tRNA sequences can assemble into a replicator

Alexandra Kühnlein[1], Simon A. Lanzmich[1] and Dieter Braun[1*]

[1] Systems Biophysics, Physics Department, Center for NanoScience,
Ludwig-Maximilians-Universität München, 80799 Munich, Germany

*corresponding author: Dieter Braun.

Email: dieter.braun@lmu.de

Abstract

Can replication and translation emerge in a single mechanism via self-assembly? The key molecule, transfer RNA (tRNA), is one of the most ancient molecules and contains the genetic code. Our experiments show how a pool of oligonucleotides, adapted with minor mutations from tRNA, spontaneously formed molecular assemblies. They replicated information autonomously using only reversible hybridization under thermal oscillations. A pool of cross-complementary hairpins self-selected by agglomeration and sedimented under gravity. The metastable DNA hairpins bound to a template, consisting of one half of the hairpin assembly, and then interconnected by hybridization. Thermal oscillations separated replicates from their templates and drove an exponential, cross-catalytic replication. The molecular assembly could encode and replicate binary sequence information and reach a fidelity of 90 % per nucleotide. This mechanism of a replicating self-assembly of tRNA-like sequences indicates that the translation to proteins could be linked closer to molecular replication than previously thought.

Main text

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

This is a complex system to set up in the first place at the emergence of life. The RNA world hypothesis proposes, that early on, the catalytic function of highly defined RNA sequences was used for self-replication (Horning & Joyce, 2016; Orgel, 2004; Turk et al., 2011). These ribozymes catalyze the ligation of RNA (Doudna et al., 1991; Mutschler et al., 2015; Paul & Joyce, 2002; M. P. Robertson et al., 2001; von Kiedrowski, 1986; Walton et al., 2020) and the addition of individual bases (Attwater et al., 2013; Horning & Joyce, 2016). These very special sequences were engineered using in vitro evolution. It is unclear how autonomous evolution of early life could have reached such levels of sequence complexity.

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

We show that replication of information can indeed be realized by the reversible base pairing interactions between tRNA-like molecules alone. The proposed mechanism is driven by an external physical non-equilibrium setting, in our case thermal oscillations. Since the process does not involve chemical ligation, it does not rely on a particular non-enzymatic or catalytic ligation chemistry (Dolinnaya et al., 1988; Engelhart et al., 2012; Patzke et al., 2014; Pino et al., 2011; Rohatgi et al., 1996; von Kiedrowski, 1986) or particular catalytically active sequences, but merely requires sequence complementarity. The advantage of reversible hybridization is the recycling of educts and products. Moreover, sequence-encoded interactions can self-select by forming agglomerates.

Nature's approach to achieve exponential growth is the usage of cross-catalysis: the replicate of a template serves as a template for the next round of replication. For short replicators, the binding between template and replicate can be weak and the strands can dissociate spontaneously (Paul & Joyce, 2002; von Kiedrowski, 1986). For longer replicates, temperature change has successfully been used to separate strands for replication, catalyzed by thermostable proteins (Barany, 1991; Saiki et al., 1985). For catalytic RNA, elevated salt concentrations disfavor strand separation by temperature and catalyze hydrolysis (Horning &

Joyce, 2016). In an interesting alternative to strand separation by temperature, Schulman et al used moderate shear flows to separate DNA tile assemblies (Schulman et al., 2012).

In the past, metastable hairpin states have been prepared in a physically separated manner. The reaction was triggered by mixing. For example, the mixing of hairpins with a trigger sequence has been shown to form long concatemers (Dirks & Pierce, 2004). With a similar logic, mixing a low entropy combination of molecules was used to create entropically driven DNA machines, including exponentially amplifying assemblies (Zhang et al., 2009). These reactions run downwards into the binding equilibrium. However, the preparation of the required initial low entropy state needs human intervention or a unique flow setting for mixing.

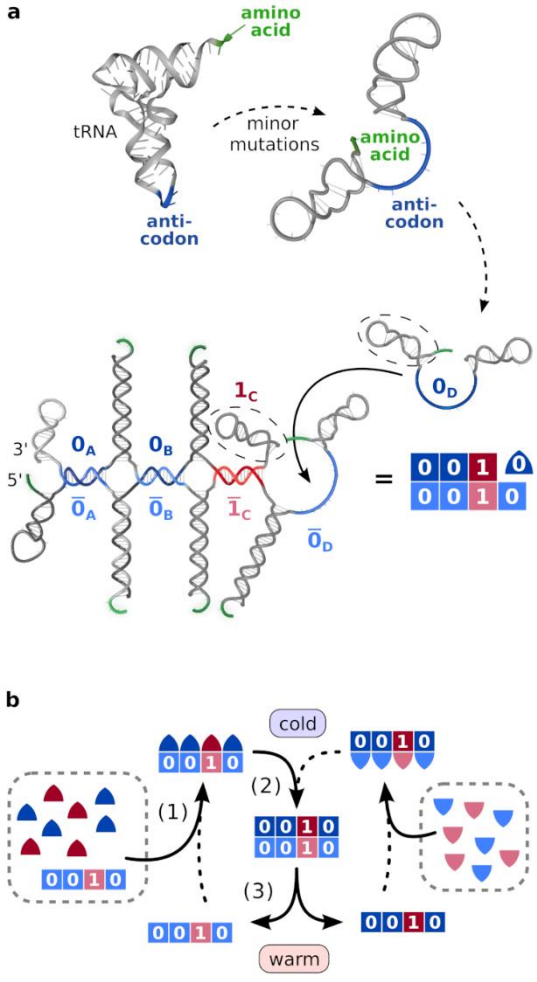

Figure 1. Heat-driven replication by hybridization using hairpin structures inspired from transfer RNA. a*, Transfer RNA folds into a double-hairpin conformation upon very few base substitutions. In that configuration, the 3'-terminal amino acid binding site (green) is close to the anticodon (blue) and a double hairpin structure forms. A set of pairwise complementary double hairpins can encode and replicate sequences of information. A binary code implemented in the position of the anti-codon, the information domain, allows to encode and replicate binary sequences (red vs blue). Each strand (82-84 nt) comprises two hairpin loops (gray) and an interjacent unpaired information domain of 15 nt length (blue/red, here: 0_D). The displayed structure of eight strands shows replication of a template corresponding to the binary code 0010. Note, that no covalent linkage is involved in the process.* ***b****, Replication is driven by thermal oscillations in four steps:*

(0) The hairpins are activated into their closed conformation by fast cooling indicated by triangles. (1) Strands with matching information domain bind to the template. (2) Fluctuations in the bound strands' hairpins facilitate the hybridization of neighboring strands. (3) Subsequent heating splits replica from template, while keeping the longer hairpin sequences connected, freeing both as templates for the next cycle.

Sequence design. We designed a set of cooperatively replicating DNA strands using the program package NUPACK (Zadeh et al., 2011). The sequences are designed to have self-complementary double hairpins and are pairwise complementary within the molecule pool, such that the 3' hairpin of one strand is complementary to the 5' hairpin of the next. Their structure resembles the secondary structure of proto-tRNAs proposed by stereochemical theories (Fig. 1a), comprising two hairpin loops that surround the anticodon with a few neighboring bases (Krammer et al., 2012). The lengths of 82-84 nt of the double hairpins are that of average tRNA molecules (Sharp et al., 1985), with stem loops consisting of 30-33 nt and the information-encoding interjacent domains of 15 nt. As the replication mechanism is based on hybridization only it is expected to perform equally well for DNA and RNA. Here, we implemented the system with DNA for practical reasons. Nevertheless, due to short heating times and very moderate magnesium concentrations, we also estimate that an RNA version can survive for weeks (Li & Breaker, 1999).

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

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

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

replication cross-catalytic in a subsequent step. Later, high temperatures spikes can unbind and recycle all molecules for new rounds of replication.

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

The core principle of this replication mechanism was previously explored in a minimal system that amplified single hairpins into dimers (Krammer et al., 2012). However, these experiments suffered from 50 % false positive amplification without template (Fig. 4c in Krammer et al., 2012). Also, significantly higher molecule concentrations required faster thermal oscillations.

Results

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

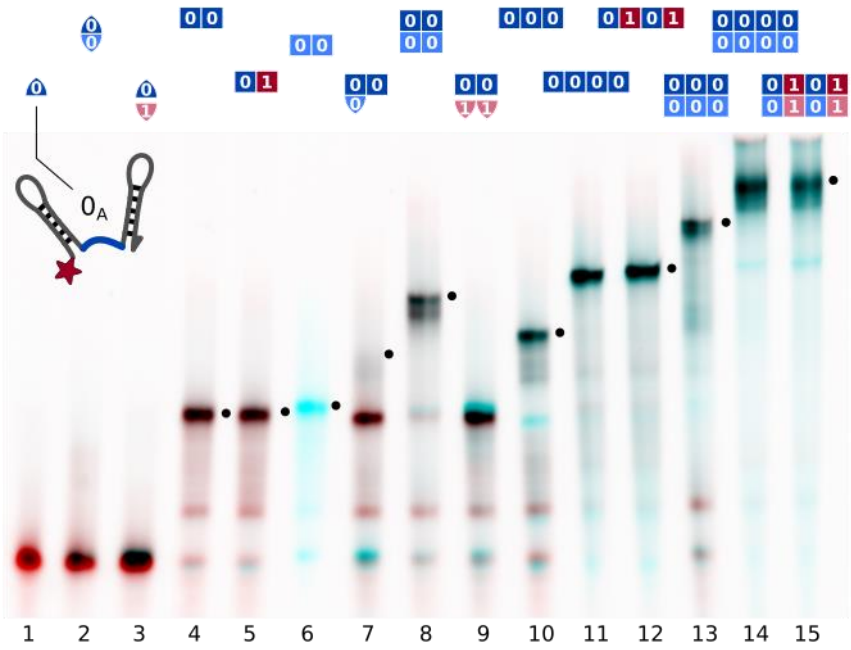

Figure 2. Assembly of different subsets of the cross-replicating system of strands observed by native gel electrophoresis. *Samples contained strands at 200 nM concentration each and were slowly annealed as described in Methods. Lane contents are indicated at the top of each lane. Comparison of different lanes allowed for the attribution of bands to complexes. Complexes incorporating all present strands are marked (•). The red channel shows the intensity* 0_A-

Cy5, the cyan channel shows SYBR Green I fluorescence. Single information domain bonds (lane 2) break during gel electrophoresis.

All complexes were formed at concentrations of 200 nM of each strand and could be resolved despite their branched tertiary structure. Friction coefficients of complexes of two to four strands were 1.6–1.8-fold higher than for linear dsDNA, and 2.4-fold higher for larger complexes (4:4 configuration, ca. 660 nt, Fig. S1). This agrees with the branched structure of the suggested strand assembly geometry (Fig. 1a). Partially assembled complexes of two or three strands bound to a four-strand template could be resolved (Fig. S3). Complexes containing single bound information domains were not stable during electrophoresis (Fig. 2, lane 2 and Fig. S3). This allowed to differentiate fully assembled complexes from those where individual strands are bound to a template but have not formed backbone duplexes. Covalent end labels and two reference lanes on each gel were used to quantify concentrations from gel intensities using image analysis as described in Methods.

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

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

The sedimentation was very selective. When only seven of the eight matching hairpins were present, sedimentation was much weaker and, in most cases, undetectable (Fig. 3b, c). For the full system the sedimentation kinetics showed to be strongly concentration dependent (Fig. S6b). Analogous experiments with random sequences (random pool of 84 nt strands) at equal concentration did not show agglomeration nor sedimentation (Fig. S6c). We have

previously found that similar hairpin molecules provided the shortest sequences capable of forming agglomerates (Morasch et al., 2016).

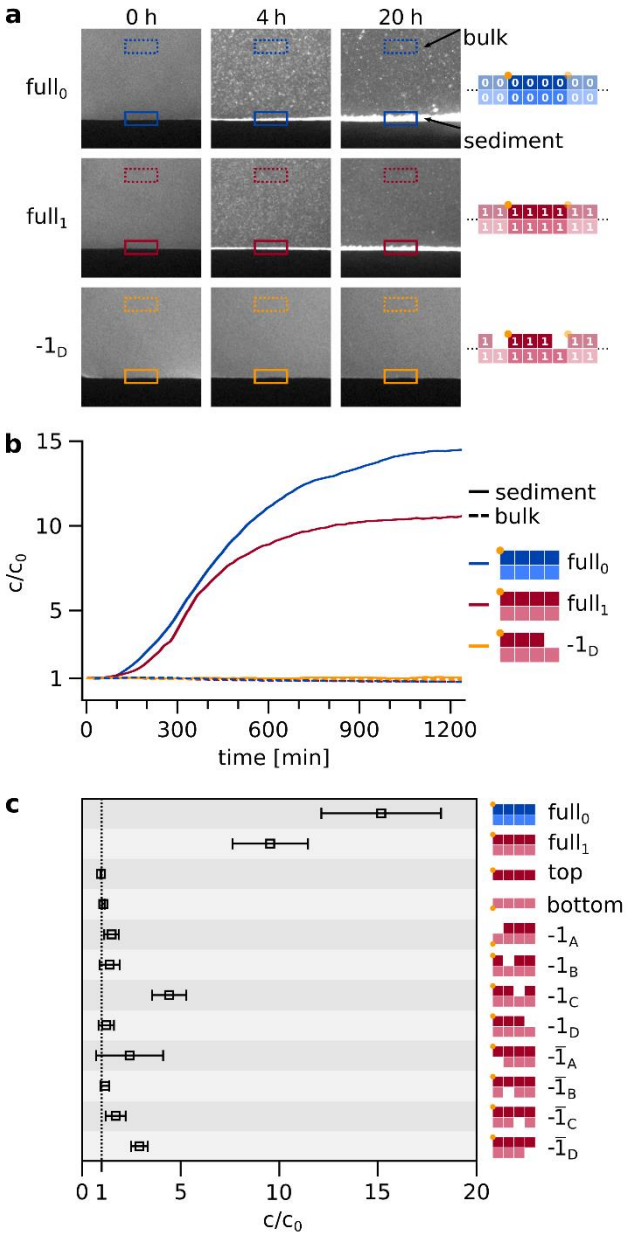

Figure 3. Spontaneous self-assembly and sedimentation of matching hairpins. a*, In a simple, sealed microfluidic chamber (Fig. S5), the hairpin strands can self-assemble into agglomerates and sediment on a timescale of hours. The sample is initially heated to 95 °C for 10 seconds to ensure an unbound initial state, then rapidly (within 30 s) cooled to 25 °C, where self-assembly and sedimentation occur. Note that agglomeration and sedimentation only occur if all eight matching hairpins are provided (top two rows) but not in the case of a knockout (-1$_D$, bottom row). For quantification, the bulk and sediment intensities are normalized by the first frame after heating. Samples contained strands at total concentration of 5 µM, about threefold higher than in Figure 2 and the following replication experiments.* ***b****, Time traces of concentration increase for sediment and bulk of different configurations, same examples as shown in a. The time traces of all further knockout permutations are shown in Figure S6b.* ***c****, Final concentration increase of sediment, relative to first frame after heating, for all configurations. The final values (N≥3) for c/c$_0$ are retrieved from fitting the*

time traces. For the full set of complementary hairpins self-assembly and sedimentation is most pronounced.

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

Templating kinetics. Hybridization between stems of neighboring hairpins (Fig. 1b, step 2) was catalyzed by the presence of already assembled complexes $\overline{0}_{\mathrm{A}}\overline{0}_{\mathrm{B}}\overline{0}_{\mathrm{C}}\overline{0}_{\mathrm{D}}$, confirming its role as a template. Assembly kinetics at 45 °C were recorded in reactions containing 200 nM of each strand for a range of template concentrations. At 120 nM template concentration, 40 % yield was achieved within 10 minutes (Fig. 4b, black line). The untemplated, spontaneous reaction proceeded significantly slower (1.4 % yield, light gray line).

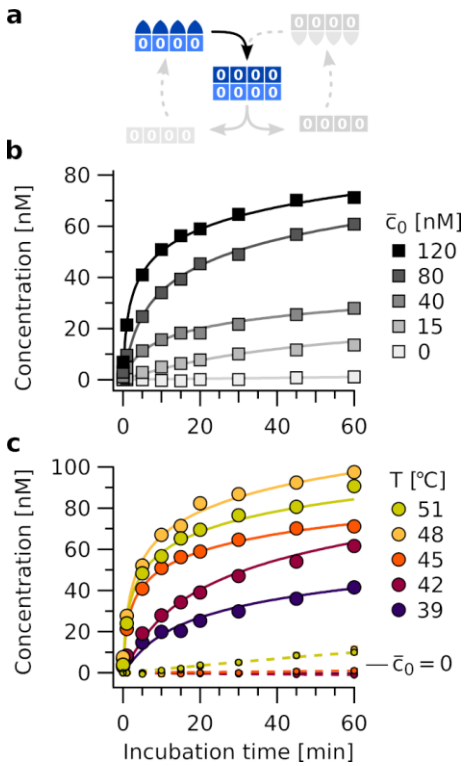

***Figure 4. Isothermal template assisted product formation. a**, Schematic representation of the templating step at constant temperature. **b**, Kinetics of tetramer formation at 45 °C with different starting concentrations of template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ ($\bar{c}_0$). Data includes concentrations of all complexes containing strands of length 4. **c**, Templating observed over a broad temperature range. Large circles show data for reactions at $\bar{c}_0 = 120$ nM of template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$, small circles show the spontaneous formation ($\bar{c}_0 = 0$). The latter increases at T > 45 °C. Above 48 °C, binding of*

monomers to the template gets weaker, slowing down the rate of template assisted formation. This is consistent with the melting temperatures of the information domains (see Fig. S2).

Assembly rates showed a strong dependence on incubation temperature (Fig. 4c). At 39 °C, the reaction proceeded significantly slower than at 42 °C or 45 °C. This is because the hairpins are predominantly in closed configuration and cannot bind to neighboring molecules in the assembly. Binding between complementary information domains still occurs, but the formation of bonds between neighboring strands becomes rate limiting. Above the melting temperature of the information domain (48 °C) (see Fig. S2), template-directed assembly becomes slower. However, the slower kinetics of template-directed product formation are partially superposed by the spontaneous product formation lacking an initial template (Fig. 4c, small circles), which becomes an additional reaction channel due to the now open hairpins.

Exponential amplification. As intermediate step towards replication, we studied amplification reactions under themal oscillations (Fig. 5). The amplification reactions only contained strands encoding for information domain "0", i.e. 0_A, $\overline{0}_\mathrm{A}$, 0_B, $\overline{0}_\mathrm{B}$, …, $\overline{0}_\mathrm{D}$. The strands were subjected to thermal oscillations between T_{base} = 45 °C and T_{peak} = 67 °C. The lower temperature was held for 20 minutes, the upper for one second with temperature ramps amounting to 20±1 seconds in each full cycle. This asymmetric shape of the temperature cycle accords with differences in kinetics of the elongation step and the melting of the information domain. It is typical for trajectories in thermal convection settings with local heating(Braun et al., 2003).

The growth of molecular assemblies with different initial concentrations of template $\overline{0}_\mathrm{A}\overline{0}_\mathrm{B}\overline{0}_\mathrm{C}\overline{0}_\mathrm{D}$ revealed an almost linear dependence of the reaction velocity on the initial amount of template (Fig. 5a,b). This confirms the exponential nature of the replication. The cross-catalytic replication kinetics can be described by a simplistic model that only considers the concentrations $c(t)$ of the template $0_\mathrm{A}0_\mathrm{B}0_\mathrm{C}0_\mathrm{D}$ and its complement $\bar{c}(t)$ of $\overline{0}_\mathrm{A}\overline{0}_\mathrm{B}\overline{0}_\mathrm{C}\overline{0}_\mathrm{D}$:

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

Here, k is the rate of cross-catalysis and k_0 the spontaneous formation rate. For $c(t) \approx \bar{c}(t)$, the model corresponds to simple exponential growth on a per-cycle basis. The model can be solved in closed form but does not account for saturation effects from the depletion of monomers. Therefore, it is not valid for concentrations similar to the total concentration of each strand. Fitting the model to the amplification reactions with 0–45 nM of template $\overline{0}_\mathrm{A}\overline{0}_\mathrm{B}\overline{0}_\mathrm{C}\overline{0}_\mathrm{D}$ revealed rate constants of k = 0.16 cycle^{-1} and k_0 = 0.4 nM cycle^{-1} (Fig. 5b). Amplification was robust with regard to the peak temperature of the oscillations. For T_{peak} below 74 °C, the

reaction remained almost unaffected (Fig. 5c). Above, the temperature is too close to the melting transitions of the hairpin-hairpin duplexes, ranging from 76 °C to 79 °C (Fig. S2).

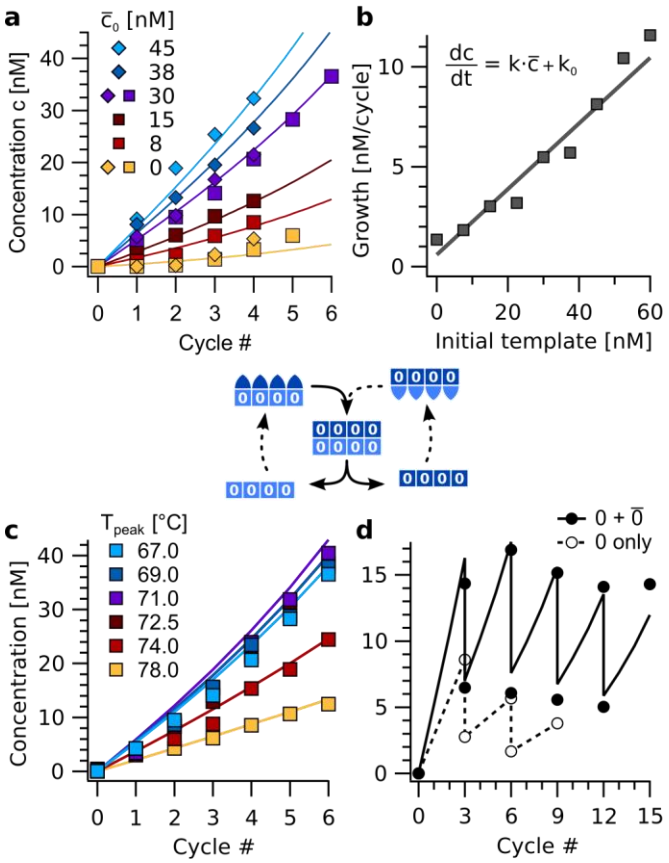

Figure 5. Exponential amplification of a restricted sequence subset with thermal oscillations. a*, Amplification time traces for concentration c for sequence* 0000 *during the first four-six cycles (*T_{peak} *= 67 °C) for template (*$\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$*) concentrations* $\bar{c}_0$ *from 0 to 45 nM. The data was fitted using the cross-catalytic model from equation (1). Strands* 0_A*,* $\bar{0}_A$*,* 0_B*, …,* $\bar{0}_D$ *were used at 200 nM concentration each. Data points show concentrations of complexes 4:4.* ***b****, Initial reaction velocity as a function of initial template concentration* $\bar{c}_0$*. The data points show good agreement with the line calculated from the fits in panel a.* ***c****, Amplification proceeded for peak temperatures below 74 °C. Above, backbone duplexes start to melt, and the complexes are no longer stable. The base temperature was 45 °C, reactions initially contained 30 nM of complex* $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$ *as template.* ***d****, Serial transfer experiment. The reaction containing strands* 0_A*,* $\bar{0}_A$*,* 0_B*, …,* $\bar{0}_D$ *(black circles) survived successive dilution by a factor of 1/2 every 3 cycles at almost constant concentration. In contrast, a reaction with the same amount of template* $\bar{0}_A\bar{0}_B\bar{0}_C\bar{0}_D$*, but lacking monomers* $\bar{0}_{A-D}$*, fades out (open circles). The solid line shows the model from Eq. (1).*

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

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

Sequence replication. The above-mentioned reactions did amplify, but not replicate actual sequence information, as they only contained strands with $0/\overline{0}$ information domains. To study the replication of arbitrary sequences of binary code, replication reactions with all 16 strands encoding for "0" and "1" were performed. To discriminate sequences encoded in equally sized complexes and deduce error rates, we compared these results to those from different reaction runs with defects, i.e. lacking one or two of the hairpin sequences required for the faithful replication of a particular template. Reference reactions contained all 16 strands (0_A, $\overline{0}_A$, 1_A, $\overline{1}_A$, 0_B, …, $\overline{1}_D$) at 100 nM each, and were run for each of three different template sequences ($\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$, $\overline{0}_A\overline{1}_B\overline{0}_C\overline{1}_D$, and $\overline{0}_A\overline{0}_B\overline{1}_C\overline{1}_D$) (Fig. 6). Yields were quantified from reaction time traces, extracted by integrating the intensities of all gel bands containing tetramers.

Leaving out a single strand (reaction label "+++−", e.g. leaving out 0_D for template $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$) reduced the yield of full-size product to 40 % (Fig. 6a, b). Instead, mostly complex $0_A0_B0_C$:$\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ (3:4) was formed, in particular during the first few cycles (Fig. S3). This was expected given the lack of strand 0_D and provides an upper limit on the error rate of the full replication. The fact that the full reaction produced almost no complexes 3:4 or 4:3 indicates that the incomplete product was indeed caused by the lack of a particular strand.

Removal of a further strand either directly next to the previous one ("++−−", missing strands $0_C/0_D$) or not ("+−+−", missing strands $0_B/0_D$) reduced the yield of tetramers even further. Replication of the other two templates $\overline{0}_A\overline{1}_B\overline{0}_C\overline{1}_D$ and $\overline{0}_A\overline{0}_B\overline{1}_C\overline{1}_D$ produced very similar results. End points after 6 cycles are given in Fig. 6c for each of the three templates as well as an average over template sequences (horizontal lines). A single defect reduced the yield of tetramer complexes to about 40 %, two missing strands to 15–20 %.

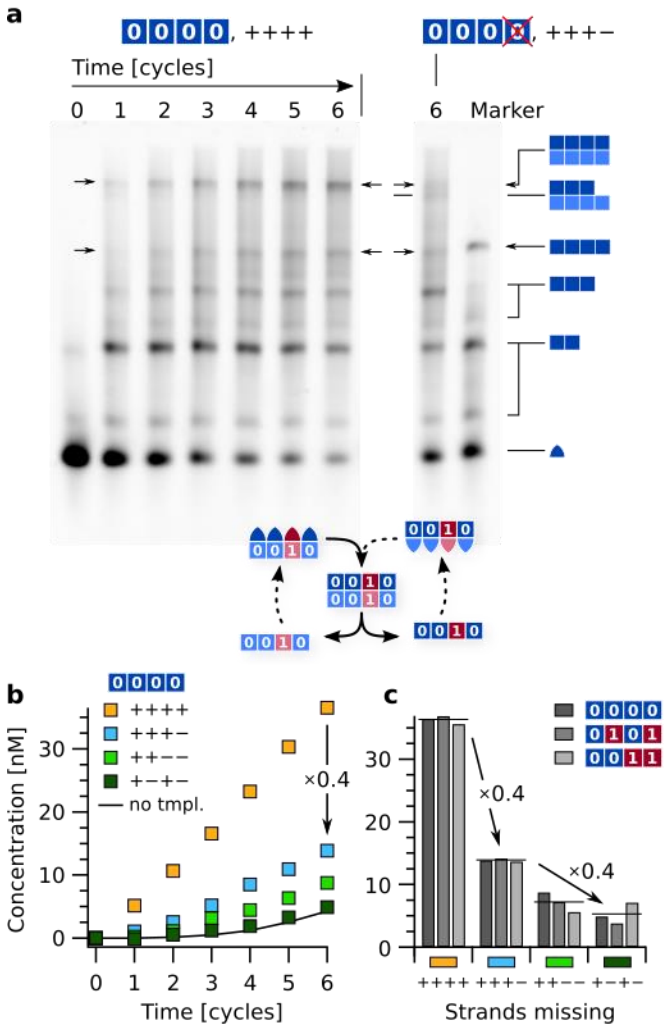

Figure 6. Sequence replication with thermal oscillations. ***a****, Replication of sequence* $0_A0_B0_C0_D$*. Native-PAGE results comparing the reaction of all 16 strands ("++++") with the reaction lacking strand* 0_D *("+++−"). Reactions were started with 15 nM initial template* $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$*. All strands were present at 100 nM each. The defective set "+++−" mostly produced 3:4 complexes instead of 4:4 complexes (see schematics on the right). The overall yield of tetramer-containing complexes was greatly reduced. As size reference, the marker lane contained complexes* $0_A0_B0_C0_D$*,* $0_A0_B0_C$*,* 0_A0_B*, and monomers* 0_A*. The complete gel is presented in Fig. S3.* ***b****, Reaction time traces of the whole sequence network (yellow) and three defective sets with missing strands. Data was integrated by quantitative image analysis from electrophoresis gels using covalent markers on the A-hairpin counting all complexes containing tetramers. Reactions lacked strands* 0_D *("+++−"),* $0_C/0_D$ *("++−−"), and* $0_B/0_D$ *("+−+−"). All reactions were initiated with 15 nM of* $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$*. The solid line shows data from reaction "++++" without template.* ***c****, End point comparison of reactions with templates* $\overline{0}_A\overline{0}_B\overline{0}_C\overline{0}_D$ *(panels a, b),* $\overline{0}_A\overline{1}_B\overline{0}_C\overline{1}_D$*, and* $\overline{0}_A\overline{0}_B\overline{1}_C\overline{1}_D$ *after 6 cycles. Horizontal lines indicate averages of the three template sequences. A single missing strand reduced product yield to 40 %, two missing strands to 15–20 %.*

Replication fidelity. The observed rate of erroneous product formation can be attributed to the spontaneous background rate (Fig. 4b, c and Fig. 6b). Reaction "+−+−" (dark green) proceeded the same as the untemplated reference reaction (solid line), as it did not contain any strands that could bind next to each other to the template and form a backbone duplex (Fig. 6b).

For the templated reactions "+++−" and "++−−", templating worked for partial sequences, producing intermediate yields.

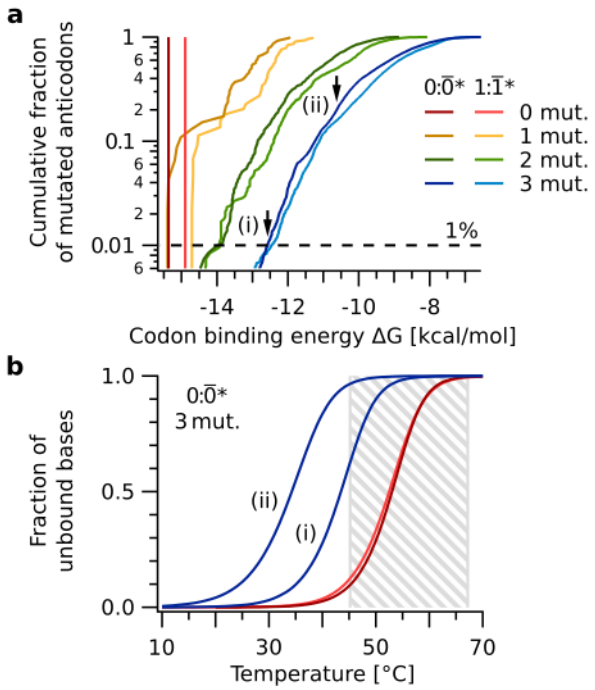

Figure 7. Sequence space analysis of information domain binding. *The binding energies quantify the ability of the replication mechanism to discriminate nucleotide mutations.* ***a****, Cumulative free energy distributions of information domain duplexes* $0{:}\overline{0}$ *(red),* $1{:}\overline{1}$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

The reduction in yield caused by a single defect (i.e. missing strand) to 40 % (and to ca. 16 % for two defects) translates into a replication fidelity per information domain of 1/(1 + 0.4) = 71 %. To compare this value to a per-nucleotide replication (i.e. polymerization) process, we estimated an equivalent per-nucleotide fidelity of the information domain replication. To do so, we compared the properties of the duplex $0{:}\overline{0}$ to duplexes $0{:}\overline{0}$*, where $\overline{0}$* differs from $\overline{0}$ by K point mutations. As the criterion to identify mutants $\overline{0}$* too different to participate in the replication, we assumed a reduction in information domain melting temperature T_m by at least 10 °C (compared to the original duplex $0{:}\overline{0}$) to be sufficient. This was inferred from the width of the melting transition of duplex $0{:}\overline{0}$ (Fig 7b), where a shift of 10 °C corresponds to an increase of the unbound fraction from 0.08 at T_{base} = 45 °C to 0.66 at 55 °C. In terms of information domain duplex free energies, this sufficient difference corresponds to

ΔG($0{:}\overline{0}$*) ≥ −12.5 kcal/mol compared to ΔG($0{:}\overline{0}$) = −15.4 kcal/mol. 99 % of all duplexes $0{:}\overline{0}$*, with $\overline{0}$* containing three point mutations, met that criterion (Fig. 7a).

To calculate the per-nucleotide fidelity p, we then, for simplicity, assumed that the replication did not differentiate between information domain $\overline{0}$ and any information domain $\overline{0}$* with less than K point mutations. The fidelity per information domain $p_K(N)$ is given by a cumulative binomial distribution

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

Here, N is the information domain length, and p the per-nucleotide replication fidelity. Using K = 3, N = 15, and the measured value of $p_3(15) = 0.71$, one finds a per-nucleotide fidelity of p = 87.5 %. In fact, mutants with a weak reduction in binding energy are those with mutations at the terminal bases. Information domains $\overline{0}$* with mutations at two internal bases all show similar properties as information domains with a total of three mutations (Fig. S4). Including this refinement, the per-nucleotide fidelity reads 92 %. This means that a per-nucleotide replication process would need a replication fidelity of 88–92 % to produce sequences with an error rate equivalent to the presented mechanism.

Discussion

The experiments indicate that a cross-catalytic replicator can be made from short sequences without covalent bonds under a simple non-equilibrium setting of periodic thermal oscillations. The replication is fast, and proceeds within a few thermal oscillations of 20 minutes each. This velocity is comparable to other replicators (Kindermann et al., 2005), cross-ligating ribozymes (Michael P. Robertson & Joyce, 2014), or autocatalytic DNA networks (Yin et al., 2008). The required thermal oscillations can be obtained by laminar convection in thermal gradients (Braun et al., 2003), which also accumulates oligonucleotides (Mast et al., 2013). Depending on the envisioned environment, the mechanism could also be driven by thermochemical oscillations (Ball & Brindley, 2014) or convection in pH gradients (Keil et al., 2017).

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

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

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

Pre-selection of nucleic acids for the presented hairpin-driven replication mechanism can be provided by highly sequence-specific gelation of DNA. This gel formation has been shown to be most efficient with double hairpin structures very similar to the tRNA-like sequences used in this study (Morasch et al., 2016). For our replication system, we have demonstrated this in Figure 3 by showing the spontaneous formation of agglomerates and their sedimentation under gravity if all molecules of the assembly were present. This self-selection shows a possible pathway how the system can emerge from random or semi-random sequences, for example in a flow or a convection system where the molecules are selected as macroscopic agglomerate (Mast et al., 2013). Another selection pressure could stem from the biased hydrolysis of double stranded nucleotide backbones, which favors assembled complexes over the initial hairpins (Obermayer et al., 2011).

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

The proposed replication mechanism of assemblies from tRNA-like sequences allows to speculate about a transition from an autonomous replication of successions of information domains to the translation of codon sequences encoded in modern mRNA (Fig. 1a). Short peptide-RNA hybrids (Griesser et al., 2017; Jauker et al., 2015), combined with specific interactions between 3’-terminal amino acids and the anticodons, could have given rise to a

primitive genetic code. The spatial arrangement of tRNA-like sequences that are replicated by the presented mechanism would translate into a spatial arrangement of the amino acids or short peptide tails that are attached to the strands in a codon-encoded manner (Schimmel & Henderson, 1994). The next stage would then be the detachment and linking of the tails to form longer peptides. Eventually, tRNA would transition to its modern role in protein translation. The mechanism thus proposes a hypothesis for the emergence of predecessors of tRNA, independent of protein translation. This is crucial for models of the evolution of translation, because it justifies the existence of tRNA before it was utilized in an early translation process.

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

Materials and Methods

Strand design. DNA double-hairpin sequences were designed using the NUPACK software package(Zadeh et al., 2011). In addition to the secondary structures of the double-hairpins, the design algorithm was constrained by all target dimers. Candidate sequences were selected for optimal homogeneity of binding energies and melting temperatures. Backbone domains connecting consecutive strands (e.g. $0_A 0_B 0_C$) had to be the most stable bonds in the system, in particular more stable than between a template and a newly formed product complex (e.g. $0_B{:}\overline{0}_B$). On the other hand, hairpin melting temperatures had to be low enough to allow for a sufficient degree of thermal fluctuations. To reconcile this with the length of the strands, mismatches were introduced in the hairpin stems. The sequences of all strands are listed in Table S1.

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

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

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

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

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

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

For the determination of reaction yields, the intensities of all gel bands containing strands of the sequence length of interest were added up. For strings of four strands, these were the single tetramer as well as its complex with di- and tri- and tetramers. Single strands separated from their complements during electrophoresis (Fig. 2 and Fig. S3).

Thermal melting curves. Thermal melting curves were measured using either UV absorbance at 260 nm wavelength in a UV/Vis spectrometer (JASCO V-650, 1 cm optical path length), via quenching of the Cy5 label at the 5'-end of strand 0_A (excitation: 620–650 nm, detection: 675–690 nm), or using fluorescence of the intercalating dye SYBR Green I (excitation: 450–490 nm, detection: 510–530 nm). Fluorescence measurements were performed in a PCR cycler (Bio-Rad C1000). Samples measured via fluorescence were at 200 nM of each strand, those measured via UV absorption contained 1 µM total DNA concentration to improve the signal-to-noise ratio. Before analysis of the melting curves (Rodbard & Chrambach, 1970), data were corrected for baseline signals from reference samples containing buffer and intercalating dye, if applicable.

Self-assembly and sedimentation analysis. The samples were mixed in the replication buffer (150 mM NaCl, 20 mM $MgCl_2$, 20 mM Tris-HCl pH 8) at a total oligomer concentration of 5 µM, i.e. varying concentration per strand depending on the number of different strands in the configuration (4, 7 or 8). The microfluidic chamber was assembled with a custom cut, 500 µm thick, Teflon foil placed between two plane sapphires (Fig. S5). Three Peltier elements (QuickCool QC-31-1.4-3.7AS, purchased from Conrad Electronics, Germany) were attached to the backside of the chamber to provide full temperature control. The chamber was initially flushed with 3M™ Novec™ 7500 (3M, Germany) to avoid bubble formation. The samples were pipetted into the microfluidic chamber through the 0.5 mm channels using microloader pipette tips (Eppendorf, Germany). The chamber was then sealed with Parafilm and heated to 95°C for 10 seconds to fully separate the strands and cooled rapidly (within 30 s) to 25°C. Assembly and sedimentation were monitored for 20 hours on a fluorescence microscope (Axiotech Vario, Zeiss, Germany) with two LEDs (490 nm and 625 nm, Thorlabs, Germany) using a 2.5 x objective (Fluar, Zeiss, Germany). The observed sedimentation was independent of the attached dye and its position (Fig. S6c). The ratio of sedimented fluorescence relative to the first frame after heating was used to quantify sedimentation (Fig. 3). The sedimentation time-traces (Fig. 3b) were fitted with a Sigmoid function to determine the final concentration increase c/c_0 (Fig. 3c). The experiment was also performed with random 84 nt DNA strands at 5 µM total concentration to exclude unspecific agglomeration (Fig. S6c).

Acknowledgments

We gratefully acknowledge financial support by the European Research Council through the ERC Starting Grant, the DFG Cluster of Excellence NIM (Nanosystems Initiative Munich), the Center for NanoScience (CeNS). We thank for funding from Quantitative Biology Munich (QBM), CRC 1032 NanoAgents and the CRC 235 Emergence of Life (Project-ID 364653263). We appreciate the fruitful discussions in the Simons Collaboration on the Origins of Life, thank for the measurements by Thomas Rind and acknowledge discussions with Tim Liedl, Christof Mast and Lorenz Keil. We thank Filiz Civril, Adriana Serrão and Thomas Matreux for comments on the manuscript.

Author Contributions

A. K. and S. A. L. performed the experiments and analyzed the data. A. K., S. A. L., and D. B. conceived and designed the experiments, and wrote the paper.

Additional information

The authors declare no competing financial interests

References

Attwater, J., Wochner, A., & Holliger, P. (2013). In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry*, *5*(12), 1011–1018. https://doi.org/10.1038/nchem.1781

Ball, R., & Brindley, J. (2014). Hydrogen peroxide thermochemical oscillator as driver for primordial RNA replication. *Journal of The Royal Society Interface*, *11*(95), 20131052. https://doi.org/10.1098/rsif.2013.1052

Barany, F. (1991). Genetic disease detection and DNA amplification using cloned thermostable ligase. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(1), 189–193. https://doi.org/10.1073/pnas.88.1.189

Braun, D., Goddard, N. L., & Libchaber, A. (2003). Exponential DNA Replication by Laminar Convection. *Physical Review Letters*, *91*(15), 158103. https://doi.org/10.1103/PhysRevLett.91.158103

Dange, V., Van Atta, R., & Hecht, S. (1990). A Mn2(+)-dependent ribozyme. *Science*, *248*(4955), 585–588. https://doi.org/10.1126/science.2185542

Dirks, R. M., & Pierce, N. A. (2004). Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(43), 15275–15278. https://doi.org/10.1073/pnas.0407024101

Dolinnaya, N. G., Sokolova, N. I., Gryaznova, O. I., & Shabarova, Z. A. (1988). Site-directed modification of DNA duplexes by chemical ligation. *Nucleic Acids Research*, *16*(9), 3721–3738.

Doudna, J. A., Couture, S., & Szostak, J. W. (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science*, *251*(5001), 1605–1608. https://doi.org/10.1126/science.1707185

Engelhart, A. E., Cafferty, B. J., Okafor, C. D., Chen, M. C., Williams, L. D., Lynn, D. G., & Hud, N. V. (2012). Nonenzymatic Ligation of DNA with a Reversible Step and a Final Linkage that Can Be Used in PCR. *ChemBioChem*, *13*(8), 1121–1124. https://doi.org/10.1002/cbic.201200167

Green, S. J., Lubrich, D., & Turberfield, A. J. (2006). DNA Hairpins: Fuel for Autonomous DNA Devices. *Biophysical Journal*, *91*(8), 2966–2975. https://doi.org/10.1529/biophysj.106.084681

Griesser, H., Bechthold, M., Tremmel, P., Kervio, E., & Richert, C. (2017). Amino Acid-Specific, Ribonucleotide-Promoted Peptide Formation in the Absence of Enzymes. *Angewandte Chemie*, *129*(5), 1244–1248. https://doi.org/10.1002/ange.201610651

Hopfield, J. J. (1978). Origin of the genetic code: a testable hypothesis based on tRNA structure, sequence, and kinetic proofreading. *Proceedings of the National Academy of*

549 *Sciences*, *75*(9), 4334–4338. https://doi.org/10.1073/pnas.75.9.4334

550 Horning, D. P., & Joyce, G. F. (2016). Amplification of RNA by an RNA polymerase ribozyme.
551 *Proceedings of the National Academy of Sciences*, *113*(35), 9786–9791.
552 https://doi.org/10.1073/pnas.1610103113

553 Jauker, M., Griesser, H., & Richert, C. (2015). Spontaneous Formation of RNA Strands,
554 Peptidyl RNA, and Cofactors. *Angewandte Chemie - International Edition*.
555 https://doi.org/10.1002/anie.201506593

556 Keil, L. M. R., Möller, F. M., Kieß, M., Kudella, P. W., & Mast, C. B. (2017). Proton gradients
557 and pH oscillations emerge from heat flow at the microscale. *Nature Communications*.
558 https://doi.org/10.1038/s41467-017-02065-3

559 Kindermann, M., Stahl, I., Reimold, M., Pankau, W. M., & von Kiedrowski, G. (2005). Systems
560 Chemistry: Kinetic and Computational Analysis of a Nearly Exponential Organic
561 Replicator. *Angewandte Chemie International Edition*, *44*(41), 6750–6755.
562 https://doi.org/10.1002/anie.200501527

563 Krammer, H., Möller, F. M., & Braun, D. (2012). Thermal, Autonomous Replicator Made from
564 Transfer RNA. *Physical Review Letters*, *108*(23), 238104.
565 https://doi.org/10.1103/PhysRevLett.108.238104

566 Li, Y., & Breaker, R. R. (1999). Kinetics of RNA degradation by specific base catalysis of
567 transesterification involving the 2γ-hydroxyl group. *Journal of the American Chemical*
568 *Society*, *121*(23), 5364–5372. https://doi.org/10.1021/ja990592p

569 Mast, C. B., Schink, S., Gerland, U., & Braun, D. (2013). Escalation of polymerization in a
570 thermal gradient. *Proceedings of the National Academy of Sciences*, *110*(20), 8030–
571 8035. https://doi.org/10.1073/pnas.1303222110

572 Morasch, M., Braun, D., & Mast, C. B. (2016). Heat-Flow-Driven Oligonucleotide Gelation
573 Separates Single-Base Differences. *Angewandte Chemie - International Edition*, *55*(23),
574 6676–6679. https://doi.org/10.1002/anie.201601886

575 Mutschler, H., Wochner, A., & Holliger, P. (2015). Freeze-thaw cycles as drivers of complex
576 ribozyme assembly. *Nature Chemistry*, *7*(6), 502–508.
577 https://doi.org/10.1038/nchem.2251

578 Nguyen, D. T., & Saleh, O. A. (2017). Tuning phase and aging of DNA hydrogels through
579 molecular design. *Soft Matter*, *13*(32), 5421–5427. https://doi.org/10.1039/c7sm00557a

580 O'Flaherty, D. K., Zhou, L., & Szostak, J. W. (2019). Nonenzymatic template-directed
581 synthesis of mixed-sequence 3′-NP-DNA up to 25 nucleotides long inside model
582 protocells. *Journal of the American Chemical Society*, *141*(26), 10481–10488.
583 https://doi.org/10.1021/jacs.9b04858

584 Obermayer, B., Krammer, H., Braun, D., & Gerland, U. (2011). Emergence of Information
585 Transmission in a Prebiotic RNA Reactor. *Physical Review Letters*, *107*(1), 018101.
586 https://doi.org/10.1103/PhysRevLett.107.018101

587 Orgel, L. E. (2004). Prebiotic Chemistry and the Origin of the RNA World. *Critical Reviews in*
588 *Biochemistry and Molecular Biology*, *39*(2), 99–123.
589 https://doi.org/10.1080/10409230490460765

590 Patzke, V., McCaskill, J. S., & von Kiedrowski, G. (2014). DNA with 3′-5′-Disulfide Links-Rapid
591 Chemical Ligation through Isosteric Replacement. *Angewandte Chemie International*
592 *Edition*, *53*(16), 4222–4226. https://doi.org/10.1002/anie.201310644

593 Paul, N., & Joyce, G. F. (2002). *A self-replicating ligase ribozyme*. *99*(20), 12733–12740.

594 Pino, S., Costanzo, G., Giorgi, A., & Di Mauro, E. (2011). Sequence Complementarity-Driven
595 Nonenzymatic Ligation of RNA. *Biochemistry*, *50*(14), 2994–3003.
596 https://doi.org/10.1021/bi101981z

597 Robertson, M. P., Hesselberth, J. R., & Ellington, A. D. (2001). Optimization and optimality of
598 a short ribozyme ligase that joins non-Watson-Crick base pairings. *Rna*, *7*(4), 513–523.
599 https://doi.org/10.1017/S1355838201002199

600 Robertson, Michael P., & Joyce, G. F. (2014). Highly Efficient Self-Replicating RNA Enzymes.
601 *Chemistry & Biology*, *21*(2), 238–245. https://doi.org/10.1016/j.chembiol.2013.12.004

602 Robertson, Michael P, & Joyce, G. F. (2012). The Origins of the RNA World. *Cold Spring*
603 *Harbor Perspectives in Biology*, *4*(5), a003608–a003608.
604 https://doi.org/10.1101/cshperspect.a003608

605 Rodbard, D., & Chrambach, A. (1970). Unified theory for gel electrophoresis and gel filtration.
606 *Proceedings of the National Academy of Sciences of the United States of America*,
607 *65*(4), 970–977. https://doi.org/10.1073/pnas.65.4.970

608 Rohatgi, R., Bartel, D. P., & Szostak, J. W. (1996). *Nonenzymatic , Template-Directed*
609 *Ligation of Oligoribonucleotides Is Highly Regioselective for the Formation of 3 ' - 5 '*
610 *Phosphodiester Bonds*. *9*, 3340–3344. https://doi.org/10.1021/ja9537134

611 Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H., & Arnheim, N. (1985).
612 Enzymatic amplification of beta-globin genomic sequences and restriction site analysis
613 for diagnosis of sickle cell anemia. *Science*, *230*(4732), 1350–1354.
614 https://doi.org/10.1126/science.2999980

615 Schimmel, P., & Henderson, B. (1994). Possible role of aminoacyl-RNA complexes in
616 noncoded peptide synthesis and origin of coded synthesis. *Proceedings of the National*
617 *Academy of Sciences*, *91*(24), 11283–11286. https://doi.org/10.1073/pnas.91.24.11283

618 Schulman, R., Yurke, B., & Winfree, E. (2012). Robust self-replication of combinatorial

information via crystal growth and scission. *Proceedings of the National Academy of Sciences*, *109*(17), 6405–6410. https://doi.org/10.1073/pnas.1117813109

Sharp, S. J., Schaack, J., Cooley, L., Burke, D. J., & Soil, D. (1985). Structure and Transcription of Eukaryotic tRNA Gene. *Critical Reviews in Biochemistry*, *19*(2), 107–144. https://doi.org/10.3109/10409238509082541

Stadlbauer, P., Šponer, J., Costanzo, G., Di Mauro, E., Pino, S., & Šponer, J. E. (2015). Tetraloop-like Geometries Could Form the Basis of the Catalytic Activity of the Most Ancient Ribooligonucleotides. *Chemistry - A European Journal*, *21*(9), 3596–3604. https://doi.org/10.1002/chem.201406140

Szathmáry, E. (2006). The origin of replicators and reproducers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1474), 1761–1776. https://doi.org/10.1098/rstb.2006.1912

Szostak, J. W. (2012). The eightfold path to non-enzymatic RNA replication. *Journal of Systems Chemistry*, *3*(1), 2. https://doi.org/10.1186/1759-2208-3-2

Turk, R. M., Illangasekare, M., & Yarus, M. (2011). Catalyzed and Spontaneous Reactions on Ribozyme Ribose. *Journal of the American Chemical Society*, *133*(15), 6044–6050. https://doi.org/10.1021/ja200275h

Vaidya, N., Manapat, M. L., Chen, I. A., Xulvi-Brunet, R., Hayden, E. J., & Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators. *Nature*, *491*(7422), 72–77. https://doi.org/10.1038/nature11549

Vlassov, A. V, Kazakov, S. A., Johnston, B. H., & Landweber, L. F. (2005). The RNA World on Ice: A New Scenario for the Emergence of RNA Information. *Journal of Molecular Evolution*, *61*(2), 264–273. https://doi.org/10.1007/s00239-004-0362-7

von Kiedrowski, G. (1986). A Self-Replicating Hexadeoxynucleotide. *Angewandte Chemie International Edition in English*, *25*(10), 932–935. https://doi.org/10.1002/anie.198609322

von Neumann, J. (1951). *The general and logical theory of automata*. 1–41.

Walton, T., DasGupta, S., Duzdevich, D., Oh, S. S., & Szostak, J. W. (2020). In vitro selection of ribozyme ligases that use prebiotically plausible 2-aminoimidazole–activated substrates. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(11), 5741–5748. https://doi.org/10.1073/pnas.1914367117

Yin, P., Choi, H. M. T., Calvert, C. R., & Pierce, N. A. (2008). Programming biomolecular self-assembly pathways. *Nature*, *451*(7176), 318–322. https://doi.org/10.1038/nature06451

Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., & Pierce, N. A. (2011). NUPACK: Analysis and design of nucleic acid systems.

654 *Journal of Computational Chemistry*, *32*(1), 170–173. https://doi.org/10.1002/jcc.21596

655 Zhang, D. Y., Turberfield, A. J., Yurke, B., & Winfree, E. (2009). *Engineering Entropy-Driven*
656 *Reactions*. *318*(November 2007), 1121–1125.

657 Zhou, L., Kim, S. C., Ho, K. H., O'Flaherty, D. K., Giurgiu, C., Wright, T. H., & Szostak, J. W.
658 (2019). Non-enzymatic primer extension with strand displacement. *ELife*, *8*, 1–14.
659 https://doi.org/10.7554/eLife.51888