with Milli-Q water, dried, sterilized by UV light for 40 min, and coated with 0.1 mg/mL poly-D-lysine (Sigma–Aldrich) in PBS (including 0.9 mM $CaCl_2$ and 1.1 mM $MgCl_2$; Biochrom) for 1 h. Slides were rinsed twice with PBS and transferred to 24-well plate chambers. HeLa cells (ATCC CCL-2; LGC Standards) were suspended in DMEM containing 10% FBS, seeded onto an object slide in 0.5 mL of medium, and cultured for 16–30 h. A slide with cells was transferred to a culture dish with PBS. After withdrawal, the fluid film was partially removed by carefully soaking with a tissue wipe at a corner of the slide. To prevent crushing of the cells during sandwiching with the coverslip, two glass filaments (Ø 7 μm, from fiber glass netting) were used as a spacer. The coverslip (18 × 25 mm) was cut from a plastic foil (Ibidi). The chamber was carefully sealed with laboratory grease (Glisseal; Borer Chemie) to prevent evaporation and mounted at the microscope.

DNA Probe and Lipofection. The 16-bp dsDNA probe consisted of the donor strand 5′-*C*AG GTT ACT ATC GTA TT*C*-3′ and the complementary strand 5′-*C*AA TAC GAT AGT AAC CT*C*-3′. The 12-bp dsDNA probe consisted of the donor strand 5′-*C*GT TAT TCC TTT AC-3′ and the complementary strand 5′-*C*TA AAG GAA TAA CC-3′. All strands were capped at both ends by a chiral cytosine (italic *C*) and labeled internally via 5-C2-amino-2′-deoxythymidine linkers (underlined *T*) with RhG and ROX, respectively. All oligos were purchased from IBA. Emission spectra were recorded with a spectrofluorometer (FluoroMax-3; Horiba Jobin Yvon). RhG showed a linear decrease over temperature with −0.6%/K; ROX fluorescence linearly increased by +0.2%/K. Before experiments, we annealed the strands in a 50 μM 1:1 mixture by a temperature ramp from 70 °C to 20 °C over 100 min. The DNA was preincubated for 18 min with lipofection agent (Roti-Fect Plus; Carl Roth) at a ratio of 50 pmol:1 μL in PBS. Between 200 and 400 pmol of DNA was added to the cells per well. After 45 min in the incubator, the medium was replaced by fresh DMEM and cells were imaged within 0.5–3 h. Labeling efficiencies were determined by UV-visible absorption to ≈70% for RhG and >95% for ROX. From the donor fluorescence at 10 °C and 50 °C, we deduced a FRET efficiency of 70% (Fig. 2*B*). The residual acceptor fluorescence at 50 °C depicted direct excitation of the acceptor. The temperature sensitivity of the dyes was determined with a fluorometer by exciting the RhG-strand/ROX-strand (each 1 μM in PBS) at 488/561 nm and averaging the fluorescence intensity over the interval from 510 to 560/575 to 720 nm. The recorded FRET spectra were corrected for the fluorescence decrease of RhG that was caused by the temperature change.

Crowding Solutions. Crowding agents were purchased from Sigma–Aldrich. The manufacturer's specification for dextrans are as follows: (*i*) 12 kDa (art. no. 31418): peak molecular weight (MP) 9890, number average molecular weight (MN) 8110, weight average molecular weight (MW) 11600, polydispersity (D) 1.43; (*ii*) 50 kDa (art. no. 31420): MP 43500, MN 35600, MP 48600, D 1.36; (*iii*) 150 kDa (art. no. 31422): MP 123600, MN 100300, MW 147600, D 1.47; and (iv) 670 kDa (art. no. 31425): 401300, MN 332800, MW 667800, D 2.01. Stock solutions of 50% (wt/vol) were prepared in distilled water and diluted to a final concentration of 20% (wt/vol). Measurement solutions further contained 15 μM Cy5 and 1× PBS without divalent ions.

Fluorescence Imaging. The filter sets (AHF Analysentechnik) used for TOOL microscopy were composed as follows: (*i*) RhG, 480/40-nm band pass, 505-nm beam splitter, 535/50-nm band pass; (*ii*) FRET, 480/40-nm band pass, 505-nm beam splitter, 575-nm long pass; and (*iii*) Cy5, 620/60-nm band pass, 650-nm beam splitter, 655-nm long pass. For confocal imaging, a blue (488 nm, 50 mW; Sapphire 488 LP; Coherent) and a green laser (561-nm, 50 mW; Jive; Cobolt) could be selected by an acousto-optical tunable filter. A dual-line beam splitter (488/561 nm) was used in combination with emission filters for RhG (525/50-nm band pass) or ROX/FRET (570-nm long pass).

Temperature Calibration of Oscillation. Temperature changes were imaged with the dye Cy5. The calibration (Fig. S2*A*) was performed between 10 °C and 96 °C in a fluorometer with excitation at 640 nm and recording within 665–720 nm. For extracellular temperature imaging, we added 0.5 μL of 15 μM Cy5 in PBS buffer during the chamber preparation. For intracellular temperature measurements, we used a ssDNA (Cy5–5′-TTT GTT TGT TTG TTT G-3′; Metabion) with the above lipofection protocol. We used the dye Cy5 as a temperature sensor because its fluorescence depends exponentially on temperature as −2.7% per Kelvin (Fig. S2*A*). Cy5-labeled ssDNA was introduced into HeLa cells by lipofection (Fig. S2*B*) (23). Amplitude and phase of the fluorescence was imaged under periodic infrared heating at different frequencies and scaled by the temperature sensitivity. The data of a single representative pixel (Fig. S2*C*) traversed a half circle in the complex plane and was well described by the transfer function (Eq. **A2**; see also Fig. S1) with a fitted time constant of ≈200 μs and a temperature amplitude of ≈2.5 K. An image of the fitted time constants confirmed a homogeneous thermal relaxation time throughout the cell (Fig. S2*D*). The observed thermal relaxation time agrees well with a finite element simulation of heat diffusion within the chamber (Fig. S2*E*). Measurements in the vicinity of the cell with Cy5 in the extracellular buffer yielded the same temperature kinetics (Fig. S4*C*). The moderate temperature amplitude allowed fitting with a linearized kinetic theory and avoided damage to the cells. Cell viability was routinely checked before and after the experiments, and no deleterious effects were observed for the chosen settings (Fig. S3).

Limits of Kinetic Range. Although the dsDNA measurements span kinetics from ≈10 to ≈300 ms, the kinetic range of the method *per se* is much wider. In principle, the kinetic range is restricted on one side by the retrieval of the phase at high frequencies and on the other side by long measurement times for very slow oscillations. As a conservative estimate for the retrieval of the phase at high frequency, let us assume that we can resolve a 90° phase lag at 2 kHz. This corresponds to a kinetic time constant of $\tau = (2\pi f)^{-1} = 80$ μs. At small frequency, the exposure time for each of the 10 images should at least comprise one oscillation period. Taking into account an initial tuning phase to the perturbation signal, the image acquisition at e.g., 0.1 Hz requires a total time of 2–3 min. Hence, reaction kinetics between ≈10

μs and 1 s can be measured, making the method especially suited for intermolecular kinetics.

Theoretical Considerations

Background Interactions with Complementary Oligonucleotides. If binding partners exist that are specific for either donor/acceptor strands and we assume the same kinetics of binding as for the probe, the concentration that will enter the probe kinetics simply can be expressed as

$$c_{total} = c_{DNA} + c_B \quad \textbf{[S1]}$$

with c_B the concentration of these additional reaction partners. This yields a time constant

$$\tau^{-1} = \sqrt{k_{off}^2 + 4k_{on}k_{off}(c_{DNA} + c_B)}. \quad \textbf{[S2]}$$

The existence of additional reaction partners will speed up the reaction as the probe more frequently meets a reaction partner and equilibrium is established faster (see Fig. S8).

Background Interactions of ssDNA with SSB Proteins. We consider a hybridization reaction between donor (D) and acceptor (A) strands according to

$$A + D \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} AD. \quad \textbf{[S3]}$$

The strand should also take part in a background reaction with binding partners B:

$$A + B \underset{k_{off}'}{\overset{k_{on}'}{\rightleftharpoons}} AB. \quad \textbf{[S4]}$$

We assume that donor and acceptor strands behave the same, namely that strand D is buffered in the same way as strand A. The concentrations of the free reactants are denoted by a, d, b and of the complexes as ad, ab, db. The conservation laws for the probe strands and the binding partners read

$$c_{DNA} = a + ab + ad. \quad \textbf{[S5]}$$

$$c_B = b + ab + db. \quad \textbf{[S6]}$$

For the assumed 1:1 mixture and equal buffering action, we also have

$$d = a \quad and \quad db = ab. \quad \textbf{[S7]}$$

In the stationary case of constant concentration ab, background binding simply reduces the probe concentration that takes part in the reaction (Eq. **S3**). We define equilibrium constants for the two reactions:

$$K = \frac{k_{on}}{k_{off}} = \frac{ad}{a \cdot d} \quad \textbf{[S8]}$$

$$K_B = \frac{k_{on}'}{k_{off}'}\frac{ab}{a \cdot b} \quad \textbf{[S9]}$$

To obtain the free probe concentration a that enters the time constant according to

$$\tau^{-1} = k_{off} + k_{on} \cdot 2a \quad \textbf{[S10]}$$

we combine Eqs. **S5**–**S9**, eliminate the variables ad, b, ab, and find the cubic equation:

$$2K_BKa^3 + (2K_B + K)a^2 + (1 + c_BK_B - 2c_{DNA}K_B)a - c_{DNA} = 0. \quad \textbf{[S11]}$$

The solutions to this equation can be determined by the method of Cardano. The choice of the root for the free probe concentration hereby depends on the values of the parameters c_{DNA},c_B,K,K_B. We implemented this solution in Mathematica and used it with Eq. **S10** to fit the experimental data (see Fig. S9).

Background Interactions of dsDNA with DSB Proteins. The derivation of free probe concentration in the case of binding partners that buffer the dsDNA is analogous to the ssDNA case. We this time describe the background reaction as

$$AD + B \underset{k_{off}'}{\overset{k_{on}'}{\rightleftharpoons}} ADB. \quad \textbf{[S12]}$$

And the conservation laws for the probe strands and the binding partners follow as

$$c_{DNA} = a + ad + adb \quad \textbf{[S13]}$$

$$c_B = b + adb. \quad \textbf{[S14]}$$

We combine Eqs. **S8**, **S9**, **S13**, and **S14**, eliminate the variables ad,b,adb, and find the equation:

$$K_BK^2a^4 + K_BKa^3 + (K + c_BK_BK - c_{DNA}K_BK)a^2 + a - c_{DNA} = 0. \quad \textbf{[S15]}$$

The solutions again were implemented in Mathematica and used with Eq. **S10** to fit the experimental data (see Fig. S9).

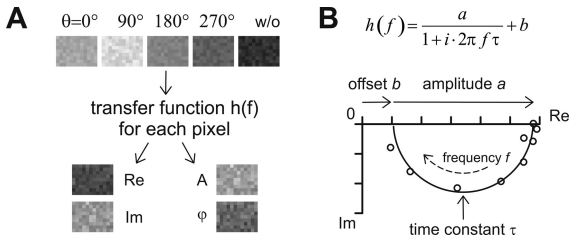

Fig. S1. Imagewise construction and graphical interpretation of the transfer function. (*A*) At a single frequency, images with four different phase lags of the illumination pattern and one background image without illumination are recorded. The complex-valued transfer function $h(f)$ is calculated from the intensity values of these five images by an analytical formula. It can be shown as two images, either divided into its real and imaginary part or, equivalently, into the absolute amplitude and a phase lag according to $h(f) = \mathrm{Re} + i\cdot\mathrm{Im} = A\cdot\exp(i\varphi)$. (*B*) Plot of the transfer function for different frequencies in the complex image plane. The transfer function clockwise traverses a half circle. Assuming small perturbations, data can be fitted by the transfer function with an amplitude *a*, a time constant τ, and a background *b*. The graphical representation allows us to directly identify these parameters in the plot: the diameter of the half circle represents the amplitude, the end of the half circle for $f\to\infty$ equals the offset, and the point of largest dissipation (where the norm of the imaginary part traverses a maximum) is reached at the frequency $f = (2\pi\tau)^{-1}$ and thus yields the typical relaxation time of the reaction.

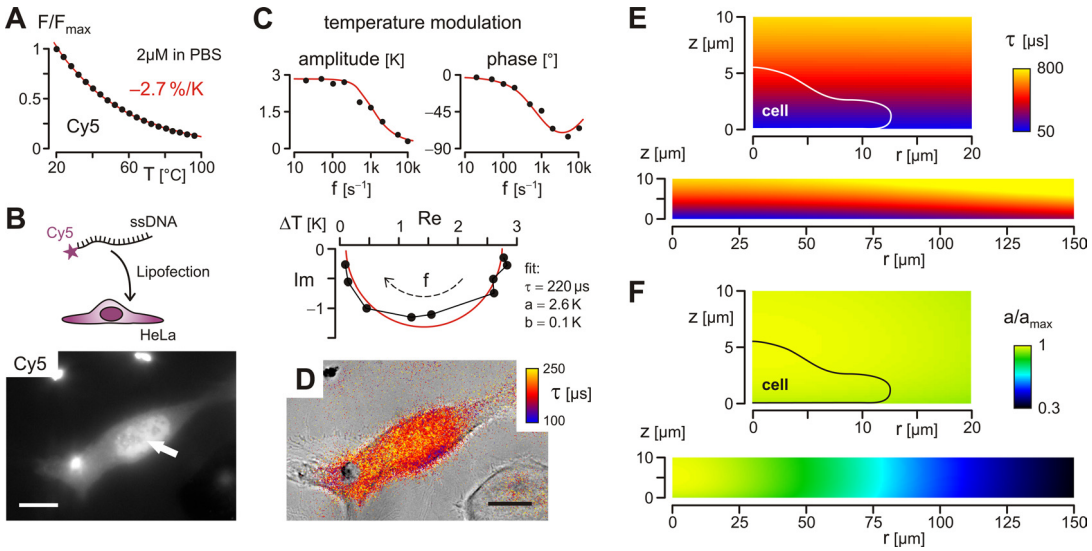

Fig. S2. Temperature oscillation. (*A*) Temperature-dependent fluorescence intensity of Cy5 dye in solution. (*B*) Lipofection of Cy5-labeled DNA into HeLa cells and epi-fluorescence micrograph of a transfected cell. (*C*) Transfer function of the temperature response at frequencies between 20 and 10 kHz. Data were taken from a representative pixel (white arrow in *B*) and fitted by Eq. **A2** (red lines). Oscillations were $\pm$ 2.5K in amplitude and sustained up to $\approx$1 kHz. (*D*) The thermal relaxation time inside the cell was homogeneous at $\approx$200 μs. (Scale bars: 10 μm.) (*E* and *F*) Finite-element simulation of the temperature characteristics inside the chamber. (*E*) Thermal relaxation time. Temperature oscillations are fastest near the substrate and damped in the upper half of the chamber by the limited velocity of heat conduction. At the site of a hypothetic cell (white silhouette) the thermal relaxation time lies between 50 and 300 μs. (*F*) Amplitude of temperature oscillations. The modulation amplitude is virtually constant in the z-direction and governed by the stationary heating profile along the radius. Throughout a hypothetic cell (black silhouette) the amplitude is nearly homogenous. Stationary solutions of the Fourier-transformed heat equations were calculated within a model of the object slide/water/plastic coverslip geometry in cylindrical coordinates using FEMLAB (Comsol AB). The heat source was located in the chromium layer and had a Gaussian profile with a FWHM of 150 μm. Solutions for different frequencies were interpolated on the points of a rectangular mesh in the water domain and fitted by the transfer function using the same routine as for the experiments.

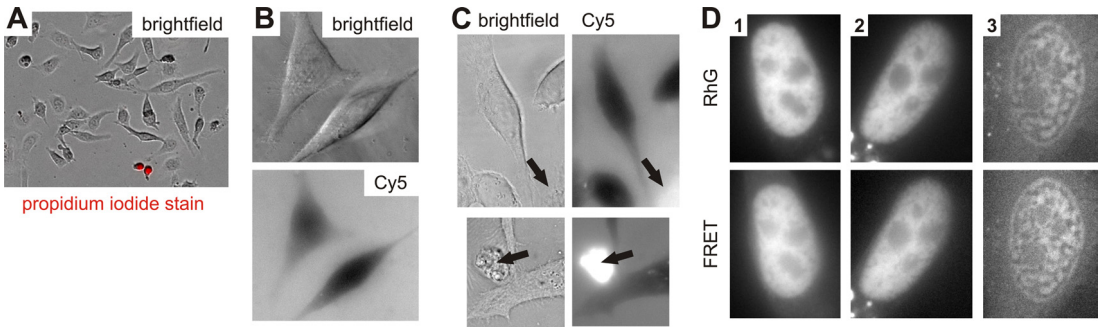

Fig. S3. Cell viability and loading artifacts. (*A*) Micrograph of HeLa cells that were cultured for 1 day on a silicon object slide coated with poly-D-lysine. The majority of cells were viable as confirmed by a propidium iodide staining of apoptotic cells. (*B*) Viable cells excluded the Cy5 dye in the extracellular buffer and showed up as darker regions in the epifluorescence image of the chamber. (*C*) Dead cells (arrows) accumulated the Cy5 dye from the buffer in their interior and were observed as bright spots in the fluorescence image. (*D*) Enhanced concentration of dsDNA in the nucleus. After lipofection, the labeled DNA was mainly accumulated inside the nucleus. Depending on the uptaken amount of DNA, the staining ranged from homogeneous (*Left*) over marmorized (*Center*) to condensated (*Right*). The latter pattern was reminiscent of hypercondensation experiments under changed osmotic conditions and such cells were excluded from analysis.

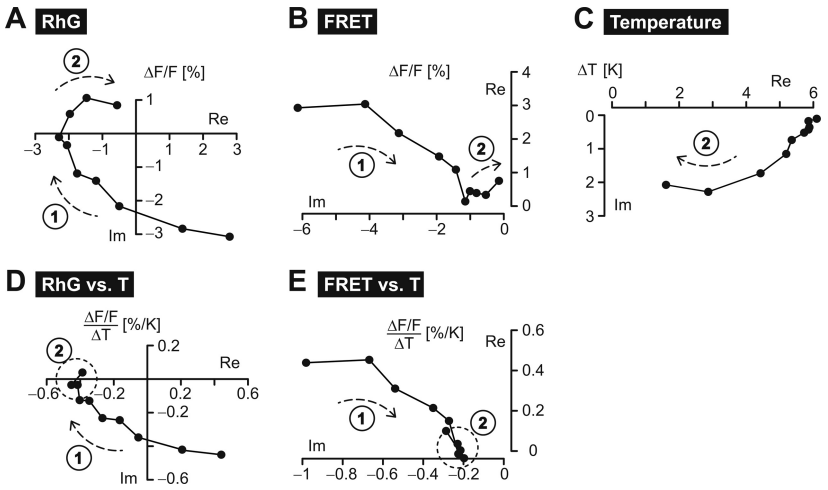

Fig. S4. Procedure for the normalization of kinetic data to the temperature response. In all shown cases, modulation frequencies lay between 1 Hz and 1 kHz. (*A*) Raw donor signal from inside a cell. Two sequential half circles were traversed: the first ① at lower frequencies had a positive amplitude and stems from the dsDNA hybridization kinetics; the second ② at higher frequencies had a negative amplitude and depicts the onset of the decrease of the temperature modulation sensed through the intrinsic temperature sensitivity of the RhG dye. (*B*) Raw FRET signal. The two sequential half circles occurred at the same frequencies as for the donor signal but the amplitude of the first transition had the opposite sign. (*C*) Temperature response derived from Cy5 signal in the buffer besides the cell. There was only one transition at high frequencies. (*D*) RhG data from *A* were divided frequencywise by the temperature data from *C*. The shape of the first transition is not affected, whereas the data at higher frequencies are compressed into a small region (dashed circle). The relative fluorescence changes are rescaled to the underlying temperature modulation. (*E*) FRET data from *B* were divided frequencywise by the temperature data from *C*. This normalization had the same effect as for the donor signal, and only the hybridization kinetics remained. Corrected data were then fitted by Eq. **A2**.

Emergence of Information Transmission in a Prebiotic RNA Reactor

Benedikt Obermayer,[1,*] Hubert Krammer,[2] Dieter Braun,[2] and Ulrich Gerland[1,†]

[1]*Arnold-Sommerfeld-Center für Theoretische Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Germany*

[2]*Systems Biophysics, Physics Department, Center for Nanoscience, Ludwig-Maximilians-Universität München, Germany*

(Received 18 March 2011; published 27 June 2011)

A poorly understood step in the transition from a chemical to a biological world is the emergence of self-replicating molecular systems. We study how a precursor for such a replicator might arise in a hydrothermal RNA reactor, which accumulates longer sequences from unbiased monomer influx and random ligation. In the reactor, intra- and intermolecular base pairing locally protects from random cleavage. By analyzing stochastic simulations, we find temporal sequence correlations that constitute a signature of information transmission, weaker but of the same form as in a true replicator.

DOI: 10.1103/PhysRevLett.107.018101 PACS numbers: 87.14.G−, 82.39.Pj, 87.15.H−, 87.23.Kg

The RNA world theory [1] posits that the first information carrying and catalytically active molecules at the origin of life were RNA-like polynucleotides [2]. This idea is empirically supported by the discovery of ribozymes, which perform many different reactions [3], among them the basic template-directed ligation and polymerization steps [4,5] necessary for replicating RNA. However, a concrete scenario of how a self-replicating RNA system could have arisen spontaneously from a pool of random polynucleotides is still lacking. Physical effects may have facilitated this step, as is believed to be the case in other transitions of prebiotic evolution [6].

From the perspective of information, an RNA replicator transmits sequence information from molecule to molecule, such that the information survives even when the original carrier molecules are degraded, for instance due to hydrolytic cleavage [7]. Rephrased in these terms, the problem of spontaneous emergence of an RNA replicator [8,9] becomes a question of a path from a short term to a lasting sequence memory. This transition occurred either as a single unlikely step or as a more gradual, multistep transition. Here, we explore a scenario of the latter type, based only on simple physicochemical processes (see Fig. 1): (i) random ligation of RNA molecules, e.g., in a hydrothermal "RNA reactor," where polynucleotides are accumulated by thermophoresis [10], (ii) folding and hybridization of RNA strands, and (iii) preferential cleavage of single- rather than double-stranded RNA segments [7]. Using extensive computer simulations and theoretical analysis, we study the behavior that emerges when these processes are combined.

Clearly, the preferential cleavage at unpaired bases effectively creates a selection pressure for base pairing in the reactor. We find that this effect increases the complexity of RNA structures in the sequence pool, which may favor the emergence of ribozymes. The underlying sequence bias also extends the expected lifetime of sequence motifs in the finite pool. Interestingly, we find that correlations between motifs persist even longer than expected. This memory effect is associated with information transmission via hybridization. Intriguingly, these correlations have the same statistical signature as templated self-replication, only weaker. In this sense, the RNA reactor could constitute a stepping-stone from which a true RNA replicator could emerge, e.g., assisted by a primitive ribozyme catalyzing template-directed synthesis.

RNA reactor.—As illustrated in Fig. 1, we envisage an open reaction volume V under nonequilibrium conditions as, e.g., inside a hydrothermal pore system where polynucleotides are strongly accumulated by a combination of convective flow and thermophoresis [10]. At any point in time, the reaction volume contains various sequences S_L of length L. The full time evolution of this pool is a stochastic process with the reactions

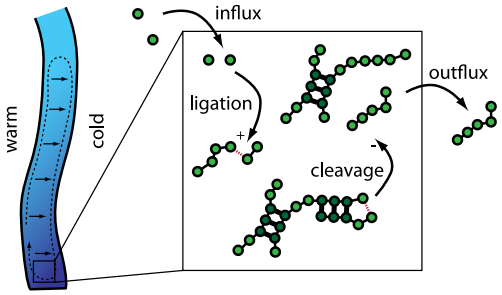

FIG. 1 (color online). Illustration of the RNA reactor. Left: Combined action of convection and thermophoresis in narrow pores subject to a temperature gradient results in strong accumulation of nucleotides, as indicated by the darker shading. Right: The region of high concentration defines an open reaction volume where nucleotides enter and bonds are formed through ligation reactions. Equilibrium base-pair formation protects bonds next to paired nucleotides (dark) from cleavage. Length-dependent outflux accounts for the preferential accumulation of long molecules.

$$\varnothing \xrightarrow{J} S_1, \qquad S_L \xrightarrow{d_L} \varnothing, \tag{1a}$$

$$S_L + S_K \xrightarrow{\alpha} S_{L+K}, \qquad S_L \xrightarrow{\beta_{L,K}} S_K + S_{L-K}. \tag{1b}$$

We assume a constant and unbiased influx of monomers (ACGU) at rate J. The effective outflux rate $d_L = d_0 e^{-(L/L_c)^{1/2}}$ accounts for the strong accumulation of nucleotides in a pore system, with a characteristic length dependence determined by the length scale L_c, which comprises parameters such as Soret coefficient, temperature gradient, and geometry [11]. Ligation of monomers or oligomers occurs at fixed rate α [12]. Finally, the most essential ingredient is a backbone cleavage process with a rate that depends on the base-pairing probability of the neighboring bases, such that double-stranded RNA is more stable than single-stranded RNA. Specifically, we calculate the cleavage rate $\beta_{L,K} = \beta_0(1 - p_{L,K})$ at backbone bond K using the average base-pairing probability $p_{L,K}$ of the two neighboring bases. We allow both intramolecular base pairs within single sequences and intermolecular base pairs within duplexes of any two molecules. RNA folding is performed by means of the Vienna package [15,16], where the partition function of the entire ensemble is calculated assuming chemical equilibrium [17], warranted by the fast hybridization kinetics [8].

We use the standard Gillespie algorithm to simulate the stochastic dynamics (1) of the sequence pool. The cleavage rate $\beta_{L,K}$, which is recalculated from the folding output for all molecules whenever necessary, effectively introduces a selection for base-pair formation. Since RNA folding depends on the temperature T and duplex formation is also concentration-dependent, we can vary the selection pressure via $p_{L,K}(T, V)$. We consider the reactions (1) under different possible conditions, with two different temperatures (a cold system at 10 °C and a hot environment at 60 °C) and concentrations (in the pM and mM range, respectively). To study the differences from a random pool, we also consider a "neutral" scenario without folding ($p_{L,K} = 0$). These scenarios are chosen mainly to highlight the effects of base pairing and not to suggest specific environmental conditions at the origin of life.

Stationary length and shape distribution.—Disregarding sequence-dependent selection, the ligation-cleavage dynamics of the RNA reactor resembles the kinetics of cluster aggregation and fragmentation. Hence, the stationary sequence length distribution shown in Fig. 2(a) corresponds to a cluster size distribution, and its moments can be obtained using established methods [16,18]. In the limit of large influx J, the average total molecule number $\langle N_{\rm tot}\rangle$ and their mean length $\langle L\rangle$ are given by

$$\langle N_{\rm tot}\rangle = \sqrt{\frac{J(d_0 + \beta_0)}{\alpha d_0}}, \qquad \langle L\rangle = \sqrt{\frac{J\alpha}{d_0(\beta_0 + d_0)}}, \tag{2}$$

where we have neglected the length dependence of the outflux ($L_c \to \infty$; a finite value for L_c shifts both $\langle N_{\rm tot}\rangle$ and $\langle L\rangle$ to larger values without strongly affecting the shape of the distribution). These analytical results readily explain why with stronger selection the total number of molecules decreases, but their mean length goes up [see Figs. 2(b) and 2(c)]: the cleavage rate β_0 is reduced as the mean base-pairing probability $\langle \bar{p}_L\rangle$ is increased especially for longer sequences [cf. Fig. 2(d)], and the distribution thus gains more weight in the tail of long sequences.

In order to characterize the structural repertoire of this RNA pool, we focused on the tail of the length distribution and analyzed the secondary structures of long sequences with $L > L^*$. We performed the analysis for $L^* = 35$ as well as $L^* = 50$ (the length of the minimal hairpin ribozyme [19]). Figure 2(f) shows the probability to observe structures within basic "shape" classes [20], such as hairpins or hammerheads [21]. We observe a significant enrichment of complex structures under selection compared to the neutral case defined above.

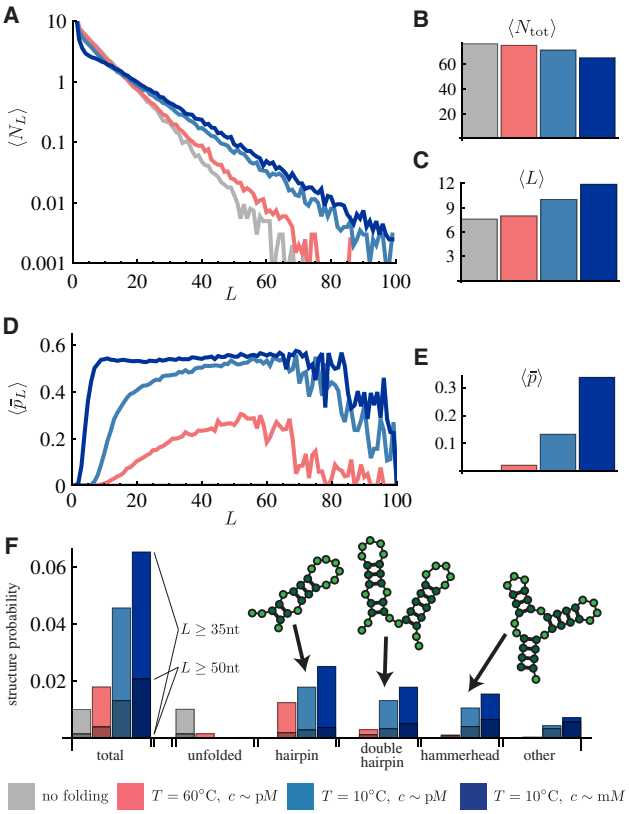

FIG. 2 (color online). Steady-state properties of the sequence pool: (a) length distribution $\langle N_L\rangle$, (b) total number $\langle N_{\rm tot}\rangle$ of molecules, and (c) their mean length $\langle L\rangle$. (d) Base-pairing probability $\langle \bar{p}_L\rangle$ averaged over sequences of length L, with mean $\langle \bar{p}\rangle$ shown in (e). (f) Structural repertoire of long sequences: steady-state probabilities for sequences longer than $L^* = 35$ (shaded parts: $L^* = 50$), which fold into a structure of similar shape as the indicated schematic drawings. Selection strength increases from light to dark color as indicated in the legend. All observables are averaged over time and 10 independent replicas. Remaining parameter values were $J = 1$, $\alpha = 0.001$, $\beta_0 = 0.01$, $d_0 = 0.005$, $L_c = 10$.

Information transmission via hybridization.—Base pairing and the ensuing correlations between sequences occur mostly within relatively short sequence regions. Therefore, we focus on the dynamics of shorter subsequences or "sequence motifs" of length ℓ, which are informational entities not tied to a specific molecule. From our simulations, we extract time trajectories for the copy numbers $n_i(t)$ of all 4^ℓ different motifs. Even for fairly small $\ell > 3$, the sequence space of motifs is not fully covered in the finite ensemble; i.e., an average motif copy number is typically $\langle n_i(t)\rangle \ll 1$. Hence, signatures of information transmission should appear as an unexpected increase in the lifetime of these motifs. Suitably averaged observables are provided by the auto- and cross-correlation functions, $C_a(t) = 4^{-\ell}\sum_i \langle n_i(t) n_i(0)\rangle$ and $C_c(t) = 4^{-\ell}\sum_i \langle n_i(t) n_i^*(0)\rangle$, respectively, where n_i^* is the copy number of a motif's (reverse) complement [21]. Figures 3(a) and 3(b) show data for these correlation functions for $\ell = 6$ and the parameter set used in Fig. 2.

The observed motif correlations can be understood in the framework of a simple stochastic process. Motifs are created when sequence ends are ligated together and destroyed by cleavage [22]. Using a mean-field-type approach, we pick an arbitrary probe motif with copy number $n(t)$. Its dynamics is described by a birth-death process, where $n(t)$ is increased with constant rate k_+ and decreased with linear rate k_- [see schema (i) in Fig. 3(c)]. The birth rate k_+ can be computed from the steady-state length distribution $\langle N_L\rangle$ by counting how many ends of long enough molecules are available for ligation. Assuming an annealed random ensemble, we obtain

$$k_+ = \frac{\alpha}{4^\ell}\sum_{k=1}^{\ell-1}\sum_{L\geq k}\langle N_L\rangle \sum_{L'\geq \ell-k}\langle N_{L'}\rangle. \tag{3}$$

The death rate k_- comprises the effects of cleavage and hybridization. A motif is cleaved with rate β_0 at any of its $\ell - 1$ bonds, but this rate is reduced by the effective base-pairing probability of its parent sequence, which in turn depends on the selection strength. On average, this reduction follows from averaging over the length and base-pairing probability distributions $\langle N_L\rangle$ and $\langle \bar{p}_L\rangle$ of parent sequences, respectively. This gives the result

$$k_- = \beta_0(\ell-1)\left[1 - \frac{\sum_{L\geq\ell}(L-\ell+1)\langle \bar{p}_L\rangle\langle N_L\rangle}{\sum_{L\geq\ell}(L-\ell+1)\langle N_L\rangle}\right]. \tag{4}$$

However, a birth-death process based on these two effective rates alone necessarily fails to describe cross-correlations between a motif and its complement [23]. The reduction in the cleavage rate of a particular motif due to hybridization is conditional on the presence of its complementary partner. Hence, we modulate the average death rate k_- with an additional factor $h(x) \leq 1$, which accounts for the probability of hybridization and depends on the number $x = n^*/n$ of available complements per motif. Since the average hybridization probability is small under the conditions considered here, it will be proportional to x. This leads us to a linear ansatz $h(x) \approx 1 - (r/k_-)x$, where the significance of the coefficient r will shortly become apparent. We find that in the "hybridization process" of Fig. 3(c), the expected copy number $\langle n\rangle$ of a motif obeys

$$\partial_t\langle n\rangle = k_+ - k_-\langle n h(n^*/n)\rangle \approx k_+ - k_-\langle n\rangle + r\langle n^*\rangle. \tag{5}$$

A symmetric equation holds for $\langle n^*\rangle$. Strikingly, this result is identical to the corresponding rate equations for a "replication process" [16], where motifs are born with rate k_+, destroyed with fixed rate k_-, and copied from their complements with rate r, as in schema (ii) of Fig. 3(c). This observation suggests that we may interpret the coefficient r as an apparent replication rate for motifs in the RNA reactor.

To validate this interpretation, and to measure the apparent replication rate in our simulations, we calculate the correlation functions of the hybridization process using the same approximation for $h(x)$ [16], yielding

$$C_{a,c}(t) = \frac{k_+^2}{(k_- - r)^2} + \frac{k_+ e^{-(k_- - r)t}}{2(k_- - r)} \pm \frac{k_+ e^{-(k_- + r)t}}{2(k_- + r)}. \tag{6}$$

In Figs. 3(a) and 3(b), we used these expressions with the rates k_+ and k_- calculated from Eqs. (3) and (4), and with r as the only free parameter fitted simultaneously to both data sets. The equivalence between the hybridization and

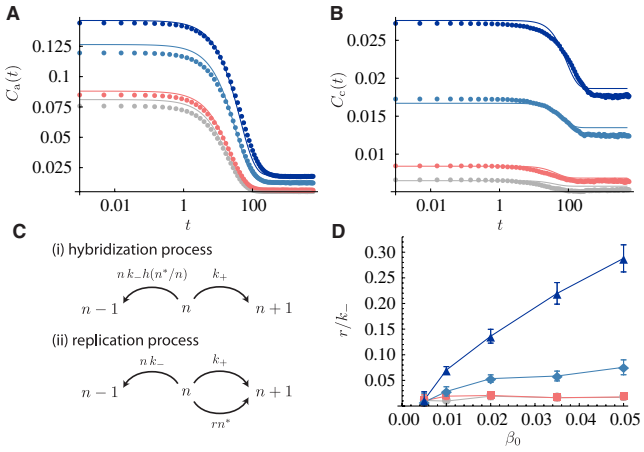

FIG. 3 (color online). Information transmission among sequence motifs. (a) and (b) Auto- and cross-correlation functions $C_{a,c}(t)$ from simulation data for $\ell = 6$ (dots) together with analytical expressions from Eq. (6) (solid lines). The rates k_- and k_+ have been computed from Eqs. (3) and (4), with r as the only fit parameter. (c) Schemata for different birth-death processes: (i) motifs are created with constant rate k_+ and destroyed with linear rate $k_- h(n^*/n)$, which is reduced by hybridization to their complements; (ii) motifs are destroyed with fixed rate k_-, but are copied from their complements with rate r. To leading order in r/k_-, both processes give rise to identical correlation functions $C_{a,c}(t)$, where a nonconstant $C_c(t)$ signifies information transmission between a motif and its complement. (d) Dependence of the replication efficiency r/k_- on the cleavage rate β_0 (error bars indicate 95% confidence intervals). Color code as in Fig. 2.

the replication processes is also exhibited by their correlation functions to leading order in r/k_- [16]. Hence, the good agreement with the simulation data indicates that the observed motif correlations are virtually indistinguishable from those expected for inefficient template-directed replication. The replication efficiency r/k_- determined by the fits is plotted in Fig. 3(d) as a function of the bare cleavage rate β_0 for the different conditions. Remarkably, it reaches levels close to 30% in the cold and highly concentrated environment, where base pairing via duplex formation is favorable. Note that a true (exponential) replicator would require that motifs are copied faster than they are degraded ($r > k_-$), while our system with $r < k_-$ is an inefficient realization.

These findings show that protection against cleavage due to folding and hybridization leads to an extended sequence memory in the RNA reactor. One global contribution to this longer motif lifetime is due to the ''protection factor'' in square brackets in Eq. (4), which renormalizes the bare cleavage rate to account for the average probability that a motif is paired. Another contribution stems from the correlation time in Eq. (6), which is increased as the apparent replication rate is subtracted from the renormalized cleavage rate, such that $C_{a,c}(t)$ decays on time scales of order $(k_- - r)^{-1}$. This specific increase occurs only when a motif and its complement mutually protect each other, and it therefore demonstrates the emergence of information transmission.

Conclusions.—We have analyzed stochastic simulations of a minimal prebiotic RNA reactor, where formation of double strands protects sequence parts from degradation. On the one hand, this selection for structure biases the resulting pool towards longer and more structured sequences, favoring the emergence of ribozymes. On the other hand, it leads to a weak apparent replication process based on ''information transmission by hybridization,'' conceptually similar to ''sequencing-by-hybridization'' techniques [24]. Together, the structural complexity and the information transmission featured in the RNA reactor suggests this type of system as plausible intermediate for the emergence of a true replicator with $r > k_-$. For instance, some of the relatively frequent simple structures observed in our simulation are similar to known ligase ribozymes [3]. This functionality in turn would facilitate the creation of more complex molecules from essential modular subunits [25]. Once ribozymes emerge, a self-replicating system could be established by template-directed ligation of suitably complementary oligomers [4]. So far, it has remained unclear how such autocatalytic RNA systems would be supplied with appropriate oligomer substrates. However, the strong cross-correlations observed in the RNA reactor demonstrate a significantly enhanced chance of finding sequences complementary to those present in the pool, including the sequence to be replicated. Thus, the RNA reactor acts as an adaptive filter to preferentially keep potentially useful substrate sequences. This adaptive selectivity would allow for the ''heritable'' propagation of small variations and thus endow the replicator with basic evolutionary potential.

This work was supported by the Nanosystems Initiative Munich (NIM), by a DAAD grant to BO, and by a DFG grant to UG.

*Present address: Department of Physics, Harvard University, Cambridge MA 02138, USA.
†gerland@lmu.de

[1] W. Gilbert, Nature (London) **319**, 618 (1986).
[2] L. Orgel, Crit. Rev. Biochem. Mol. Biol. **39**, 99 (2004).
[3] J. Doudna and T. Cech, Nature (London) **418**, 222 (2002).
[4] N. Paul and G. F. Joyce, Proc. Natl. Acad. Sci. U.S.A. **99**, 12 733 (2002).
[5] W. Johnston *et al.*, Science **292**, 1319 (2001).
[6] I. Chen, R. Roberts, and J. Szostak, Science **305**, 1474 (2004).
[7] D. Usher and A. Mchale, Proc. Natl. Acad. Sci. U.S.A. **73**, 1149 (1976).
[8] C. Fernando, G. von Kiedrowski, and E. Szathmáry, J. Mol. Evol. **64**, 572 (2007).
[9] M. Nowak and H. Ohtsuki, Proc. Natl. Acad. Sci. U.S.A. **105**, 14 924 (2008).
[10] P. Baaske *et al.*, Proc. Natl. Acad. Sci. U.S.A. **104**, 9346 (2007).
[11] S. Duhr and D. Braun, Proc. Natl. Acad. Sci. U.S.A. **103**, 19 678 (2006).
[12] While nontemplated ligation occurs spontaneously [13] or via inorganic catalysis [14], we neglect template-directed reactions, which are less plausible in early prebiotic chemistry in the absence of ribozymes [2].
[13] S. Pino, F. Ciciriello, G. Costanzo, and E. Di Mauro, J. Biol. Chem. **283**, 36 494 (2008).
[14] J. Ferris and G. Ertem, J. Am. Chem. Soc. **115**, 12 270 (1993).
[15] I. Hofacker *et al.*, Monatsh. Chem. **125**, 167 (1994).
[16] See supplemental material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.018101 for more details on the algorithm and the calculations, as well as supplementary results regarding GU pairs, self-complementarity, and shorter motifs.
[17] S. H. Bernhart *et al.*, Algorithms Mol. Biol. **1**, 3 (2006).
[18] R. Li, B. J. McCoy, and R. B. Diemer, J. Colloid Interface Sci. **291**, 375 (2005).
[19] A. Hampel and R. Tritz, Biochemistry **28**, 4929 (1989).
[20] R. Giegerich, B. Voss, and M. Rehmsmeier, Nucleic Acids Res. **32**, 4843 (2004).
[21] Results shown in Figs. 2 and 3 were obtained disallowing ambiguous GU wobble pairs. See [16] for the length and shape distribution including GU pairs.
[22] Since most motifs live on long sequences, we can neglect outflux reactions $\propto d_0 e^{-\sqrt{L/L_c}}$ against cleavage $\propto \beta_0 L$.
[23] The presence of self-complementary sequences in a finite ensemble, which obey different statistics inherited by the corresponding motifs, leads to small cross-correlations even in the neutral case. See [16] for more details.
[24] R. Drmanac *et al.*, Advances in Biochemical Engineering/Biotechnology **77**, 75 (2002).
[25] C. Briones, M. Stich, and S. C. Manrubia, RNA **15**, 743 (2009).

Emergence of information transmission in a prebiotic RNA reactor – Supplementary Information

Benedikt Obermayer,[1,*] Hubert Krammer,[2] Dieter Braun,[2] and Ulrich Gerland[1,†]

[1] *Arnold-Sommerfeld-Center für Theoretische Physik and Center for NanoScience, Ludwig-Maximilians-Universität München, Theresienstr. 37, 80333 München, Germany*

[2] *Systems Biophysics, Physics Department, Center for Nanoscience, Ludwig-Maximilians-Universität München, Amalienstr. 54, 80799 München, Germany*

[*] Present address: Department of Physics, Harvard University, Cambridge MA 02138, USA.

[†] gerland@lmu.de

I. SUPPLEMENTARY FIGURES

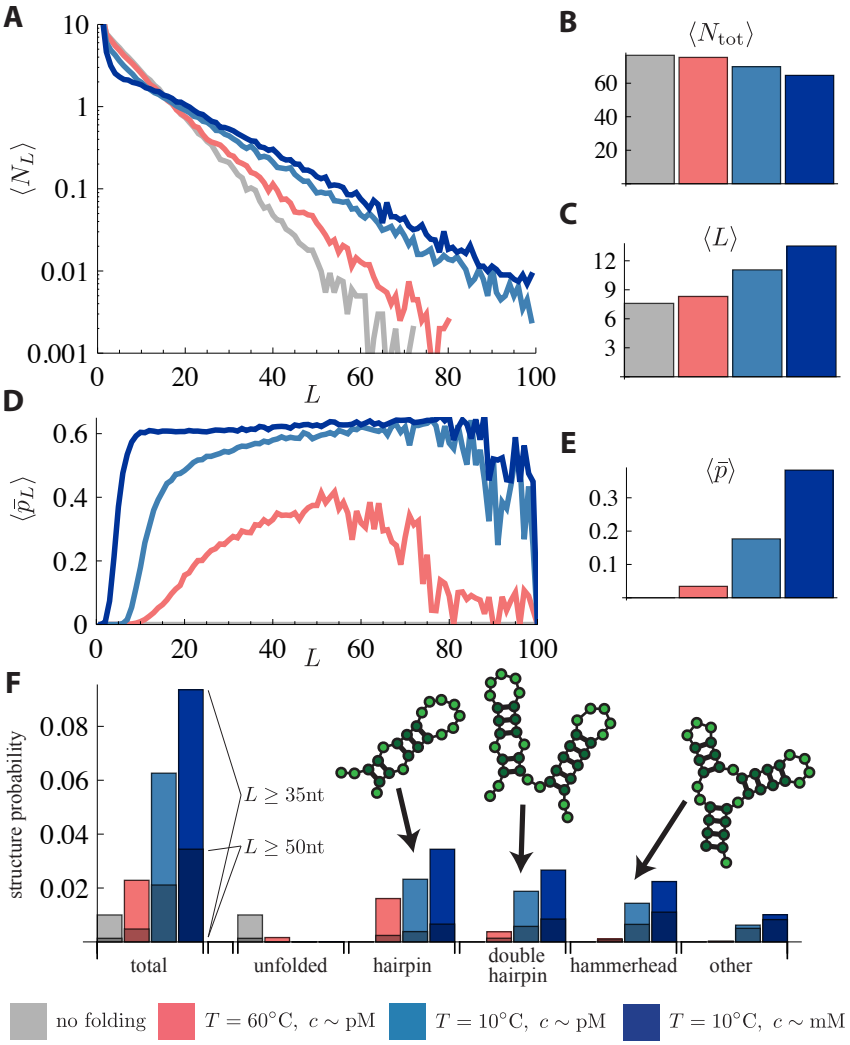

FIG. S1. GU pairs. Properties of the steady-state ensemble as in Fig. (1) of the main text, but in a simulation including GU wobble pairs. (a) length distribution $\langle N_L \rangle$, (b) total number $\langle N_{\rm tot} \rangle$ of molecules, and (c) their mean length $\langle L \rangle$. (d) base pairing probability $\langle \bar{p}_L \rangle$ averaged over sequences of length L, with mean $\langle \bar{p} \rangle$ shown in (e). (f) Structural repertoire of long sequences. While the differences to the results without GU pairs shown in Fig. 2 in the main text are comparably small, we observe that this additional pairing mode provides additional stability especially for longer RNA and thus further increases the chances of finding structured molecules in random pools.

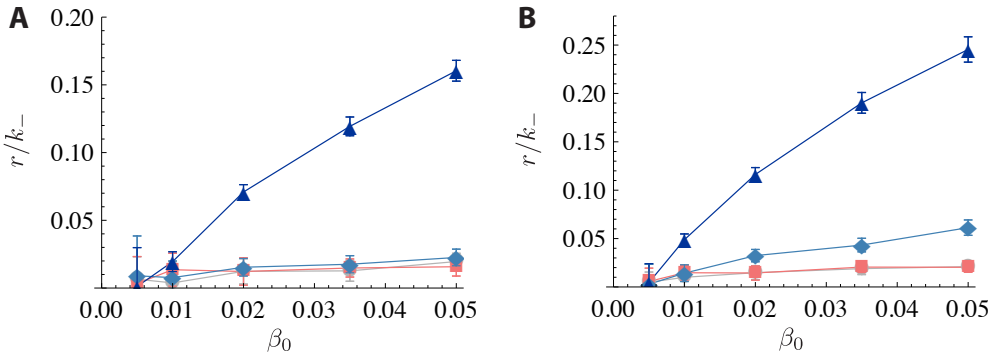

FIG. S2. Shorter motifs. Dependence of the replication efficiency r/k_- on the bare cleavage rate β_0 as in Fig. (3c) in the main text, but for shorter motifs of length $\ell = 4$ (a) and $\ell = 5$ (b).

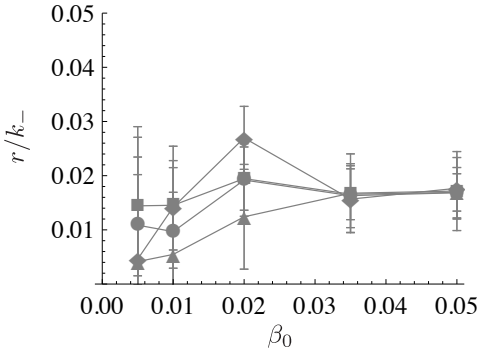

FIG. S3. Analysis of the influence of self-complementary sequences. Self-complementary sequences in the pool give rise to different motif statistics. To test this effect, we ran control simulations without RNA folding but fixed sequence-independent base pairing probabilities $\langle \bar{p}_L \rangle$ chosen from the distribution measured in the full simulation (cf. Fig. 1(d) in the main text). This leads to almost identical length statistics in the sequence ensemble, but motif correlations due to hybridization are absent. Shown is the dependence of the apparent replication efficiency r/k_- on the bare cleavage rate β_0 as in Fig. (3c) in the main text. Self-complementarity gives rise to subdominant cross-correlations resulting in small non-zero values for r largely independent of the "selection strength" (note the different scale on the ordinate).

II. IMPLEMENTATION DETAILS: CALCULATION OF BASE PAIRING PROBABILITIES

Our code is based on the Gillespie algorithm [1] for the stochastic simulation of chemical reactions. At each time step, we compute the propensities for each of the four possible reactions involving sequences $S_{L,i}$ of length L_i that are present in $N_{L,i}$ copies:

1. influx of a monomer with propensity J (monomers are chosen randomly among the four nucleotides A,C,G, and U);

2. outflux of one of $N_{L,i}$ copies of $S_{L,i}$ with propensity $d_0 N_{L,i} \mathrm{e}^{-\sqrt{L_i/L_c}}$;

3. ligation of two sequences $S_{L,i}$ and $S_{L,j}$ to a combined sequence of length $L_i + L_j \leq L_{\max}$ with propensity $\alpha N_{L,i}(N_{L,j} - \delta_{ij})$. We scale out the volume dependence of ligation to facilitate comparison of different scenarios, which operate at different concentrations. Also, we restrict ligation to sequences with combined length smaller than $L_{\max} = 100$ to limit computationally expensive RNA folding.

4. cleavage of one of $N_{L,i}$ copies of a sequence $S_{L,i}$ at position K with propensity $\beta_{L_i,K} = N_{L,i}\beta_0(1 - p_{L_i,K}(T,V))$.

One event is randomly chosen according to its relative propensity, and time is updated by a time interval drawn from an exponential distribution with a mean equal to the inverse of the sum of the propensities.

While the first three steps are straightforwardly implemented, the cleavage reaction involves the sequence-specific, temperature- and concentration-dependent probability $p_{L_i,K}(T,V)$ that the nucleotides next to bond K are paired. The calculation is done by means of the Vienna package for RNA secondary structure folding [2]. We allow both intra- and intermolecular base pairs in complexes involving at most two sequences. To simplify the following argument, we omit the length index on the sequences S_i. For each sequence S_i, we calculate the simplex partition sum Z_i for all possible secondary structures of that sequence, and the corresponding duplex partition sums Z_{ij} that result from folding a duplex involving two molecules S_i and S_j. Note that duplex formation is concentration dependent, and we therefore need to calculate the partition sum $\mathcal{Z}$ of the ensemble of sequences [3–5]. If each sequence is initially present in n_i^0 copies, and the ensemble after hybridization will contain

n_i simplex structures and n_{ij} duplex structures, this partition sum can be written as:

$$\mathcal{Z} = \prod_i \frac{n_i^0!}{n_i! \prod_{j \le i} n_{ij}!} Z_i^{n_i} \prod_{j \le i} Z_{ij}^{n_{ij}}, \tag{1}$$

under the mass conservation constraint that each sequence be part of at most one complex at the same time:

$$n_i + 2n_{ii} + \sum_{j \neq i} n_{ij} = n_i^0. \tag{2}$$

The chemical equilibrium is obtained by minimizing the ensemble free energy $\mathcal{F} = -k_\mathrm{B}T \ln \mathcal{Z}$ with respect to the variables n_i and n_{ij}, under the constraint Eq. (2). Even though in our relatively small system these variables are all small numbers, we can efficiently perform this calculation only in the thermodynamic limit, assuming rapid chemical equilibration due to the very fast hybridization kinetics [6] and the convective flow cycles encountered in the thermal trap. Hence, we switch to concentration variables $c_i = n_i/V$ in a volume V (correspondingly for c_i^0 and c_{ij}).

Following Ref. [5], we now introduce Lagrange multipliers λ_i (which are chemical potentials measured in units of $k_\mathrm{B}T$), and minimize $\mathcal{L} = \mathcal{F}/k_\mathrm{B}T + \sum_i \lambda_i (c_i^0 - c_i - 2c_{ii} - \sum_{j \neq i} c_{ij})$ instead. Using Stirling's formula, this requires finding the minimum of

$$\begin{aligned} \mathcal{L}(c, \lambda) = \sum_i \Big[& c_i^0 (1 - \ln c_i^0 + \lambda_i) - c_i (1 - \ln c_i + \ln Z_i + \lambda_i) \\ & - \sum_{j \le i} c_{ij} (1 - \ln c_{ij} + \ln Z_{ij} + \lambda_i + \lambda_j) \Big]. \end{aligned} \tag{3}$$

The minimum is given by

$$c_i^* = Z_i \mathrm{e}^{\lambda_i^*}, \quad c_{ij}^* = Z_{ij} \mathrm{e}^{\lambda_i^* + \lambda_j^*}, \tag{4}$$

where the stationary values λ^* for the chemical potentials are obtained from minimizing

$$f(\lambda) = -\mathcal{L}(c^*, \lambda) = \sum_i \left[c_i^0 (\ln c_i^0 - 1 - \lambda_i) + Z_i \mathrm{e}^{\lambda_i} + \frac{1}{2} \sum_j (1 + \delta_{ij}) Z_{ij} \mathrm{e}^{\lambda_i + \lambda_j} \right]. \tag{5}$$

Even though this lower-dimensional problem is in principle not ill-conditioned [5], the minimization becomes numerically unstable for large systems on the order of 100 molecules with possibly very different hybridization energies. A stable code was obtained by using the L-BFGS library [7] implementing the limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm [8], to obtain equilibrium values of c^* that obey the mass conservation Eq. (2) within a relativ error of at most 10^{-4}.

The probability $p_{L_i,K} = \frac{1}{2}[p_{i,K}+p_{i,K+1}]$ that enters the cleavage rate of bond K of sequence S_i involves the probabilities $p_{i,K,j,K'}$ that nucleotide K is paired with another nucleotide at position K' of sequence S_j, and therefore the probability c_{ij}/c_i^0 that sequence S_i is actually part of the corresponding duplex:

$$p_{i,K} = \sum_{K'} p_{i,K,i,K'} \frac{c_i}{c_i^0} + \sum_{j,K'} p_{i,K,j,K'} \frac{c_{ij}}{c_i^0}. \tag{6}$$

III. DERIVATION OF THE STEADY-STATE LENGTH DISTRIBUTION

In the absence of sequence-specific cleavage rates, the sequence length distribution is identical to the cluster size distribution obtained in a simple aggregation-fragmentation process with a mass-independent aggregation rate α and a fragmentation rate βL that is proportional to cluster size L, with random binary breakage. As a variation on the standard problems discussed in the literature, we also include monomer influx with rate J and a length-dependent outflux d_L. For our parameter regimes, we expect that the aggregation-fragmentation dynamics results in a nonequilibrium steady state length distribution N_L (here we omit the angle brackets). It is obtained as the stationary solution of the following mass-balance equation:

$$\dot{N}_L = \alpha \sum_{K=1}^{L-1} N_K N_{L-K} - 2\alpha N_L \sum_{K=1}^{\infty} N_K - \beta(L-1)N_L + 2\beta \sum_{K=L+1}^{\infty} N_K + J_0\delta_{L,1} - d_0 N_L \mathrm{e}^{-\sqrt{L/L_\mathrm{c}}}. \tag{7}$$

The first term on the right hand side describes the creation of a sequence of length L from two fragments of sizes K and $L-K$, while the next term models the ligation of sequence S_L to any other sequence (the factor of 2 accounts for the correct counting of same-mass clusters). The third term corresponds to the breakage of sequence S_L at any of its $L-1$ bonds, and the next term the production of a sequence of length L as one of the two cleavage fragments of a longer sequence. The fifth and sixth terms, respectively, are monomer influx and the length-dependent outflux, where the square-root dependence in the exponential stems from the specific thermodiffusive behavior of polynucleotides in a thermal trap, and the crossover scale L_c combines parameters such as the Soret coefficient, the trap geometry and the temperature difference across the trap. Because this complicated length dependence inhibits further analysis, we will set $L_\mathrm{c} \to \infty$ in the following.

To proceed with the analysis, we perform the continuum limit $N_L \to N(L)$ in the rate equations Eq. (7):

$$\begin{aligned}\dot{N}(L) = \alpha \int_0^L \mathrm{d}K\, N(K)N(L-K) - 2\alpha N(L) \int_0^\infty \mathrm{d}K\, N(K) - \beta L N(L) \\ + 2\beta \int_L^\infty \mathrm{d}K\, N(K) + J\delta(L) - d_0 N(L).\end{aligned} \tag{8}$$

Now we introduce the moments

$$M_n = \int_0^\infty \mathrm{d}L\, L^n N_L, \tag{9}$$

where $M_0 = \langle N_{\mathrm{tot}} \rangle$ is the number of molecules, $M_1 = \langle L \rangle \langle N_{\mathrm{tot}} \rangle$ is the total mass, and so forth. The rate equations for the moments are given by

$$\dot{M}_n = \alpha \sum_{k=0}^{n} \binom{n}{k} M_k M_{n-k} - 2\alpha M_0 M_n + \beta \left(\frac{2}{n+1} - 1 \right) M_{n+1} + J_0 - d_0 M_n. \tag{10}$$

Even though the hierarchy of rate equations for the moments is not closed due to the fragmentation term, the equations for the first two moments decouple from the rest:

$$\dot{M}_0 = -\alpha M_0^2 + \beta M_1 + J - d_0 M_0 \tag{11}$$

$$\dot{M}_1 = J - d_0 M_1. \tag{12}$$

Hence, we can easily obtain stationary solutions for the total number of molecules $\langle N_{\mathrm{tot}} \rangle$ and their mean length $\langle L \rangle$:

$$\langle N_{\mathrm{tot}} \rangle = \sqrt{\frac{d_0^3 + 4\alpha J(\beta + d_0)}{4\alpha^2 d_0}} - \frac{d_0}{2\alpha} \approx \sqrt{\frac{J(\beta + d_0)}{\alpha d_0}} \quad \text{if } J \gg d_0^3/(\alpha(\beta + d_0)), \tag{13}$$

$$\langle L \rangle = \frac{J}{d_0 \langle N_{\mathrm{tot}} \rangle} \approx \sqrt{\frac{J\alpha}{d_0(\beta + d_0)}}. \tag{14}$$

Numerical studies indicate that the resulting length distribution is very similar to a Γ-distribution, which can be used in a moment closure approximation to compute higher moments [9]. For our parameter regime, the distribution is in fact close to exponential ($\langle \Delta L^2 \rangle \approx \langle L \rangle^2$).

We find that the thermal trap, through an outflux rate that drops with the exponential of the square root of sequence length, serves mainly to shift the distribution towards longer sequences. It does not, however, significantly affect the shape of tail of the distribution, because the dynamics of the longer molecules is mostly determined by their cleavage rate, which scales linearly with sequence length and thus quickly beats the outflux.

IV. DERIVATION OF THE AUTO- AND CROSS-CORRELATION FUNCTION

The master equation for the production and destruction of motifs of length ℓ and their complements is given by:

$$\begin{aligned}\partial_t p_{n,n^*} = {} & k_+[p_{n-1,n^*} + p_{n,n^*-1}] + k_-[(n+1)h(\tfrac{n^*}{n+1})p_{n+1,n^*} + (n^*+1)h(\tfrac{n}{n^*+1})p_{n,n^*+1}] \\ & - [2k_+ + k_-(nh(\tfrac{n^*}{n}) + n^*h(\tfrac{n}{n^*})]p_{n,n^*},\end{aligned} \tag{15}$$

where n and n^* are the copy number of a motif and its complement, respectively, k_+ and k_- are its birth and death rates, and $h(x)$ is the "hybridization function", which describes the decrease in the death rate of a motif in terms of the probability of hybridization, which in turn is proportional to the number $x = n^*/n$ of complements per motif that are available for base pairing.

The dynamics of the mean $\langle n(t)\rangle = \sum_{n,n^*} np_{n,n*}(t)$ follows as

$$\partial_t \langle n\rangle = k_+ - k_- \langle nh(n^*/n)\rangle \,. \tag{16}$$

Assuming that $h(x)$ decreases only slowly from unity due to a small hybridization probability, we write

$$h(\tfrac{n^*}{n}) \approx 1 - |h'(0)|\, \tfrac{n^*}{n}, \tag{17}$$

which gives

$$\partial_t \langle n\rangle \approx k_+ - k_- \langle n\rangle + r \langle n^*\rangle \,, \tag{18}$$

where $r = k_- |h'(0)|$ is the apparent replication rate. Note that Eq. (15) is symmetric with respect to n and n^*, and we can therefore directly infer the corresponding equation for $\langle n^*\rangle$. Conditional on the initial conditions $\langle n(0)\rangle = n_0$ and $\langle n^*(0)\rangle = n_0^*$, the solution of these two equations reads:

$$\langle n(t)\rangle_{n_0,n_0^*} = \frac{k_+}{k_- - r}\left(1 - \mathrm{e}^{-(k_- - r)t}\right) + \frac{1}{2}(n_0 - n_0^*)\mathrm{e}^{-(k_- + r)t} + \frac{1}{2}(n_0 + n_0^*)\mathrm{e}^{-(k_- - r)t}. \tag{19}$$

The correlation functions $C_\mathrm{a}(t)$ and $C_\mathrm{c}(t)$ are defined as

$$C_\mathrm{a}(t) = \langle n(t)n(0)\rangle = \sum_{n_0,n_0^*} n_0 \langle n(t)\rangle_{n_0,n_0^*} p^0_{n_0,n_0^*}, \tag{20a}$$

$$C_\mathrm{c}(t) = \langle n(t)n^*(0)\rangle = \sum_{n_0,n_0^*} n_0^* \langle n(t)\rangle_{n_0,n_0^*} p^0_{n_0,n_0^*}, \tag{20b}$$

where p^0_{n,n^*} is the steady-state solution of Eq. (15). All we actually need are the three steady-state averages $\langle n_0\rangle = \langle n_0^*\rangle$, $\langle n_0^2\rangle = \langle n_0^{*2}\rangle$ and $\langle n_0 n_0^*\rangle$, which are obtained from Eq. (15) by expanding the hybridization function as in Eq. (17):

$$\langle n_0\rangle = \langle n_0^*\rangle = \frac{k_+}{k_- - r} \tag{21}$$

$$\left\langle n_0^2\right\rangle = \left\langle n^{*2}\right\rangle = \frac{k_+}{k_- - r}\frac{k_-(k_- + k_+) + (k_+ - k_-)r}{k_-^2 - r^2} \tag{22}$$

$$\langle n_0 n_0^*\rangle = \frac{k_+}{k_- - r}\frac{k_-(k_+ + r) + (k_+ - r)r}{k_-^2 - r^2}. \tag{23}$$

Evaluating Eq. (20) gives Eq. (6) in the main text.

For a scenario with actual replication according to the reaction $n \stackrel{rn^*}{\to} n+1$, the master equation reads:

$$\begin{aligned}\partial_t p_{n,n^*} = {} & k_+[p_{n-1,n^*} + p_{n,n^*-1}] + k_-[(n+1)p_{n+1,n^*} + (n^*+1)p_{n,n^*+1}] \\ & + r[n^* p_{n-1,n^*} + n p_{n,n^*-1}] \ - [2k_+ + (k_- + r)(n+n^*)]p_{n,n^*}.\end{aligned} \tag{24}$$

It is easy to check that this equation gives rise to the same expression Eq. (19) for $\langle n(t)\rangle$ as Eq. (15). However, the stationary second moments $\langle n_0^2\rangle$ and $\langle n_0 n_0^*\rangle$ are slightly different:

$$\left\langle n_0^2\right\rangle = \left\langle n_0^{*2}\right\rangle = \frac{k_+}{k_- - r}\frac{k_-(k_+ + k_-) + k_+ r}{k_-^2 - r^2} \tag{25}$$

$$\langle n_0 n_0^*\rangle = \frac{k_+}{k_- - r}\frac{k_-(k_+ + r) + k_+ r}{k_-^2 - r^2}. \tag{26}$$

The resulting correlation functions read:

$$C_{\mathrm{a/c}}(t) = \frac{k_+^2}{(k_- - r)^2} + \frac{k_- k_+ \mathrm{e}^{-(k_- - r)t}}{2(k_- - r)^2} \pm \frac{k_- k_+ \mathrm{e}^{-(k_- + r)t}}{2(k_-^2 - r^2)}. \tag{27}$$

The time dependence, given through Eq. (19), is clearly the same as that of the correlation functions of Eq. (15), and the amplitudes are identical to those of Eq. (6) in the main text to first nonzero order in r:

$$C_{\mathrm{a}}(0) - C_{\mathrm{a}}(\infty) = \frac{k_+}{k_-} + \mathcal{O}(r), \tag{28}$$

$$C_{\mathrm{c}}(0) - C_{\mathrm{c}}(\infty) = \frac{k_+ r}{k_-^2} + \mathcal{O}(r^2). \tag{29}$$

---

[1] D. Gillespie, J Phys Chem **81**, 2340 (1977).

[2] I. Hofacker *et al.*, Monatsh Chem **125**, 167 (1994).

[3] S. H. Bernhart *et al.*, Algorithm Mol Biol **1**, 3 (2006).

[4] R. Dimitrov and M. Zuker, Biophys J **87**, 215 (2004).

[5] R. M. Dirks *et al.*, Siam Rev **49**, 65 (2007).

[6] C. Fernando, G. von Kiedrowski, and E. Száthmary, J Mol Evol **64**, 572 (2007).

[7] http://www.chokkan.org/software/liblbfgs/index.html.

[8] D. Liu and J. Nocedal, Math Prog B **45**, 503 (1989).

[9] R. Li, B. J. McCoy, and R. B. Diemer, J Colloid Interf Sci **291**, 375 (2005).

Accepted in Physical Review Letters

A thermal, autonomous Replicator made from Transfer RNA

Hubert Krammer, Friederike M. Möller and Dieter Braun

Systems Biophysics, Physics Department, Center for Nanoscience,
Ludwig Maximilians Universität München, Amalienstrasse 54, 80799 München, Germany

Abstract. *Evolving systems rely on the storage and replication of genetic information. Here we present an autonomous, purely thermally driven replication mechanism. A pool of hairpin molecules, derived from transfer RNA (tRNA) replicates the succession of a two letter-code. Energy is first stored thermally in metastable hairpins. Thereafter, energy is released by a highly specific and exponential replication with a duplication time of 30 seconds which is much faster than the tendency to produce false positives in the absence of template. Our experiments propose a physical rather than a chemical scenario for autonomous replication of protein encoding information in a disequilibrium setting.*

Introduction. In modern biology, an RNA-dominated machinery encodes proteins and proteins replicate genetic information. Before this interlinked machinery evolved, early life most probably replicated genes using a minimal pool of rather short RNA sequences[1]. RNA world[2,3] approaches advocate chemical base-by-base replication[4,5] under high salt conditions[6]. However, the translation of proteins relies only on a redundant three-letter code of codons. It is thus suggestive to replicate a succession of codons rather than single bases. Here we explore a physical, not chemical replication mechanism for codon successions.

As a consequence of the second law of thermodynamics, replicators must be driven by disequilibrium conditions. The demonstrated replication is solely driven by waste-less thermal energy which could be provided by thermal convection for example. The replicator is built from four halves of transfer RNA (tRNA), forming hairpins with a sequence encoding toehold (Figure 1). tRNA molecules define the genetic code as they translate codons into amino acids in the translation process of proteins. Arguably, it is one of the most ancient molecules of biology[7,8,9].

In a sense, we create a physically driven version of the chemical ligation chain reaction[10] that replicates a two-letter code cross-catalytically[11,12,13]. Instead of joining the matching strands by chemical backbone ligation, the substrates are joined within seconds by physical base pairing. Our replication is driven by two modes of thermal oscillation. The binding energy is first stored using an initial high temperature step (95°C) where tRNA molecules are quenched on ice water into monomolecular hairpin states. Subsequent moderate temperature oscillations between 10°C and 40°C connect the hairpins exponentially in a cross-catalytic reaction (Figure 1a). The approach is inspired by isothermal DNA machines[14-17] and non-autonomous replicators[18]. As discussed below, no highly specialized RNA sequence was required. We randomly chose half of a tRNA sequence to create the replicator. The thermal approach is compatible with hydrothermal molecule traps[19-21], thermal microconvection[22] and might have been selected by asymmetric hydrolysis of the strand backbone[23].

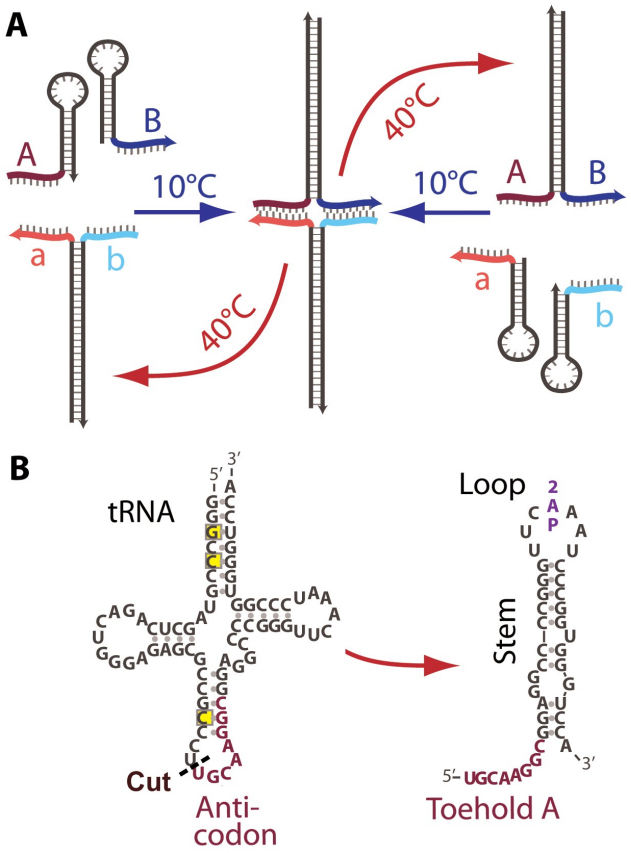

Figure 1. Thermal replication. *(a) Crosscatalytic replication reaction. A purely thermally driven exponential replication of sequence succession* ***ab*** *is expected from metastable hairpins and microscale convection. (b) As fuel, a typical tRNA is cut next to the anticodon, forming a hairpin with a toehold sequence* ***A****.*

Materials and Methods. Alanine-TGC tRNA of Methanobacterium Thermoautotrophicum was cut left of the anticodon and was predicted[24] to fold into a hairpin with a toehold consisting of anticodon and upstream bases (Figure 1b). Inside the loop, we replaced an adenine base by the UV-fluorescent 2-aminopurine (2-AP) which is quenched in the double stranded configuration and does not perturb the RNA structure significantly[25]. From this hairpin **A**, hairpins **B**, **a** and **b** were constructed. The stem-loop region of **B** is complementary to the stem-loop of **A**, but has a different sequence at the toehold. The 8-base long toehold regions of **a** and **b** are complementary to toehold regions of hairpins **A** and **B.** Both **a** and **b** have again complementary stem-loop regions. To prevent binding crosstalk and keep the reaction symmetric, the backward read sequence of **A** and **B** was chosen. The resulting sequences are 5'-UGC AAG GCG GAG GCC CCG GGU UCAP AAU CCC GGU GGG UCC A-3' for hairpin **A** with AP=2-Aminopurine, 5'-U GGA CCC ACC GGG AUU UGA ACC CGG GGC CUC CAC GUU CCG-3' for hairpin **B**, 5'-G GGC CAC CCA GGU AAG UUU ACC UCC GGG GCC CGC CUU GCA-3' for hairpin **a** and 5'-FAM-CGG AAC GUG GGC CCC GGA GGU AAA CUU ACC UGG GUG GCC C-3' for hairpin **b** with a fluorescent FAM modification at the 5' end for gel electrophoresis. RNA hairpins (IBA Göttingen) at 2.5 µM in buffer (150 mM NaCl, 20 mM Tris pH 7, 500 nM $MgCl_2$, Carl Roth) were heated to 95 °C and quenched by ice-water to form the hairpin conformation. Electrophoresis was performed in 3.5 % high-resolution agarose gel (2.5 g Agarose high resolution #K297.2, Carl Roth, 70 ml TB-Buffer) by the FAM-labelled hairpin b. TB buffer was 10-fold diluted from a stock of 54 g TRIS, 27.5 g Boric acid, 500 µM $MgCl_2$ at pH 8 in one litre of pure water.

Setup. Nucleotides, pipette tips and reaction tubes were cooled on ice. Hairpins and annealed template duplex were mixed on ice within 35 seconds and put on a 2x2 array of Peltier-Elements (MPC-D701, Micropelt GmbH, Freiburg) covered by silicon and prepared with a 500µm thin silicone rubber sheet (KU-TCS50, Kunze, Oberhaching) with a hole of 4mm in diameter. The 10 µl reaction chamber was covered with a sapphire slide for UV transparency. Fluorescence imaging was provided by a microscope (AxioTech Vario, Zeiss) through a 20x, NA=0.65 quartz objective (Partec, Germany) using a standard Tryptophane filter set (F36-300, AHF Tübingen), and illuminated with a UV-LED (LED285W, Thorlabs). Emission was detected with a photomultiplier tube (MP 973, Perkin-Elmer Optoelectronics). Temperature is increased from 10°C to 40°C every 27s for 3s. Temperature was measured inside the chamber using the temperature dependent fluorescence of 5µM BCECF (mixed isomers, B-1151, Invitrogen) in 10 mM TRIS before reaction.

Theory and Calibration. Concentrations of all constituents were predicted over time based on published kinetic data[18,26,27] (supplementary material) and transformed to fluorescence F(t) by linear superposition of calibrated AP-fluorescence contributions. Dimer **ab** and quadruple **ABab** were created by slow annealing of 1 μM of the corresponding hairpins, cooling them from 95 °C to 5 °C within 2 hours. By comparing gel electrophoresis (Figure 2b) with 2-AP-fluorescence in buffer at 10 °C and 40 °C, the relative 2-AP-fluorescence contributions {hairpin, duplex, quadruple}={1,0.3,0.5} were inferred (supplementary material). Melting temperatures of annealed species were measured with microscale thermophoresis[28].

Results. Let us assume an RNA molecule can either form high energy hairpins with itself or complexes at lower energies with other RNA molecules. When a heated molecule is cooled rapidly, the kinetics to form a hairpin is typically faster than finding a partner RNA with more paired bases and lower free energy. As a consequence, RNA hairpins can store metastable thermal energy which we use to drive a replication reaction.

The following replication is cross-catalytic under a moderate temperature oscillation (Figure 1a). Four partially complementary hairpins **a**, **b**, **A** and **B** were derived from half a tRNA molecule (Figure 1b) and prepared to be in a metastable state. The toehold succession **ab**, both memorized in the duplex **ab** or **AB**, is replicated. At 10°C, the duplexes **ab** or **AB** bring the hairpins with complementary toeholds in mutual proximity and enhance the kinetics to form the quadruple **ABab**, releasing the energy from the hairpin loop. Subsequent moderate heating (40°C) separates **ABab** into duplexes **AB** and **ab**, which both replicate again with hairpins **a,b** and **A,B**, respectively. The duplex amount doubles in each temperature cycle and a purely thermally fuelled and driven exponential replication of sequence succession **ab** is expected. Temperature oscillations could be provided by convection in millimeter-sized rock pores or clefts.[19,22]

The replication was monitored by time-resolved fluorescence (Figure 2a). At time = 0 s, all four hairpin species **A**, **B**, **a** and **b** were present at 500nM concentration, together with 35nM of annealed duplexes **ab**. Hybridisation of the hairpins to duplexes and quadruples reduces the time dependent fluorescence $F_{(t)}$ of the UV-fluorescent base analogue 2-Aminopurine[25] in the loop of hairpin **A**. As a result, the quenching $Q = 1 - F_{(t)} / F_{(t=0)}$ rised as the replication progresses and saturated in less than 10 minutes, indicating a fast progression of the reaction (Figure 2a, black). In contrast a reaction of hairpins **A** and **B** with template

ab resulted in no quenching increase (Figure 2a, red) and hence no replication.

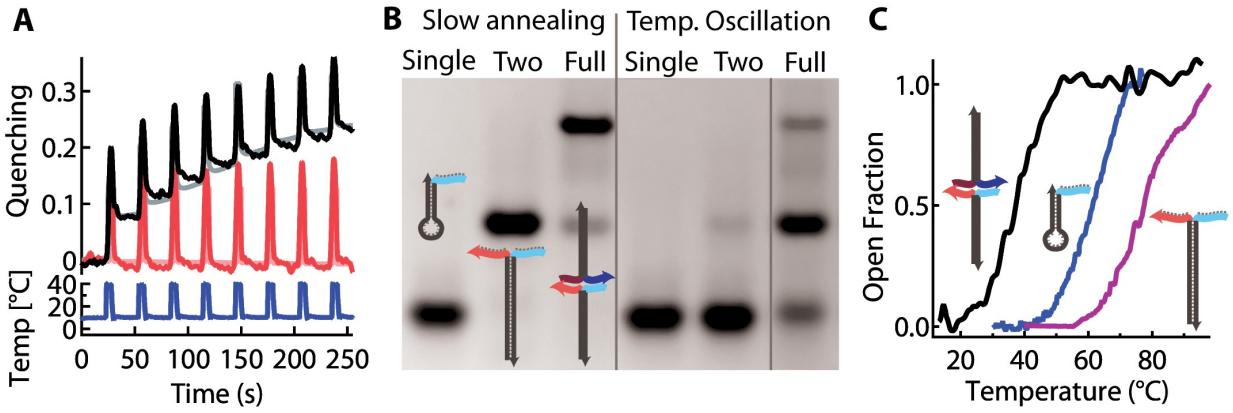

Figure 2. Replication results. *(a) Quenching of 2-Aminopurine raises as duplexes are formed in the full reaction with 500nM of four hairpins* ***a,b,A,B*** *and 35nM template duplex* ***ab*** *(black). The non-cross-catalytic mixture of two hairpins* ***A,B*** *with template* ***ab*** *shows no signal (red). An explicit numerical model fits the fluorescence quenching accurately in both cases (shaded colours). (b) Gel electrophoresis of the resulting products. Slow annealing of single, two or four hairpins leads to equilibrium expectation of hairpins, duplexes and quadruples. Under thermal oscillation, only the full reaction transforms four hairpins into duplexes and quadruples.* ***(C)*** *Melting curves of quadruple (*37±3 °C*, black), hairpin (*62±2 °C *, blue) and duplex (>75°C, purple).*

The thermal cycling times were chosen to exceed the on-rates of toehold hybridization[18,26,27] $k_{on} = 1.3\,(s * \mu M)^{-1}$ at the initial hairpin concentration of 500 nM. The opening-closing kinetics of the hairpins were faster (0.1-1 ms)[26] than the average toehold binding time constant (1 s)[18,27], enabling the catalysis of hairpin opening. In each cycle, the temperature was set to 10°C for 27s, followed by 40°C for 3s (Figure 2a, blue). The periodic peaks in fluorescence report the separation of quadruple **ABab** into duplexes **AB** and **ab** at 40°C.

The choice of temperatures is confirmed by melting curves (Figure 2c), revealing clearly separated melting transitions of hairpins, duplexes and quadruples. Running the reactions between 10°C and 40°C ensures a robust replication without opening of hairpins or separation of duplexes.

To substantiate the replication mechanism, the concentrations of all reaction species have been modelled (supplementary material). The simulated concentrations were converted into quenching and showed good agreement for all reactions (Figure 2a, shaded colors). In Figure 2b, the products of the

reaction were analyzed with high resolution agarose gel electrophoresis. Slow annealing from 95°C to 5°C within 2 hours of single hairpins, two hairpins and all hairpins resulted in clearly distinguishable bands, revealing the equilibriums of the reactions, which correspond to single hairpins, duplexes and quadruples, respectively. After 15 temperature cycles the full reaction yielded up to 80±5 % of information carrying constituents, while in the reactions with one or two hairpins the high energy hairpin conformation was maintained (Figure 2b). These results affirmed the finding presented in Figure 2a.

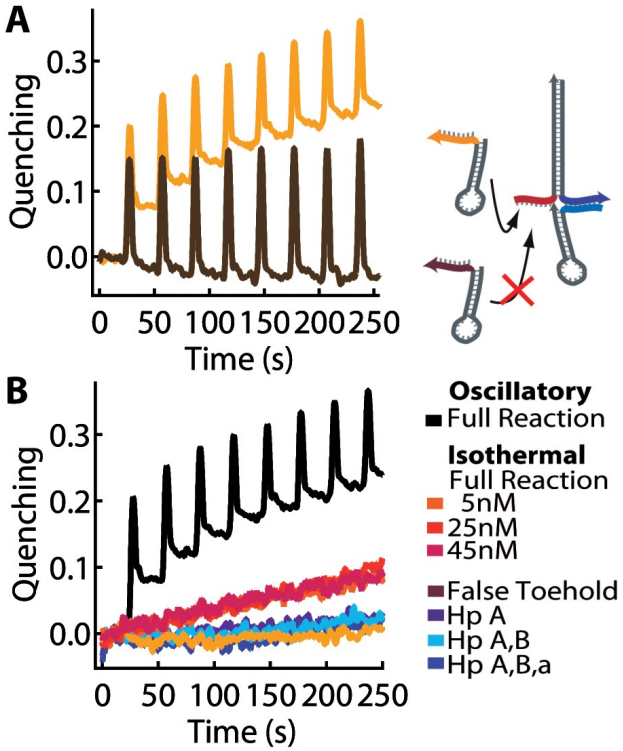

Figure 3. Specificity. *(a) No replication was found for a mutated toehold sequence (yellow). (b) With isothermal driving, the complete reaction replicated slowly and showed no dependence the on template concentration ([template duplex]={5nM, 25nM and 45nM}, [hairpin]=500nM). Isothermal reactions using a non-matching toehold sequence (yellow) or without hairpins (blue) showed no signature of replication.*

Specificity was tested by replacing the toehold of one of the four hairpins with a random sequence (Figure 3a). Although duplexes could still be formed in equilibrium, no replication was observed, reproducing results from only two hairpins (Figure 2a, red). In the absence of matching toehold sequences, the hairpins cannot enter a cross-catalytic cycle, demonstrating the high specificity of the replication.

Without thermal cycling, quadruples do not open actively and cannot serve as templates for cross-catalysis. As expected, isothermal reactions were dominated by the background rate and independent of the initial template concentration (Figure 3b, shades of red). A further removal of one or more hairpins from the reaction, again similar with an introduction of a false toehold, further reduced the background rate (Figure 3b, shades of blue and yellow). No significant levels of duplexes were found without thermal activation of cross-catalysis.

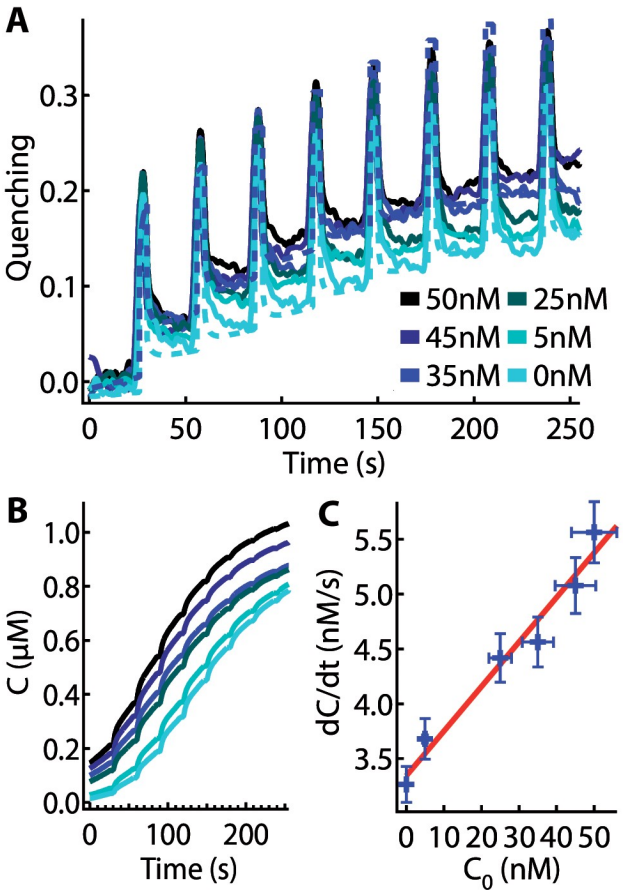

Figure 4. Exponential Replication. *(a) Measured (solid) and simulated (dashed) quenching of complete reactions with different initial template concentration are in agreement. (b) The increase of the replicated product raised with increasing template concentration. (c) Time derivative of modelled template concentration dC/dt within the first three temperature cycles plotted versus the initial template concentration. A fit by equation (1) revealed a doubling time* $\tau = 28 \pm 5s$*, matching the thermal cycling time of 30 s.*

An exponential replicator with doubling time τ and false positive background rate k_0 enhances the concentration of replica *C* according to

$$dC/dt = C * ln(2)/\tau + k_0 \qquad (1)$$

We tested the concentration dependence of dC/dt. In Figure 4a, replications were performed with increasing template concentration **AB** between 5 nM and 50 nM. The quenching was modelled (Figure 4a, dashed curves) and the time derivative dC/dt of the information carrying molecules $C = [AB]_{calc} + [ab]_{calc} + 2[ABab]_{calc}$ was averaged within the first three replication cycles. In the beginning, the hairpin concentration was not yet significantly depleted and replication did not yet saturate. The fit with equation (1) revealed a duplication time $\tau = 28 \pm 5s$ together with a background rate $k_0 = 3.0 \pm 0.3 \mathrm{nM}/s$. Error bars were determined from three independent experimental runs. The duplication time matched the total cycling time $\tau_{cycle} = 30s$, indicating that the sequence information was doubled with each temperature oscillation.

Discussion.

To maintain information against chemical degradation and physical removal, fast exponential replication is desirable if not necessary.[29,30] The shown replicator provides exponential growth rates on the time scale of seconds to minutes due to the fast kinetics of RNA hybridization and the efficient template separation by thermal cycling. In comparison with chemical, protein-free replicators,[11-13] the hairpin replicator is more than 2 orders of magnitude faster.

Does the hybridization of toehold and hairpin loops suffice to gain free energy in the reaction A+B+a+b $\leftrightarrow$ ABab? Transforming two hairpins into a duplex yields[24] $\Delta G_{duplex} = -45kcal/mol$ and the hybridization of one toehold $\Delta G_{codon} = -13kcal/mol$, resulting in a total gain from base pairing of $\Delta G_{rep} = 2 * \Delta G_{duplex} + 2 * \Delta G_{codon} = -116kcal/mol$. In contrast, the entropic penalty[16] $\Delta G_{entropic} = RTln(Q)$, estimated from $Q = ([A][B][a][b]/c_0^4)/([ABab]/c_0)$ at an intermediate reaction point with $[A][B][a][b] = [ABab] = c = 250nM$ is more than fourfold smaller ($26kcal/mol$), allowing the reaction to proceed.

Microconvection provides thermal cycling that is able to drive protein-catalyzed DNA replication.[22,31] Together with thermophoresis in elongated clefts, molecules are trapped[19]. Both can be implemented in the same chamber[21]. Notably, such settings can accumulate prebiotic lipids to form cell-like vesicles[20]. One can envision that hairpins are recycled and quenched in the outer flow lines of laminar convection in order to fuel replication in the inner more moderate convection regions. Replication then becomes fully reversible and does not create waste products.

The starting pool of RNA oligomers could be produced by non-templated RNA polymerization from surfaces[32] or other precursors[33] and accumulated by thermophoretic trapping.[19] Recent models indicate that the combined trapping and base-pair selective degradation leads to a pre-replication stabilization of complementary sequence motifs where hairpin conformations dominate.[23] Over time, hairpins with the shown dynamics could be selected since replication settles with fast kinetics in double-stranded structures (e.g. Figure 3a), saving the replicating molecules from replication. Thermophoretic trapping of the larger complex could further select participating hairpins.[19,21]

Conclusion. We have shown that a two letter code can be selectively and exponentially replicated by physical base-pairing from a pool of thermally quenched RNA hairpins. Replication is driven by moderate temperature oscillation and - besides a background rate - doubles in each thermal oscillation. It had not escaped our attention that the demonstrated replication with half a tRNA suggests a possible physico-chemical replication-translation mechanism[34] from the concatenation of multi-letter codes using complete tRNA molecules.

Acknowledgments. We thank Zan Luthey-Schulten for discussions and suggesting 2-AP fluorescence. Michael Nash, Michael Hartmann, Simon Lanzmich, Carolin Leonhardt, Ulrich Gerland, Benedikt Obermayer, Moritz Kreysing and Michael Russell gave valuable comments on the manuscript at various stages. Financial support from the NanoSystems Initiative Munich, the LMU Initiative Functional Nanosystems and the ERC Starting Grant is gratefully acknowledged.

References

1. Koonin E.V., and A.S. Novozhilov, Origin and Evolution of the Genetic Code: The Universal Enigma. IUBMB Life 61, 99-111 (2009).
2. Gilbert, W. The RNA world. Nature, 390, 618 (1986).
3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell, 35, 849 (1986).
4. Deck, C., Jauker, M., Richert, C., Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. Nature Chemistry, 3, 603-608 (2011).
5. Rajamani, S., et al., Effect of stalling after mismatches on the error catastrophe in non-enzymatic nucleic acid replication. J. Am. Chem. Soc. 132, 5880-5885 (2010).
6. Johnston, W.K., et al., RNA-Catalyzed RNA Polymerization: Accurate and General RNA-Templated Primer Extension. Science 292, 1319-1325 (2001).
7. Eigen, M., et al., How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 244, 673-679 (1989).
8. Widmann, J., Harris, J.K., Lozupone, C., Wolfson, A., Knight, R., Stable tRNA-based phylogenies using only 76 nucleotides. RNA, 16, 1469-1477 (2010).
9. Woese, C.R., Kandler, O., Wheelis, Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. PNAS, 87, 4576-4579 (1990).
10. Barany, F. Genetic disease detection and DNA amplification using cloned thermostable ligase. PNAS 88, 189-93 (1991).
11. von Kiedrowsik, G. A Self-replicating Hexadeoxynucleotide. Angewandte Chemie, 25, 932-935, (1986).
12. Zielinski WS, Orgel LE. Autocatalytic synthesis of a tetranucleotide analogue. Nature 327: 346-347 (1987).
13. Lincoln, T.A., Joyce, G.F, Self-Sustained Replication. Science, 323, 1229 (2009)
14. Dirks, RM, Pierce, NA, Triggered amplification by hybridization chain reaction. PNAS, 101 15275–15278 (2004).
15. Barish, R.D., Schulman R, Rothemund P.W.K and Winfree E., An information-bearing seed for nucleating algorithmic self-assembly. PNAS, 106, 6054-6059 (2009)
16. Zhang, D.Y., Turberfield, A. J., Yurke, B., Winfree, E. Engeneering Entropy-driven Reactions and Networks Catalyzed by DNA. Science, 318, 1121-1125 (2010).
17. Green, S.J., Lubrich, D. Turberfield, A.J., DNA Hairpins: Fuel for autonomous DNA Devices. Biophysical Journal, 91, 2966-2975 (2006).

18. Wang, T et al., Self-replication of information-bearing nanoscale patterns. Nature, 478, 225-228 (2011).
19. Baaske et al. Extreme accumulation of nucleotides in simulated hydrothermal pore systems. PNAS,104, 9346-9351 (2007).
20. Budin, I., Bruckner, R.J., Szostak, J.W. Formation of Protocell-like Vesicles in a Thermal Diffusion column. Jacs, 131, 9628-9629 (2009).
21. Mast, C.B., Braun, D., Thermal trap for DNA Replication. PRL, 104, 188102 (2010).
22. Braun, D., Goddard, L. N., Libchaber, A, Exponential DNA Replication by Laminar Convection. PRL 91, 158103 (2003).
23. Obermayer, B., Krammer, H., Braun, D., Gerland, U. Emergence of Information Transmission in a Prebiotic RNA reactor. PRL, 107, 018101 (2011).
24. Hofacker, I. Vienna RNA secondary structure server. Nucleid acid research, 31, 3429-3431 (2003).
25. Evans, K., Xu, D., Kim, Y., Nordlund, T.M. 2-Aminopurine Optical spectra: Solvent, Pentose Ring and DNA Helix melting dependence. Journal of fluorescence. 2, 209-216 (1992).
26. Bonnet G, Krichevsky O Libchaber A, Kinetics of conformational fluctuations in DNA hairpin-loops. PNAS 95, 8602-8606 (1998).
27. Howorka, S., Movileanu, M., Braha, O., Bayley, H. Kinetics of duplex formation for individual DNA strands within a single protein nanopore. PNAS, 98, 12996-13001 (2001).
28. Wienken, C.J., Baaske, P., Duhr, S., Braun, D. Thermophoretic melting curves quantify the conformation and stability of RNA and DNA, Nucleotide acid research, 39 e 52, 1-10, (2011).
29. von Kiedrowski, G., Szathmary, E. Selection versus Coexistence of Parabolic Replicators Spreading on Surfaces. Selection, 1, 173-180 (2001).
30. Kovac, L., Nosek, J., Tomaska, L., An overlooked riddle of life's origins: energy-dependent nucleic acid unzipping. Journal of Molecular Evolution, 57, S182–S189 (2003).
31. Krishnan, M., Ugaz, V.M., Burns, M.A., PCR in a Rayleigh-Bénard Convection Cell, Science, 298 793 (2002).
32. Ferris, J.P., Hill, A.R.Jr., Liu, R., Orgel, L.E. Synthesis of long prebiotic oligomers on mineral surfaces. Nature, 381, 59-61 (1996).
33. Constanzo, G., Pino, S., Cicirello, F., Di Mauro, E. Generation of long RNA chains in water. Journal of Biological Chemistry, 284, 33206-33216 (2009).
34. Kuhn H, Waser J, Evolution of early mechanisms of translation of genetic information into polypeptides. Nature, 298, 585 - 586 (1982).

Supporting Information (SI)

Instrument setup and baseline correction

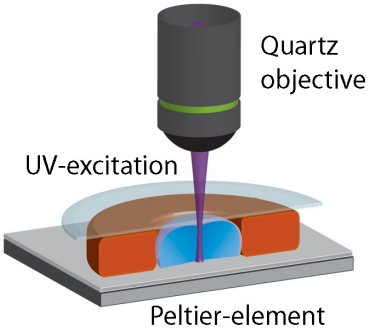

Figure SI 1: Instrument geometry. The sample chamber is defined by a Si-covered Peltier-element on the bottom, a sapphire glass on top and a rubber sheet as spacer. It allows fast, simultaneous temperature control and UV-fluorescence detection.

A sketch of the instrument geometry for simultaneous fluorescence measurements and thermal cycling is given in Figure *SI 1*. Temperature oscillations were realized by fast-cooling Peltier elements. The sample chamber was defined by a pure silicon plate at the bottom and a sapphire glass on top. As a spacer, we used a 0.5 mm thick rubber sheet with a hole of 4 mm in diameter. Finite element simulations show that the reaction chamber reaches thermal equilibrium faster than 500ms, keeping a homogeneous temperature distribution at any time.

While the whole sample was exposed to temperature cycles, only about 1% of the sample in the focus of the objective was exposed to UV-excitation light and thereby suffered from enhanced degradation. Since high-resolution gel electrophoresis and simulations cover the complete sample volume, degradation could be neglected. In contrast, all molecules contributing to fluorescence signals suffer from degradation. All quenching curves are corrected for this degradation before comparison to the model. The correction function is an exponential fit to the fluorescence of non-interacting single stranded RNA molecules (Figure SI 2, green).

Reactions and reaction rates

A numeric model was used to simulate quenching rates in order to compare them to measurements and to determine time resolved concentrations. Table SI 3 shows the considered partial reactions under consideration as well as the kinetic rates in use. For any sub-reaction, we consider a chemical equation of the form

$$X+Y \leftrightarrow XY$$

with k_{off} rates from the left to the right and k_{on} rates towards the right. For any of the reactions in column 1 of Table SI 3, we derived the change of concentration with a differential equation of the form

$$d[X]/dt = -[X]*[Y]*k_{on} + [XY]*k_{off}$$

$$d[Y]/dt = -[X]*[Y]*k_{on} + [XY]*k_{off}$$

$$d[XY]/dt = +[X]*[Y]*k_{on} - [XY]*k_{off}$$

and added up all equations with identical left sides.

A cash-carp algorithm solves the resulting set of coupled differential equations over time, taking into account the different rates for hot and cold temperatures. In all rate estimations, symmetry of the reaction pathways catalysed by duplex **AB** or duplex **ab** was assumed (left and right cycle in Figure 1a). Replication measurements with both catalysts verified this assumption.

Reactions dominated by the binding of a single toehold region to its complement were considered to have equal k_{on} rates. Since the duplex formation by opening of hairpins is considered to be orders of magnitude faster than the opening of toeholds, k_{off} rates are low for reactions including duplex formation and high for non-stabilized reactions.

Rates for spontaneous formation of duplexes are verified by cycling a mixture of hairpins (A) and (B). Measurements are shown in Figure SI 2 in green, rates are given in Table SI 1.

Any structure that is capable of forming hairpins is also capable of forming homo-dimers. Therefore the reactions of the type $X+X \leftrightarrow XX$ are included in the simulations. To avoid these reactions as far as possible, all hairpin samples were quenched from 95°C on ice-water before the experiments to enforce the hairpin conformation. Rates for this type of reaction are comparably slow. They were verified by measurements with only hairpins (A) in a full reaction at cold temperatures and hot temperatures. The homo-dimer concentration after 15 cycles is so low that detection on high-resolution agarose gels is not possible (see Figure 2b lane 4). The lower four reactions in Table SI 3 describe the closing of open hairpins (Footnote “oh” for open hairpins).

After these rates are fixed, the only remaining crucial rate is the binding of the toehold to the template at cold temperatures, represented in the first four equations in Table SI 3. In comparison with measurements and fitting results, this rate was fixed at $k_{on}^{(10°)} = 1.3(s*\mu M)^{-1}$ which in good agreement with literature values of similar toeholds.[22]

Reaction	k_{off} T= 10°C [1/s]	k_{on} T=10°C [(s*µM)$^{-1}$]	k_{off} T=40°C [1/s]	k_{on} T=40°C [(s*µM)$^{-1}$]
AB+a ↔ ABa	0.2	1.3	200	1.3
AB+b ↔ ABb	0.2	1.3	200	1.3
A+ab ↔ Aab	0.2	1.3	200	1.3
B+ab ↔ Bab	0.2	1.3	200	1.3
ABa+b ↔ ABab	1e-09	1.3	1e-06	1.3
ABb+a ↔ ABab	1e-09	1.3	1e-06	1.3
A+Bab ↔ ABab	1e-09	1.3	1e-06	1.3
B+Aab ↔ ABab	1e-09	1.3	1e-06	1.3
AB+ab ↔ ABab	0.0025	25	100	25
A+B ↔ AB	1e-06	0.00015	0.00015	0.00015
a+b ↔ ab	1e-06	0.00015	0.00015	0.00015
A+a ↔ Aa	0.1	0.57	120	0.57
B+b ↔ Bb	0.1	0.57	120	0.57
A+A ↔ AA	0.00001	0.00015	0.001	0.00015
B+B ↔ BB	0.00001	0.00015	0.001	0.00015
a+a ↔ aa	0.00001	0.00015	0.001	0.00015
b+b ↔ bb	0.00001	0.00015	0.001	0.00015
Aa+B ↔ AaB	1e-06	0.0015	0.00015	0.0015
Aa+B ↔ AaB	1e-06	0.0015	0.00015	0.0015
Bb+A ↔ ABb	1e-06	0.0015	0.00015	0.0015
Bb+a ↔ Bab	1e-06	0.0015	0.00015	0.0015
Aa+Bb ↔ ABab	1e-09	0.00015	1e-06	0.00015
Aoh ↔ A	1e-09	0,001	1e-09	0.5
Boh ↔ B	1e-09	0,001	1e-09	0.5
aoh ↔ a	1e-09	0,001	1e-09	0.5
boh ↔ b	1e-09	0,001	1e-09	0.5

Table SI 1: Rates for a detailed simulation of Quenching curves and concentrations.

Quenching and fluorescence calibration

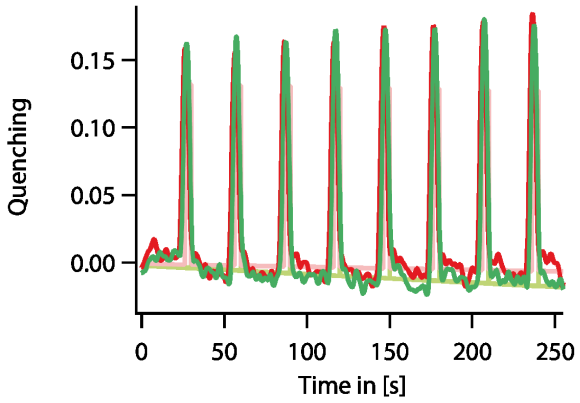

*Figure SI 2: Control measurements with hairpin **A** only (red) allow correction for degradation. A mixture of hairpin **A** and hairpin **B** confirmed the rates for spontaneous duplex formation (green).*

To calculate quenching from calculated concentrations, we measured the fluorescence intensities of hairpins, duplexes and quadruples, that were prepared as described in the materials section and normalized to the fluorescence of hairpins, yielding the multiplication factors of relative fluorescence:

A	AB	ABab	Aa	Aab	ABa	ABb	AA	A_{oh}
1	0.3	0.5	1.05	1.05	0.3	0.3	2.5	1.32

Those factors were multiplied with calculated concentrations and summed up to a calculated fluorescence which was transformed to a calculated quenching and the compared with measurements. The temperature dependent fluorescence of 2-AP in the strands was recorded and fixed to 0.5 %/K for hairpins and 0,16 %/K for paired conformations (duplexes and quadruples).

Exponential replication

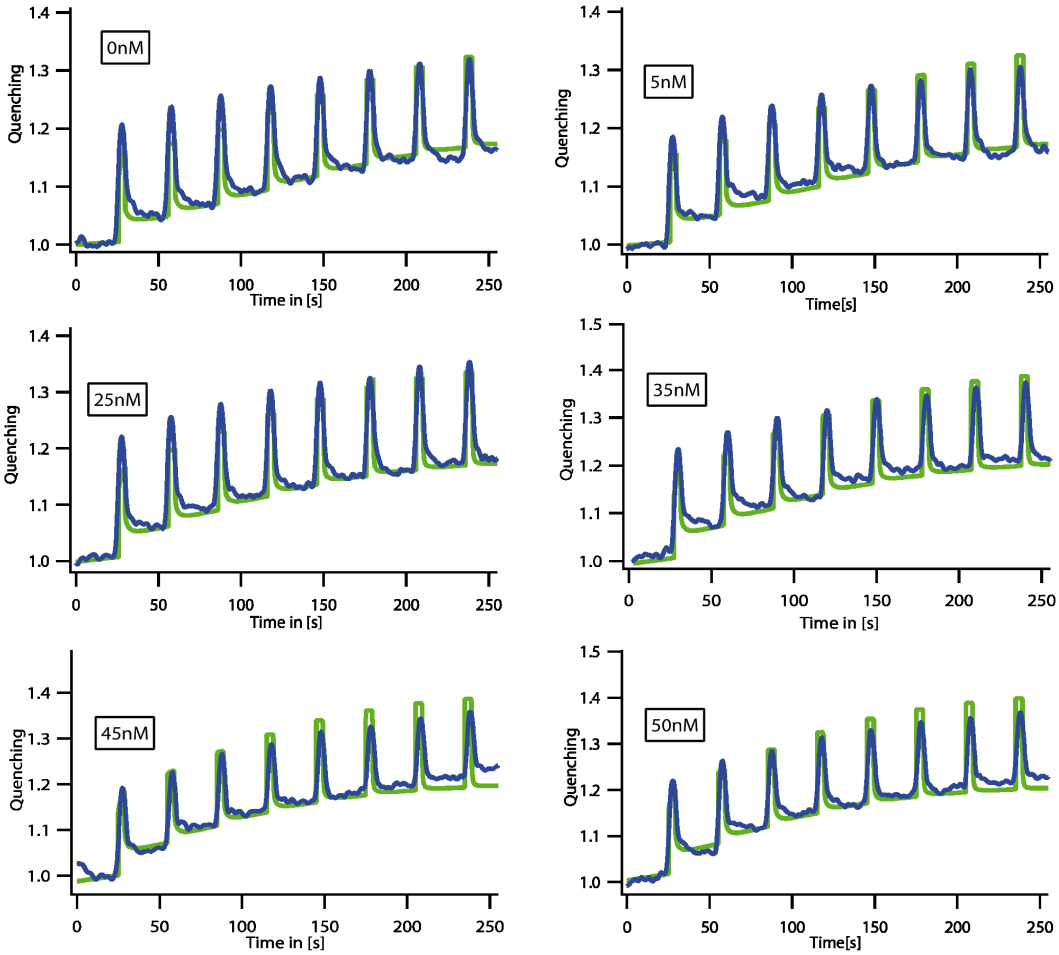

Figure SI 3: Comparison of simulated and measured quenching curves. For calculations in Figure 4c, the first three cycles (90s) were taken into account.

Every measured quenching curve was individually compared to a simulated quenching. Only two parameters are changed between the fitting runs, the preparation time, accounting for a fraction of open hairpins, and the initial catalyst concentration. Both parameters are under experimental control. Figure SI 4 shows each quenching measurement included in the calculation in Figure 4a,b and its model in a separate graph. The good agreement between measurement and simulation is evident.

Therefore, the concentration of information carrying molecules C over time can be extracted from the simulations as plotted in Figure 4. We calculated the concentration as:

$$C = [duplex\ AB] + [duplex\ ab] + 2 * [quadruple\ ABab]$$

Melting curve measurement with micro-scale thermophoresis

It is notoriously difficult to determine the melting transition of RNA double strands by classical means of UV-absorption. Even though the absorption of light with a wavelength of 250nm is modified by a melting transition, RNA often causes non linear baselines and intermediate folding that smear out melting curves so much that a fit with two-level models is not satisfactory. We used micro-scale thermophoresis to determine melting transitions of RNA constructs. Melting of duplexes (purple) and hairpin opening (blue) in Figure 2 was directly determined by the changed quenching of 2-AP in case of hybridization. To record separation of template and product with temperature (Figure 2c, black), the Soret-coefficient of the part including hairpin b was recorded. During separation, the thermodiffusion of the labeled part changes in a two level system. Both records were normalized to a fraction bound and melting temperatures were determined using a sigmoid fit.

Solving equation (1)

The differential equation (1) is solved by an exponential function of the form:

$$C_{(t)} = d * exp\left(ln2 / \tau * t\right) - k_0 * \tau / ln2$$

Thereby, the integration constant d is determined by the concentration of information carriers C_0 at time t = 0 and the background rate k_0 as:

$$d = C_0 + k_0 * \tau / ln2$$

Quantitative gel analysis

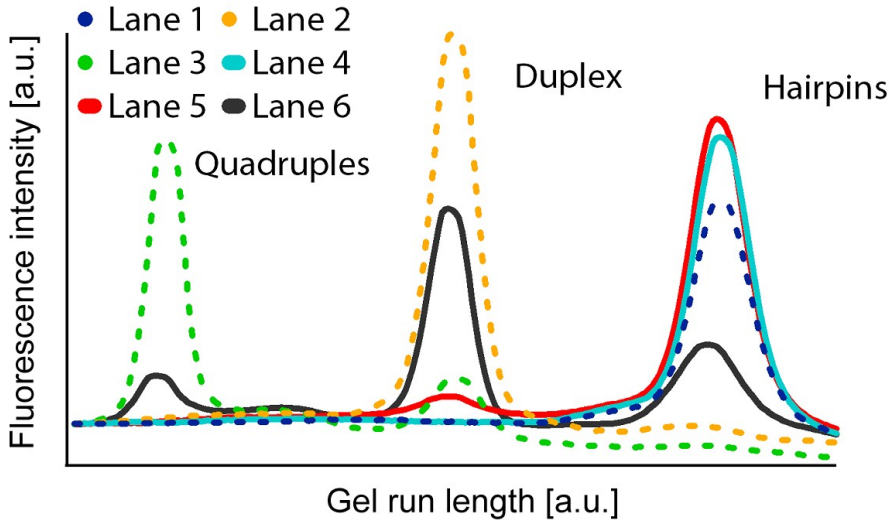

Figure SI 4: Quantitative Analysis of high resolution agarose gels. Intensities were integrated perpendicular and plotted along the electric field.

High resolution gel analysis was done in the absence of intercalating dyes. The fluorescence of the 5-FAM label at hairpin (b) is independent of the size of the molecule complex. Therefore, intensity accounts in first approximation for concentration of molecules and allows for a quantitative comparison. Figure SI 5 shows the integrated intensity distribution of the gel shown in Figure 2b. Dashed lines represent the ladder in lane 1-3. Solid lines are products of 15 temperature cycles in lanes 4-6.