technische universität dortmund

Probing RNA catalysis in protocellular settings relevant to the origin of life

Dissertation

Submitted for the degree of

Doctor of Natural Science

(Dr. rer. nat.)

by

Elia Salibi

TU Dortmund University

Department of Chemistry and Chemical Biology

1st Examiner: Prof. Dr. Hannes Mutschler

2nd Examiner: Prof. Dr. Daniel Summerer

If I have seen further, it is by standing on the shoulders of giants.

Isaac Newton

Preface

When I first started brainstorming how I wanted to write this thesis, I also became interested in attempting to trace historical achievements and present an encompassing review on origin of life research. I quickly realised that this goal was perhaps too ambitious. So much has been done, so many leads have been chased, so many hypotheses postulated. To stay true to my initial motivations, I have therefore selected those which, in my opinion, seemed to have expanded knowledge on the origin and evolution of life. The finished work is structured in the following manner: I first discuss what I deemed to be historical landmarks in origin of life research up until the late 1980s. Next, I discuss the history of our Earth from a geological perspective as well as the hypothesised geochemical niches in which life could have emerged, because to truly understand how we came to be, I believe we must first understand where is it we came to be. I then transition into exposing the efforts made to support a prominent hypothesis that has been put forward, reviewing the feats of prebiotic chemists, biochemists and biologists alike. Finally, I introduce different methods of compartmentalisation that potentially pre-date modern membrane system as they are of high relevance to the topic of this work. In the following sections, I present some of the work that we have done within the last few years to contribute to the field and attempt to place it in context. My hope is that the reader will appreciate how it all connects to the research undertook during my stay with Hannes, and that they leave this essay with a somewhat comprehensive view of the interdisciplinary research that deals with the emergence and evolution of life.

Work undertaken in this thesis has resulted in articles that have been published in peer-reviewed journals: (# equal contribution; * corresponding author)

Co-first author publications:

- **Salibi[#] E.**, Peter B.[#], Schwille P.*, Mutschler H.* Periodic temperature changes drive the proliferation of self-replicating RNAs in vesicle populations. *Nat Commun* 14, 1222 (**2023**). https://doi.org/10.1038/s41467-023-36940-z
- Le Vay K.[#]*, **Salibi E**.[#], Ghosh B., Tang D. TY.*, Mutschler H.* Ribozyme activity modulates the physical properties of RNA–peptide coacervates. *eLife* 12:e83543 (**2023**). https://doi.org/10.7554/eLife.83543

Contributing author publications:

- Salditt A., Karr L., **Salibi E.**, Le Vay K., Braun D.*, Mutschler H.* Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment. *Nat Commun* 14, 1495 (**2023**). https://doi.org/10.1038/s41467-023-37206-4
- Matreux[#] T., Le Vay[#] K., Schmid A., Aikkila P., Belohlavek L., Çalışkanoğlu A.Z., **Salibi E.,** Kühnlein A., Springsklee C., Scheu B., Dingwell D.B., Braun D., Mutschler H.*, Mast C.B.*. Heat flows in rock cracks naturally optimize salt compositions for ribozymes. *Nat. Chem.* 13, 1038–1045 (**2021**). https://doi.org/10.1038/s41557-021-00772-5
- Donau C., Späth F., Sosson M., Kriebisch B., Schnitter F., Tena-Solsona M., Kang H.S., **Salibi E.**, Sattler M., Mutschler H., Boekhoven J.*. Active coacervate droplets as a model for membraneless organelles and protocells. *Nat Commun* 11, 5167 (**2020**). https://doi.org/10.1038/s41467-020-18815-9
- Le Vay K.*, **Salibi[#] E.**, Song[#] E.Y., Mutschler H.* Nucleic Acid Catalysis under Potential Prebiotic Conditions. *Chem. Asian J.* 15, 214-230. (**2019**). https://doi.org/10.1002/asia.201901205

Summary

The emergence of life has long been a topic of high interest. Of the many hypotheses proposed for the emergence of early cellular predecessors of our last universal common ancestor (LUCA), the RNA World hypothesis is a prime candidate. It posits a time before the evolution of modern DNA / protein-based life, where both roles of information storage and processing were played by RNA, while DNA and complex proteins had not yet evolved. With the advent of *in vitro* evolution, selection experiments have yielded a variety of different catalytic RNAs (ribozymes) able to catalyse a range of chemical reactions in a sequence-dependent manner. Nonetheless, a number of outstanding problems remain, such as the dissemination of genetic information among protocell populations in the absence of complex cell division machinery, genotype-phenotype coupling in RNA-peptide based membrane-less compartments and finally, the general sequence-independent templating abilities of ribozymes. In this work, cycles of freezing and thawing were employed to drive the proliferation of encapsulated RNA replicators among feedstock vesicles, allowing their survival following multiple rounds of dilution via self-amplification, thereby enabling a primitive form of lateral gene transfer among protocell populations and eliminating the need for sophisticated cell division apparatus. In the following section, a ribozyme ligase was re-engineered to perform concatenated ligation of an RNA substrate leading to a drastic increase in length. Ribozyme activity was subsequently characterised in the presence of short lysine homopeptides that allow phase separation and RNA partitioning. Moreover, ribozyme activity was found to modulate the physical properties of coacervates, demonstrating a primitive form of genotype-phenotype linkage in membraneless model protocells. Finally, a group I intron-derived ribozyme ligase that performs templated ligation reaction was employed as a general templated RNA ligase and characterised for the synthesis of a variety of functional RNA sequences. The system was subsequently adopted by collaborators as a starting point for the development of environmentally-driven RNA replication systems. Taken together, the data presented here offer insight into the possible mechanisms for persistence and proliferation of prebiotically relevant protocell models in environmental settings as well as an overview of the 'proto-biosphere' on early Earth and potential transitional stages of chemical and biological evolution leading to the emergence of the first cellular systems as we know them.

Zusammenfassung

Die Entstehung des Lebens ist seit langem ein Thema von großem Interesse. Von den vielen Hypothesen, die für die Entstehung früher zellulärer Vorläufer unseres letzten gemeinsamen Vorfahren vorgeschlagen wurden, ist die RNA-Welt-Hypothese ein Hauptkandidat. Sie geht von einer Zeit vor der Evolution des modernen DNA-/Protein-basierten Lebens aus, in der die RNA sowohl die Rolle der Informationsspeicherung als auch die der Informationsverarbeitung übernahm, während sich die DNA und komplexe Proteine noch nicht entwickelt hatten. Mit dem Aufkommen der In-vitro-Evolution haben Selektionsexperimente eine Vielzahl verschiedener katalytischer RNAs (Ribozyme) hervorgebracht, die in der Lage sind, eine Reihe chemischer Reaktionen in sequenzabhängiger Weise zu katalysieren. Dennoch gibt es noch eine Reihe offener Probleme, wie z. B. die Verbreitung genetischer Informationen zwischen den Populationen von Protozellen in Abwesenheit einer komplexen Zellteilungsmaschinerie, die Genotyp-Phänotyp-Kopplung in membranlosen Kompartimenten auf RNA-Peptidbasis und schließlich die allgemeinen sequenzunabhängigen Template-Fähigkeiten von Ribozymen. In dieser Arbeit wurden Zyklen des Einfrierens und Auftauens eingesetzt, um die Vermehrung eingekapselter RNA-Replikatoren in den Ausgangsvesikeln anzutreiben und ihr Überleben nach mehreren Verdünnungsrunden durch Selbstamplifikation zu ermöglichen, wodurch eine primitive Form des lateralen Gentransfers zwischen Protozellpopulationen ermöglicht wird und die Notwendigkeit eines hochentwickelten Zellteilungsapparats entfällt. Im folgenden Abschnitt wurde eine Ribozym-Ligase so umgestaltet, dass sie eine verkettete Ligation eines RNA-Substrats durchführt, was zu einer drastischen Zunahme der Länge führt. Die Ribozymaktivität wurde anschließend in Gegenwart von kurzen Polylysin-Polymeren charakterisiert, die eine Phasentrennung und RNA-Verteilung ermöglichen. Darüber hinaus wurde festgestellt, dass die Ribozymaktivität die physikalischen Eigenschaften von Koazervaten moduliert, was eine primitive Form der Genotyp-Phänotyp-Verknüpfung in membranlosen Modell-Protozellen demonstriert. Schließlich wurde eine von einem Gruppe-I-Intron abgeleitete Ribozym-Ligase, die eine matrizenabhängige Ligationsreaktion durchführt, als allgemeine RNA-Ligase eingesetzt und für die Synthese verschiedener funktioneller RNA-Sequenzen charakterisiert. Das System wurde anschließend von Kollaborationspartnern als Ausgangspunkt für die Entwicklung von umweltgesteuerten RNA-Replikationssystemen übernommen. Zusammengenommen bieten die hier vorgestellten Daten einen Einblick in mögliche Persistenz- und Vermehrungsstrategien präbiotisch relevanter Protozellmodellen

sowie einen Überblick über eine mögliche "Proto-Biosphäre" auf der frühen Erde und potentielle evolutionäre Zwischenstufen, die zur Entstehung der ersten zellulären Systeme geführt haben könnten.

Table of Contents

1. Introduction

1.1. Historical sketch on origin of life research

1.1.1. Vitalism and matter

In the 1800s, vitalism was the prevailing hypothesis that set matter apart, claiming that living and dead matter were fundamentally different. This stark contrast was due to the presence of vital molecules, or organic molecules that are exclusively found in living matter, which can only be produced by living organisms that possess a vital force. However, in 1828, Friedrich Wöhler published a paper demonstrating that when a solution potassium cyanate and ammonium chloride was heated, a crystalline molecule would precipitate out of solution. He analysed the crystals and determined them to be urea, a vital molecule that, until then, was only found in urine (*1*, *2*). Soon after came a seminal synthesis reaction known today as Strecker-synthesis, after Adolf Strecker, who uncovered that a molecule known then to be essential to life, the amino acid alanine, was an intermediate during the organic synthesis of lactic acid. The production of alanine can proceed via the reaction of acetaldehyde and hydrogen cyanide in aqueous ammonia, and its subsequent reaction with metal-nitrates produces lactic acid (*3*).

In 1854, Hermann Kolbe promoted the hypothesis that organic molecules could be synthesised from inorganic sources and tested it by producing acetic acid from carbon disulphide. He praised Wöhler's experiments as the inspiration that ultimately gave birth to the field of organic chemistry (*4*). Yet another significant synthesis feat of the 19th Century that was underappreciated until the basic structure of nucleotides was resolved is the formose reaction, discovered by Aleksandr Butlerov in 1861 (*5*, *6*). The reaction uses formaldehyde as a carbon source to produce sugars in the presence of a base and a divalent metal ion. After many successful syntheses of vital molecules from inorganic sources, scientists were becoming convinced that life relied on ordinary chemistry, but that somehow these organic molecules would be used to bring about the survival and reproduction of a living entity.

1.1.2. Spontaneous generation and Darwinian evolution: friends or foes?

The year 1859 proved to be one of scientific revelations. The first breakthrough was when Louis Pasteur decisively demonstrated that only life begets life, thereby disproving the theory of spontaneous generation of complex life (*7*). His success is attributed to his ingenious experimental design of the swan-necked flask, which allows the passage of

air but not of bacteria that ultimately settle at the bottom of the neck **(Figure 1)**. With this experiment, Pasteur was able to put the nails in the coffin of spontaneous generation, convincing the world that any form of extant life that was detected must have undoubtably come from a predecessor.

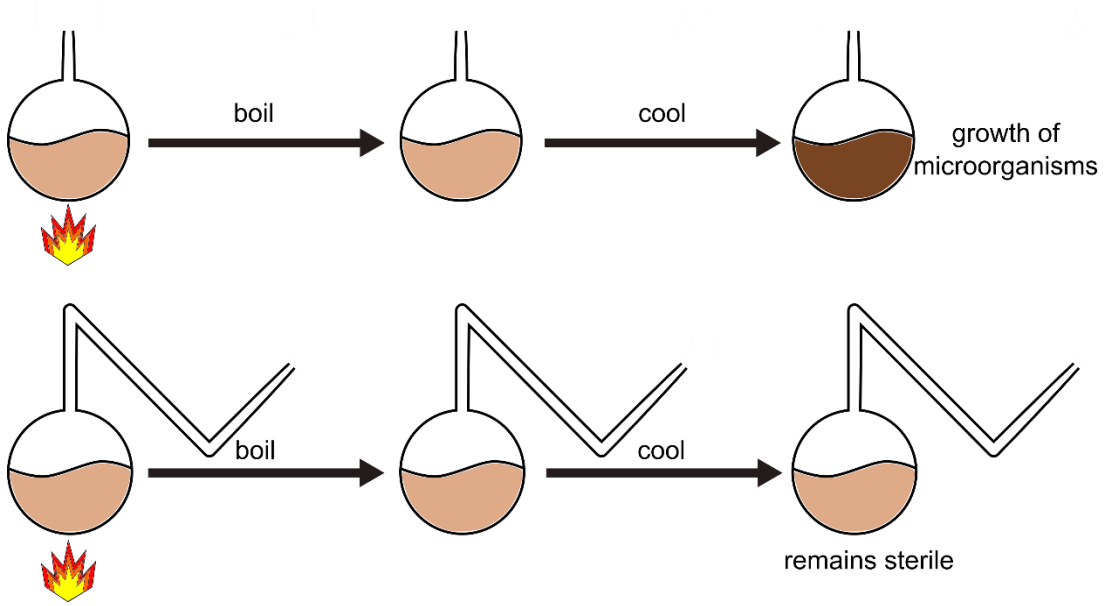


Figure 1: Pasteur's swan necked flask experiment that falsified the hypothesis of spontaneous generation. *A boiled broth in contact with air quickly exhibits signs of fermentation and hence, microorganism growth, whereas the same broth in a flask that prevents microorganisms from reaching it remains sterile.*

Whereas Pasteur was preaching about how life can only come from previous life, Charles Darwin published his book *On the Origin of Species* in which he concludes that the only hypothesis that agrees with the data regarding the geographical distribution of plants and animals, is that they are all descendants of a common ancestor which over large amounts of time diverged to give the species we see today, or in other words, descent with modification (*8*). He states that individuals of a species vary slightly according to some then unknown laws, and if those slight variations prove to be useful to the individual under the current conditions of life, they are more likely to survive and procreate, and the variation will be retained or even enriched over generations. Thus, all the different species observed in nature could have descended from a common progenitor, that gradually diverged by accumulated profitable variations, eventually differing greatly from their original parental form such as to merit classification as a new species. The ideas put forth by Charles Darwin at the time revolutionised the scientific field, particularly due to his powerful statements: *'all organic beings which have ever lived on this Earth may be descended from some primordial form'* and *'light will be thrown on the origin of man and his history'*. However,

although Darwin proposed a theory on how life was shaped by natural selection to become what it is today, he shines no light as to how life emerged in the first place. In fact, he insisted that the thought of pondering the origin of organic life was, with the current understanding of natural sciences at the time, a moot point. Nonetheless, in the very famous private letter to Joseph Dalton Hooker in 1871 (Darwin Correspondence Project, "Letter no. 7471"), he does conceive of a '*warm little pond'* that given the right conditions could give rise to protein molecules that can replicate and evolve. Perhaps coincidentally, in the same letter, he also mentions the experiments of Pasteur that were being scrutinized by Benjamin Lowne, who observed that *Penicillium* spores retained their ability to form mycelial filaments even after boiling. He adds to the observation that some spores are resistant to boiling: *'... but then how on Earth is the absence of all living things in Pasteur's experiments to be accounted for?'.* This marked the initial "turtles all the way down" scenario (infinite regress problem) in the origin of life, since Darwin was championing a primordial universal ancestor to all life that should have emerged from non-life, while Pasteur had just proved that life cannot arise from nothing. Thus, the question of how this primordial being sprang to life in the first place arises.

In the years shortly after Darwin's publication, Friedrich Rolle and Ernst Häckel, both loyal proponents of Darwin's theory of evolution, were kickstarting the idea for a material origin of life (*9*). In 1866, building on the earlier works of Wöhler, Strecker and Kolbe, Häckel stressed that there is no qualitative difference between animate and inanimate chemistries and that, consequently, there is a natural continuous flow from one to the other (*10*). In other words, since the chemistries of the living and non-living are not of a different nature, then a direct link between them can be formed. This idea has become part of the modern view of life and is known as the continuity principle, and will come to play a very central role in the hypotheses for how life emerged.

1.1.3. Establishing a scenario: reducing early Earth atmosphere

The first attempts to provide a detailed chemical process that directly leads to the emergence of life from inorganic molecules were undertaken in the mid-1920s by Aleksander Oparin (1926 edition in Russian, and later in English in 1936) and John B.S. Haldane (1929). They independently attempted to resolve the paradox of how the first living entities emerged in the absence of any ancestors, fully convinced of the non-existence of spontaneous generation (*10–14*). The Oparin-Haldane hypothesis postulates that, prior to the emergence of the first living being, there was a period of prebiotic chemical evolution

during which abiotic processes produced organic compounds that gradually accumulated and increased in complexity, eventually giving rise to the first life-like system. They essentially propose decoupling the synthesis of organic compounds from the growth and replication of a protocell, such that the primitive cellular metabolism was carried out by processes that occur in the environment, analogous to an *ex vivo* metabolism. While initial ideas of the first living cell revolved around an autotrophic cell capable of producing its own food, Oparin suggested that the first cells were more likely to have been heterotrophs that acquired food from the environment. This conclusion was based on his observations that extant autotrophic cells are structurally and biochemically more complex that heterotrophic cells, then according to evolutionary theory, the latter must have preceded the former. Another important idea proposed was the presence of a reducing atmosphere on early Earth, a necessary condition to drive the synthesis of organic molecules by the hydrogenation of carbon dioxide. Based on the then current approximations of chemical abundances of elements in the universe, models of planetary formation and atmospheric evolution, Oparin suggested that the early Earth atmosphere was composed of methane, ammonia, hydrogen gas and water vapor. To tie the story together from the abiotic synthesis of organic compounds to the emergence of the first cells, Oparin suggested that they initially were colloidal droplets that spontaneously coalesced when components crossed a certain concentration threshold. These droplets would then act as basic evolutionary units subject to the laws of variation and selection. Oparin hypothesised that such primordial cells comprised metabolic networks that could support their growth and division. In this manner, the synthesis of cellular components and growth and replication become coupled, and networks that have faster synthesis rates or slower decay rates are enriched over time. As such, it represents a 'metabolism-first' approach to the origin of life, in which metabolic networks that produce coacervate components are subsequently compartmentalised in the coacervates which themselves undergo cycles of growth, division, variation and selection. The end of this metabolic phase of life would be marked by the emergence and evolution of the canonical genetic apparatus.

1.1.4. Watson-Crick and Miller-Urey

In 1953, two seminal articles were published in the journals *Nature* and *Science*. Firstly, Francis Crick and James Watson had proposed a structure for DNA (*15*), basing their design on DNA X-ray diffraction analyses. We now know in great detail that DNA is

formed from two helical, antiparallel strands bound together via hydrogen bonding of nitrogenous bases facing inwards, which are connected to a backbone of deoxyribose sugars linked by phosphate groups facing outward, making them easily accessible to cations. A base residue is present on each strand every 3.4 Angstroms, and if a pyrimidine base is on the first strand, then the opposite strand must have a purine base. This hydrogen bonding scheme (Adenine to Thymine and Guanine to Cytosine) was proposed based on experimental data produced by Erwin Chargaff, who meticulously demonstrated that the molar ratio of purines to pyrimidines in DNA of various species tended towards one (*16*). The structure of DNA had clear implications for how it would be copied, since base complementarity meant that one strand already had the information to produce the other. Solving the structure of DNA was undoubtably a great accomplishment, but that was not all that had occurred that year. On the other side of the story is the experiment that transformed origin of life from a philosophical field of research to a scientific one that can produce empirically testable hypotheses. Stanley Miller had abiotically synthesised organic molecules, of which some were amino acids, from an anoxic, reducing atmosphere composed of methane, ammonia, hydrogen gas and water vapor **(Figure 2)** (*17*, *18*).

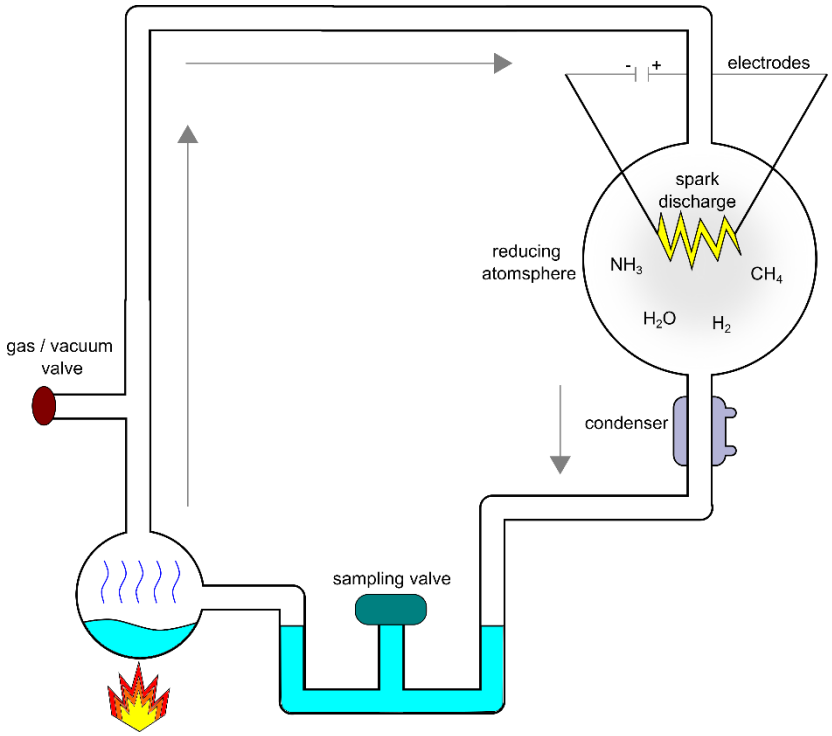


Figure 2: The experimental setup for the Miller-Urey experiment. *A reduced atmosphere is generated in a closed system by boiling the liquid phase constituents after generating a vacuum. Lighting strikes are mimicked by spark discharges within the reducing atmosphere. The products that are synthesised can condense into the lower portion of the apparatus and be sampled.*

This composition was initially proposed by Oparin as discussed above, and further suggested by his then doctoral supervisor Harold Urey to be the atmospheric composition of early abiotic Earth (*19*). The experimental setup consisted of an apparatus that circulates the reducing atmospheric gasses past an electrical discharge that is supposed to mimic lightning, a plausible source of free energy from the environment. The resulting material can condense into a heated 'primordial sea' mimic. After a few days of spark discharges, the solution turned brown and sludgy and liquid chromatographic analysis of the products revealed the production of the amino acid glycine in substantial amounts, as well as other amino acids such as alanine and isoleucine in lesser amounts along with other oily organic compounds. This experiment, famously known as the Miller-Urey experiment, placed in the perspectives of Darwinian evolution and Oparin's proposed sequence of events for abiogenesis, almost instantly transformed the origin of life into a testable scientific hypothesis and paved the way for the nascent field of prebiotic chemistry.

1.1.5. The advance of prebiotic chemistry

Although Miller's paper was published a few weeks after Watson and Crick had published the structure for the DNA double helix, there existed no link between the two emerging fields of molecular biology and prebiotic chemistry. That is, until 1960-1961, when Juan Oró demonstrated that, under a similar reducing atmosphere as that employed by Miller, a DNA nucleobase could be synthesised inorganically from hydrogen cyanide (*20*, *21*). Specifically, he was able to synthesise significant amounts of Adenine in an aqueous solution of ammonium cyanide, via the sequential cyclisation of five hydrogen cyanide molecules. At more or less the same time, Sidney Fox and Kaoru Harada published their work that demonstrated the synthesis of the nucleobase Uracil, starting from malic acid and urea in the presence of polyphosphoric acid (*22*). Fox and Harada argued for the prebiotic plausibility of the reaction, since urea can be formed by heating carbon dioxide and ammonia, while malic acid could be produced from acetic acid, which in turn can be condensed from hydrogen, methane and water. Furthermore, others had already devised reaction schemes for the non-enzymatic polymerisation of nucleotides by subjecting them to non-aqueous solvents and strong dehydrating reagents, such as carbodiimides (*23*, *24*). Surprisingly, Schramm *et al.* (*24*) were able to synthesise nucleic acid homopolymers with molecular weights ranging from 15,000 – 50,000 Daltons (approximately 50 – 150 nucleotides in length). These findings, along with the work of Meselson and Stahl that uncovered the semi-conservative nature of DNA replication (*25*) and the derivation of the central dogma

of biology by Crick (*26*), reshaped the ideas on the origin of life, transforming the field from a 'protein-centred, metabolism first' approach as heralded by Oparin and Haldane years before, to a 'nucleic acid-centred' approach supported by molecular biologists and geneticists, which promotes a 'replication first' scenario for the origin of life.

1.1.6. Evolution in a test tube

The 1960s were an exciting time not only for biochemists and biologists, but also for origin of life research, following the publication of a series of articles from the laboratory of Sol Spiegelman. When studying the life cycles of RNA viruses, specifically the MS2 (now *Emesvirus zinderi*) and Qβ bacteriophages, both of which infect *Escherichia coli*, the overarching question revolved around the ability of a viral RNA genome to replicate and reproduce infectious viral units without going through a DNA intermediate (*27*). It was already established that some oncogenic viruses went through a DNA intermediate stage for genome replication, however no cDNA copies of the viral genome could ever be detected during bacterial infections by the MS2 or Qβ bacteriophages. The hypothesis that was formulated based on these observations was that an RNA-dependent RNA polymerase (i.e., RNA replicase, or simply replicase) was replicating the RNA genome directly from RNA. Spiegelman was later able to isolate both the RNA genome and the purified enzyme, demonstrating that a viral-specific replicase could recognise and replicate its cognate RNA genome extracellularly (i.e., cell-free *in vitro* RNA replication), but not unrelated RNA genomes (*27*, *28*). The newly acquired capability of replicating viral RNA genomes *in vitro* prompted Spiegelman to set up a model evolutionary experiment in his lab. By combining RNA replicase, viral RNA genome, free ribonucleotide triphosphates (rNTPS) and some salts, the replicase was able to synthesise *de novo* viral RNA genomes. After some time, he transferred the RNA to a fresh solution containing all the requirements to make more RNA. After 74 generations of serial transfers that were time adjusted to select for the fastest replicators, the initial RNA genome had lost 83% of its original molecular weight (estimated around 170 kDa) essentially retaining only the replicase recognition sites required for its replication (*29*, *30*). Furthermore, the variant RNA showed a 2.6-fold increase in replication rate and had a 4-fold decrease in lag time compared to the original RNA. These experiments were the first attempts at an *in vitro* directed evolution experiment and reveal the consequences of a selection pressure for speed on the self-replicating system.

1.1.7. Chicken or Egg?

A well-known problem that emerged in the origin of life field was a paradox analogous to the classical chicken and egg dilemma: since DNA is needed to synthesise proteins, and proteins are needed to replicate DNA, which biomolecule necessarily preceded the other? The first notions of a solution came from Alexander Rich, who proposed polynucleotides as the origin of living systems (*31*). He describes a primitive living system as one possessing an autocatalytic replicating information-bearing polymer, utilising building blocks provided by the environment to produce replicas of itself and subject to modifications influenced by the environment. Rich argues against a proteinaceous origin of life due to the apparent coupling between protein synthesis (translation) and the information content of nucleic acids, which he believes cannot be accomplished by the random polymerisation of amino acids. Rich proposes that the first molecule was a nucleic acid polymer such as RNA or an RNA-like molecule, since it is able to carry information and as well as organise amino acids into a specific sequence. He continues by postulating that a primordial RNA strand had the ability to replicate itself via complementarity much the same way DNA is replicated in extant cells. A subsequent transition to DNA-based information storage may have been spurred by the reduced reactivity and therefore increased stability of DNA, likely resulting from the loss of the 2'-hydroxyl group on the sugar component, thus providing a considerable selective advantage for early cells, by having one nucleic acid specialised for protein synthesis (RNA) and another less metabolically active variant specialised in self-replication (DNA).

The ideas put forth by Alex Rich were quickly adopted in the late 1960s by renowned scientists Carl Woese, Francis Crick and Lesley Orgel (*32–35*). Together, they proposed that early life forms were based on RNA genes and RNA enzymes rather than DNA genes and protein enzymes, as is the case for life as we know it. Resting their ideas on the single-stranded nature of RNA and its ability to adopt tertiary structures (such as the clover-leaf structure of tRNA), they hypothesised that RNA polymers could have acted as carriers (genotype) of genetic information as well as catalysts (phenotype) for maintenance and replication. In this frame of reference, the paradox is rendered obsolete as no requisitely coupled pair of biomolecules essentially predates the other. Thus, the emergence of life from the perspective of 'information / replication first' now rested on the prebiotic synthesis of nitrogenous bases and sugars to be used as nucleotide building blocks, their assembly into activated nucleotide building blocks, their random or non-random polymerisation and finally the emergence of function, most notably self-replication **(Figure 3)**.

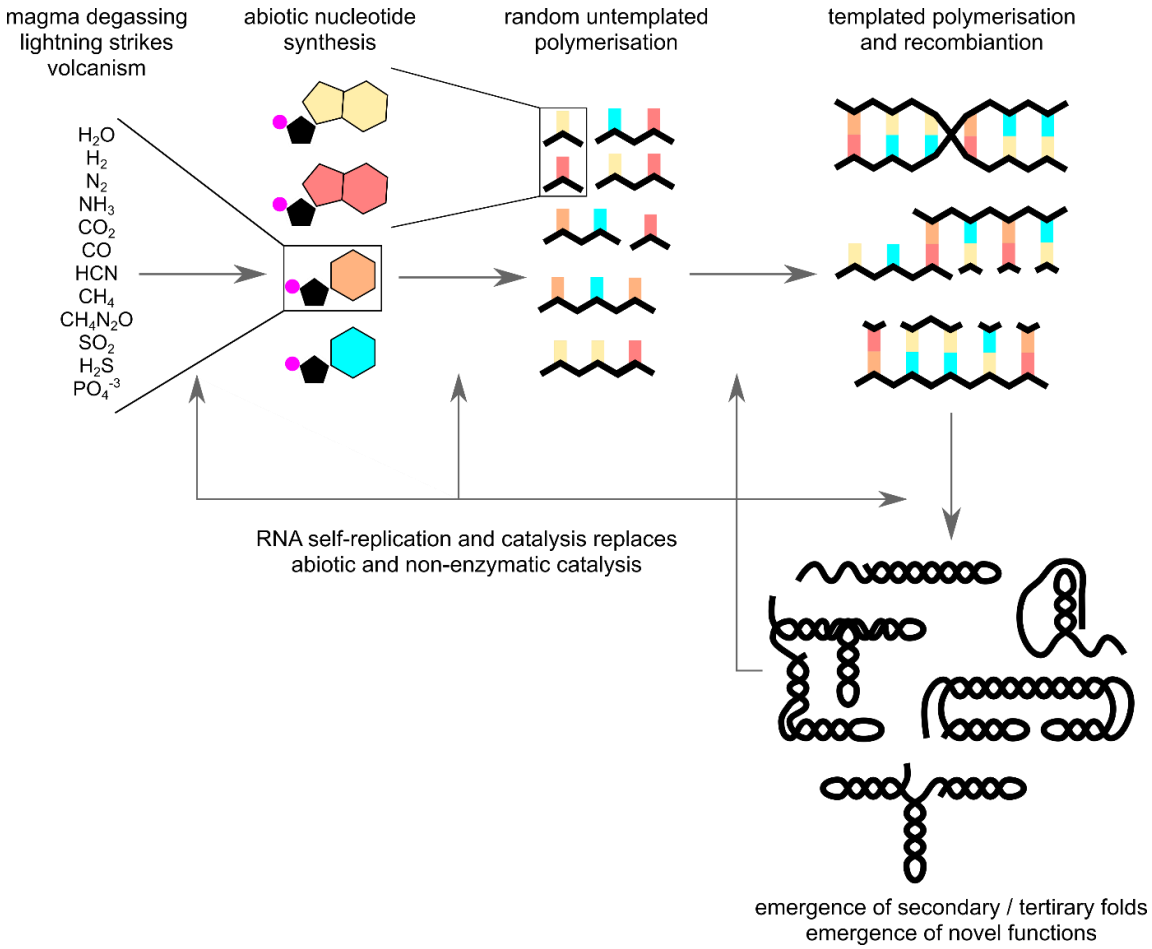


Figure 3: Schematic depicting the steps required for the emergence of an RNA world. *Geological and meteorological conditions on early Earth provide the basic elements and simple molecules required to form nucleotide building blocks under favourable conditions. The building blocks initially polymerise randomly, potentially aided by environmental factors such as wet-dry cycling or mineral surfaces, then transition to template dependent polymerisation and recombination thereby increasing polymer length and pool complexity, until the emergence of function from single-stranded RNAs that adopt tertiary folds enabling them to catalyse reactions. Compartmentalisation occurring at any stage links RNA genotype to ribozyme phenotype and, in turn, protocell phenotype allowing selection and evolution of protocells under early Earth environmental conditions.*

1.2. The 'RNA World' hypothesis

Nowadays, the RNA world hypothesis has evolved to encompass more diverse functions as catalysts for RNAs than simply self-replication. However, a plausible scenario that provides substantial amounts of RNA precursors from abiotic processes still remains a challenge, since the general structure of nucleotides comprises three distinct components that require synthesis under differing conditions: the nitrogenous base, the pentose sugar and the phosphate linker. For this reason, others who still advocate a 'replication-first' origin of life but view the synthesis of enantiopure activated RNA nucleotides virtually impossible have opted for the emergence of proto-RNA polymers that over time chemically

evolved to become RNA (*36*, *37*). In fact, recent evidence has uncovered that functional RNAs could be assembled with mixed phospho-ester and boronate ester linkages (*38*). As Darwin already recognized over 150 years ago, any organism, even a network of functional RNA catalysts, would advance by one general law, namely, *'multiply, vary, let the strongest live and weakest die.'* Compartmentalisation could have taken place at any stage of the above processes, a necessary step to protect and concentrate reactants while coupling the genotype and phenotype, thereby generating the first heterotrophic replicative cell capable of undergoing Darwinian evolution. Note that such a cell is still far from what is thought to be the Last Universal Common Ancestor (LUCA) to Bacteria and Archaea, as it still lacks the evolutionary coupling RNA sequence to amino acid identity, i.e., emergence of coded protein synthesis, as well as the transition from an RNA-based genetics and metabolism to a protein / DNA dominated architecture, with specification of DNA for information storage and proteins for metabolism.

1.2.1. Prebiotic paths to RNA monomers

Many have shown how the different components of nucleotides can be synthesised non-enzymatically from simple precursor molecules thought to have been abundant, or at least present, on the early Earth, such as formaldehyde, hydrogen cyanide, ammonia and derivatives thereof, thereby providing an abiotic source of carbon, nitrogen and oxygen. The formose reaction can produce a variety of sugars from the polymerisation of formaldehyde in the presence of glycolaldehyde, including ribose (*5*, *39*). Purines, on the other hand, can be synthesised from aqueous hydrogen cyanide via a hydrogen cyanide tetramer intermediate (diaminomaleonitril; DAMN), pioneered by Orgel and Ferris (*40*). The hydrogen cyanide tetramer can be cyclised to become aminoimidazolecarbonitrile (AICN) and subsequently react with a fifth hydrogen cyanide to yield adenine; its hydration product on the other hand, aminoimidazolecarboxamide (AICA), can react with cyanogen to produce guanine (**Figure 4a**). Similarly, the hydrolysis of cyanoacetylene to cyanoacetaldehyde and its subsequent condensation with urea produces cytosine, which can undergo hydrolytic deamination producing uracil (**Figure 4b**) (*41*).

By the end of the 20th Century, although scientists were able to independently synthesise the components of nucleotides, the question as to how they could be combined in the correct form to obtain activated nucleosides remained unresolved, particularly because the addition of nucleobases to ribose was inefficient for purines (*42*) and next to impossible for pyrimidines(*39*). Moreover, these reactions, while producing the canonical

nucleobases and sugars, produce a wide range of non-canonical bases and sugars as well (*43*, *44*). Ribose can also adopt several conformations (furanose and pyranose) in aqueous solvents (*45*), further thwarting the synthesis of canonical nucleosides and their activated forms.

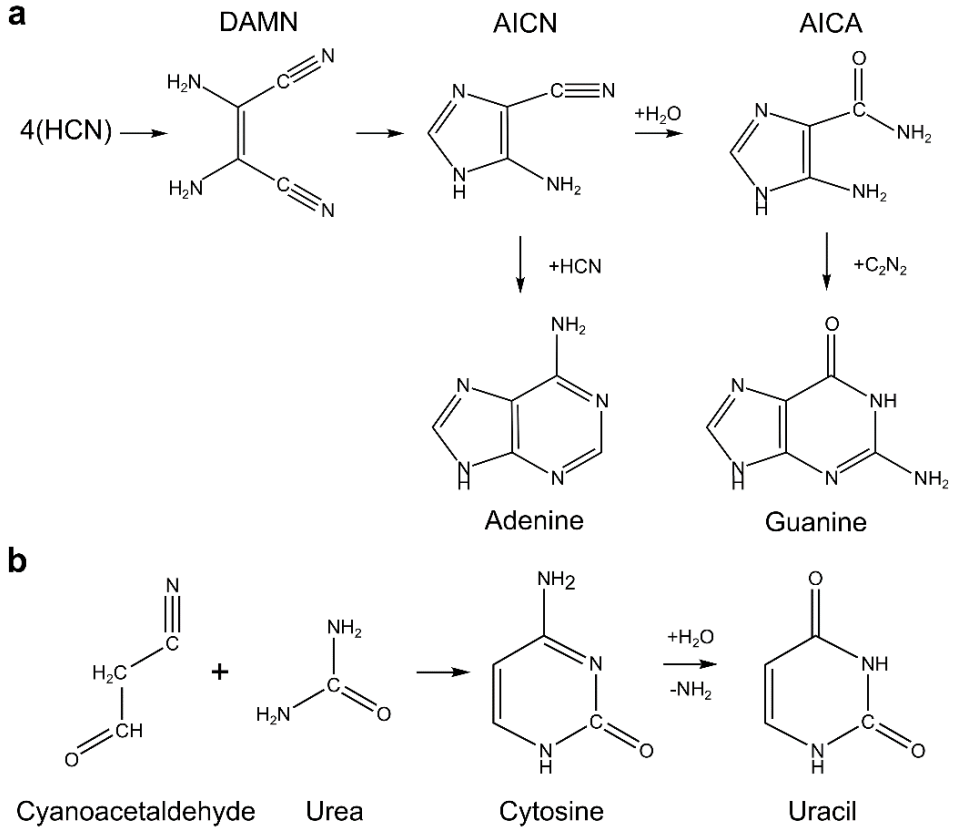


Figure 4: Synthesis pathways from hydrogen cyanide and its derivatives*. Synthesis of* ***a)*** *purines and* ***b)*** *pyrimidines. The work was pioneered by Ferris and Orgel* (*40*)*. Shown here are only the canonical base products, however, many other non-canonical bases are also synthesised, such as hypoxanthine and diaminopurine.*

A breakthrough was achieved in 2009, spearheaded by Powner, Gerland and Sutherland, by adopting a different approach (*46*). Rather than assuming that ribonucleotides were produced by coupling independently synthesised nitrogenous base and sugar, they achieved the synthesis of cytosine-2',3'-cyclic phosphate staring from a prebiotically plausible feedstock of cyanamide, cyanoacetylene, glycolaldehyde, glyceraldehyde and inorganic phosphate via arabinose amino-oxazoline and anhydronucleoside intermediates **(Figure 5)**. The authors note that the photoanomerization from alpha- to beta-ribose is rate-limiting, but demonstrate in a follow-up article that this inefficient step can be greatly hastened by using hydrogen sulphide as a reducing agent, thus going through thiocytidine intermediate (*47*).

Interestingly, the group also show that precursors to RNA, proteins and lipids can be generated from the hydrogen sulphide mediated reductive anabolism of hydrogen cyanide and its derivatives, hypothesising an early metabolism based in cyano-sulphide world and supporting the idea of the chemical co-evolution of biomolecules on early Earth (*48*). While the synthesis of 2',3'-cyclic phosphate activated pyrimidine nucleotides was a great milestone for prebiotic chemistry, it was driven by the stepwise reaction and purification of compounds, necessitating a rather elaborate scenario on the early Earth such that the exact mechanism of synthesis be achieved. Still missing was the ability to generate activated nucleotides in a one-pot synthesis that excludes repeated exogenous intervention by the experimenter.

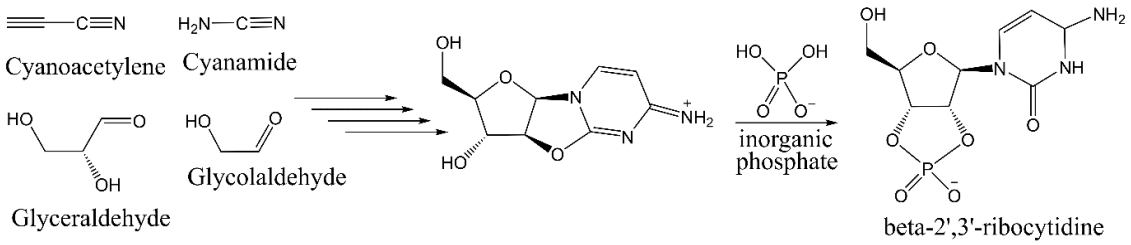


Figure 5: *Multi-step synthesis of activated pyrimidines developed by Powner and Sutherland* (*46*) *bypasses the requisite sugar and base coupling by going through an anhydrous nucleoside intermediate step. The resulting product is a 2',3'-cyclic phosphate cytosine, which, as we have seen before, can deaminate to yield uracil.*

As for purine synthesis, the Carell group developed a high-yielding regioselective purine nucleoside synthesis pathway that is different from that developed by Orgel years before through the hydrogen cyanide tetramer. The synthesis route they propose utilises prebiotically available organic compounds, where pyrimidine synthesis proceeds via an isoxazolylurea intermediate whereas purine synthesis goes through formamidopyrimidine (FaPy) intermediate, based on the earlier work from Albert Eschenmoser (*49*), both of which are regioselectively coupled to linear ribose and cyclised to form the respective nucleoside **(Figure 6)** (*50*). Interestingly, FaPy derivatives occur in DNA when oxidative stress and free radicals react with purine bases, suggesting that their proposed pathway is not very far from extant biochemistry (*51*). The group later investigate a plausible route for the one-pot production of pyrimidine bases from a cyanoacetylene derivative in a continuous manner in water as the solvent, although the matter of coupling the base to a ribose sugar was not treated (*52*). The challenge remained unifying the conditions of purine and

pyrimidine nucleoside synthesis into a common synthetic pathway. In a seminal paper published in 2019, a different synthetic pathway for the production of pyrimidines was ideated, one that necessitates conditions similar to those required for purine synthesis. By applying a similar strategy to the purine synthesis, they were able to produce canonical mono- and diphosphorylated RNA nucleosides in a one-pot prebiotic reaction (*53*).

In summary, we have seen how the building blocks of RNA can be synthesised from simple precursors by geochemical means rather than by living cells, starting from reactants that are thought to have been readily available on the early Earth. To ask which reaction pathway is 'correct' is a moot point, as there is no way (yet?) of accurately predicting the conditions under which life arose. Nonetheless, it is clear that the most likely path to the origin of life can only be teased out by a collaboration between Earth scientists and prebiotic chemists such that the parameters that prevailed on early Earth are constrained and used to guide the synthesis pathways of increasingly complex organic molecules.

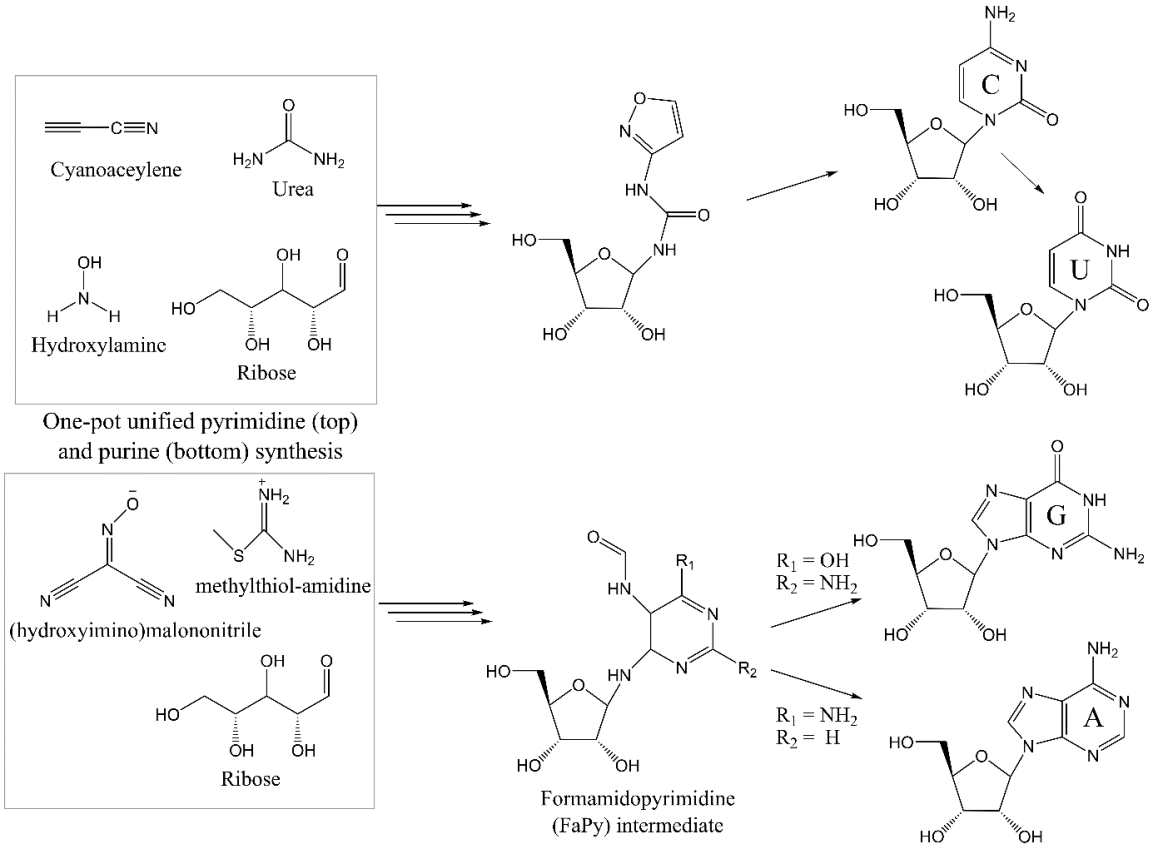


Figure 6: One-pot synthesis of canonical RNA nucleosides. *Methods were developed by the Carell group* (*52*). *Pyrimidine synthesis proceeds through via an isoxazolylurea intermediate while purine formation proceeds via a formamidopyrimidine intermediate. The resulting nucleosides were also shown to undergo autophosphorylation reactions leading to the formation of both mono- and diphosphate nucleosides, potentially capable of undergoing non-enzymatic random polymerisation forming an initial pool of short length RNA polymers.*

1.2.2. Non-enzymatic polymerisation

The next phase for a plausible path to an RNA world would be the polymerisation of monomers into oligonucleotides, given a sustainable influx of activated monomers. This facet has been treated both experimentally and theoretically, the results of which are quite convincing for the likelihood of non-enzymatic polymerisation to occur. First, there has to be a distinction between non-templated and templated polymerisation, since the former requisitely preceded the latter, otherwise no template would be available on which to polymerise. The problem with polymerisation, irrespective of whether starting from amino acids or nucleotides, is that it is a reversible dehydration reaction that is thermodynamically uphill, especially in aqueous solvents. Nonetheless, some strategies previously alluded to emerged that shift the equilibrium toward oligomerisation and promote nucleotide condensation, for example by either modifying the leaving group or subjecting the monomer solution to dehydrating conditions, to name a few.

In fact, early on studies have shown that poly-adenine oligonucleotides up to ~150 residues can form under conditions that were deemed prebiotically plausible (*24*). Moreover, a solution of imidazole-phosphate (Imp) activated adenine nucleotides can polymerise to form short poly-adenine oligonucleotides. Although this happens quite slowly in solution, it can be accelerated by the addition of divalent metal ions such as lead (*54*). Surprisingly, a montmorillonite clay, which readily forms from rehydrated volcanic ash and therefore should have been common on the early Earth, was shown to adsorb Imp-activated RNA nucleotides and promote their assembly into long chains of up to 50 nucleotides (*55–57*). More recently, prebiotic rock glasses that form from the rapid cooling of minerals after impacts or volcanic activity, allowed the formation of poly-ribonucleotides with an average length ranging from 100 – 300 nucleotides starting from a mixture of triphosphorylated monomers (*58*). These oligomers had a substantial proportion of 3’,5’-linkages along with 2’,5’-linkages and were found to be stable for months under the conditions tested.

Dehydration has also proved to be effective at promoting the non-templated polymerisation of RNA. Eutectic ice phases that mimic dehydrating conditions by selectively crystallising water molecules in ice were shown to be permissive to the condensation of mixed Imp-activated ribonucleotides up to 17-mers, with 6- to 9-mer products being the most abundant (*59*). Similarly, wet-dry cycling at elevated temperatures (approximately 80 °C during the hot cycle) produced long RNA oligomers as well as circular RNA products that were observed under an atomic force microscope, although no investigation into

linkage stereochemistry was performed (*60*). The major obstacle that arises from the non-enzymatic, non-templated polymerisation reactions is that the majority of linking phosphodiester bonds are found to be non-canonical 2',5'-linkages rather than 3',5'-linkages. Despite the mixed linkage products, solutions of activated RNA nucleotides have the potential to provide short oligomers that act as templates or primers for the continued elongation of the oligomer pool. Notably, computer simulations of diverse sequence oligomer templates have shown that the continued synthesis long RNA is promoted in populations of random sequences, as opposed to homotypic templates that prevent subsequent copying cycles due to the inability of complementary base pairing (*61*). These simulations suggest that sequence diversity is essential to obtain longer polymers over time and that solutions of single nucleotides are unlikely to produce longer strands which in turn increase pool complexity.

Templated polymerisation and ligation have been more extensively investigated than non-templated polymerisation. An intriguing phase state that was found to promote templated strand ligation is that of liquid crystals (*62*). Liquid crystal phases composed of complementary RNA oligomers form spontaneously at high concentrations of RNA at room temperature (≥250 mg/mL), due to the ordered arrangement of double stranded nucleic acids upon base stacking at high concentrations. These liquid crystal states were shown to allow up to 8 ligations of blunt-end 12-mer RNA duplexes as well as ligation in duplexes with overhangs.

Under less exotic conditions, aminoimidazole-activated ribonucleotides were amenable substrates for primer extension reactions (*63*). The system was also shown to support templated synthesis through double-stranded regions by introducing invading oligomers that help in separating partially double stranded regions via a strand invasion mechanism. Primer extension reactions have also been carried out concomitantly with non-enzymatic monomer activation in buffer conditions supposedly compatible with prebiotic synthesis (*64*). The study demonstrates that nucleotide monophosphates can be activated with an aminoimidazole leaving group by isocyanide-based activation chemistry and subsequently be incorporated in a growing primer via intermediate 5',5'-aminoimidazolium-bridged dinucleotides (NpAIpN). Nucleotide activation and primer extension was also found to occur under freeze-thaw cycling conditions with improved rates of activation and nucleotide incorporation, probably resulting from the increase in local concentration of substrates and reactants that takes place upon eutectic ice formation (*65*). Primer extension proceeds at the 3'-end via the nucleophilic attack of the deprotonated hydroxyl group on

the ribose, however, the slow catalytic rate warranted investigations for a potentially faster reaction cycle. To this end, the authors carried out the same primer extension reactions but with a terminal 3'-amino,2'-deoxynucleotide, reasoning that the 3'-amino moiety is a better nucleophile than a hydroxyl group (*66*). The change in chemistry led to a 100-fold increase in primer extension kinetics, suggesting that, similarly to nucleotide binding, nucleophilic attack can also be a limiting factor. They also show that by setting up a repetitive splinted ligation using this chemistry, a pool of 4-mer substrates was extended by up to 200 nucleotides, well above the length threshold for nucleic acid folding and catalysis (rough estimations are 30-60 nucleotides). Following up on this work, they achieved the splinted ligation of a minimal ligase ribozyme under frozen conditions and coupled the synthesis of the ribozyme with its cognate substrate ligation upon heating (*67*). So far, this represents one of the only successful attempts at coupling non-enzymatic and enzymatic ligation, bypassing problems of catalytic rate and strand-separation by exploiting increased nucleophilicity of the invading functional group and by introducing non-canonical base pairing (G•U) in the short splints employed at the ligation junctions to decrease the resulting duplex's free energy, thus facilitating strand separation at temperatures not too high such that RNA is degraded rather than denatured.

One tantalising piece of evidence for the emergence of an RNA world at some point during early evolution is the observation that primer extension on chimeric templates, starting from a mixture of canonical and non-canonical nucleotides, predominantly yields RNA products, suggesting that, after multiple rounds of non-enzymatic templated replication, RNA products will outcompete other non-canonical nucleic acids, eventually accumulating RNA oligonucleotides and depleting non-canonical nucleic acid arrangements (*68*). This idea also ties with the previous section regarding the abiotic synthesis of nucleotides, since it seems rather tolerant on the production of side-products which would otherwise inhibit polymerisation reactions.

1.2.3. Ribozyme mediated RNA replication

Initial suspicion and subsequent proof that RNA held the potential for catalysis was significant, but the development of in vitro selection and evolution of truly pushed the limit to what could be achieved with nucleic acids. Systematic Evolution of Ligands by Exponential Enrichment (SELEX) and *in vitro* selection were developed in 1990 and de-

scribe a process in which through multiple rounds of randomisation, selection and amplification, a class of structural and / or functional sequences can be obtained from a quasi-random starting pool **(Figure 7)** (*69*, *70*).

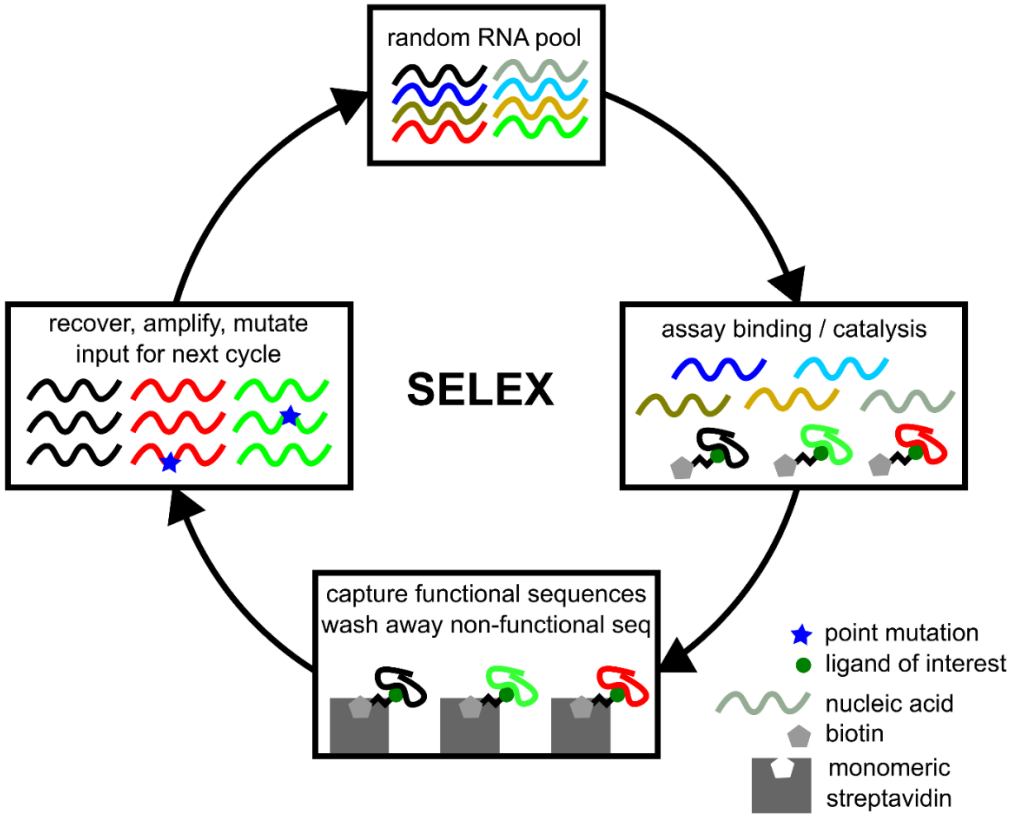


Figure 7: Cartoon describing the general methodology for the systematic evolution of ligands by exponential enrichment (SELEX). *The process begins with a random-sequence pool of oligomers. The pool is allowed to bind to a molecule of interest or catalyse a reaction of interest. Next, the bound oligomers (or the reacted oligomers) are recovered from the random pool while the non-binders are washed away. The selection step shown in the schematic relies on the high-affinity binding between streptavidin and biotin; the ligand of interest can be functionalised with a biotin via a short carbon chain linker, for example, whereas magnetic beads can be functionalised with streptavidin, such that they recover only those oligomers that successfully bind the ligand of interest. The final step before initiating the next cycle is to amplify the recovered sequences and optionally introduce random mutations to increase the pool diversity.*

Ever since, a wide range of functional classes of RNA have been described. For example, aptamers are RNA sequences able to bind a ligand molecule, while ribozymes are sequences with catalytic activity. Aptazymes are a combination of aptamer and ribozyme, which functions only in the presence of the aptamer ligand. Light-up aptamers are able to bind a specific ligand molecule and greatly increase its fluorescence output; biosensors are a combination of a light-up aptamer and a small molecule aptamer, such that a fluorescence output is produced upon small molecule detection. Even within the class of ribozymes, a plethora of catalytic functions and secondary structures have been discovered **(Figure 8).**

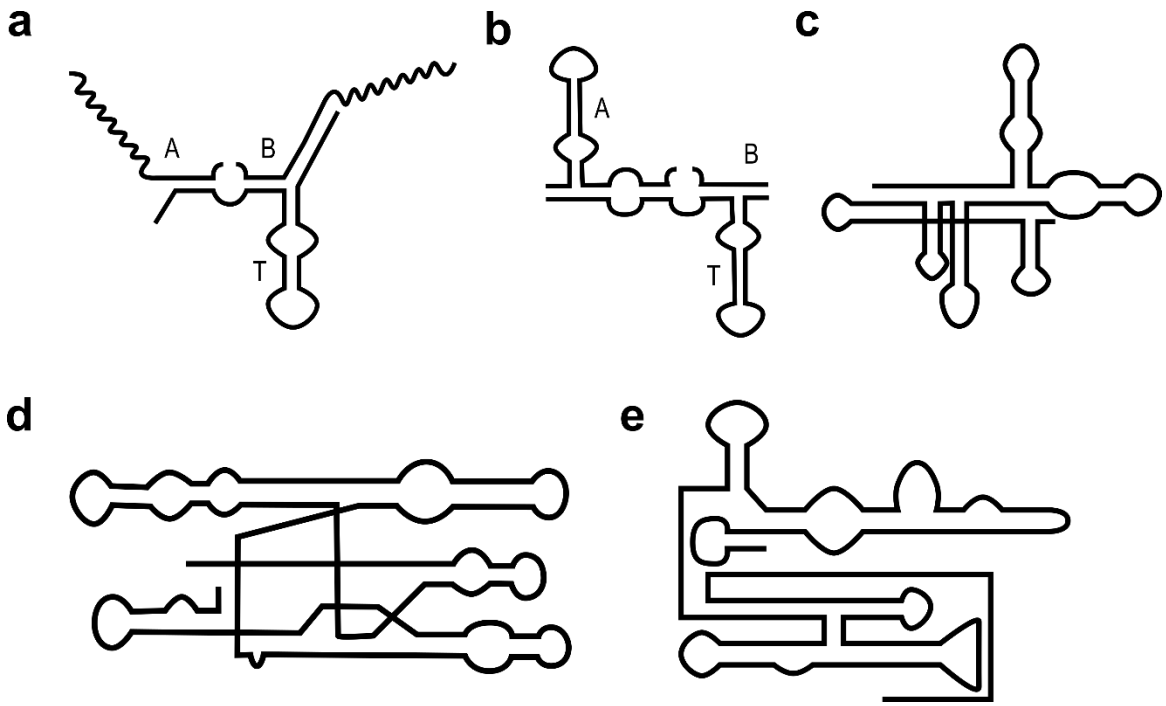


Figure 8: Illustration showing secondary structures of various ribozymes used as model systems in origin of life studies. a) *Hairpin-derived self-replicating ribozyme (T) that accepts two substrates (A and B) to produce another T molecule* (*71*). *The reaction chemistry relies on 2',3'-cyclic phosphates.* ***b)*** *R3C-derived self-replicating ribozyme (T)* (*72*). *Different from the hairpin-derived ribozyme, the reaction chemistry here necessitates 5'-triphosphates.* ***c)*** *Group I intron-derived Azoarcus ribozyme* (*73*). *It uses an internal guide sequence to direct ligation of two RNA strands.* ***d)*** *Group I intron-derived sunY ribozyme* (*74*). *It performs templated RNA ligation reactions.* ***e)*** *RNA polymerase ribozyme (tC19Z) consisting of the class I ligase core and an accessory domain* (*75*). *It performs templated RNA polymerisation from triphosphate activated nucleotides.*

Attempts at the replication of RNA by catalytic RNAs have focussed, on the one hand, on re-engineering bacterial group I introns that are capable of undergoing two trans-esterification reactions resulting in their excision from the messenger RNA strand and ligation of neighbouring exons (*76*). Through structural remodelling, group I intron-derived ribozymes can act as template-dependent RNA ligases by promoting the nucleophilic attack of a 3'-hydroxyl to the phosphodiester bond between a free 5' guanosine and any other templated nucleotide (*74*). The product is a stable double-stranded RNA helix, which is often very difficult to separate and thus non-productive for continuous cycles of replication (*77*, *78*). Duplex inhibition is exemplified in the attempts at self-replication of a group 1 intron-derived ribozyme called *sunY*, which was incapable of ligating together another full copy of itself when provided with short oligomers and templates required for its own replication, in part due to substrate inhibition and strand separation problem (*74*, *79*, *80*). An-

other example is the *Azoarcus* ribozyme, which was exploited for its ability to utilise internal guide sequences (IGS) to align two exons and catalyse their ligation *in trans* (*81*). The ribozyme was subsequently fragmented into 4 (and later 5) shorter oligomers and re-engineered such that these fragments assemble noncovalently and catalyse their own ligation in an autocatalytic manner (*73*, *82*, *83*). Another naturally occurring ribozyme that was re-engineered as a tool for RNA recombination and other functions is the *hairpin* ribozyme (Hpz), which can catalyse the reversible cleavage / ligation of 2',3'-cyclic phosphate and 5'-hydroxyl substrates (*84*). Interestingly, the Hpz was utilised to assemble a longer polymerase ribozyme in eutectic ice phases as well as coacervate phases (*85*, *86*). Moreover, Hpz variants were generated to perform splicing reactions that can perform a congruent exon ligation step of splicing based on sequence design (*87*, *88*). For these reasons, re-engineered group 1 introns and hairpin ribozyme variants remain a powerful tool for sequence dependent RNA assembly and potentially recombination.

On the other hand, there are synthetic ligase ribozyme systems that have been selected for their ability to ligate using triphosphate chemistry in hopes of attaining faster ligation rates, since they are high energy bonds and mimic the activation chemistry in extant biology. Indeed, several different ribozyme ligases were selected and evolved for faster rates and their potential for autocatalysis. The L1 ligase, characterised by the Robertson and Ellington, was evolved to catalyse its ligation reaction with a maximal rate of 1 min^{-1} with a length of only 130 nucleotides (*89*, *90*). Minimal versions (~60 nucleotides) have also been engineered based on the parent structure by shortening stems, but they also suffer from decreased activity, as is usually the case with most minimal variant of ribozymes. The Joyce laboratory were able to select for a self-replicating ligase based on the R3C ligase ribozyme, which can ligate two substrates in an autocatalytic manner with a catalytic rate of 0.0011 min^{-1} after the first selection experiment (*72*). Further rounds of sequence randomisation and selection led to an enhanced ribozyme capable of autocatalytic exponential amplification with a rate 0.14 min^{-1}, corresponding to a doubling time of only 5 minutes and a ~127-fold increase in catalytic rate compared to the parent ribozyme (*91*). The Inoue laboratory selected a ribozyme (DSL ligase) that can ligate a substrate *in cis* (i.e., to itself) with a rate constant of 0.12 min^{-1} under optimal conditions (*92*, *93*). The most active of ligases is the class I ligase selected by Bartel and Szostak, with some versions of ribozyme possessing a catalytic rate of 100 min^{-1} under physiological pH and even reaching up to 360 min^{-1} with multiple-turnover for *trans*-acting versions under optimal conditions (pH 9),

surpassing most natural ribozymes and approaching the rates of proteinaceous enzymes (*94–96*). As a reference, one of the most active protein enzymes known is carbonic anhydrase, limited *in vivo* by the diffusion rate of its substrates, with an estimated rate constant *in vitro* of 660 ms^{-1}, or approximately 396×10^5 min^{-1} (*97*, *98*).

In 2001, Bartel and colleagues appended a random region to the 3'-end of the class I ligase they had evolved and selected for the ability to incorporate nucleotide triphosphates to a primer, leading to the discovery of the first polymerase ribozyme, consisting of the class I ligase core and an accessory domain that is thought to endow the ribozyme with single nucleotide extension capabilities (*99*). Today, three different cutting-edge RNA polymerase ribozymes have been evolved and described in the literature. The Holliger group selected for a triplet polymerase ribozyme, meaning that it incorporates tri-nucleotide triphosphates to the growing RNA chain (*100*). Inspired by the mechanism of translation, this method of RNA synthesis is easily connected with that of codon ascription, potentially facilitating the transition from a random to a coded peptide / protein synthesis pathway. Moreover, the triplet polymerase is unique in function since it requires no primer, proceeds in both 3'–5' and 5'–3' directions and is able to template through short double-stranded stem structures, in part due to the strand invasion capabilities of trinucleotide substrates.

The Joyce lab developed their own version of a polymerase ribozyme through three generations of evolution starting from the one Bartel had found. During the first generation, they applied selective pressures for copying complex sequences with double stranded structures, resulting in the 24-3 polymerase, named after selection round 24, clone 3 (*101*). This polymerase ribozyme was shown to synthesise 2 aptamers, the R3C ligase ribozyme and a yeast tRNA, proving that it could very well polymerise through complex secondary structures. The next generation polymerase was obtained by mutating the 24-3 and applying a modified selection procedure such that the activity of the newly-synthesised ribozyme is also included in the selection step, pushing evolution of the ribozyme towards the synthesis of functional sequences (fidelity). This resulted in the 38-6 polymerase, which was able to synthesise an active fragmented version of the class I ligase (*102*). The activity of the newly synthesised class I ligase was assayed but found to be 8000-fold slower than that synthesised by T7 RNA polymerase, suggesting that, while the 38-6 polymerase can synthesise large complex RNAs, the error rate is too high and the de novo synthesised ribozyme sequence diverges from the original thus suffering in catalysis. Consequently, the most recent 52-2 polymerase was evolved under conditions requiring the synthesis of functional hammerhead ribozyme and under sub-optimal magnesium chloride concentrations,

essentially selecting further for ribozyme fidelity (*103*). Deep sequencing of the hammerhead ribozyme and class I ligase synthesised by the 52-2 polymerase revealed that the average fidelity per nucleotide was 91.7% and 84.1%, respectively, providing an explanation for the reduced activity of RNA-synthesised as opposed to protein-synthesised ribozymes.

The most structurally complex polymerase evolved to date is the clamping polymerase (CP) ribozyme from the Unrau laboratory (*104*). Drawing inspiration from modern polymerases in efforts to improve processivity, they sought to develop a polymerase that can undergo a conformational change upon binding such that it becomes clamped on to the template. Similar to extant polymerases, the CP ribozyme can search for a promoter sequence with the help of a specificity strand by forming a holopolymerase complex and through conformational changes rearrange into a template-bound processive elongation form. The CP ribozyme was shown to synthesize duplexes ranging from ~50 – 107 bp in size, with more than an order of magnitude increase in extension compared to unclamped versions.

The development of RNA polymerase ribozymes has offered an unprecedented view of the potential for early RNA-based replicative systems, bringing us tantalizingly close to the development of an evolving replicative system *in vitro*. Nonetheless, many outstanding challenges to produce such a system based on ribozyme polymerases remain, including enhancing polymerisation rate, fidelity, and strand displacement ability.

1.3. Constraints on the time for the origin of life

The attempt to recreate life in the lab is the ultimate goal of the field of origin of life research, as succinctly stated by Albert Eschenmoser (*105*): *'the origin of life cannot be discovered; it has to be re-invented'*. However, to best exemplify an abiotic origin of life, the initial conditions under which it first arose must be identified or, at least, constrained, in order to guide prebiotic chemistry towards the conditions that favour the synthesis of the relevant organic molecules. Consequently, investigating the Earth's formation and development is necessary to understand the evolution of environmental conditions under which life first arose over the course of Earth's formation. Information dating back billions of years has been indirectly inferred from crystal minerals and ancient rocks by determining decay states of radioactive isotopes and measuring ratios of different stable isotopes to piece together a timeline for the birth and subsequent maturation of our solar system and planet. Unfortunately, with plate tectonics operating on Earth, there are no direct rock archives that date back to the first 500 million years of Earth's history more than

4 billion years ago (Ga), during what is referred to as the Hadean Eon, leading to a paucity of information regarding the early stages of the planet (*106*). Nonetheless, the discovery of detrital zircon minerals (zirconium silicate, $ZrSiO_4$) has allowed a unique glimpse of the early years of planetary formation owing in part to the hardiness of zircons, which can persist almost indefinitely and be recycled through several crustal rock sediments. What allows zircons to be used as geological clocks is their crystal lattice, which acts as a uranium-238 / lead-206 ($^{238}U/^{206}Pb$) chronometer, by permitting the selective incorporation of uranium-238 into the crystal lattice and excluding all lead atoms. Uranium-238 decays into lead-206 with a half-life of 4.47 billion years, meaning that any lead-206 atoms detected in zircon crystals are a direct result of uranium-238 decay, which in turn can be used to date the time of zircon crystal formation. Consequently, based on radiometric dating and the discovery and analysis of microfossils, a time for the likely emergence of life can be constrained **(Figure 9)**.

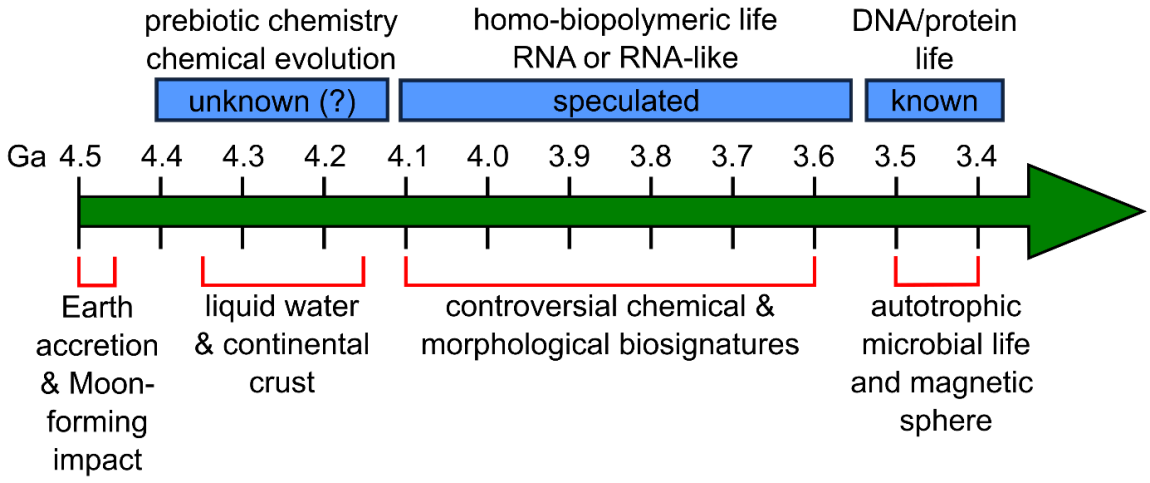


Figure 9: Hypothetical timeline for an early origin of life. *Predictions are based on zircon analyses as well as models for planetary and atmospheric evolution. The timeline takes into consideration proto-cellular evolution during which a single biopolymer was used for both information storage and catalysis, whether RNA itself or a non-canonical nucleic acid polymer that subsequently evolved into RNA remains a matter of debate.*

1.3.1. *Bona fide* evidence of life on Earth

The earliest consolidated and agreed upon evidence for life on Earth are the ~3.5 Ga stromatolites in Western Australia, that are believed to have formed from the deposition of layers of photosynthetic bacteria resembling microbial mats of modern-day cyanobacteria (*107*, *108*). Others have also noted early forms of life preserved in 3.5 Ga hot springs deposit (*109*). More recently, evidence was uncovered for the presence of diverse communities of both bacteria and archaea in 3.5 – 3.3 Ga microbial mats in South Africa

(*110*). As such, these findings indicate that by 3.5 Ga autotrophic cellular life was already widespread, hinting at a likely presence of life even earlier.

1.3.2. Disputable evidence of life on early Earth

Due to the labile nature if biomolecules, the search for fossils older than the time under which they are stable must be one of an indirect nature, that is, searching for remnants of biological activity rather than biological beings. Chemical biosignatures for early life are based on the biological fractionation of carbon isotopes, namely carbon-12 and carbon-13. This is possible because abiotic carbon fixation has no bias whatsoever for either isotope, whereas biology slightly prefers carbon-12 over carbon-13. Thus, carbon fixation from a biogenic source will enrich carbon-12, leading to a higher ratio of $^{12}C/^{13}C$ over time and, by comparing this ratio to modern biological fractionation ratios of carbon isotopes, the biogenicity of the carbonaceous inclusions can be either confirmed or refuted (*111*). That being said, analysis of carbon isotope signatures in a 4.1 Ga Hadean zircon crystal revealed a $^{12}C/^{13}C$ ratio consistent with that of biological fractionation (*112*). Others have also noted Eoarchean (older than ~3.6 Ga) rocks depleted of ^{13}C, further strengthening the case for an early inhabited Earth (*113*, *114*).

The oldest morphological biosignatures, on the other hand, are much more challenging to decipher for two reasons (*106*). First, rocks can be dated but fossils cannot, therefore, a relationship between the purported microfossil and its host rock must first be established. Second, many abiotic processes that often occur on crustal rocks mimic life-like processes and thus can be easily mistaken evidence of life (*115*). In fact, a range of self-organising structures that naturally arise from the physico-chemical conditions have been described to leave traces suggestive of microfossils and potential remnants of early life (*116*). For these reasons, any morphological description of life-like structures in ancient rocks must be approached with the highest degree of scepticism to avoid making inferences based on flawed reasoning. The diversity and resolution of techniques available for the characterisation of potential biosignatures are ever better, but are only meaningful if interpreted in their geological context and corroborated with other approaches. As Pierre-Simon Laplace put it: *'the weight of evidence for an extraordinary claim must be proportioned to its strangeness', or paraphrased by Carl Sagan: 'extraordinary claims require extraordinary evidence'.*

1.3.3. A habitable Earth

The Hadean Eon (~4.5 – 4.0 Ga) was named after Hades – the Greek mythical god of the underworld – because scientists initially assumed, based on the then current models of solar system and planetary formation, that the early atmosphere and surface of Earth was hellish. This has been a topic of heated debate among geologists and planetary scientists alike, but seemingly varying opinions have funnelled into a generally agreed upon consensus for the initiation of habitable conditions on the early Earth. In fact, atmospheric modelling combined with zircon analyses of oxygen and silicon isotopes as well as uranium-lead dating have provided evidence for the existence of both liquid water and continental crust during the Hadean Eon (*117–121*). However, the validity of these claims must be approached with caution, since they are only as accurate as the radiometric dating method of zircon crystals employed to estimate the age of the rocks in which they were found. When estimating dates from 500 million years ago, an error of tens of millions of years is acceptable, but for estimating dates that go back 4 billion years, one must be content with errors that range in the hundreds of millions of years. While hundreds of millions of years can be considered short time period in a geological sense, from an evolutionary perspective it is a long time period.

Oxygen isotopes are a good measure of the temperatures at which minerals form because oxygen-16 is lighter than oxygen-18, causing water molecules composed of oxygen-16 to evaporate more frequently than water composed of oxygen-18. Thus, the ratio between oxygen isotopes can be used to predict the relative environmental temperature at which the geological evidence formed. An enrichment of oxygen-18 is suggestive of higher temperatures whereas an equivalent amount of oxygen-16 or 18 isotopes trapped in minerals is indicative of lower temperatures.

The moon-forming giant impact released large amounts of CO_2 into the air, effectively creating a runaway greenhouse effect thereby maintaining high surface temperatures well above the boiling point of water. Therefore, the solidification of the magma ocean and subsequent condensation of water vapor to form the oceans must have followed three heat producing phenomena that kept surface temperatures on early Earth high, namely, the decay of highly radioactive elements in the Earth's core (such as hafnium-182), the outgassing of CO_2 to outer space reducing the greenhouse effect, and finally the decrease in gravitational tidal forces from the newly formed and slowly receding moon (*122–124*). Knowing the timescales at which these phenomena occur allows an estimation into when the conditions on early Earth allowed the precipitation of liquid water. As such, many

predict that surface conditions of the early Earth were clement already during the Hadean Eon, optimistically by around 4.35 Ga. The presence of liquid water at that time has not yet been empirically confirmed, however, definitive evidence for the presence of liquid water on the surface has been deduced from oxygen isotopes in zircon crystals dated to 4.15 Ga using the previously introduced oxygen isotope analysis. Therefore, at least prebiotic chemistry was very likely already in play during the end of the Hadean Eon and perhaps, under favourable conditions, the formation, accumulation and polymerisation of primitive nucleic acids and peptides was already underway.

1.4. Plausible locations for the origin of life

Numerous environments have been hypothesised as cradles for the emergence of life and / or life-like systems **(Figure 10)**. The different environments proposed vary greatly in their physical and chemical parameters, however they all possess common characteristics that are thought to be essential for abiogenesis. First, they should have a source of free energy, be it thermal, solar or chemical. The free-energy input is critical since it is exploited by chemical systems to synthesise monomers, activate them, polymerise them, etc. It is commonly associated with gradients of all types, including redox gradients, solute concentration, temperature, pH, density, etc. Second, they must have been common and / or persistent during early Earth evolution, such that the prebiotic processes occurring would have enough time to accumulate important precursors for biomolecular anabolism. I will here list the most appreciated environments postulated to have been the locations hosting the emergence of the first cells.

1.4.1. Surface springs associated with hydrothermal environments

Hot springs have been recently proposed as suitable locations for the origin of life (*125*, *126*). These environments are generally associated to underground hydrothermal activity, and the physico-chemical parameters can vary greatly over hundreds of individual vent pools in a common locality (*127*). Thus, an interconnected population of pools varying in temperature, pH, ionic strength and elemental composition (depending on what types of clays form from the upwelling of hydrothermal fluids) can act as the individual reactors for the multistep prebiotic chemical synthesis reactions. These types of environments were also predicted to last for timescales that are probably permissive to accrue organic compounds, in the range of 10 to 100 million years (*128*). Much evidence for the viability of hot spring environment as locations for chemical and biological evolution has surfaced in recent years.

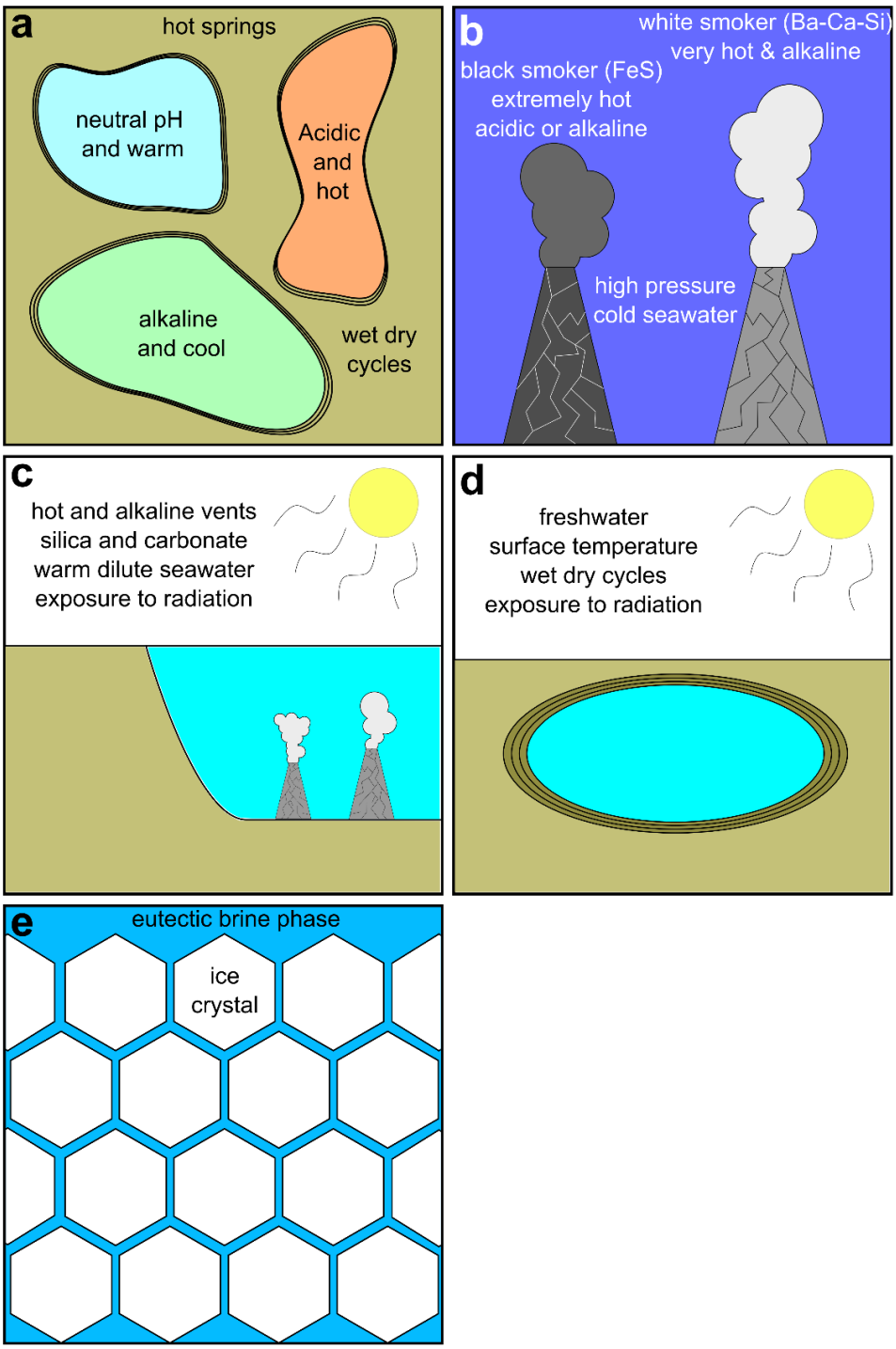


Figure 10: Locations for the emergence of life. *Cartoon showing the different locations postulated to have supported the origin of life.* ***a)*** *Subaerial hot springs can consist of different pools in close proximity but harbouring greatly varying conditions of pH, salts, and temperature.* ***b)*** *Deep sea hydrothermal vents can be black smokers or white smokers, depending on the type of minerals that form the chimneys: iron sulphide (black) or calcium / barium silicate (white) precipitate from hydrothermal fluids upon contact with cold seawater.* ***c)*** *Shallow hydrothermal vent settings occur in shallow basins experiencing hydrothermal activity and comprise both freshwater and seawater runoff.* ***d)*** *Freshwater ponds and lakes formed from impact craters host wet-dry cycles at their boundaries and better accumulate organics compared to open sea.* ***e)*** *Pure ice crystals sequester water from aqueous solutions such that the interstitial brine effectively concentrates solutes and ions.*

For example, short- and single-chain amphiphiles (10-carbon and 12-carbon chains) were able to self-assemble into vesicles when resuspended in hydrothermal fluid procured directly from the hot springs at Yellowstone National Park, in the United States (*129*). In contrast, the same mixture of amphiphiles failed to produce any membrane compartments when resuspended in seawater. This observation suggests that seawater has too high an ionic strength to allow vesicle formation from short-chain fatty acids or their glycerol derivatives, whereas hydrothermal fluids provide a more adequate chemical milieu for vesicle content, such as wet-dry or freeze-thaw cycling, the polymerisation of available monomers is promoted due to the inherent evaporation of water molecules, since polymerisation is essentially a dehydration step. Indeed, many have shown how hydrothermal fluids, especially acidic ones, are conducive to the synthesis of both canonical and non-canonical nucleosides and nucleotides, further strengthening the hypothesis that the first cells emerged in hot spring pools (*46*, *50*, *130*). These kinds of environments make it possible to concatenate synthesis reactions across a wide range of chemical conditions, as model aerosols have also been shown to carry cargo and potentially become vesicles if landing in a solvent with amphiphilic solutes (*131*).

1.4.2. Deep sea hydrothermal vents

Deep sea hydrothermal vents typically form around mid-ocean ridges, where continental plates of the oceanic crust meet (*128*). In these regions, cold oceanic water percolates downwards through cracks in the ocean floor, eventually reaching the very high temperatures of the mantle before being ejected back into the ocean. The superheated seawater, however, returns with a wealth of solutes that were dissolved during its path through the ocean floor (*132*). The solutes vary depending on the exact chemistry of the vent and its location relative to the ridge axes (and therefore, its temperature), such that different minerals precipitate out of the hydrothermal water producing the vents (*124*). Black smokers form close to the ridge axis, are rich in mineral sulphides and characterised by very high temperatures and extreme pH values both acidic and alkaline. White smokers, on the other hand, form farther away from the fault lines. They are carbonate- and silicate-rich vents characterised by lower temperatures and strictly alkaline pH that gained much fame during the early days of origin of life research due to their elementary composition and microenvironment. Such microenvironments were first predicted by geologists to exist and host the right redox chemistry to reduce carbon dioxide with hydrogen gas; the discovery of the Lost City Hydrothermal Field sparked high interest in these settings after confirmation of

chemosynthetic carbon fixation taking place (*133–136*). It was (and maybe still be) unproductive to attempt to define life in a way that encompasses all the variety of different organisms that have evolved, leading Michael Russel to attempt to define via a reductionist approach of what life does rather than what it is: *'Life hydrogenates carbon dioxide'*. The finding that alkaline hydrothermal vents are a non-biological sources of carbon fixation, whose mechanism closely resembles that of modern chemosynthetic bacteria, was considered by Russel to be the 'smoking gun' indicating that this is where life must have emerged. While these environments do provide a redox gradient for the hydrogenation of carbon dioxide, the extreme conditions of pressure, temperature and pH, along with the vastness of the ocean, would effectively prevent the accumulation of organic molecules simply due to the fast degradation rates spurred by these extreme conditions and dilution factors of oceanic volumes. In fact, this particular environment is advocated by proponents of a metabolic origin of life scenario, such as Nick Lane, since the chemical gradients originating from abiotic geological phenomena are unnervingly similar to the proton motive force employed by extant cells for chemiosmotic energy production (*134*).

1.4.3. Shallow water hydrothermal environments

Shallow hydrothermal vents, on the other hand, host environmental conditions that are much more clement, possibly proving amicable to the stability of biomolecules. At 5 – 15 m below the surface where light can still penetrate on today's Earth, they could have also occurred at higher depths during the Hadean / Paleoarchean as a result of more common storms and stronger tidal effects (*137*). These environments were likely present on volcanic islands where freshwater from rain would flow into shallow sea basins producing a diluted seawater rich in hydrothermal fluids and rare Earth elements (*110*). What is remarkable in these shallow vents as opposed to their deep-sea counterparts is that the hydrothermal activity is exclusively characterised by low-temperatures relative to deep sea hydrothermal vents (circa 100 °C) and that they are rich in silica (*138*). Since they are found in shallow regions, they could potentially benefit from Sun's free energy. Their dependence on volcanic activity and the formation of islands suggests that these environments were probably transient, forming in an episodic manner comprising of approximately 100 years of active volcanism followed by an equally long time of quiescence. Thus, the ability of these systems to dilute ocean water with fresh water in combination with their rich elemental diversity makes them attractive locations to explore prebiotic chemistry and the emergence of life.

1.4.4. Freshwater lakes and 'warm little ponds'

The first setting proposed for the emergence of life, to the best of my knowledge, is Charles Darwin's 'warm little pond' (or lake), essentially describing a body of freshwater in a crater on surface exposed land, in which key organic species can be up-concentrated via drying on the banks as a result tidal effects from the moon or oscillations in temperature driven by day-night or seasonal cycles (*139*). Geochemical systems such as these continuously generate fluctuations not only in the concentration of organic molecules, but also in salts and pH, which can have several orders of magnitude difference between wet and dry states. Many prebiotically relevant reactions benefit from an increased concentration of reactants, since abiotic condensation reactions are typically only activated at high reactant concentrations, such as the condensation of hydrogen cyanide into its tetrameric form, an intermediate in a synthesis pathway for nucleobases (*42*, *140*). Other reactions that are favoured under dry conditions include the phosphorylation of prebiotic molecules and non-enzymatic RNA polymerisation (*60*, *141*). Intriguingly, a theoretical model for the onset of the RNA world in muddy pools utilising experimentally calculated polymerisation rates as input parameters has revealed that the spontaneous emergence of catalytic RNA and tRNA-like structures occurs favourably under conditions of environmental cycling through distinct hydration phases of dry, semi-wet and wet (*142*). In spite of these attractive characteristics, these locations have been called into question particularly due to the poor stability of nucleobases irradiated under dry conditions, as well as the dilution of highly concentrated reactants into lake or pond water upon rehydration, effectively preventing their build-up (*142*).

1.4.5. Eutectic ice phases and the faint young Sun paradox

While the Hadean atmosphere is more often associated with slightly higher temperatures compared to today's atmosphere, it is nonetheless probable that ice was a common medium, at least on exposed landmasses or at the poles (*123*, *143*). The hypothesis for a frozen early Earth during the Hadean is based on observations and subsequent calculations that led to non-zero probability for its occurrence. The argument proceeds as follows: firstly, the stellar formation and evolution models that emerged for main-sequence stars, such as our Sun, predict that luminosity of a star gradually increases after it begins nuclear fusion, such that the Sun's luminosity output during the Hadean Eon ~4.5 – 4.0 Ga was approximately 25% – 30% lower than today (*144*, *145*). The second line of reasoning is based on a recent model that also takes into consideration extra-terrestrial bodies that, upon

impact with Earth, result in the release of cations from silicate weathering which easily react with carbonic acid, thereby sequestering carbon and decreasing the overall partial pressure of carbon dioxide in the atmosphere, diminishing the greenhouse effect (*146*). This in turn indirectly affects water pH and atmospheric temperatures as dissolved carbon dioxide in water exists as an equilibrium between carbonate and carbonic acid (*147*), while carbon dioxide gas is a strong greenhouse gas (*148*), which were predicted to deviate significantly from the traditional hellish Hadean environment, suggesting sub-zero surface temperatures and a slightly alkaline ocean, as opposed to a hot and acidic ocean. Although the initial predictions of the Hadean atmosphere centred around a reducing atmosphere, it is becoming clearer that this was not the case, urging planetary scientists and prebiotic chemists to explain where the reducing power needed to synthesise organic molecules originated. A possible solution came from an investigation into the role of impactors in providing reducing power to an otherwise oxidizing atmosphere, encouraging a reassessment of the timeline and timescales for the emergence of organic molecules and their biological anabolic products, concluding that non-planet sterilizing impacts could have played an important role in driving the abiotic synthesis of biomolecular precursors (*149*). Such impacts in a frozen Earth scenario would result in the transient melting of ice over a defined area that graduates in the relative amount of liquid to crystal water, potentially producing zones of partially frozen water and a reducing atmosphere. The subsequent cooling after impact to low temperatures would also favour the retention and stability of the produced molecules until the next burst of prebiotic synthesis.

Further support is lent to this hypothesis because as solute-rich aqueous solutions begin to freeze, the solutes are up-concentrated in interstitial brines while pure water freezes to form ice crystals, generating oscillations in solute concentration and pH as ice crystals form and melt (*150, 151*). This phenomenon has made liquid brine phases in eutectic ice matrices an attractive, potentially prebiotic environmental niche since they would favour reactions that are otherwise incompatible in dilute aqueous phases. Experimental evidence has revealed that partially frozen eutectic ice phases are conducive to abiotic synthesis, non-enzymatic polymerisation, as well as nucleic acid catalysis, with some cases of enhanced or altered function reported (*85*, *152–155*). For example, UV-irradiation of urea solutions followed by freeze-thaw (FT) cycles enables the prebiotic synthesis of purines and pyrimidines, and a plausible pathway for the for the synthesis of peptide-RNA species is promoted by eutectic ice formation and the accompanying change in pH (*65*, *156*). Interestingly, FT cycles have been shown to drive the exchange of content among membrane

vesicles without requiring complex protein machinery tailored to content exchange (*157*). For instance, the liquid nitrogen-based freezing and subsequent thawing of pelleted lipid vesicles enables the fusion and fission of liposomes and concomitant viral RNA replication when replicase and genome are initially independently encapsulated, leading to the dissemination of the viral RNA genomes within the vesicle population in response to freezing and thawing. Similarly, temperature cycling of giant unilamellar vesicles (GUVs) between -80 °C and room temperature promotes content exchange of labelled DNA fragments among densely packed GUVs independently from fusion and fission (*158*).

In summary, both theoretical and experimental data discussed in this section highlight the potential of geological settings with transient eutectic ice formation to support both chemical and biological evolution. As such, both the hypothesis of a frozen early Earth and geological settings exhibiting freezing conditions should not be discounted as plausible locations for the emergence and evolution of chemical and biological systems.

1.5. Compartments at the origin of cellular life

As mentioned earlier, the emergence of a cellular compartment is an essential feature that must have occurred at some point during biochemical evolution. Compartmentalisation is required to isolate and protect intracellular systems from environmental perturbations and parasitic replicators as well as establish genotype-phenotype coupling (*159–162*). It also serves the purpose of containing all biomolecules and metabolites thereby preventing their loss via dilution. Natural selection does not exclusively act on compartmentalised systems, otherwise the *in vitro* selection of specific RNA sequences from random pools would not have had such success. However, uncompartmentalized systems are unable to select for more complex functions, such as one which includes feedback mechanisms and / or network motifs (*163*). In these cases, compartmentalisation becomes essential; for example, a GTP-synthase ribozyme was recently selected by coupling nucleotide activation with ribozyme polymerisation in microemulsions (*164*). In the absence of a compartment, activated nucleosides do not offer a fitness advantage in bulk solution since they will simply diffuse away. The absence of a fitness advantage signifies that there is no basis for which selection can favour certain sequences, and in turn, Darwinian evolution cannot proceed. This is exemplified in the literature, since many of the *in vitro* selection and evolution experiments aiming for complex functions are performed in cell-like compartments, such as water-in-oil emulsions by employing microfluidic platforms (*164–166*). Over the years,

different types of compartments have been proposed as potential ancestors to those of modern cells, each having its own series of advantages and disadvantages.

1.5.1. Membrane-bound compartments

Membrane compartments composed of amphiphilic molecules are by far the most common model membrane systems, owing to their close relationship with modern cellular membranes (*163*, *167*). A large proportion of protocell research has been dedicated to membranes composed of simple single chain amphiphiles, such as fatty alcohols and fatty acids, as opposed to phospholipid-based membranes **(Figure 11)**. This reasoning stems from the more facile non-enzymatic synthesis of fatty acids, such as in the Fischer-Tropsch type reactions where carbon monoxide is reduced into a mixture of hydrocarbons with a terminal oxygen (*168*). These reactions were also postulated to occur around deep-sea hydrothermal vents as the high temperature and pressure at these depths drive the reaction of carbon monoxide and hydrogen gas (*169*). Moreover, if considering a heterotrophic protocell with a membrane-bound replication system, semi-permeability of the membrane to allow the flux of building blocks and waste becomes an essential characteristic. The permeability of fatty acid vesicles to nucleotides and ions have made them a more attractive protocell membrane model as opposed to phospholipids and their impermeable membranes, which unlike modern phospholipid membranes, are not endowed with semi-permeability due to the presence of highly evolved protein pumps, channels and transporters (*170*, *171*). Indeed, the non-enzymatic templated polymerisation of RNA inside simple lipid vesicles has already been demonstrated (*172*, *173*), as well as the compatibility of a nucleolytic ribozyme (RNA cleavage by the hammerhead ribozyme) in the intra-vesicular milieu (*171*). Efforts to increase the complexity of the ribozyme systems in fatty acid-based vesicles resulted in inefficient systems, particularly due to the low stability of fatty acid vesicles in the presence of divalent metal ions, which are at most times pre-requisites for nucleic acid catalysis (*174*, *175*). These studies from the Danelon laboratory attempt to find suitable reaction conditions that reconcile ribozyme ligase activity and fatty acid vesicle stability, namely the L1 ligase and myristoleic acid (C14:1) vesicles. However, reaction conditions that were permissive to vesicle formation only slightly mediated ligation activity, whereas conditions that favoured catalysis were inhibitory to vesicle formation. Strategies to overcome the issue of membrane stability towards salts include membranes with heterogenous composition, such as those combining fatty acids with their fatty alcohol and glycerol derivatives (*176*). Mono-acyl cyclophospholipids were also shown to stabilise fatty acids membranes in the

presence of increasing divalent metal ion concentrations (*177*). Even fatty acid-phospholipid blended membranes were proposed as a transition stage between protocell membrane and extant cell membranes due to their improved stability and robustness to fluctuating environmental factors (*178*). A particularly interesting strategy developed involves chelating the magnesium ions with citrate, which allowed the non-enzymatic copying of RNA to proceed inside fatty acid vesicles with only a minor reduction in catalytic rate (*173*). While this approach enabled non-enzymatic RNA primer extension within fatty acid vesicles, it does have its drawbacks since any chemistry that requires the direct involvement of a magnesium ion in the active site i.e., a catalytic metal ion, such as those catalysed by many ribozymes (*179*), will suffer greatly from the absence of free magnesium ions in solution.

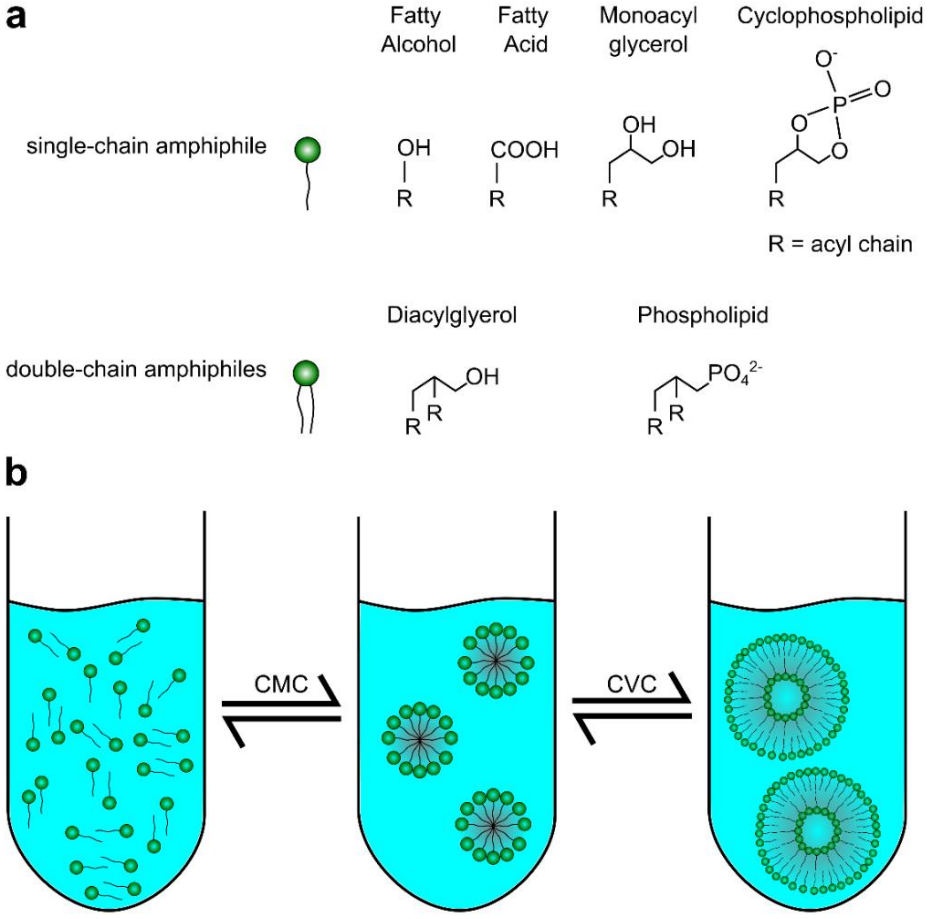


Figure 11: Structure and self-assembly of amphiphiles in solution. a) *Amphiphiles can be classified according to the number and length of their acyl chains (R) as well as the nature of their head group.* ***b)*** *Amphiphiles in aqueous solutions self-assemble into micelles after lipid concentration exceeds the critical micelle concentration (CMC) and subsequently into vesicles after crossing the critical vesicle concentration (CVC) threshold. The type of lipid determines the threshold concentrations for self-assembly. Generally, phospholipids have a lower threshold concentration and their membranes are impermeable to charged solutes as opposed to fatty acids which require higher concentrations for spontaneous assembly while their membranes allow the diffusion of ions and nucleotides.*

Notwithstanding the accessibility of fatty acid synthesis under prebiotic conditions, phospholipids have been gaining traction as the potential membrane component of early protocells owing to their more robust membrane characteristics. The first objection to the plausibility of phospholipids as early protocell components is one of availability, since they are structurally more complex than their fatty acid counterparts and, in extant cells, are synthesised exclusively by protein enzymes. However, recent advancements in organic synthesis have revealed three plausible pathways for the non-enzymatic synthesis of phospholipids (*180*): one involves the phosphorylation of diacyl glycerol to give a mixture of phosphorylated racemic lipid amphiphiles able to self-assemble into stable primitive compartments (*181*). Another comprises the selective alkylation of phosphoglycerol, which produces mono- and diacyl phospholipids that self-assemble into compartments and are stable over a substantial pH and temperature range (*182*). Finally, a water-based enzyme-free synthesis pathway occurs through the alkylation of an acyl phospholipid precursor (lysophospholipid) with an acyl thioester reactant at alkaline pH (*183*). Intriguingly, the synthesis reaction was also carried out in field-recovered solvents, including hydrothermal vent fluid (Lost City Hydrothermal Field (*184*)) and lake water from Mono Lake (a Californian soda lake) that have pH values of 9.1 and 10, producing phospholipid yields of 42% and 72%, respectively. The reaction products were able to form phospholipid membranes *in situ,* and maintained a proton gradient over a timescale of hours. These results suggest that phospholipids on early Earth might not have been as scarce as once thought.

The second major objection to phospholipid-based protocells at the origin of life is their permeability. Sharing of genetic information, import of precursors and export of waste products is a pre-requisite for a species' ability to proliferate. While modern cells are gifted with the ability to transfer genetic information vertically through cell division, the purported absence of complex protein machineries in early protocells suggests that they must have enabled the dissemination of genetic information via a different route, possibly horizontal transfer mediated by repetitive physico-chemical processes that could alter membrane permeability (*35*, *185*). In fact, such processes have already been shown to support both encapsulation of biomolecules as well as dissemination of genetic information. Wet-dry cycling promoted vesicle formation from amphiphilic solutes and encapsulation of relevant biomolecules, whereas temperature oscillations (freeze-thawing or heating-cooling cycles) enabled the exchange of genetic content among protocell populations, as discussed above as well (*157*, *158*, *166*, *186–188*). Taken together, the evidence adduced

strengthen the hypothesis that early protocells may have also had membranes consisting of simple phospholipids, warranting further investigations as a potential component of predecessors to modern cell membranes.

1.5.2. Membrane-less compartments

Coacervates were initially proposed by Oparin in the 1920s as the first compartments during the emergence of life (*10–12*). Complex coacervates form spontaneously above a concentration threshold by the coalescence of oppositely charged polyelectrolytes into a separated phase, driven by the entropically favoured release of water and / or other solutes **(Figure 12a)** (*189*). Typically, the macroscopic behaviour of two oppositely-charged polyelectrolytes follows a phase diagram exhibiting a binodal curve, where the area under the curve designates a two-phase system while a single phase governs the area above it **(Figure 12b)** (*190*). For example, for the horizontal grey arrow shown in **Figure 12b**, the system initially in a single phase exhibits a phase transition when salt concentration is kept constant but the length of a polyelectrolyte polymer is increased, crossing the binodal curve and into the two-phase region. Conversely, the vertical grey arrow starts with a two-phase system. The salt concentration can be increased such that the system crosses the binodal curve and transitions back into a single phase, an effect of polyelectrolyte charge screening by the dissolved salt ions. It is important to note that the phase diagram of system does vary with components type, polymer lengths and charge-density (*191–193*).

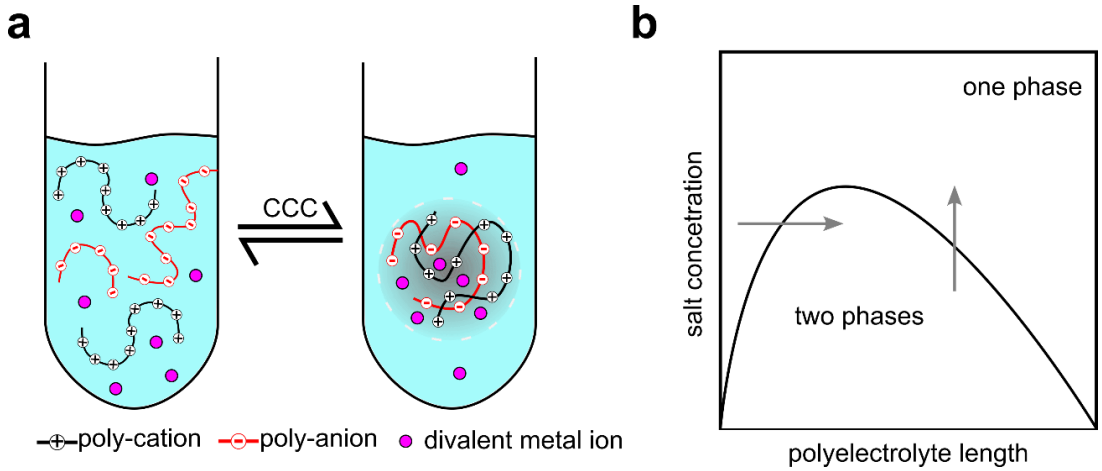


Figure 12: Schematic depicting the formation of coacervate microdroplets in aqueous solutions. ***a)*** *The critical coacervate concentration (CCC) is the threshold concentration of the components above which coacervates spontaneously coalesce. The resulting aqueous two-phase system consists of a polymer-poor phase and a polymer rich phase, the latter being much smaller in volume relative to the former thereby increasing the local concentration of metal ions and other relevant biomolecules according to their partitioning coefficients and polymer types.* ***b)*** *Example of phase diagram showing binodal curve as a function of salt and polyelectrolyte length. Arrows show phase transitions of the solution by increasing / decreasing either of the two variable parameters.*

Their ability to selectively partition biomolecules and ionic species and to support a variety of enzymatic processes have made them attractive models for protocell models in the origin of life research. For example, liquid-liquid phase separated droplets formed from ATP and polyallylamine (PAH) exhibited at least 100-fold increase of magnesium ions and a ~10,000-fold increase of random 54-mer RNAs in the coacervate phase (*194*). ATP was strongly partitioned, indicating that coacervate phases could have hosted non-enzymatic RNA polymerisation. In addition, coacervates composed of carboxymethyl-dextran and poly-L-lysine were shown to be compatible with *in vitro* gene expression systems, demonstrating that the highly orchestrated components of transcription and translation could be contained in the condensed phase, assemble and remain functional (*195*).

Of particular interest to origin of life studies are coacervate phases that are formed from peptide-nucleic acid interactions and / or support nucleic acid catalysis and other reaction pertinent to the field (*189*, *191*, *196*). Many studies have shown that RNA-peptide interactions can have mutualistic or commensal relationships; a correlated increase in stability of both RNA and proto-peptides and their polymerisation when combined under wet-dry cycling has been reported (*197*), whereas others found that short peptides with sequence homology to ribosomal peptides bolster RNA polymerase ribozyme activity (*198*). Coacervates were also shown to promote the evolution of folded domains of the peptides they host through sequential events of duplication and fusion (*199*). Oddly, the activity of nucleic acid catalysts in condensed phases has only recently been demonstrated, with some cases showing catalytic rate enhancements or equilibriums shifted. A landmark paper in 2018 was the first to demonstrate that coacervate phases support the catalysis of a minimal nucleolytic hammerhead ribozyme, and shorty after came non-enzymatic templated RNA polymerisation within coacervates (*200*, *201*). More recent evidence surfaced showing that coacervates formed from RNA and Arginine-Glycine-Glycine tetrameric repeat $(RGG)_4$ peptide, which has positively charged residues interspaced with non-polar amino acids effectively decreasing peptide charge density, were able to partition the R3C ligase ribozyme and its substrate, while retaining functionality of the former (*202*). A study on the properties of poly-L-lysine RNA coacervates revealed that a partitioned minimal hairpin ribozyme, which reversibly cleaves RNA in solution under physiological conditions, possessed a shifted reaction equilibrium towards ligation when compartmentalised (*86*). In short, the ability of liquid-liquid phase separation to selectively partition a variety of biomolecules depending on the composition of the condensed phase polymers, makes them convincing protocellular compartments.

1.5.3. Non-canonical compartments

Other than the abovementioned types of compartments, environments that provide similar advantages of compartmentalisation should also be considered. As we have seen in the different potential geological locations for the origin of life, the interaction of solvated charged mineral surfaces with organic biomolecules is inevitable after the condensation of water took place on the early Earth. A wide variety of minerals and clays cover the walls of hydrothermal vents and of some porous rocks, as well as form the sediments in freshwater ponds and lakes. Clay minerals in particular can be considered as primitive compartment systems owing to their characteristically large adsorption capacity, their ordered arrangement, and their ability to shield nucleic acids from environmental pressures and thus fulfil some of the benefits of being in a compartment (*203*) . In particular, positively-charged clay minerals interact strongly with nucleic acids offering a variety of prebiotically relevant effects and likely playing a major role in chemical and molecular evolution(*204*). Borate minerals were shown to stabilize ribose sugars whereas phosphate minerals act as phosphate donors enabling the generation of phosphorylated nucleosides(*205*, *206*). Both DNA and RNA polymerisation can proceed on montmorillonite clays and hydroxyapatite, which are calcium and phosphorus rich minerals (*56*, *207*, *208*), while montmorillonite has specifically been shown to preferentially accumulate long RNA molecules as well as protect a model ribozyme from degradation by UV irradiation (*209*, *210*).

Differentially heated rock pores can also be treated as primitive compartments since they inherently host non-equilibrium dynamics that could benefit nucleic acids. Initial theoretical and experimental evidence demonstrated that a thermal gradient of 10 °C across 5 cm distance (2 °C cm^{-1}) enhanced DNA polymerisation from short duplexes with overhangs. The spatial accumulation of oligonucleotide substrates originating from thermophoresis and convection currents in the pore boosted sequence length and enabled the selection of oligonucleotides with increasing lengths (*211*, *212*). Thermo-convection also supported RNA catalysis by naturally optimizing salt concentrations (e.g., ratio of K^+/Na^+) made available from leaching minerals (*213*). These porous rock systems can trap gas bubbles arising from mantle degassing or through the incorporation of atmospheric air, thereby producing an air-water interface at a differentially heated surface (*214*, *215*). This triggers microscale water cycles (wet-dry) due to the evaporation of water at the hotter side and condensation of dew droplets on the cooler side, which invariably fall back into the aqueous reservoir. Studies have shown that these geologically plausible compartments promote a variety of prebiotically relevant reactions and phenomena: when carbon dioxide gas was

used to mimic early atmospheric composition, the drop in pH resulting from the dissolution of carbon dioxide in condensing water droplets free from ionic species allowed the melting of dsDNA at temperatures up to 20 °C lower than those in bulk solution (*214*). The water cycles also supported a polymerase mediated PCR-mimic reaction without the need for thermocycling, suggesting that the setup innately allowed denaturation, promoter annealing and extension, leading to an overall increase in pool length (*215*). Finally, the heated air-water interfaces promoted phosphorylation of nucleosides, RNA polymerisation and encapsulation of DNA in lipid vesicles, as well as fusion, fission and fragmentation of model membrane-less coacervate compartments (*216–218*). For these reasons, such systems are very attractive models as primitive compartments that have the potential to support a wide range of physico-chemical parameters likely beneficial to early replicating systems.

2. Aims and Objectives

The overarching motivation of this thesis is to explore RNA catalysis under non-standard conditions, with the aim of mimicking life-like characteristics that emerge from interactions between RNA catalysts and their environment. To this end, re-engineered ribozymes were characterised and employed in prebiotically plausible microenvironments, namely, catalysis in vesicles, in coacervates and at air-water interfaces in differentially heated systems. In each of the three following sections, an outstanding problem in the emergence and evolution of life is addressed. The first section attempts to determine a mode of heredity among early protocells, which are thought to have been void of complex proteins, raising the question as to how genetic information was transmitted among early life-like systems. Lateral gene transfer driven by environmental temperature oscillations were investigated as a means to disseminate genetic information among a population of membrane-bound protocells. The semi-continuous self-replication of an autocatalytic RNA replicator compartmentalised in phospholipid vesicles was enabled by cycles of freezing and thawing, thereby allowing RNA species to evade death and maintain a stable intra-vesicular concentration despite concurrent serial dilution in feedstock vesicles. Next, the challenge of genotype-phenotype linkage in membrane-less compartments was confronted, as it remains unclear how membrane-less compartments potentially generated fitness advantages that could be exploited by natural selection for the co-evolution of simple peptides and RNA. Thus, the role of complex coacervation in providing an adequate milieu for ribozyme catalysis was explored and an attempt to couple coacervate genotype and phenotype by utilising a ribozyme that alters the length of a coacervate component polymer was undertaken, which showed that indeed coacervate phenotype can be modulated by ribozyme activity and *vice versa*. Finally, a major obstacle concerning templated replication, namely, the strand separation / re-annealing problem was confronted. To achieve complete cycles of templated replication, the newly synthesised strand must be released from its template, a feat that is nearly insurmountable under standard laboratory conditions due to the high thermal stability of RNA duplexes. A re-engineered group I intron was employed to synthesise RNAs from short oligomers by templated ligation and subsequently used as a model general replicase system. The system was used to demonstrate that far-from-equilibrium conditions, achieved by microfluidic mimics of hydrothermal porous rocks exposed to CO_2-rich atmosphere fabricated by biophysicist collaborators, are compatible with and even en-

hance ribozyme-mediated templated RNA ligation. The separation of newly ligated products from their template and their subsequent catalysis was observed in a one-pot setting, effectively bypassing purification steps and recovery of single-stranded RNA. In summary, the results strongly support the hypothesis that bottom-up synthetic systems can be designed to emulate life-like characteristics without the need to invoke highly complex protein enzyme activities and provide several models for the study of RNA protocells from an origin of life perspective.

3. Results and Discussion

3.1. RNA autocatalysis driven by environmental temperature fluctuations

3.1.1. Background

Modern cells proliferate by replicating their genome and subsequently dividing themselves equally into two daughter cells, each with a copy of the mother cell genome. The mechanism of cell division is orchestrated by highly evolved proteins that tightly regulate cellular processes and govern cell state, ultimately leading to the generation of progeny. In contrast, protocells are often modelled as a membrane-bound replicating informational system, preceding the emergence of coded protein synthesis and therefore lacking the requisite enzymes for cell division. Hence, the dissemination of genetic information in protocells cannot have been driven solely by endogenous processes, rather in combination with an exogenous stimulus perhaps originating from the environment.

Preliminary inquiries have revealed that freezing at -80 °C followed by passive thawing at room temperature promoted the exchange of short fluorescently-labelled DNA strands amongst densely packed giant unilamellar vesicles (GUVs) (*158*). Consequently, freezing and thawing of packed GUV populations was investigated for possibly allowing both the exchange of RNA as well as its subsequent catalysis within the GUVs. The main objective was to design a protocellular system capable of compartmentalised self-replication and proliferation enabled by periodic temperature fluctuations, mimicking the proliferation of an early replicative system in a plausible early Earth geochemical environment **(Figure 13)**.

With that vision in mind, the parameters to produce adequate numbers of giant unilamellar vesicles containing encapsulated RNA were first optimised. Next, a minimal hammerhead ribozyme that cleaves its cognate substrate with high efficiency was utilised to demonstrate that RNA catalysis can be rescued by freezing and thawing if the ribozyme and substrate are separately encapsulated. Transitioning to a different ribozyme system, namely, the R3C ligase, which is capable of autocatalytic ligation and exponential amplification in solution, revealed the ability of freeze-thaw (FT) cycles to enable autocatalytic amplification of the ribozyme. The autocatalytic replicator was further challenged with repeated dilution in feedstock vesicles and demonstrated an ability to survive despite loss via dilution and degradation (*219*).

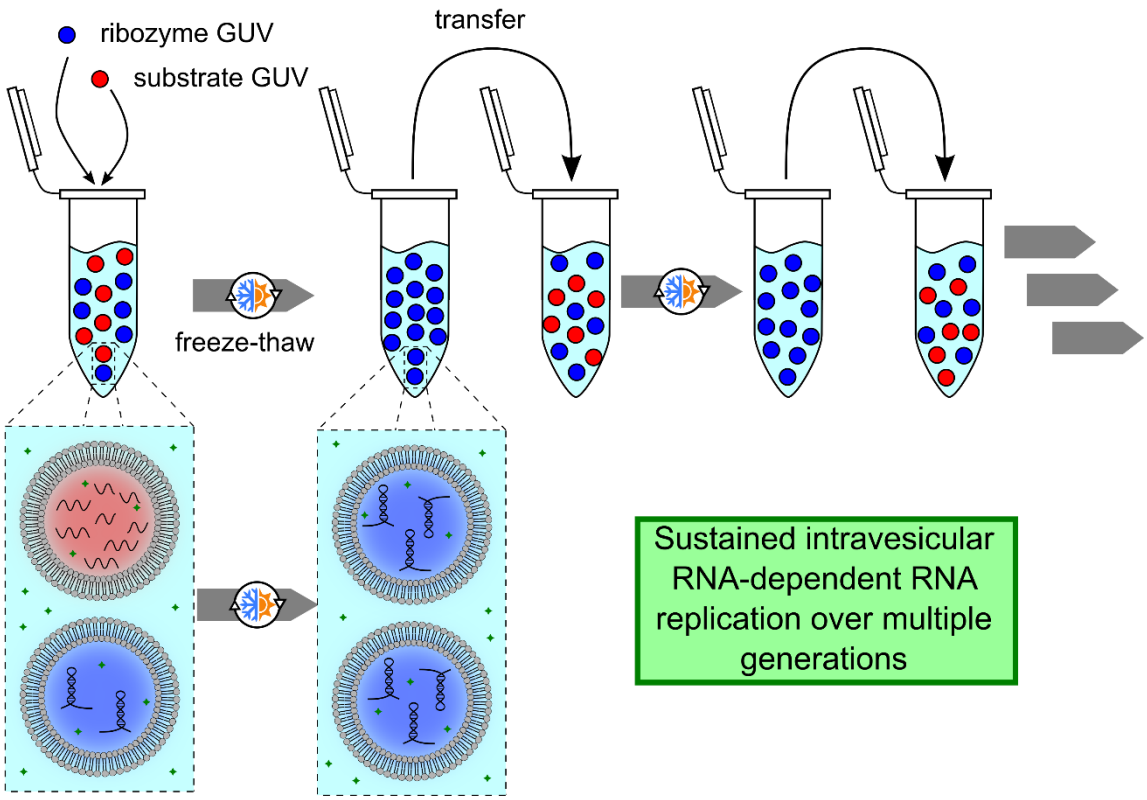


Figure 13: Sustained RNA replication in model protocells. *Schematic showing the experimental goal and expected outcome of serial transfer experiment of giant unilamellar vesicles (GUVs) encapsulating an autocatalytic RNA replicator. Guvs allowed to react are transferred to fresh substrate-containing vesicles and subjected to another round of freezing and thawing, thereby promoting content exchange and autocatalysis giving rise to a new generation of ribozyme vesicles.*

3.1.2. Inverted emulsion transfer optimization

Inverted emulsion transfer was selected as the method of choice for GUV production due to its relatively quick preparation and scalable vesicle production capacity **(Figure 14)**. The lipid of choice was 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine, 16:0-18:1 PC (POPC), since the phase transition temperature lies at -2 °C and thus reliably permits content exchange upon freezing and thawing (*158*).

There was a need to first scale and optimise the inverted emulsion transfer to enable the production of large quantities of giant unilamellar vesicles for subsequent manipulations. The parameters affecting vesicle preparation include carrier oil, which determines the lipid solubility and therefore lipid concentration, emulsion preparation method and ratio of inner phase to oil phase, which affects the size and number of inverse micelles generated, density gradient between inner and outer phase that affects centrifugation times and speed, interface surface area that determines how many vesicles can successfully form before the lipids are depleted from the interface, and finally, vesicle recovery step (*220*).

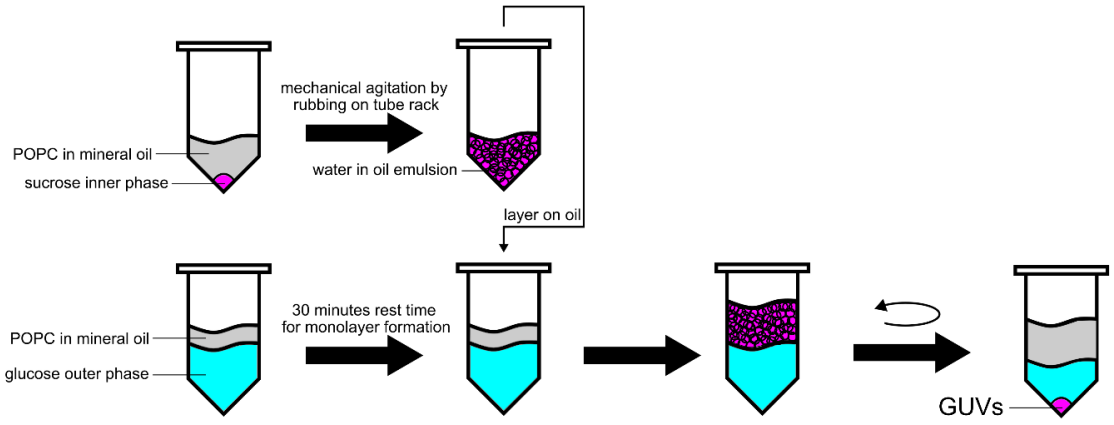


Figure 14: Illustration depicting the inverted emulsion transfer method. *Briefly, reverse micelles are prepared by combining lipids dissolved in a carrier oil with an aqueous inner phase solution in a 1.5 mL tube and vigorously rubbing it over a tube rack to create water-in-oil droplets. The resulting emulsion is layered onto an outer phase solution and centrifuged to push the reverse micelles through the lipid-water interphase, acquire a second leaflet and become vesicles.*

Of those, the ratio of inner phase to oil phase and centrifugation speed and time were systematically screened, while maximising the interface surface area by working in 1.5 mL tubes with an outer phase volume (600 µL) reaching above the inflection point of the tube. Initially, two carrier oils were tested, silicone oil and mineral oil. The latter was chosen due to its superior solvent capabilities despite having to work under a low humidity environment (<10% relative humidity) as mineral oil is hygroscopic and this seriously decreases vesicle yields at high humidity levels. At 400 µM POPC concentration, a clear, homogenous solution of lipids in mineral oil was obtained by first drying the lipids in glass tube under vacuum, resuspending them in an adequate volume of mineral oil, heating the mixture 5 minutes at 60 °C, vigorously vortex the solution until the lipids dissolve (~3-5 minutes) and finally sonicating for 1 hour to disperse any lipid aggregates. The density gradient required for centrifugation was realised by equimolar concentrations of sucrose and glucose, due to their inertness when combined with nucleic acids. A high concentration of sugars (900 mM) was achieved to obtain a steep density gradient, such that we could generate vesicles at low centrifugation speeds and times. Keeping these parameters constant, we varied the ratio of lipids in oil to inner phase when preparing the inverse emulsion from 35 (400 µL oil to 11.2 µL of inner phase) to 16 (400 µL oil to 24 µL inner phase) as well as the centrifugation speed (200-300 g) and time (2-5 minutes). The inner phase consisted of a calcein fluorescent reporter dye (Ex/Em = 495/515 nm) and pH buffered at 7.5 to visualize vesicle production via fluorescence microscopy and assess GUV

production capacity **(Figure 15)**. As expected, higher inner phase volumes and centrifugation times generally yielded more and higher quality GUVs. A higher inner phase volume shifted produced more and larger vesicles on average, as can be observed in the 24 µL inner phase images acquired. Conversely, further increasing inner phase volumes was unsuccessful in generating vesicles and rather resulted in micellar aggregates, while higher centrifugation speeds and time resulted in fewer vesicles of much larger diameter, suggesting that membrane fusion was taking place during centrifugation, probably driven by the centrifugal force. Hence, all subsequent preparations were performed with 24 µL inner phase volume and a 5-minute centrifugation at 300 g as these conditions produced the highest quantity and quality of GUVs.

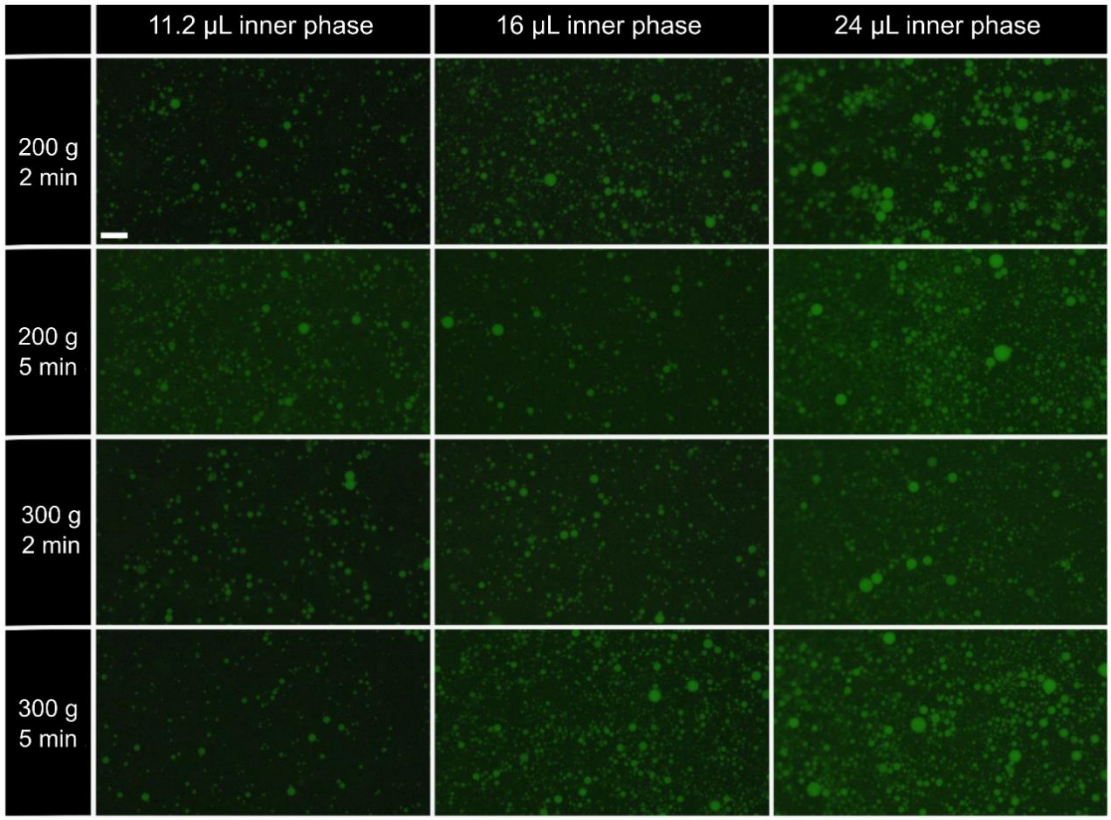


Figure 15: Parameter optimization for inverted emulsion transfer. *Maintaining lipid and sugar concentration constant at 400 µM and 900 mM, respectively, the ratio of lipids in oil to inner and the time and speed of centrifugation were screened by vesicle preparation and fluorescence microscopy. After production, 10 µL of vesicles were transferred to a passivated glass slide and cover slip and sealed with transparent nail polish. Images were acquired within 15 minutes of slide preparation using a dry 40x objective. Scale bar = 20 µm.*

3.1.3. Compatibility of GUVs and catalytic RNA

Next, a catalytic RNA was assayed under the conditions compatible with GUV production to assess activity. To this end, a minimal *trans*-acting version of the *hammerhead* ribozyme (HH-min) was adopted that, under standard *in vitro* conditions, catalyses the reversible cleavage of a substrate RNA strand to which it binds non-covalently by com-

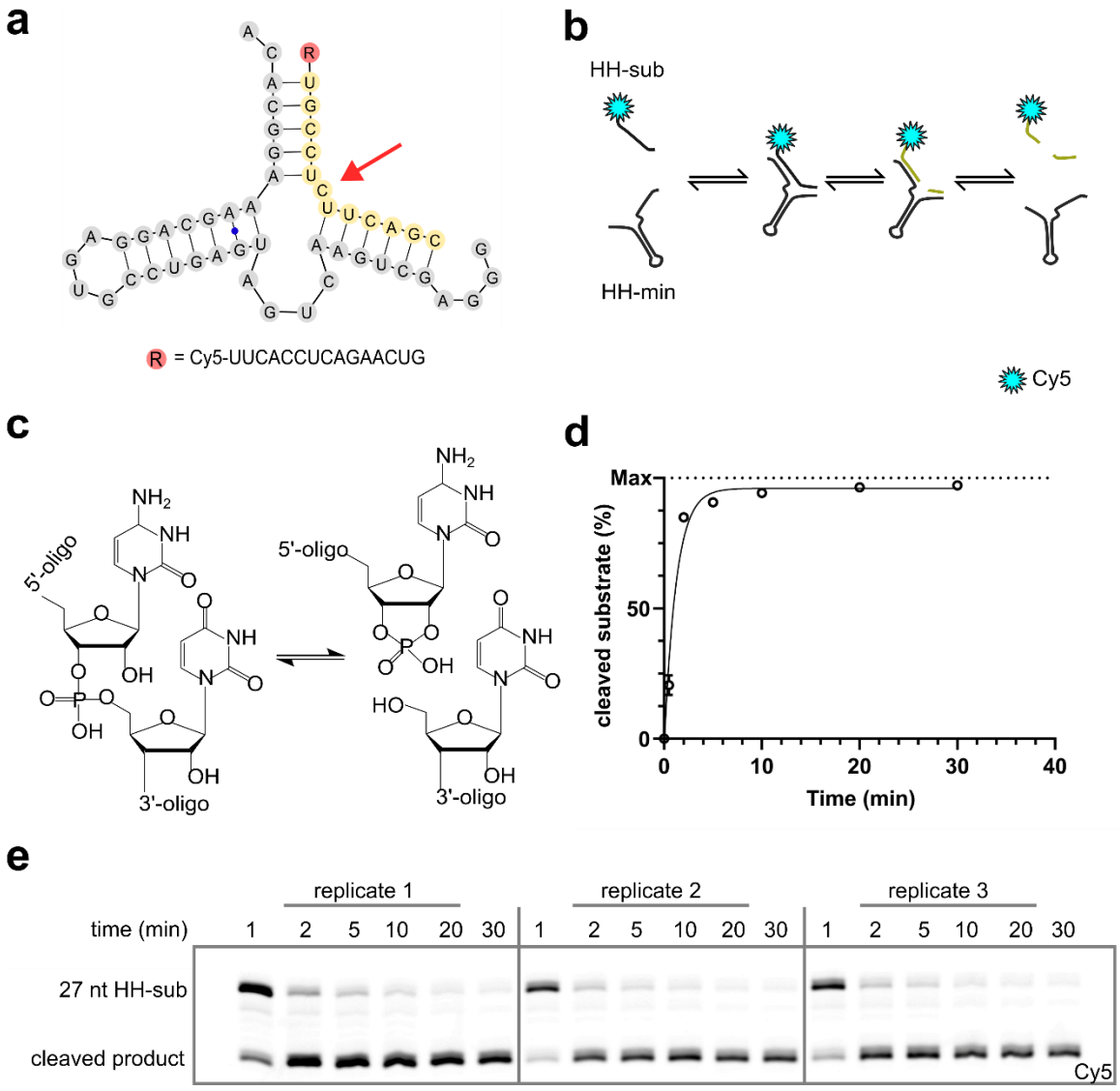


Figure 16: Hammerhead secondary structure and function under GUV-compatible conditions. ***a)*** *Secondary structure of minimal hammerhead ribozyme (HH-min; grey) bound to its cognate substrate (HH-sub; yellow). The red arrow indicates the substrate cleavage site. The catalytic core consists of the conserved CUGA central loop.* ***b)*** *The hammerhead ribozyme catalyses the reversible cleavage of a substrate strand.* ***c)*** *General mechanism of substrate cleavage site is shown. Deprotonation of 3'-hydroxyl leads to the rearrangement of phosphodiester bonds, forming a 2',3'-cyclic phosphate in place of 3',5'-phosphodiester bond, and cleavage of the substrate strand.* ***d)*** *Scatter plot showing the yield of cleaved hammerhead substrate over time. 5 µM of HH-min were incubated with 2.5 µM HH-sub in 20 mM Tris-HCl pH 8.3 and 4 mM $MgCl_2$ at 37 °C. Data are presented as mean ± standard deviation (n=3) and are fit with an exponential plateau constraining Y_0 to 0.* ***e)*** *20% denaturing PAGE analysis of the data presented in* ***d)***.

plementarity in the binding arms **(Figure 16a-b)** (*221*). The reaction proceeds via an acid-base catalysis: base-catalysed in-line nucleophilic attack deprotonates the 3'-hydroxyl at the scissile bond, followed by the nucleophilic attack of the 3'-hydroxide on the phosphate linker and synchronised acid-catalysed protonation of the 5'-oxygen on the downstream base, with a trigonal bipyramidal intermediated, leaving the 5'-oligo substrate with a 2',3'-cyclic monophosphate and the 3'-oligo substrate with a 5'-hydroxyl **(Figure 16c)** (*222*). Kinetic assays were performed under conditions expected to be compatible with GUV production and optimal cleavage activity of hammerhead ribozyme under single turnover conditions. RNA activity was analysed by electromobility shift assay of a fluorescently tagged substrate on denaturing polyacrylamide gels. At pH 8.3 and in the presence of 4 mM $MgCl_2$ and 900 mM sucrose, the ribozyme was able to cleave almost all available substrate within 10 minutes of the start of the reaction **(Figure 16d-e)**. The results obtained here indicate that HH-min is active under the relatively low magnesium concentration typically required for ribozyme catalysis *in vitro* (*179*). Moreover, since the cleavage activity observed in solution was not inhibited by the presence of high concentrations of sucrose (*223*), a similar activity was expected when the system would be encapsulated in GUVs.

3.1.4. Freeze-thaw enabled encapsulated ribozyme catalysis

To assess the ability of freeze-thaw (FT) cycles to drive the exchange of catalytic RNAs between independent vesicles without impairing ribozyme activity, HH-min or HH-sub were separately encapsulated in POPC-based GUVs and combined equally before being subjected to a cycle of freezing and thawing **(Figure 17a)**. The encapsulation process was not hampered by the addition of either RNA or the hammerhead buffer containing 4 mM $MgCl_2$, except for some elliptical shaped vesicles probably caused by osmotic or mechanical stress **(Figure 17b)**. Also apparent from the bright field image are light-diffracting bubbles that are attributed to residual oil / solvent from the vesicle production and recovery process. Intravesicular RNA activity was analysed by PAGE as before, but only after outer phase buffer exchange to eliminate any non-compartmentalised catalysis bias. When the vesicle populations each containing either ribozyme or substrate were combined, a substantial increase in cleaved substrate (approximately 80%) was observed after just one FT cycle and 1-hour incubation at 37 °C, suggesting that content exchange had taken place and that the ribozyme encountered the substrate and catalysed its cleavage **(Figure 17c-d).** In the absence of a FT cycle, approximately 20% cleaved substrate was detected, which might stem from membrane defects that occur during vesicle handling and centrifugation

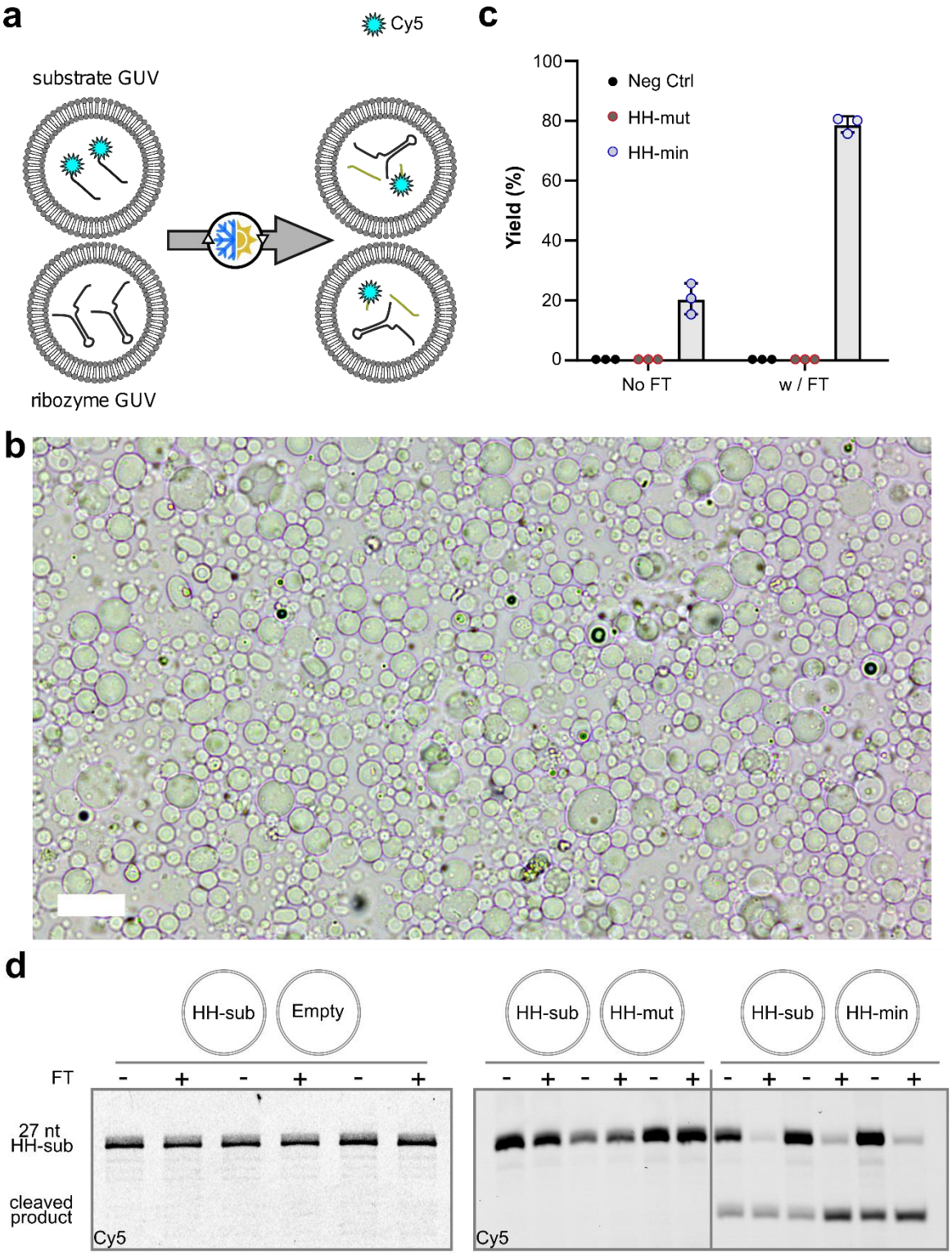


Figure 17: Freezing and thawing drives the exchange of catalytic RNA species between two vesicle populations thereby enabling catalysis. a) *Schematic of experimental setup and expected results.* ***b)*** *Bright field contrast adjusted image of combined vesicle populations directly after production. Scale bar = 20 µm.* ***c)*** *Bar graph showing cleavage yields of encapsulated HH-sub: GUVs encapsulating 2.5 µM HH-sub were combined 1:1 with GUVs encapsulating 5 µM HH-min or HH-mut, or empty vesicles for the negative control, and analysed before and after a FT cycle in 4 mM $MgCl_2$ at pH 8.3 followed by incubation period of 1 hour at 37 °C. Data are presented as mean ± standard deviation (n=3).* ***d)*** *20% denaturing PAGE analysis of the data presented in* ***c)****.*

steps. Confirmation of site-specific ribozyme-mediated cleavage was obtained by co-incubation of substrate containing vesicles with either empty or catalytically inactive ribozyme (HH-mut; G13A and G30A) before being subjected to identical conditions. The control experiments showed no detectable substrate cleavage in the absence of HH-min ribozyme.

Total yields between catalysis in solution and in vesicles were similar although not identical. Specifically, a lower maximum was reached when HH-min was compartmentalised as opposed to being free in solution, even after a prolonged incubation time of 1 hour. This effect can stem from various causes, the first of which would be the adhesion of RNA to the phospholipid membrane, thus reducing the pool of free and active RNA species. Another cause might stem from the loss of content induced by the FT cycle. The content of GUVs that burst is released into the outer phase and subsequently washed away before analysis, further decreasing the pool available ribozyme and substrate. A final cause of the incomplete catalysis in GUVs concerns the initial distribution of ribozyme and substrate vesicles prior to freezing and thawing. Content exchange in this case is stochastic and diffusion mediated. Although GUVs are expected to be equal in number upon mixing, it is unlikely that a perfect and thorough mixing is achieved. The result is a non-homogenous distribution of ribozyme GUVs, such that active ribozymes can only effect substrate cleavage within a certain volume around the GUV, which is determined by their rate of diffusion during thawing. In principle, if the latter cause is assumed to be true, total ribozyme activity should be rescued by subjecting the samples to further cycles of freezing and thawing. However, the problem encountered by supplemental FT cycles is content loss to the outer phase and equilibration of the sugar gradient necessary for centrifugation. In short, there seems to be a trade-off between homogenous exchange among vesicle populations and RNA catalysis analysis, which warrants further investigation into the molecular mechanisms occurring during thawing and content exchange.

To summarize, it was shown that the inhibition of catalysis by spatial separation of ribozyme and substrate in separate vesicle populations could be rescued by subjecting the GUV populations to a FT cycle. In other words, FT cycling enabled content exchange and concomitant ribozyme catalysis between two vesicle populations, validating the hypothesis that functional RNAs can be exchanged upon FT cycling and retain function. These results served as motivation to proceed with a more interesting ribozyme system, one of autocatalytic potential.

3.1.5. Establishing an autocatalytic RNA compatible with GUVs

A literature search identified an RNA ligase that functions as a self-replicator by ligating two substrates to produce a copy of itself. The R3C autocatalytic ligase F1, developed in the Joyce laboratory in 2014 (*91*), is the highly evolved product of several rounds of *in vitro* selection and directed evolution of the R3C ligase, which culminated in a highly efficient 66 nucleotide ligase ribozyme whose autocatalytic nature is enabled by its diagonal symmetry **(Figure 18a-b)**. The ribozyme catalyses the nucleophilic attack of the terminal 3'-hydroxyl of the A substrate to the alpha-phosphate of the triphosphate-activated B substrate, resulting in phosphodiester bond formation between A and B thereby producing another F1 copy and releasing pyrophosphate. Due to the presence of wobble base pairs and intervening bulges in the substrate-binding arms of the ribozyme, the resulting stability of the ribozyme-ribozyme complex post ligation is low, allowing product release and autocatalysis. Product inhibition is further mitigated by the high incubation temperatures employed for kinetic measurements, facilitating dissociation and exponential self-amplification.

The reported optimal conditions of the self-replicator were found to be rather harsh with respect to the success of the inverted emulsion transfer method for GUV preparation, as the ribozyme requires 25 mM $MgCl_2$ at pH 8.5 and an incubation temperature of 48 °C for optimal activity. A reduction in activity brought about by a decrease in both magnesium ions and temperature would not significantly reduce yields if given more time to react. Accordingly, kinetic assays were performed with the ribozyme under conditions posited to be suitable with efficient vesicle production, but without excessively sacrificing reaction rate. Keeping the pH level equivalent, $MgCl_2$ concentration and incubation temperature were decreased to 20 mM and 42 °C, respectively, and ligation activity was screened under multiple turnover conditions with tenfold excess substrates. When A (10% Cy5 labelled) and B substrates were seeded with F1 under the abovementioned conditions, high ligation activity both with and without sucrose present was observed, reaching a plateau of approximately 90% yields already within 30 minutes of incubation, confirming the hypothesis that ribozyme activity would not greatly suffer under sub-optimal conditions **(Figure 18c-d)**. Moreover, the close correlation of ribozyme kinetics between the samples with and without sucrose also corroborate the previous conclusions with the HH-min ribozyme on the inertness of sugars with respect to ribozyme catalysis.

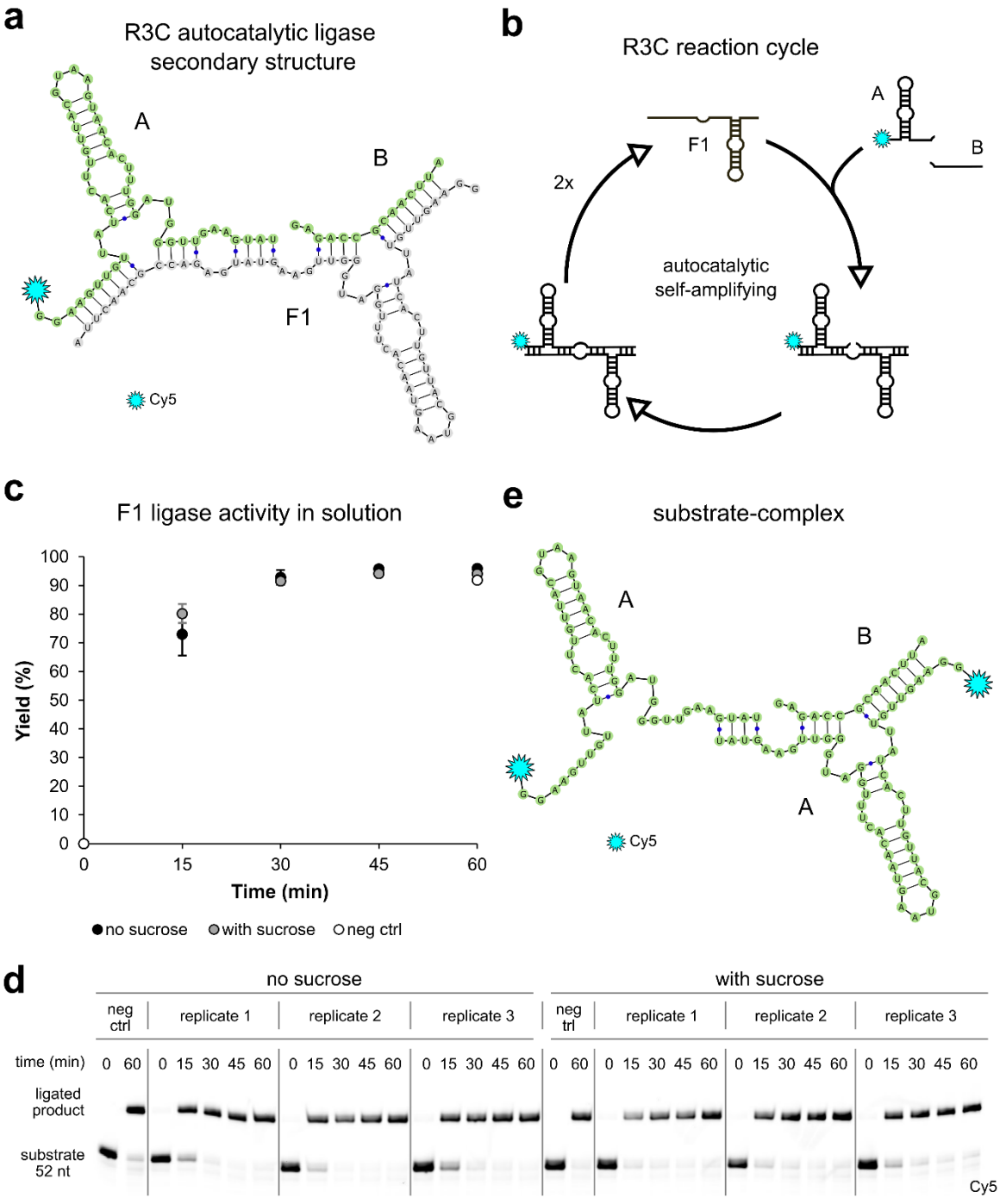


Figure 18: Establishing F1 ligation activity under conditions permissive to GUV production. a) *Secondary structure of the autocatalytic F1 ribozyme developed by Robertson and Joyce[176] complexed with its substrates A and B.* ***b)*** *Schematic of autocatalytic reaction cycle of the F1 ligase self-replicator.* ***c)*** *Bar graph showing ligation F1 ligation yields over the course of 1 hour: 5 µM A (10% labelled) and 12 µM B were incubated with 0.5 µM F1 ligase in 20 mM MgCl2, 50 mM EPPS pH 8.5, with or without 900 mM sucrose at 42 °C. Samples were quenched at the indicated timepoints and resolved on a denaturing PAGE. Negative control was not seeded with F1 ribozyme. Data are presented as mean ± standard deviation (n=3).* ***d)*** *12% denaturing PAGE analysis of the data presented in* ***c)****.* ***e)*** *Secondary structure of substrate-only complex A:B:A, predicted to be the reactive complex that allows substrate-mediated background ligation activity.*

Unexpectedly, the negative control reaction to which no F1 ribozyme was added whatsoever was positive for ligation after the 60-minute incubation period **(Figure 18c-d)**. The result suggests that a species in the reaction mix retains some background activity. Closer inspection of the substrates revealed that the longer A substrate is essentially a truncated version of the F1 ligase, preserving a weak binding affinity to both substrates, albeit with much lower affinity than F1, but nonetheless binds complementarily **(Figure 18e)**. Since the A substrate contains the active site of the ribozyme and is able to bind both substrates, it was assumed that the A substrate was partially catalytically active. To prevent substrate mediated ligation that might obscure F1-catalysed ligation, an inactive, fluorescently tagged A substrate was designed. Comparing sequences of the *in vitro* evolved F1 ligase to its parent sequence (E1 ribozyme), which was the initial starting point for randomisation and selection experiments, revealed that the F1 ribozyme, the most active of 31 clones of the SELEX experiment, differs by just six point substitution mutations from the E1 ribozyme. Conspicuously, only one was pervasive in all of the 31 sequenced clones, suggesting that this specific base mutation confers a great fitness advantage compared to the E1 wild-type sequence. Consequently, a reporter substrate carrying the G38A point mutation and a Cy5 tag at the 5'-end was designed and termed Cy5-A. On the other hand, the original A substrate was left as is, and hereafter referred to it as Hyper-A **(Figure 19a)**. Interestingly, the secondary structure shows that this substitution is affects a single base pair at the 3-way junction, altering a non-canonical Wobble base pair U•G to a canonical Watson-Crick U**-**A base pair. In contrast to U**-**A base pairs, U•G base pairs do not project the exocyclic primary amine that does not participate in hydrogen bonding with uracil into the major groove, leading to a more negatively charged and electrostatically dense major groove that potentially better accommodates divalent metal cations(*224*). This interpretation might explain the strong dependence of Hyper-A activity on the G•U wobble base pair and the inactivity of the substituted Cy5-A substrate.

Confirmation of inactivity of Cy5-A substrate and its ligation product Cy5-F1 was obtained by subjecting the reaction to serial dilutions in feedstock solutions containing the required reaction components but no functional enzyme. Rationally, after enough serial dilution of the original input ribozyme, if no ligation occurs, we could conclude that the ligated product is inactive as the active input ribozyme has been thoroughly diluted. Conversely, if no initial F1 ribozyme is seeded but the substrate and its cognate product are active, then serially diluting the reaction should have no effect on the yields as the active ribozyme is regenerated after every incubation, producing the characteristic sawtooth graph

of dilution and regeneration experiments. In fact, this is exactly what was observed when Cy5-A was used as the sole A substrate for F1 mediated ligation and was followed by a 10-fold dilution post-incubation into fresh substrate-containing solution. After just 2 dilutions (1:100 dilution of the 1 µM input F1 i.e., 10 nM of remaining input F1), no ligated product could be detected **(Figure 19b)**. The same is true for the generations that follow, confirming the inactivity of both Cy5-A substrate and its ligation product. In contrast, when Hyper-A (spiked with 10% Cy5-A) and B were used as the substrate pool, the opposite trend was

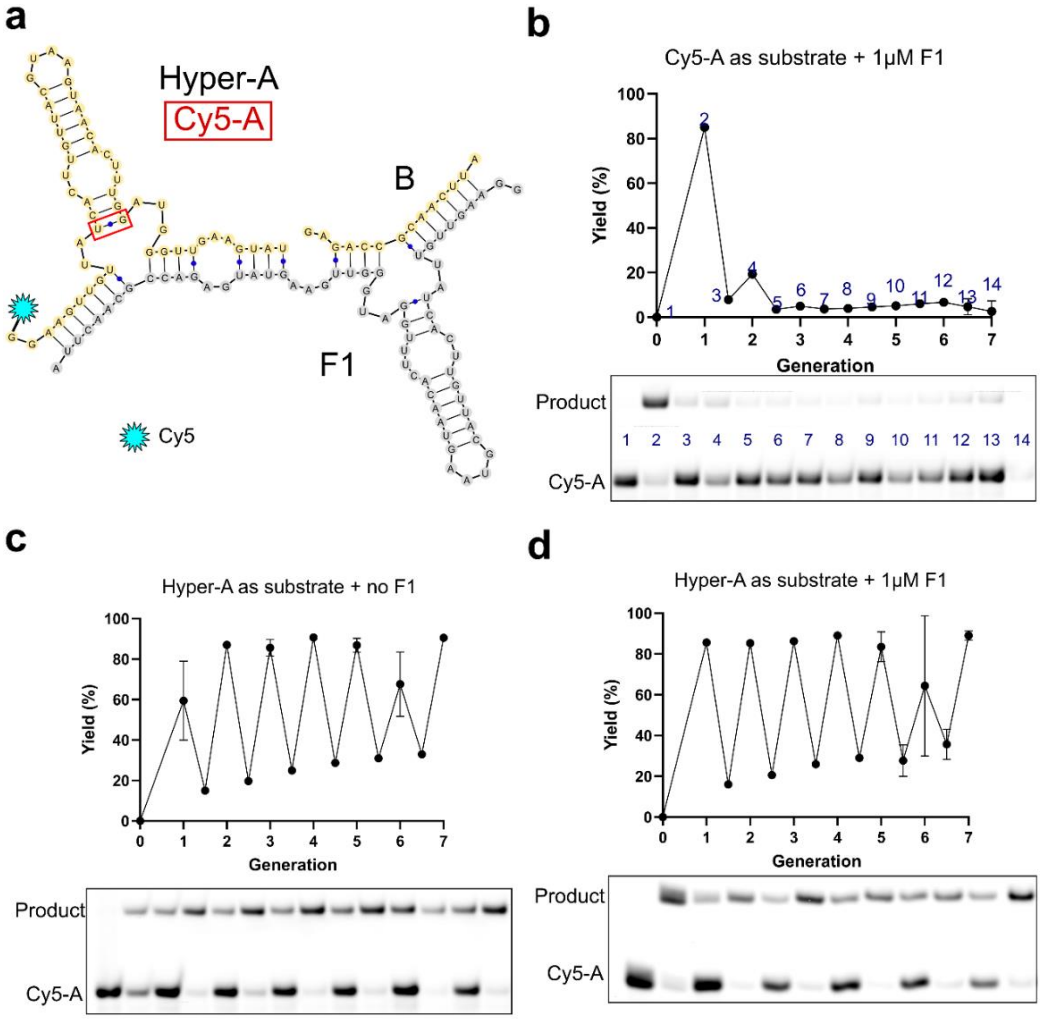


Figure 19: Design and validation of catalytically inactive reporter substrate Cy5-A. a) *Secondary structure of the original A substrate (Hyper-A; no modifications) and the designed inactive reporter substrate termed Cy5-A, which carries a Cy5 fluorophore at the 5'-end and an inactivating mutation (G38A; red box), substituting a U•G wobble base pair with a canonical U-A Watson-Crick base pair.* ***b-d)*** *Serial dilution in fresh reaction components of an initial reaction containing* ***b)*** *5 µM Cy5-A, 6 µM B and 1 µM F1,* ***c)*** *4.5 µM Hyper-A, 0.5 µM Cy5-A and 6 µM B,* ***d)*** *4.5 µM Hyper-A, 0.5 µM Cy5-A, 6 µM B and 1 µM F1. All reactions were carried out in 50 mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM sucrose. Each generation involves a 30-minute incubation at 42 °C before 1:10 dilution into fresh substrate containing solution. Samples were taken at t_0, after incubation and after dilution and resolved on a 12% denaturing PAGE. Data are presented as mean ± standard deviation (n=3).*

observed irrespective of whether F1 ribozyme had been seeded or not despite subjecting the systems to identical conditions **(Figure 19c, d)**. Together, the data indicate that Hyper-A is capable of background catalysis and that the system as a whole is indeed autocatalytic and can survive serial dilution in solution.

The above results provided a working system to characterise the activity of Hyper-A as a catalyst using the newly introduced inactive Cy5-A as a reporter substrate and compare it with F1 activity under the same conditions, to gain a better understanding of how to design the experimental setup for encapsulating the system. To this end, a simple multi-turnover ligation with varying Hyper-A or F1 concentrations was performed, using only Cy5-A as an inactive reporter substrate **(Figure 20a, b)**. Under the pre-established conditions for catalysis (specifically, pH 8.5 and 20 mM $MgCl_2$ at 42 °C) and without positive feedback or autocatalysis, the F1 ribozyme exhibited high activity at concentrations above or equal to 0.5 µM, with the reaction almost attaining total conversion of substrate within 60 minutes **(Figure 20c; F1).** Expectedly, lower F1 concentrations of 0.1 and 0.25 µM yielded less ligated substrate after 60 minutes of incubation, reaching approximately 40% and 70% conversion, respectively. In contrast, at sub-micromolar concentrations of Hyper-A, no product could be reliably detected after 60 minutes. For a concentration range from 1 to 2.5 µM inclusive with 0.5 µM increments, a proportional increase in yield from approximately 15% to 60% was observed **(Figure 20c; Hyper-A)**. Importantly, the negative control produced no product throughout the course of the reaction, corroborating previous data regarding substrate inactivity. To better visualize activity between F1 and Hyper-A, the yields after 60 minutes of incubation as a function of initial catalyst concentration was plotted **(Figure 20d)**. Here, the exponential increase in yield exhibited by F1 as opposed to a linear increase for Hyper-A yield was clear. Indeed, at 0.5-1 µM seed concentration, yields for F1 were 5- to 10-fold higher than for Hyper-A.

The data obtained highlight a number of interesting observations. First, the stark dependence of catalyst activity on a single G•U base pair suggests the direct involvement of these residues in the catalytic centre of the ribozyme when adopting its tertiary conformation. As mentioned before, the conformational features conferred to RNA helices by wobble base pairs is quite different from canonical base pairs, which likely provides a denser electronegative niche for metal ion coordination. The fact that the R3C ligase is actually a metallo-ribozyme, that is, it hosts a magnesium ion in its active site that is directly involved in catalysis, lends further support to the hypothesis that the metal ion binding in

the active site is coordinated by the G•U wobble base pair at the three-way junction. Second, by comparing yields of data points at t = 15 mins and [F1] = 1 µM in **Figures 18c** and **20c**, it is apparent that the highly active F1 species remains just as active under non-auto-

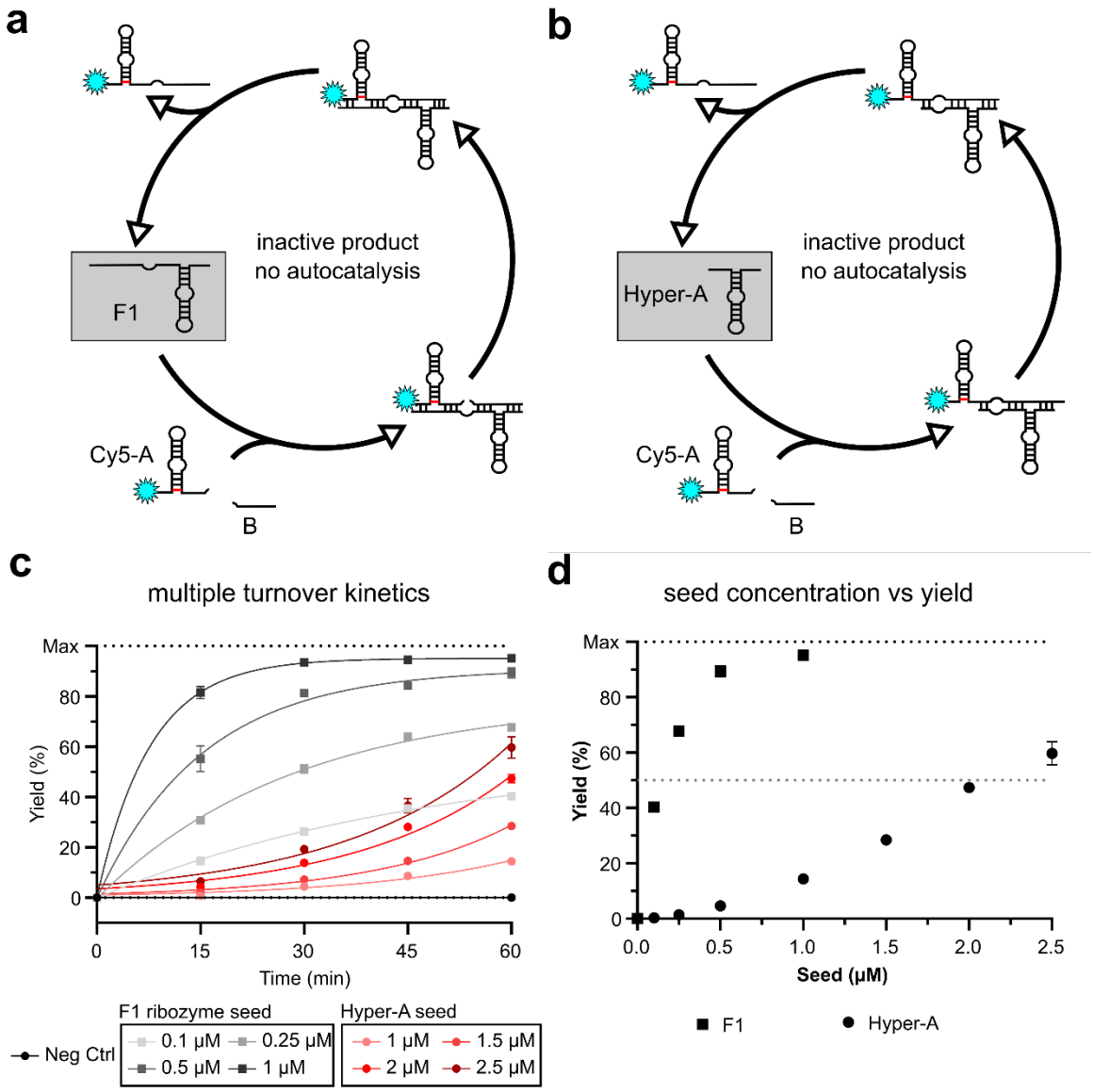


Figure 20: Comparison of F1 and Hyper-A activity in solution. *Schematic of reaction cycle for* ***a)*** *F1 and* ***b)*** *Hyper-A with inactive Cy5-A and B as the substrates.* ***c)*** *Plot showing kinetic measurements of varying seed concentrations of either F1 or Hyper-A with 5 µM Cy5-A and 6 µM B in 50 mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM sucrose at 42 °C. Negative control was not seeded with either catalytic species. Samples were quenched at the indicated timepoints and resolved on a 12% denaturing PAGE. Data are presented as mean ± standard deviation (n=3) and fitted with an exponential growth equation for descriptive purposes.* ***d)*** *Endpoint data from* ***c)*** *plotted as a function of seed concentration of F1 or Hyper-A. Data are presented as mean ± standard deviation (n=3).*

catalytic conditions, in both cases reaching 80% substrate conversion. Hence, although autocatalysis is a necessary requirement for serial transfer / dilution the reaction as was demonstrated in Figure 17, the rate of ligation *per se* is not directly dependent on the exponentially increasing concentration of ribozyme, at least for starting concentrations in the low micromolar range. Finally, in hindsight, the catalytic activity of Hyper-A no longer appears to be surprising given that it still holds both the catalytic centre of the F1 ribozyme and, to a lesser degree, the substrate binding arms.

It would be interesting to understand whether the decrease in rate accompanied by the deletion of the 3'-end substrate binding site in Hyper-A is related to the decrease in substrate binding capacity with Cy5-A. The predicted ΔG for the complementary bases between F1 and Cy5-A binding (substrate binding arms) using Vienna RNAcofold (*225–227*) webserver was -17.05 kcal/mol, whereas that of Hyper-A and Cy5-A binding was -0.14 kcal/mol. The relative changes between activity and binding obviously do not correlate linearly i.e., a ~10-fold reduction in rate and ~100-fold reduction in binding energy, suggesting that the complete active site and 3'-terminal of the ribozyme also contribute to proper folding and stability. Nonetheless, having characterised a ribozyme reporter system that self-amplifies in conditions amenable to GUV production, the next step was to test whether autocatalysis can be rescued by freezing and thawing, similar to what was done with the hammerhead ribozyme.

3.1.6. Intra-vesicular RNA replication enabled by freezing and thawing

The reliance on the environment to drive inter-protocellular exchange of the genetic material greatly simplifies the process of proliferation, as no exceedingly complex process must be invoked, one that was very likely unavailable to such early protocells. A hypothesis was made that freeze-thaw (FT) cycles enable the horizontal transfer of the autocatalytic replicating RNAs and thus, their replication, mimicking the proliferation of RNA sequences in protocellular populations, a fundamental process for early cellular life. To this end, encapsulated F1 self-replication was probed under varying numbers of cycles and with a supply of substrate containing GUVs. To explore if the F1 replicator could propagate in GUV populations by infecting substrate-containing vesicles during freezing and thawing, the autocatalytic reporter system was employed with both autocatalytic and reporter product cycles **(Figure 21a)**. The inverted emulsion transfer was successful despite the presence of a higher magnesium concentration in the buffer as opposed to the hammerhead buffer **(Figure 21b)**. Although traces of left-over solvent were observed, since it was

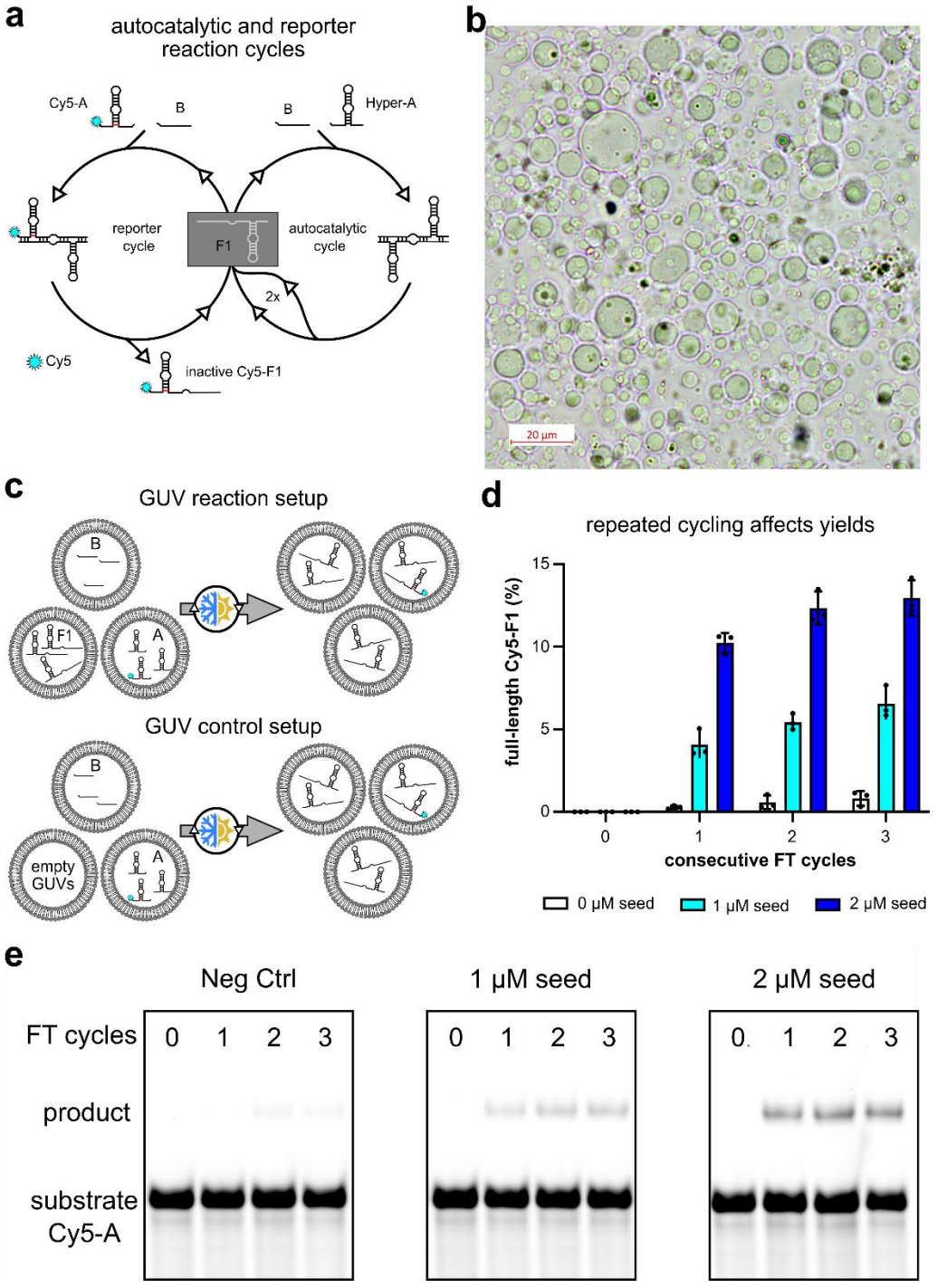


Figure 21: Temperature-driven content exchange and autocatalysis of GUV-encapsulated RNA. ***a)*** *Schematic of double reaction cycle implemented for compartmentalised catalysis.* ***b)*** *Bright field contrast-adjusted image of combined vesicle populations directly after vesicle production. Scale bar = 20 µm.* ***c)*** *Schematic representation of the. GUV components and experimental setup.* ***d)*** *Bar graph showing product yields when varying both initial seed concentration and number of freeze-thaw (FT) cycles. Substrate GUVs contained either 10 µM Hyper-A and 2.5 µM Cy5-A or 15 µM B. Ribozyme GUVs contained either 0, 1 or 2 µM F1, as indicated. Buffer consisted of 50mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM sucrose. One freeze-thaw cycle consisted of fast freezing at -80 °C followed by passive thawing at room temperature. FT-cycling was followed by a 60-minute incubation at 42 °C. Data are presented as mean ± standard deviation (n=3).* ***e)*** *12% PAGE of the reactions data shown in* ***d)****.*

not problematic beforehand, the overall production process was deemed functional. Three GUV populations were separately prepared, containing either 0-1-2 µM F1 ribozyme, 10 µM Hyper-A and 2.5 µM Cy5-A or 15 µM B in 50 mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM sucrose / glucose for inner /outer phase, respectively. The selected Hyper-A concentration was estimated to provide enough substrate to active product conversion despite concurrent loss brought about by FT-cycling. The A and B substrates were separately encapsulated and the final vesicle population was comprised from a 1:1:1 mixture to prevent premature Hyper-A ligation activity within vesicles before freezing and thawing **(Figure 21c)**. As expected, co-incubation of the three populations for 1 hour at 42 °C prior to freeze-thawing resulted in no detectable ligation, irrespective of whether F1-containing GUVs were present or not.

On the other hand, subjecting the GUVs to multiple successive FT cycles resulted in a linear increase in the levels of Cy5-F1 product **(Figure 21d-e)**. GUVs containing 1 µM F1 showed the formation of approximately 4% after one FT cycle, reaching a maximum of 6% after three successive FT cycles. Starting with 2 µM encapsulated F1, an increase in product yield to 10% after one FT cycle and 13% after three consecutive cycles was observed. When empty vesicles were provided instead of ribozyme containing vesicles, less than 1% of full-length Cy5-F1 was detected, which was likely catalysed by the background activity of Hyper-A. The apparent correlation between ligation yields and initial encapsulated F1 concentrations indicates that initial concentration is a strong determinant of the eventual product yields despite the equivalent supply of GUVs. Similarly, the correlation between ligation yields and number of FT cycles suggests that increased cycling results in improved content exchange which, in turn, permits a higher fraction of substrate to react notwithstanding the concurrent loss of content that occurs during freezing and thawing.

In efforts to rule out any activity not mediated by the F1 ligase, various control experiments were undertaken aimed at elucidating the effect of both FT-cycling and the presence of ordered membranes on ribozyme activity in solution. Consequently, the substrates were first subjected to varying number of FT cycles with or without F1 ribozyme in solution, and subsequently incubated at 42 °C **(Figure 22a)**. When the reactions were only cycled between -80 °C and room temperature without an incubation at 42 °C, only a slight increase in product formation was measured, reaching a final yield of approximately 20% for both reactions after three consecutive FT cycles **(Figure 22b)**. This suggests that some ligation was occurring, probably during passive thawing when the solution passes through a eutectic ice phase. It follows that product yields should scale proportionally to the time

the system spends in eutectic phase, which coincided with the observations upon increased cycling. On the other hand, subjecting the reaction to a 42 °C incubation period boosted yields to approximately 60% for all cases, suggesting that slightly elevated temperatures are required for efficient catalysis. The data indicate that ribozyme catalysis occurs only under unfrozen conditions at elevated temperatures, and refute the idea that ribozyme catalysis had occurred during freezing or thawing.

Secondly, to investigated the effect of ordered membrane lipids on ribozyme catalysis, F1 ligation activity was screened under single turnover conditions (2-fold excess Cy5-A substrate; 4-fold excess B substrate) in bulk solution and with increasing volumetric fractions of independently prepared empty GUVs **(Figure 22c)**. Generally, the effects of RNA-lipid interactions remain somewhat ambiguous, owing to the large repertoire of diverse lipid molecules and combination of lipids available for membrane assembly. Nonetheless, some factors have been shown to play a role in modulating the interactions, such as sequence composition and length, hybridization state and membrane phase and charge (*228*). RNA-lipid interactions can therefore be deduced *a posteriori* from first principles, but they remain difficult to predict *a priori,* in part due to the lack of a sufficient amount of data describing the possible interactions among different RNAs and lipid membranes. Perhaps surprisingly, we noted a slight enhancement in the product yields between catalysis in bulk solution with no GUVs on the one hand and catalysis in the presence of either 0.25 or 0.5 v/v GUVs on the other **(Figure 22d)**. After 60 minutes of incubation at 42 °C, the bulk reaction showed yields varying from approximately 10% to 33%, while the reaction with 25% of its volume supplemented with GUVs (1:4 GUV dilution) showed a mean product yield of approximately 38%. Increasing the GUV volume fraction to 0.5 led to a 2% increase in yield, reaching and slightly surpassing 40%. These results clearly demonstrate that the F1 ligase indeed does remain active in the presence of ordered lipid membranes, at times even benefiting from slight rate enhancements. The rate increase might stem from concentration effects brought about by the zwitterionic head groups of the lipid, namely, the negatively charged phosphate and the positively charged choline moiety. Since the choline moiety is terminal, it could in principle attract the negatively charged RNA backbone, effectively increasing RNA concentration near the membrane. However, more data is needed to better elucidate the nature of the interaction in this particular system and draw incontrovertible conclusions.

These above experiments support the claim that the ligation product observed during the FT-cycling of mixed GUV populations is strictly catalysed by the F1 ribozyme

under encapsulated aqueous conditions at elevated temperatures. This effectively decouples the transfer genetic information, which in this case is RNA diffusion during freezing and thawing, and replication of the genetic material, which again here is the self-replication of F1. The results obtained here were used as a starting point employ the system to mimic a life-like behaviour that draws analogy from virology, as viruses are relatively simple in relation to extant life and a good example of species teetering on the edge of living and dead.

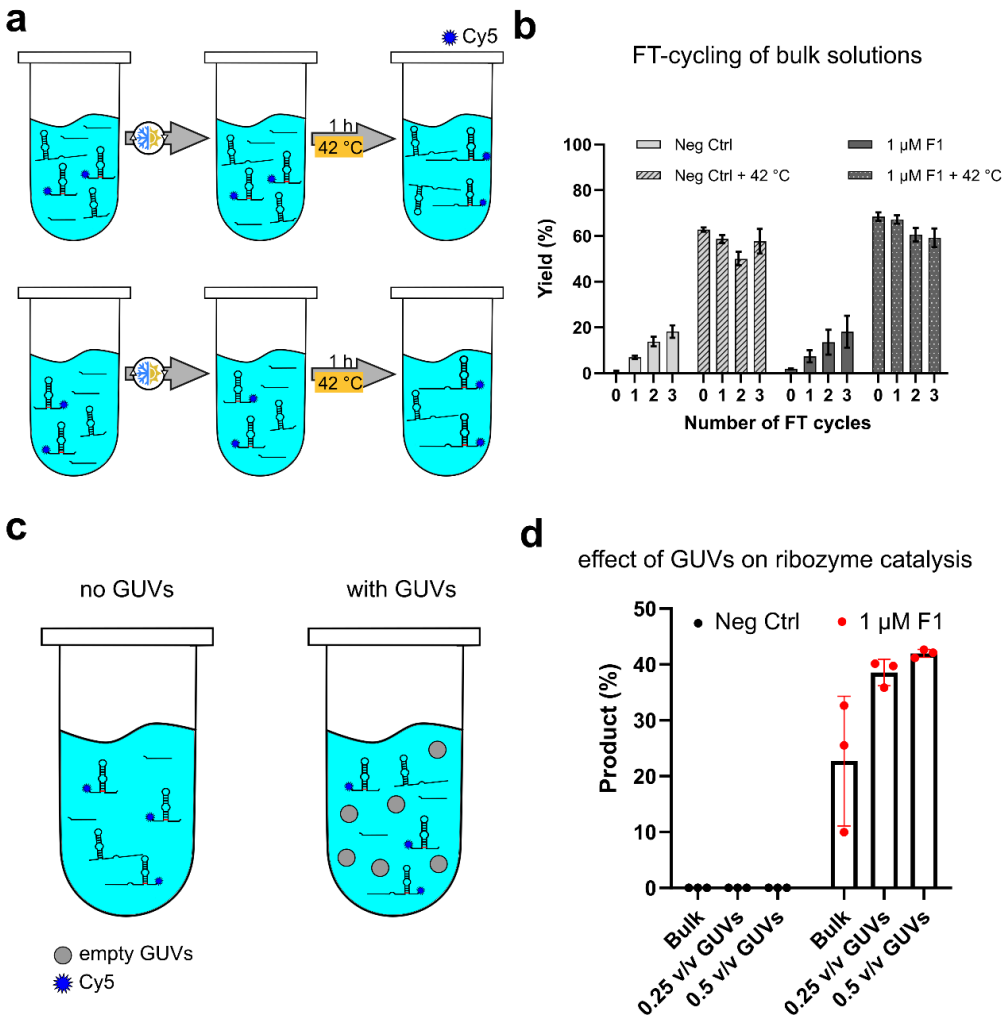


Figure 22: Modulation of R3C ligation activity by temperature and ordered lipids. a) *Schematic depicting the experimental setup to test the activity of the autocatalytic system under cycling conditions.* ***b)*** *Bar graph showing ligation yields obtained from reactions containing 10 µM Hyper-A (1 µM Cy5-A) and 15 µM B, either seeded with 1 µM F1 or not, and varying the number of FT cycles and a 42 °C incubation post-cycling in 50 mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM sucrose. Data are presented as mean ± standard deviation (n=3).* ***c)*** *Schematic depicting the experimental setup to test for modulation of activity by vesicles in solution. d) Bar graph showing product yields from 0.5 µM Cy5-A and 1 µM B, with or without (Neg Ctrl) a 1 µM F1 seed and an increasing volumetric fraction of GUVs. Reactions were carried out in 50 mM EPPS pH 8.5, 20 mM $MgCl_2$ and 900 mM glucose and incubated at 42 °C for 60 minutes. Data are presented as mean ± standard deviation (n=3).*

Having identified the optimal conditions for encapsulated ribozyme catalysis (10-fold excess substrate, three FT cycles followed by a 60-minute incubation at 42 °C), a very basic selective pressure that an early autocatalytic system can succumb to was implemented, namely, death by dilution. The selective pressure can be overcome only by a sufficient amount of autocatalysis (*159*), otherwise functional replicators are quickly washed away and lost. The ability of encapsulated F1 self-replicator to persist in GUV populations during serial transfer experiments by escaping the selective pressure of dilution in an open system through the repeated infection of substrate-containing GUVs driven by periodic temperature changes was investigated. To this end, GUVs encapsulating an initial functional F1 replicator at 1 or 2 µM (F1 seed) were combined with GUVs encapsulating either 15 µM B or 10 µM A (supplemented with 2.5 µM Cy5-A) in a 1:1:1 ratio. To capitalise on content exchange, the tripartite GUV population was subjected to three consecutive FT cycles and sampled before (generation 0) and after (generation 1: G_1) a 60-minute incubation period at 42 °C. This was followed by equally combining half of the resulting GUV population with fresh substrate-containing feed vesicles and the whole cycle of freeze-thaw and incubation was repeated for several generations, but without any additional F1 seed **(Figure 23a)**. The supply of fresh substrate GUVs, or feeding, was expected to provide sufficient substrates for self-replication during the next generation. For 1 µM F1 seed, approximately 5% full-length Cy5-F1 product was observed after first generation **(Figure 23b-c)**. The ligation yields endured somewhat stably for the following 6 generations, indicating that F1 autocatalysis was sufficient to persist at low levels notwithstanding ongoing dilution and content loss brought about by consecutive cycles of freezing and thawing. It was hypothesised that a higher initial seed concentration would allow for improved F1 self-replication similar to what was previously observed, and, to this end, the experiment was repeated with an increased initial F1 seed concentration of 2 µM. Here, the ligation yields after generation 1 were already at approximately 10%, which reached a stable plateau in the following 6 generations at 12%. Ligation mediated by the Hyper-A substrate present in substantial amounts was excluded as control samples with empty GUVs (0 µM F1) showed background levels of ligation that readily plateaued below 1.8% throughout all sampled generations. By measuring total fluorescence intensity per lane of the resolved PAGE, content loss following three repeated FT cycles was calculated to be approximately 45%. Thus, taking the relative concentrations of substrate added initially and after each generation, the molar concentration of newly synthesised F1 RNA could be estimated. By subtracting 45% from 4.0 / 4.2 µM (for a 1:4 and 1:3 dilution, respectively) from a total of 12.5 µM A sub-

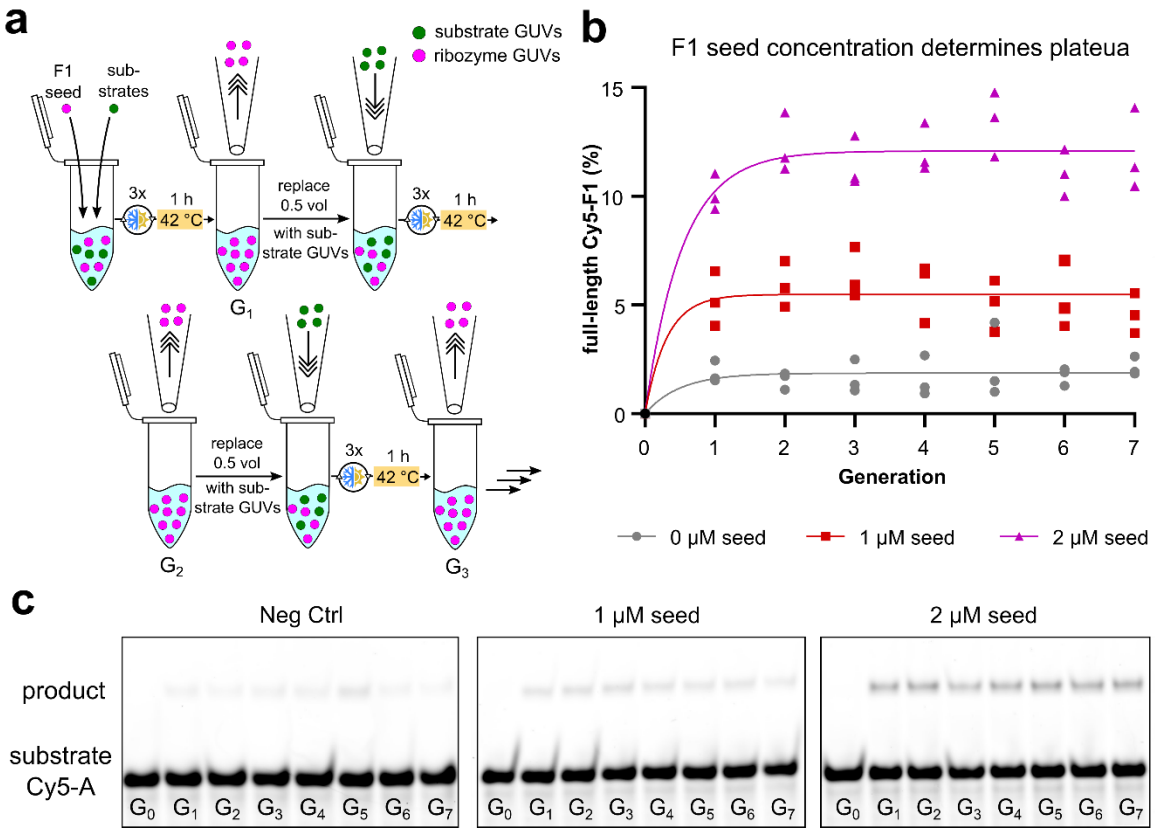


Figure 23: Serial dilution of GUV-encapsulated self-replicating F1. a) *Schematic representation of the experimental workflow for the serial transfer of GUVs encapsulating autocatalytic RNAs.* ***b)*** *Ligation yields of Cy5-labelled F1 after each generation (G_n). GUVs containing either 0, 1 or 2 µM unlabelled F1 ribozyme as initial seed, 10 µM Hyper-A with 2.5 µM Cy5-A, or 15 µM B were combined in 1:1:1 ratio as shown in* ***Fig 18c****. PAGE samples were taken before (G_0) and after (G_1) FT-induced content mixing and self-replication (42 °C for 1 h). In all further generations, PAGE samples were taken after 1:1 serial dilution of the previous generation with fresh substrate GUVs followed by three FT-7cycles and the 1 h incubation at 42 °C as shown in* ***a****. Individual data points are shown (n=3) and are fitted with a single exponential for descriptive purposes.* ***c)*** *12% PAGE of the data shown in* ***b)****.*

strate, then 2.2 / 2.3 µM substrate was available for ligation. This points to an estimated F1 concentration of 0.22 / 0.23 µM after generation 1 for the sample seeded with 2 µM F1, readily levelling at approximately 0.3 µM of *de novo* total produced F1 in the following generations. For samples seeded with 1 µM F1-containing GUVs, *de novo* synthesised F1 stabilised at approximately 0.11 µM, while the control with empty GUVs remained at approximately 0.04 µM throughout. Therefore, the apparent main determinant for the pseudo-steady state persistence of F1 ligation levels observed during serial transfer was the initial F1 seed concentration. These findings suggest that the FT-cycling and transfer protocol resulted in sub-saturating concentrations of F1 and a quasi-linear dependence of ligation yields on initial F1 seed, similar to what we observed for sub-saturating conditions of the F1 catalysis in bulk solution **(Figure 20c-d)**. Altogether, the experimental model system

allowed us to generate a persistently infective encapsulated RNA self-replicator capable of sustained survival over multiple generations of successive dilutions via a horizontal RNA transfer among vesicle populations driven by recurring oscillations of environmental conditions.

3.1.7. Summary and conclusions

In this section, vesicle-encapsulated catalytic RNAs, especially autocatalytic replicators, were examined to assess whether such systems can benefit from temperature induced defects in their vesicle membranes by enabling their propagation within the available vesicle populations. First, optimized the conditions for vesicle preparation were determined and the activity of a simple nucleolytic ribozyme (HH-min) under conditions that were compatible with production was screened. Then, encapsulated ribozyme activity was triggered by enabling the exchange of RNA between vesicle populations encapsulating either ribozyme or substrate, thereby demonstrating the feasibility of the system to indirectly induce ribozyme activity. Next, an autocatalytic self-replicating system was adopted and characterised *in vitro* to confirm its activity under conditions amicable to GUV production and stability. Finally, the autocatalytic system was shown to be capable of sustained intravesicular amplification within vesicle populations during serial transfers driven by FT cycling when combined with feeding of substrate-filled GUVs, notwithstanding the concurrent loss and dilution.

Based on earlier investigations (*158*), periodic FT cycles were chosen as the drivers of content exchange in an attempt to mimic processes and phenomena that might have occurred on the early Earth. In the current system, high vesicle densities were employed when conducting FT cycles to minimise content loss and maximise content exchange. In an open world scenario, this necessitates a mechanism that supports a flux of vesicles and promotes their continuous enrichment. A suitable environment could have been surface-exposed freshwater streams flowing through microporous rocks that act as vesicle traps in which GUVs are accumulated before being exposed to FT cycles. The porous rocks would act much in the same way as a microfluidic device that has been engineered to contain pockets that successfully trap and accumulate GUVs (*229*). Since vesicle accumulation was observed under controlled laboratory conditions, it lends credence that such a phenomenon could have occurred under environmental conditions on early Earth. Supporting data could be obtained by employing a natural porous rock sample rather than a synthetic polydimethylsiloxane (PDMS) microfluidic chip for vesicle trapping.

On a similar note, a limitation of this study is that the timescales and temperature gradients adopted clearly diverge from what one expects in a natural environment. In this setup, cooling to -80 °C ensured homogeneous freezing and prevented supercooled solutions, while by passive thawing at room temperature allowed the regeneration of bilayer membranes (approximately 100 °C difference). Future experiments might therefore attempt to build a natural environmental mimic to empirically test this hypothesis under more plausible environmental conditions that better emulate the current extrapolated conditions on early Earth. Moreover, since content exchange relies on the membrane instability during the phase transition temperature of lipids, a system capable of content exchange at temperatures remaining above the freezing point of water could potentially be designed by utilising phospholipids with higher melting temperatures. For example, 1,2-dimyristoyl-sn-glycero-3-phosphocholine, DMPC, possesses a phase transition temperature of 24 °C (*230*). Different temperature regimes can also lead to the same desired effect of content exchange or genetic reshuffling among protocells, as thermal cycling between 25-95 °C was shown to reversibly drive fatty acid phase transitions resulting in release and re-encapsulation of mixed contents, rescuing fluorescence of a split Broccoli aptamer (*231*).

The findings reported here support the argument that temperature fluctuations, including those involving eutectic ice formation, provide a suitable environment to drive a variety of reactions relevant to the emergence and evolution of life, potentially supporting a continuous path from the prebiotic chemical synthesis of monomers, their subsequent activation and non-enzymatic polymerisation to the point where they allow tertiary folds and associated catalysis including, but not limited to, autocatalytic replication. The simultaneous synthesis of nucleic acid precursors and phospholipids in an environment supporting the cyclical freezing and thawing of water could have resulted in compartment formation and encapsulation of molecules (*188*), while exchanging resources with the outer environment and other protocells by means of the temperature-induced membrane permeabilization.

3.2. Coacervates as cradles for the co-evolution of RNA and peptides

3.2.1. Background

Complex coacervates, driven by electrostatic liquid-liquid phase separation, could potentially provide the benefits of a compartment for the genetic information that it holds. However, this depends on how well the genetic information is retained and how stable and robust the compartment is to external pressures. The first difficulty can be addressed by imposing the genetic information as a necessary component of the coacervate compartment. In this way, a compartment would not form unless it contains genetic information that can be in the form of nucleic acids. This compartment design allows a nucleic acid polymer to act as both the polyanion required for phase separation and the genetic information contained within. While this type of genotype-phenotype linkage is easily implemented *in vitro*, it is still far from what a true linkage is, namely, one that provides natural selection with the ability to act at higher levels of complexity rather than directly on the genetic information. In other words, natural selection favours phenotypes that are themselves mainly determined by the genotype. The linkage between genotype and phenotype is therefore a dynamic one and a good target to establish a nucleic acid-peptide based coacervate system that would be characterised by a conspicuous effect of the genotype on the compartment phenotype and *vice versa* **(Figure 24)**.

Although compartment phenotype has already been shown to undergo dynamic changes in response to the action of proteinaceous enzymes hosted within, compartments with altered phenotypes in response to nucleic acid catalysis have not been previously reported, in part due to the limited catalytic repertoire of these catalysts as opposed to their proteinaceous counterparts (*232*). Beyond environmental factors such as solution pH and

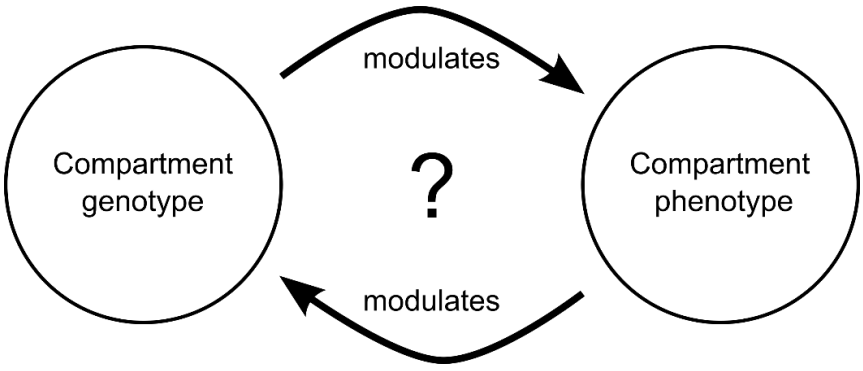


Figure 24: Genotype-phenotype linkage in artificial protocell models. *An important aspect to enable Darwinian evolution of adaptive complexity is the linking of genotype and phenotype such that genotypic variants that manifest phenotypic fitness advantages can be favoured by natural selection.*

salt concentration (*192*), the physical properties and association behaviour of coacervates are determined by internal parameters such as poly-ion chain length and charge density (*233*, *234*). Poly-anion chain length could potentially be addressed by the action of a nucleolytic or ligase ribozyme, if RNA was the coacervate poly-anion component. It follows that in a coacervate composed of RNA poly-anions and a poly-cation, for instance short poly-lysine oligopeptides, ribozyme catalysed RNA cleavage or ligation might alter the physical properties or association behaviour of the system, given that a sufficiently large change in average RNA length within the compartment is achieved. In particular, a hypothesis was devised that compartmentalisation would enhance ribozyme ligation activity, leading to the elongation of the RNA component, which in turn causes the formation coacervates with altered physical properties **(Figure 25)**. To this end, a ribozyme capable of concatenating RNA and significantly increase the average RNA length in a coacervate system was developed. Once established, ribozyme activity was assessed in solution and in RNA-peptide coacervates while material properties of coacervate systems with encapsulated ribozyme were investigated (*235*).

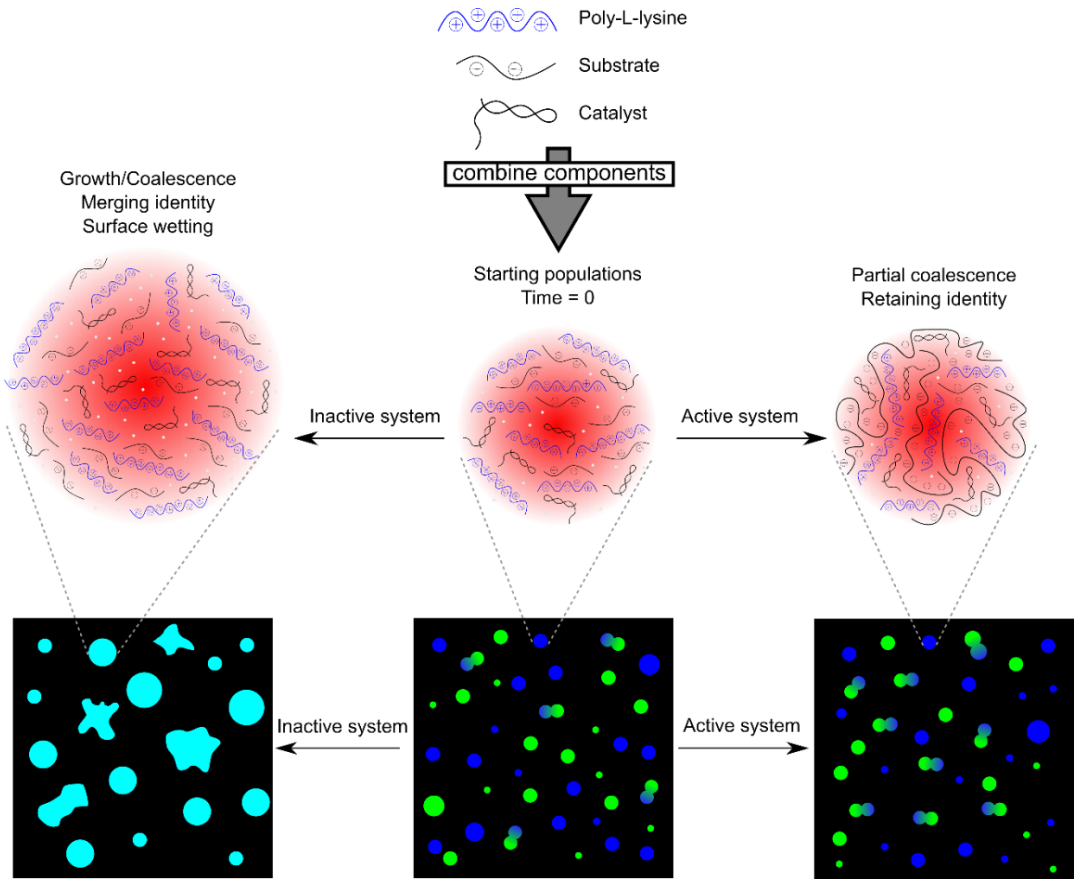


Figure 25: Experimental design investigating how ribozyme activity potentially modulates coacervate properties. *Schematic showing the experimental goal and expected outcome of coacervates composed of catalytically active versus inactive RNA and short poly-L-lysine oligopeptides* (*235*).

3.2.2. Design and characterisation of laddering ribozyme

Although ribozyme systems that concatenate a short substrate have been previously reported, these do so via a reversible cyclic-phosphate mediated mechanism leading to a rapid decrease in the concentration of higher length products (*85*). The R3C autocatalytic ligase ribozyme (F1) used in the previous section was thus an attractive starting point for the development of a laddering system, since the 5'-triphosphate chemistry required by the ligase is irreversible unless hydrolysed and the ribozyme is already optimised for high rate and yield (*91*). Recall that the F1 autocatalytic ribozyme ligates substrate A and B to form another F1 ribozyme. To engineer a laddering system, first, three nucleotides (GGA) from the 5'-end of A were deleted and the B substrate was appended to the 5'-end. Secondly, the stem containing the catalytic loop was deleted and the fragments connected, resulting in a 31-nucleotide substrate. Finally, to reduce substrate homodimer formation and intramolecular folding thereby promoting concatenation by forming functional substrate:ribozyme complexes, five point mutations were introduced to substrate strand (G16C, U18G, G19C, G22U, U24G), the corresponding five bases on the ribozyme were mutated accordingly and the final base was deleted (G55U, C57A, C60G, A61C, C63G, 66del), yielding the 31-nucleotide substrate S1 and the 65-nucleotide laddering ribozyme v2F **(Figure 26a).** The newly designed v2F ribozyme binds and ligates two S1 substrates together, resulting in a product pool of increasing length depending on the number of ligations catalysed **(Figure 26b)**.

Initial screening of reaction conditions was performed in 10 mM $MgCl_2$ at pH 8.6 with varying temperatures between 30, 37 and 45 °C under single turnover conditions (equimolar substrate and ribozyme) or multi-turnover with two-fold or four-fold excess substrate **(Figure 26c)**. To maintain a constant amount of RNA among the different conditions, total RNA monomer charge concentration, which is the sum of molar concentration multiplied by length for each RNA species, was fixed at 1 mM. Ribozyme activity directly proportional to temperature and inversely proportional to substrate:ribozyme ratios was observed. The latter behaviour was simply attributed to the higher probability of ribozyme binding the much higher concentration of unligated S1 substrates rather than already growing concatemers. The results in all cases demonstrate the ability of the v2F ribozyme to convert a pool of short oligonucleotide substrates into one with substantially longer product lengths, indeed reaching lengths of ≥ 500 nucleotides, signifying 16 ligation events for these particular products. Moreover, RNAs with altered electrophoretic mobility for bands above approximately 900 nucleotides were observed under many conditions tested. This

effect is amplified for reaction that were incubated at higher temperatures (e.g., 45 °C), which also showed the prominent appearance of irregularly spaced bands at slightly above 50 nucleotides, and at approximately 230 and 500 nucleotides, similarly suggesting an altered electrophoretic mobility. In efforts to minimise hydrolytic RNA degradation of RNA while maximising RNA activity, all subsequent experiments were performed at 30 °C with an equimolar concentration (1:1) of substrate and ribozyme at a total RNA charge concentration of 1 mM, equalling to 10.5 µM of each.

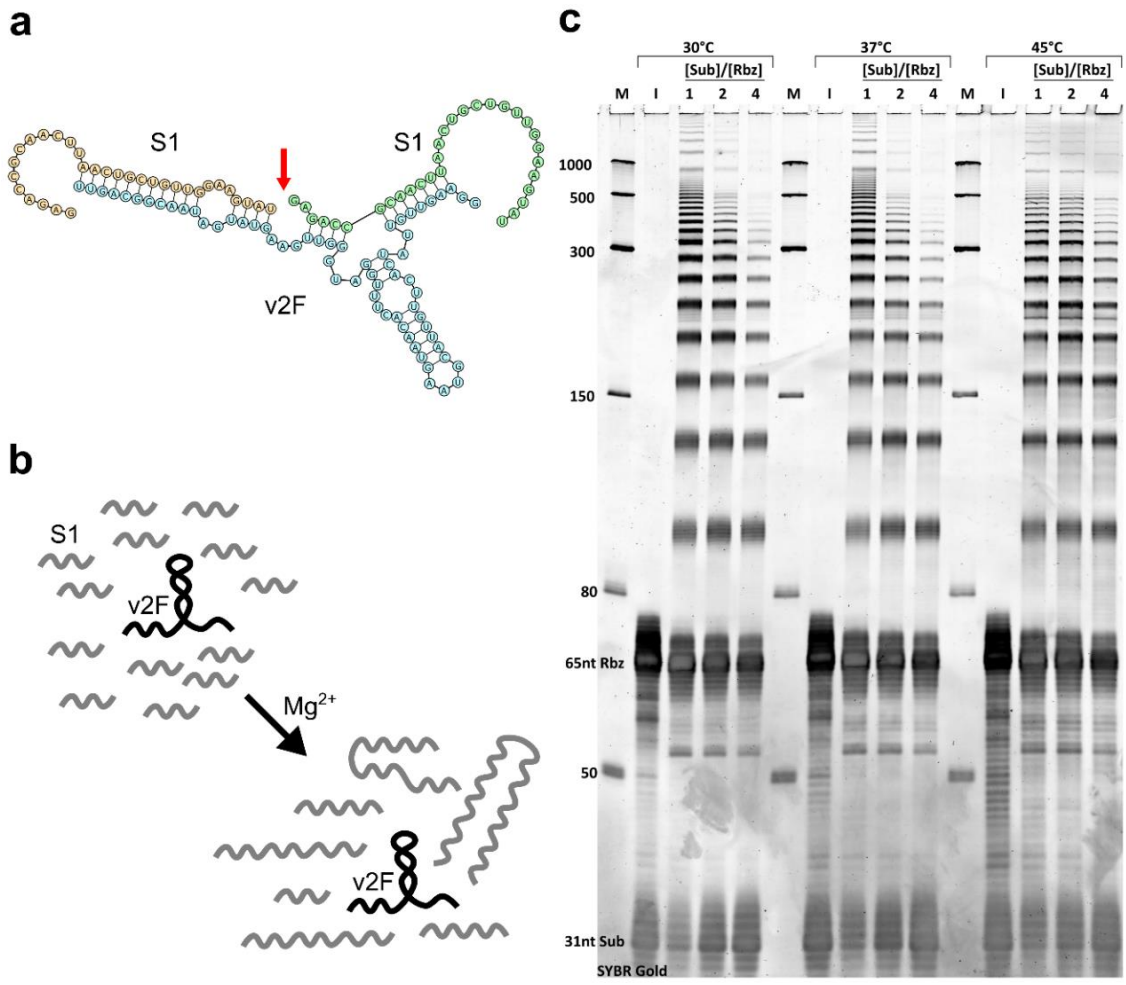


Figure 26: Design and characterisation of the laddering ribozyme v2F at various temperatures and substrate:ribozyme ratios. a) *Sequence and secondary structure of the ribozyme (v2F) bound to two copies of its substrate (S1). The red arrow indicates ligation site. The ribozyme requires a triphosphate activation chemistry at the 5'-end of S1.* ***b)*** *Schematic showing the proposed activity of v2F in solution Note that catalysis is magnesium dependent.* ***c)*** *PAGE analysis screening the activity of the ribozyme at 30, 37, or 45 °C for 2 hours with either equimolar (1), two-fold (2) or four-fold (4) substrate concentration with respect to ribozyme concentration. The reaction buffer contained 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$. Marker lane (M) is the low range ssRNA ladder from NEB (nucleotides; nt).*

3.2.3. Determination of critical coacervate concentration

Having established conditions under which extended products are produced and accumulated, ribozyme activity was first assessed inside coacervates and the system was subsequently screened for the optimal ratio of poly-ions such that spherical droplets form while maintaining ribozyme ligation activity. To establish a model coacervate system, poly-L-lysine oligopeptides with different polymer length distributions were utilised as the positively charged component, chiefly because coacervate systems with poly-lysine have already been described and characterised (*191*, *196*), providing a good starting point for the realisation of an internally active phase-separated model system.

Ribozyme behaviour was screened in the presence of increasing poly-lysine with ranges of either 5-24 residues ($(Lys)_{5\text{-}24}$; 1-5 kDa poly-L-lysine) or 19-72 residues ($(Lys)_{19\text{-}72}$; 4-15 kDa) into a fixed concentration of equimolar substrate:ribozyme RNA with total monomer (nucleotide residue) concentration of 1 mM. The reported monomer concentrations for $(Lys)_n$ and RNA are calculated for monomer unit concentration and therefore, also represent the total charge ratio (positive to negative charges) of the solution. When titrating $(Lys)_{5\text{-}24}$, no ribozyme inhibition could be detected, even at $(Lys)_{5\text{-}24}$ monomer concentrations four-fold in excess **(Figure 27a)**. However, altered band mobility at ratios greater than or equal to 2 was observed, specifically, the disappearance of the band pattern above ~500 nucleotides. In contrast, the longer lysine polymers $(Lys)_{19\text{-}72}$ reduced ribozyme activity already at a ratio of 1.25, which was further exacerbated by the addition of more peptide until complete inhibition of ribozyme activity at excess $(Lys)_{19\text{-}72}$ peptides over RNA. The altered mobility of the slow migrating bands was also observed as with the addition of the shorter peptide, but now the transition occurs at a ratio of 0.5 $(Lys)_{19\text{-}72}$:RNA. The data suggest that, above certain ratios of peptide to RNA, lysine peptides are able to modulate the activity of the laddering ribozyme system, which is likely driven by their interaction in solution: the inhibition of activity by the longer $(Lys)_{19\text{-}72}$ under conditions of excess peptide indicates that the interactions between peptide and ribozyme are quite strong, potentially leading to the unfolding of the catalytic structure of the ribozyme and a concomitant drop in activity, or perhaps a reduced diffusive coefficient effectively preventing the formation of functional ribozyme:substrate complexes. The absence of inhibition by the shorter $(Lys)_{5\text{-}24}$ suggests weaker interactions that still allow ribozyme folding and catalysis even at higher ratios.

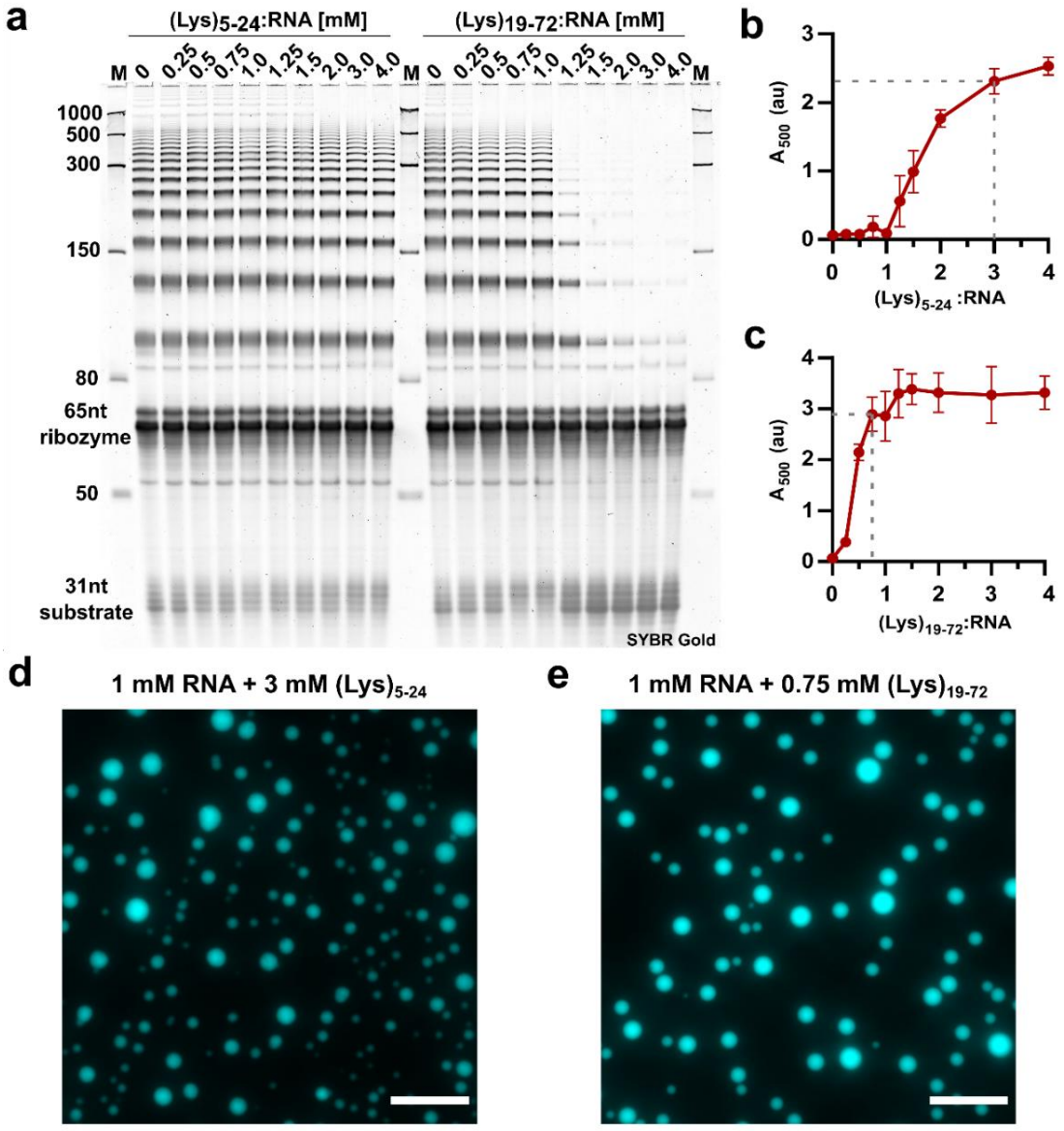


Figure 27: Characterisation of ribozyme activity with varying (Lys)$_n$ oligopeptides. a) *An 8% denaturing PAGE showing the S1 products of the v2F laddering system in solution at varying ratios of either short (Lys)$_{5-24}$ or long (Lys)$_{19-72}$ poly-lysine to RNA (total monomer concentration = 1 mM) after a 2 h reaction at 30 °C, with 10 mM $MgCl_2$ in 50 mM Tris-HCl at pH 8.6.* ***b)*** *Absorbance spectrum at 500 nm of varying ratios of (Lys)$_{5-24}$ to RNA before ligation. Data are presented as mean ± standard deviation (n=3).* ***c)*** *Absorbance spectrum at 500 nm of varying ratios of (Lys)$_{19-72}$ to RNA before ligation. Data are presented as mean ± standard deviation (n=3).* ***d)*** *Fluorescence microscopy image coacervates of (Lys)$_{5-24}$:RNA at a ratio of 3:1 (Lys)$_{5-24}$:RNA, imaged using 10% 5'Cy5-tagged S1substrate strand. Scale bar = 20 µm.* ***e)*** *Fluorescence microscopy image coacervates of (Lys)$_{19-72}$:RNA at a ratio of 0.75:1 (Lys)$_{19-72}$:RNA, imaged using 10% 5'Cy5-tagged S1substrate strand. Scale bar = 20 µm. Fluorescence microscopy images shown in* ***(d-e)*** *were acquired by Kristian Le Vay and reproduced from* (*235*).

Next, the phase separation behaviours of the ribozyme system with (Lys)$_n$ peptides were investigated by repeating the previous experiment but analysing the resulting solution by measuring optical density at 500 nm (A_{500}). Coacervation turns a solution turbid

which, consequently, differentially diffracts light; as such, there is a direct correlation between optical density of a solution and droplet formation (*236*). Indeed, the addition of $(Lys)_n$ peptide led to a gradual increase in turbidity as a result of phase separation in both cases **(Figure 27b-c)**. As expected, turbidity increase with $(Lys)_{19-72}$ preceded that with $(Lys)_{5-24}$ owing to the higher polymer length and therefore, higher association strength. The onset of coacervation, or critical coacervate concentration (CCC), for lysine peptides was estimated to be approximately 0.14:1 and 0.93:1 for $(Lys)_{19-72}$:RNA and $(Lys)_{5-24}$:RNA, respectively. Based on the data obtained here and in efforts to realise a similar coacervate compartment system with both poly-lysine species, charge ratios of 0.75:1 for $(Lys)_{19-72}$:RNA and 3:1 for $(Lys)_{5-24}$:RNA were selected for further investigation as both ratios occur shortly before their respective absorbance maxima and modulated ribozyme activity but did not inhibit it, *vis-à-vis* the disappearance of irregularly-spaced and slowly-migrating bands. While optical density can be measure of coacervation, it does not unquestionably demonstrate the formation of liquid-like coacervates, since any particle type including macromolecular aggregates also result in an increased absorbance. Thus, to elucidate the nature of particles that were forming upon peptide addition, the abovementioned ratios were subjected to fluorescence microscopy with 10% Cy5-tagged S1 substrate. The resulting images confirmed the formation of phase separated coacervates that strongly partition the Cy5-tagged S1 substrate **(Figure 27d-e)**. The smooth, round conformation and absence of aggregates suggested that these coacervates are liquid-like, containing a fair amount of water molecules and magnesium ions within and allowing ribozyme catalysis to proceed under compartmentalised conditions. The non-overlapping conditions of activity and inhibition between $(Lys)_{5-24}$ and $(Lys)_{19-72}$ concentrations meant that a direct comparison of the effect of peptide length as a function of RNA length, by fixing equal monomer concentrations of all polymers, could not be made. However, the positive findings encouraged further experimentation with these ratios, unless specifically mentioned, despite the non-equivalent concentration of the different lysine peptides. Having established working conditions that support both coacervation and ribozyme activity, the next step was to uncover what types of effects arise from the interaction of ribozyme and peptide in the environment of a charge-rich phase-separated complex coacervate compartment.

3.2.4. Effects of coacervation on ribozyme activity

Returning to the anomalous irregularly-spaced bands that suggested different electrophoretic mobility of RNA products, others have shown that a re-engineered intron-

derived ribozyme capable of RNA concatenation also produced circular RNA products in addition to linear concatemers (*237*), leading to the hypothesis that the abovementioned product bands that are apparent when the reaction was carried out in solution are circular and therefore possess an altered electrophoretic mobility during PAGE, whereas the reaction that occurred in the presence of coacervate droplets altered ribozyme behaviour by reducing or inhibiting circular product formation altogether. To test this hypothesis, the purified RNA products of laddering reactions in solution or with either lysine oligopeptides were subjected to an RNase R digest followed by denaturing PAGE analysis. RNase R is a 3' to 5' exoribonuclease, effectively digesting linear species while leaving circular products and lariats or structured RNAs with short 3'-overhangs intact. Thus, treatment of the reaction products with RNase R should enrich for circular products that can subsequently be resolved and identified, if any. As expected, the RNase R digest revealed that circular products were indeed forming as a result of ribozyme activity in the absence of lysine peptides **(Figure 28a)**. A decrease in circular product formation was observed when the reaction was incubated in the presence of 2 mM $(Lys)_{5\text{-}24}$, whereas the incubation in the presence of 1 mM of $(Lys)_{19\text{-}72}$ completely supressed the formation of circular products. The absence of circular products essentially means that the ends of the same substrate no longer find themselves bound to the same ribozyme. This downregulation or outright suppression can be attributed to a variety of factors, such as a decrease in RNA mobility within the condensed phase, an increased stiffness of the RNA polymer, increased substrate:ribozyme stability preventing the release of substrate and multi-turnover catalysis, or any combination of these effects to varying degrees. While the modulation of ribozyme circularisation activity by coacervation is obvious, more data is needed to unequivocally determine the exact mechanism of this inhibition in this specific case. Nonetheless, we can revise our model for v2F activity: concatenation of RNA and formation of circular products in solution; various degrees of suppression of circular products in coacervates **(Figure 28b-c)**.

The proceeding experiments aimed to explore whether coacervation was able, at least to some degree, to modulate the rate of ribozyme catalysis compared to catalysis in solution. Since condensed coacervate phases were known to up-concentrate relevant ions and biomolecules, perhaps the ribozyme could benefit an enhanced catalytic rate in the presence of coacervate droplets. To this end, kinetic analyses were performed for v2F ribozyme without $(Lys)_n$, with 3 mM $(Lys)_{5\text{-}24}$ or with 0.75 mM $(Lys)_{19\text{-}72}$ in 1 mM total RNA monomer charge with equimolar substrate and ribozyme **(Figure 29a)**. In solution, the average substrate length gradually increases over time, reaching a plateau after 180 minutes

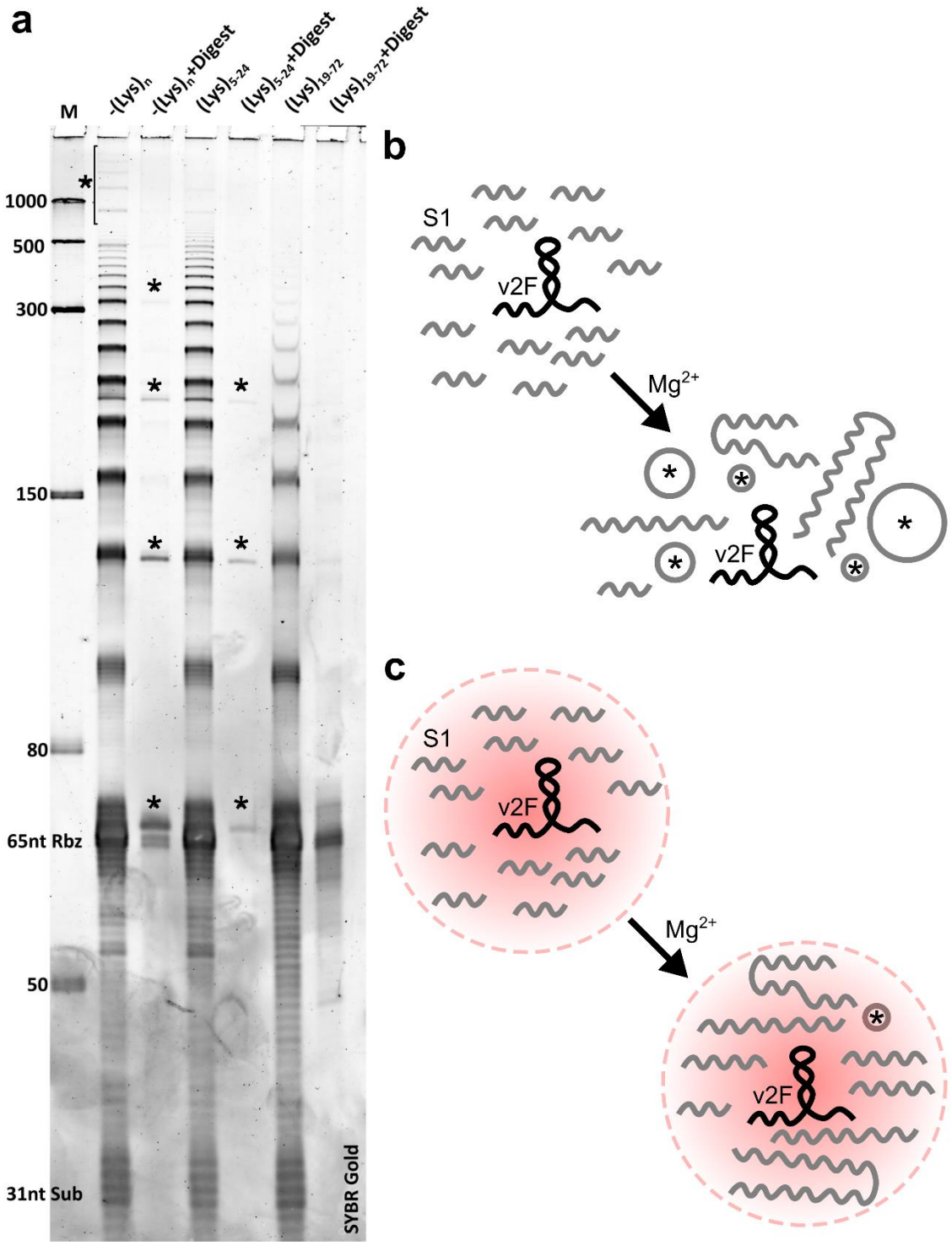


Figure 28: Formation of circular products by v2F. a) *An 8% denaturing PAGE verifying the formation of circular products (marked with an asterisk) after RNase R digestion. Equimolar concentrations (10.5 µM; 1mM total RNA) of substrate and ribozyme were incubated in 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$ in solution, with 2 mM $(Lys)_{5-24}$ or 1 mM $(Lys)_{19-72}$ for 3 hours at 45 °C. The RNA was subsequently purified and treated with 30 U of RNase R (digest lane) for 3.5 hours at 37 °C before PAGE analysis.* ***b-c)*** *Revised schematics showing v2F activity* ***(b)*** *in solution or* ***(c)*** *in coacervate droplets.*

at 52 nucleotides **(Figure 29b)**. In stark contrast to catalysis in bulk solution, for both cases where lysine peptide was added to form coacervates, the average substrate length increases sharply within the first 30 minutes of initiation, followed by a slow increase and eventual plateau for the remaining time sampled. The estimated average substrate lengths after 240 minutes were 79 and 85 nucleotides for the reactions supplemented with 0.75 mM $(Lys)_{19-72}$ and 3 mM $(Lys)_{5-24}$, respectively. The kinetics of substrate elongation in solution were best described using a first order kinetic model yielding an apparent rate constant of $0.015 \pm 0.001\ min^{-1}$, whereas both condensed phase kinetics were approximated using a two-phase model yielding an initially fast and subsequently slow apparent rate constant, k_{fast} and k_{slow}, respectively **(Figure 29c)**. The approximate apparent rate constants for the reaction with a ratio of 0.75:1 $(Lys)_{19-72}$:RNA were $k_{fast} = 0.77 \pm 0.11\ min^{-1}$ and $k_{slow} = 0.066 \pm 0.013\ min^{-1}$, and the reaction with a ratio of 3:1 $(Lys)_{5-24}$:RNA exhibited similar rate constants estimated to be $k_{fast} = 0.75 \pm 0.16\ min^{-1}$ and $k_{slow} = 0.057 \pm 0.034\ min^{-1}$. Thus, although not clearly discernible from the initial endpoint sampling as in the lysine titration experiment in **Figure 27a**, the time resolved quantitative analysis revealed that addition of either $(Lys)_n$ peptide led to an estimated 50-fold increase in v2F rate of catalysis and a concomitant 30 nucleotide increase in final average substrate length when compared with catalysis in bulk solution. These data indicate a marked enhancement of catalytic abilities of the laddering ribozyme in the coacervate environment, at the level of both type of products formed as well as rate of catalysis, confirming the earlier hypothesis for environmentally-driven modulation of ribozyme activity.

The above experiments demonstrated enhanced ribozyme catalysis in coacervate phases; however, they shine no light on the cause of this enhancement. A possible motive is that increased RNA and divalent ion concentration in the crowded environment of the condensed phase supported ribozyme catalysis. Consequently, various RNA concentrations **(Figure 30a, e)** and magnesium ion concentrations **(Figure 30b, f)** were screened, and reactions were also supplemented with a commonly utilised molecular crowder poly-ethylene glycol (PEG) 8000 **(Figure 30c, g)**. Since the kinetic experiments revealed that the most notable difference in catalytic rate was observed during the first tens of minutes after initiation, it was best to capture any rate modulation by incubating the reactions for only 10 minutes at 30 °C before PAGE analysis. Halving (0.5x; 0.5 mM total RNA monomer concentration) or doubling (2x; 2 mM total RNA monomer concentration) the total RNA concentration produced a predictable effect, with a slight decrease or slight increase in the average substrate length compared to the standard conditions previously established. Doub-

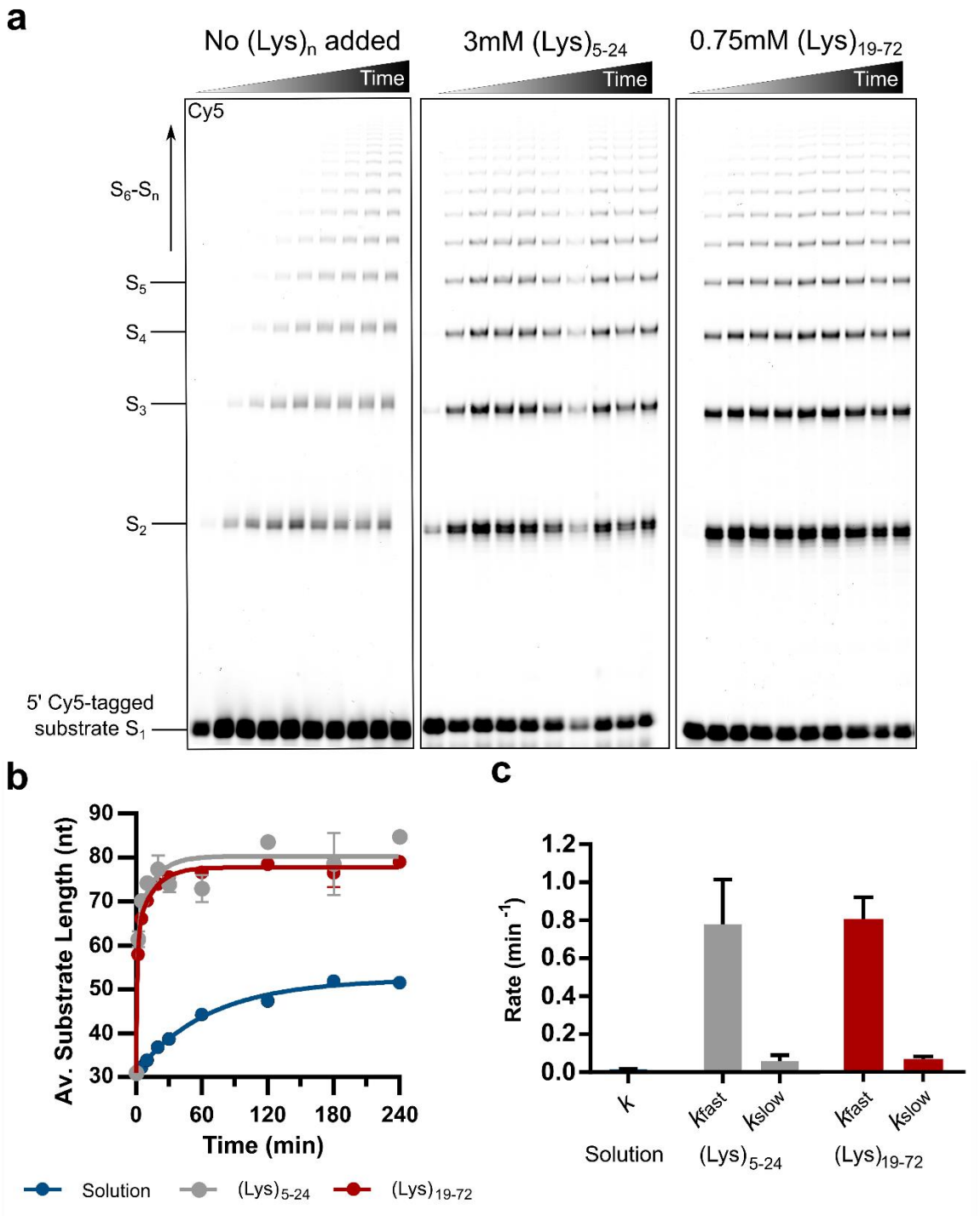


Figure 29: Kinetics of v2F ligation in solution and in the presence of lysine peptides. a) *8% denaturing PAGE analyses showing exclusively the substrate and products of the v2F laddering system in solution, with 3 mM $(Lys)_{5-24}$ or 0.75 mM $(Lys)_{19-72}$ over the course of 4 hours at 30 °C in 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$ with 10.5 µM of each v2F ribozyme and S1 substrate, of which 10% was Cy5-tagged S1 substrate for improved visualization and quantification.* ***b)*** *Plot of reaction kinetics from* ***a)*** *showing the change in average substrate length over time for bulk solution in blue, with 3 mM $(Lys)_{5-24}$ in grey or with 0.75 mM $(Lys)_{19-72}$ in red. The data points were fitted with a first-order model in the case of bulk solution, whereas data points for the coacervate phases were fitted with a second-order model.* ***c)*** *Calculated rate constants for RNA concatenation by the v2F ribozyme extracted from* ***b)****. All data are presented as mean ± standard deviation (n=3).*

ling the total RNA concentration increased the average substrate length from 35 nucleotides at 1x concentration to approximately 40 nucleotides, corresponding to a 15% increase in average substrate length. Upon titration of magnesium chloride, a concentration-dependent response was observed, reaching an average substrate length of 57 nucleotides at maximum dose (100 mM), corresponding to a 45% increase at ten-fold higher magnesium chloride concentration. Perhaps coincidentally, at 20 mM magnesium chloride, or, double the ion concentration, average substrate length was estimated to be 41 nucleotides, almost equal to what was observed upon doubling the RNA concentration. As expected, a similar trend was seen upon titration of PEG 8000, although reaching an average substrate length of 40 and 52 nucleotides at 10% and 20% v/v PEG 8000, respectively, only slightly lower than that observed with 100 mM magnesium chloride. In short, none of the tested conditions alone was able to replicate the activity enhancement observed in the presence of lysine peptides, suggesting that the straightforward up-concentration of a single reaction component alone is not sufficient to explain the rate enhancement otherwise observed in coacervates. Indeed, this was found to be the case: when increasing magnesium chloride concentrations (25, 50 and 75 mM) were added to reactions supplemented 10% v/v PEG 8000, a more prominent increase in average substrate length was observed reaching 72 nucleotides and almost rivalling that observed in reactions with coacervate compartments, which obtained an average substrate length of 74 and 76 nucleotides upon addition of 3 mM $(Lys)_{5-24}$ and 0.75 mM $(Lys)_{19-72}$, respectively **(Figure 30d, h)**. While these results cannot conclusively demonstrate the mechanism of rate enhancement occurring in coacervates, they do provide some clarity as to the possible roles that a complex coacervate compartments can offer, by showing that similar activity enhancements can be achieved using macromolecular crowders and higher divalent metal ion concentrations. The increased activity in coacervates might also stem from the interactions between peptide and RNA that are absent with simply increasing crowding and ion concentration. In fact, nucleotide and polyamines such as spermine or spermidine were shown to phase-separate and provide strong compartmentalisation of RNA and Mg^{2+} ions (*194*). Surprisingly, when titrating increasing magnesium chloride into reactions supplemented with 20% v/v PEG 8000, an opposite response was obtained, that is, decreasing average substrate lengths with increasing magnesium chloride concentrations. This effect can be attributed to substrate inhibition by formation of unproductive 2:1 substrate:ribozyme complexes that cannot be recycled for multi-turnover due to the high stability of the ligated S2:v2F complex under conditions of high Mg^{2+} ions and PEG 8000. This inhibition of activity is reminiscent of that observed with excess $(Lys)_{19-72}$ peptides,

although, as explained earlier, peptide-mediated inhibition is more likely driven by a different mechanism of strong RNA binding and subsequent unfolding of catalytically active structures rather than by substrate inhibition.

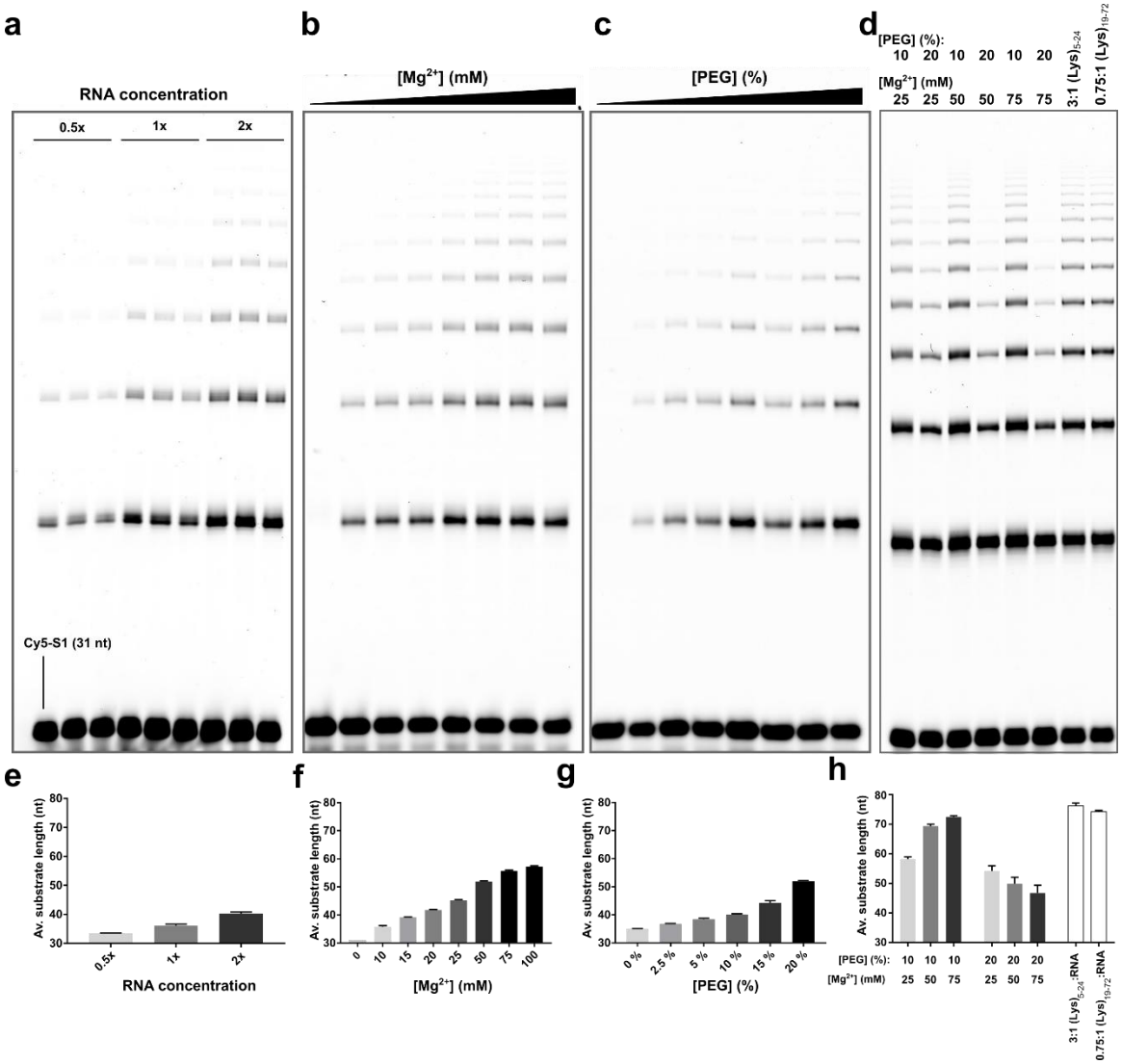


Figure 30: Effect of RNA concentration, $MgCl_2$ or PEG 8000 on ribozyme activity. a) *An 8% denaturing PAGE showing the products of reactions carried out with half (0.5x; 5.25 µM), normal (1x; 10.5 µM each) or double (2x; 21 µM each) the concentration of each ribozyme and substrate after 10 minutes at 30 °C in 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$. Triplicates are shown here side by side in PAGE.* ***b)*** *An 8% denaturing PAGE showing the products of reactions containing increasing concentrations of magnesium chloride. Reactions were carried out in 50 mM Tris-HCl pH 8.6 and contained equimolar (10.5 µM) concentrations of ribozyme and substrate. A negative control with 0 mM $MgCl_2$ is included.* ***c)*** *An 8% denaturing PAGE showing the products of standard reactions supplemented with increasing volume percent of PEG 8000. Reactions contained 10.5 µM each of ribozyme and substrate in 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$ and were quenched after a 10-minute incubation at 30 °C.* ***d)*** *An 8% PAGE showing the products of a combinatorial screen of different magnesium chloride concentrations and PEG 8000 volumetric percentages. As positive controls, the last two lanes are reactions carried out in the presence of either 3 mM $(Lys)_{5\text{-}24}$ or 0.75 mM $(Lys)_{19\text{-}72}$.* ***e-h)*** *Bar plots showing the quantification of average substrate length for the respective PAGE gels in* ***a-d)****. All data are presented as mean ± standard deviation (n=3).*

Taken together, the data presented thus far provide substantial evidence that the compartmentalisation of a catalytic RNA and its substrate through complex coacervation with cationic oligopeptides endows it with enhanced catalytic rate and, to a lesser degree, alters its activity by hindering the formation of circular products. The obvious next question was whether elongating the poly-anionic component, i.e., S1 RNA substrate, of peptide-RNA coacervates through ribozyme-mediated ligation could produce phenotypic changes in terms of droplet behaviour or morphology.

3.2.5. Effect of ribozyme activity on coacervate compartment

The inquiry into the potential effect exerted on coacervate phenotype by the catalytic activity of its genetic component necessitated the development of an inactive version of the ribozyme such that a comparison can be made between coacervates hosting either catalytically active or inactive ribozyme, thereby revealing the role of both length of anionic and cationic components in determining droplet properties. A catalytically inactive v2F ribozyme, hereon referred to as "dead" v2F or dv2F, was obtained by mutating 11 bases in the active sequence, effectively closing the bulge in the ribozyme stem and substituting residues in the proximity of the catalytic site with uracil bases (U16C, A25C, A26C, C33G, U34C, U35G, G46U, A47U, A48U, G53U, A54U). The substrate binding arms were not affected by these mutations such that the inactive ribozyme could still bind the substrate but not catalyse ligation. The co-incubation of inactive ribozyme and substrate showed no ligation activity after 2 hours at 30 °C, irrespective of the presence of $(Lys)_n$ peptides, confirming the inactivity of the mutated ribozyme sequence **(Figure 31)**. In contrast, the active variant showed reproducible activity indicating that the conditions tested were amenable to catalysis and that inactivity is a direct consequence of the interruption of the ribozyme active site, which could not be rescued via compartmentalisation by coacervation. The development of an inactive ribozyme allowed forming coacervates from RNA without catalytic activity, and thus sparked interest into whether ligation activity could be localised to within the condensed coacervate phase. To elaborate, ribozyme catalysis could just as well occur only in solution even in the presence of $(Lys)_n$-based coacervates and it could be that the compartmentalised RNA is inactive. Since microscopic localisation studies require a fluorescent output, a variant of the R3C ligase ribozyme that permits real-time detection of ligation by fluorescence increase was employed **(Figure 32a)**. The FQ-B substrate enables detectable fluorescence increase upon ligation, since a fluorophore (TAMRA) is attached

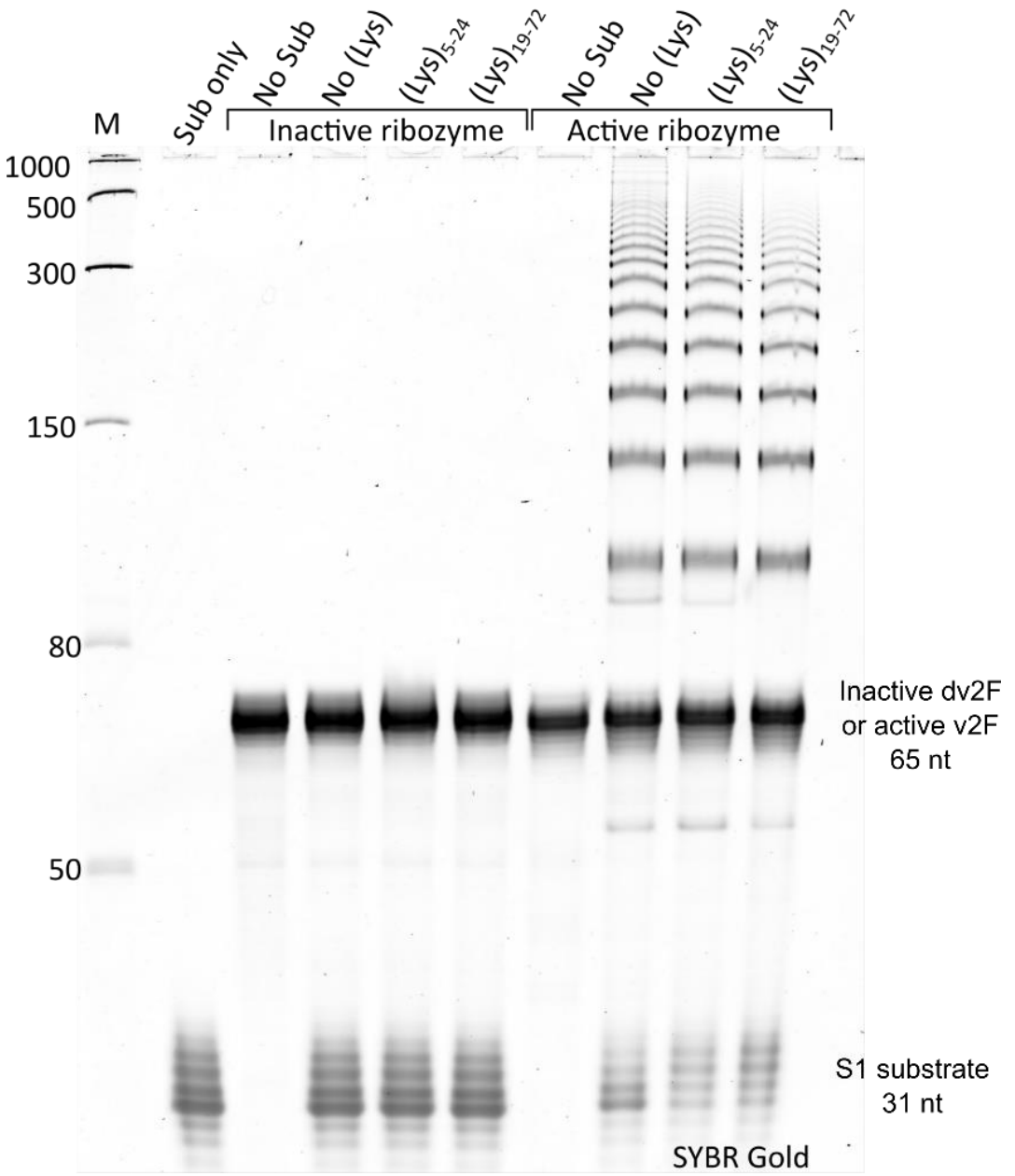


Figure 31: Verification of inactivating mutations in dead v2F (dv2F) inactive ribozyme system. *An 8% PAGE showing the products of active (v2F) and inactive (dv2F) versions of the laddering ribozyme either in solution, with 0.75 mM $(Lys)_{19-72}$ or 3 mM $(Lys)_{5-24}$ in buffer containing 10 mM $MgCl_2$ and 50 mM Tris-HCl pH 8.6, at equimolar concentration of ribozyme and substrate (10.5 µM) for a 1 mM total RNA monomer concentration. The reactions were quenched and resolved on an 8% denaturing PAGE after incubation for 2 hours at 30 °C.*

to the gamma-phosphate on the 5'-end of FQ-B substrate via a 6-carbon linker, whereas a quencher is installed at the 3'-end such that fluorescence is initially quenched, but upon ligation, the pyrophosphate-linker-fluorophore is released thus escaping quenching and resulting in a fluorescence signal.

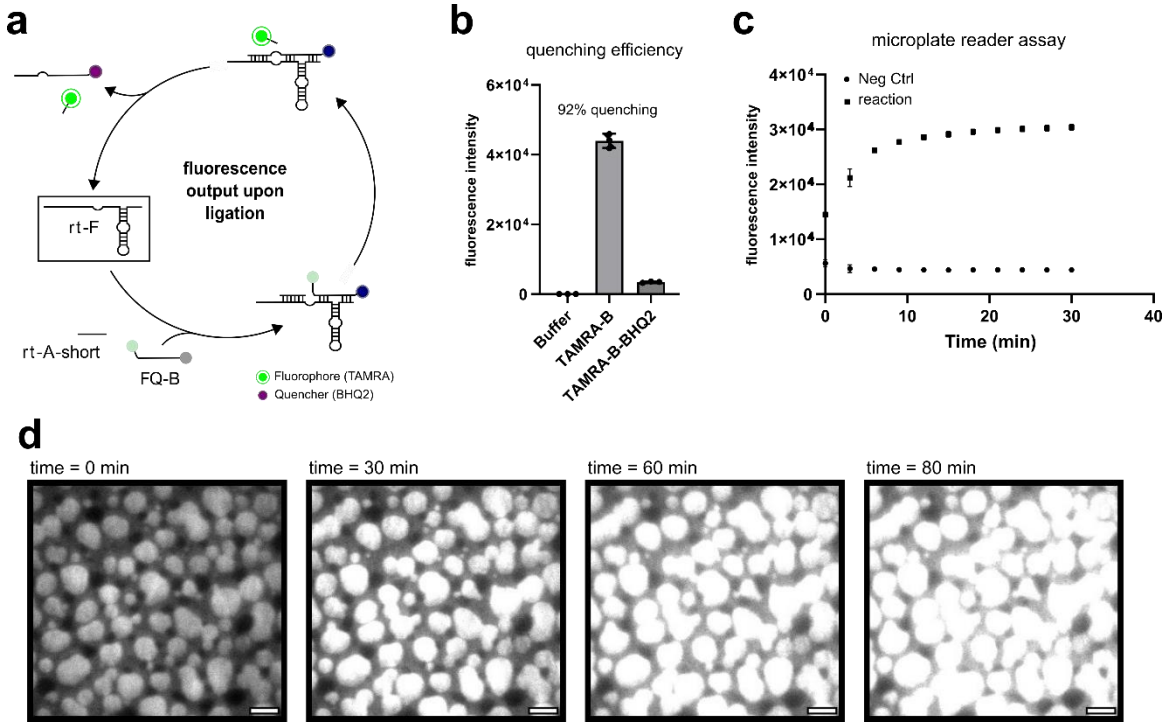


Figure 32: A modified R3C ligase ribozyme activity occurs within complex coacervates. a) *Schematic showing the catalytic activity of the real-time R3C ligase system. Ligation results in the release of fluorophore from the RNA-quencher strand and an increase in fluorescence.* ***b)*** *Determination of quenching efficiency upon installation of quencher at the 3'-end of the RNA substrate. Endpoint measurements of 1 µM TAMRA-B or TAMRA-B-BHQ2 (FQ-B)* ***c)*** *Fluorescence readout ligation assay performed in microplate reader with 2 µM FQ-B, 5 µM of each rt-A-short and rt-F in 20 mM $MgCl_2$ 50 mM EPPS pH 8.5 and 900 mM sucrose. The negative control lacked rt-A-short and rt-F. The reaction was monitored for 30 minutes at 37 °C. The initial increase in fluorescence is not measured due to the lag time between transferring the samples to the microplate and starting the measurement in the microplate reader.* ***d)*** *Time series of TIRF microscopy of real-time R3C reaction shown in* ***a)*** *with 200 nM FQ-B, 5 µM rt-A-short and 1 µM rt-F in the presence of coacervates prepared with 2 mM $(Lys)_{5-24}$ and 1 mM RNA monomer concentration (equimolar S1 substrate and inactive ribozyme) in 10 mM $MgCl_2$ and 50 mM Tris-HCl pH 8.6 and incubated for 2 hours at 30 °C. All data are presented as mean ± standard deviation (n=3). Scale bars = 5 µm.*

Synthesis of the substrate proceeded via a primed *in vitro* transcription with a γ-(6-aminohexyl)-GTP initiator nucleotide. The resulting transcript with a terminal 5'-primary amine was reacted with N-hydroxysuccinimidyl (NHS) ester modified TAMRA fluorophore, forming a stable amide bond between the 5'-end of the substrate and the fluorophore. Finally, the quencher was added by the RNA ligase 2 mediated reaction of a 5'-phosphorylated pentamer carrying a 3'-end quencher with the substrate on a complementary DNA splint. A 92% quenching efficiency was estimated after ligation of the quencher-pentamer to the 3'-end of the fluorophore substrate (TAMRA-B) to yield the fluorophore-quencher substrate FQ-B **(Figure 32b)**. Incubation of the reaction components i.e., substrate rt-A-short, ribozyme rt-F and FQ-B, in an adequate buffer and temperature regime

resulted in a time-dependant increase of fluorescence **(Figure 32c)**. Importantly, excluding the ribozyme in the same buffer and temperature regime showed only basal background fluorescence and no change over time, with the background fluorescence being attributed to the incomplete quenching. The results suggested that ribozyme-mediated substrate ligation was coupled to the increase in fluorescence, which could therefore be utilised as a reliable indicator for ligation. To assess whether the coacervate environment could support ribozyme catalysis and possibly co-localise coacervation and ligation, the fluorescent R3C reporter system was deployed in reactions containing inactive coacervates comprised of 2:1 mM $(Lys)_{5-24}$:RNA supplemented with 200 nM of FQ-B substrate and 5 µM of rt-A-short in 10 mM $MgCl_2$ and 50 mM Tris-HCl pH 8.6. The reactions were monitored for 2 hours with total internal reflection (TIRF) microscopy **(Figure 32d)**. Before addition of ribozyme to the coacervate and substrate mix (time = 0 min), background substrate fluorescence showed that they localise within the coacervates. Upon the addition of ribozyme, we observe a steady increase in fluorescence over time up to time = 80 minutes within the coacervates and not in the surrounding bulk volume, suggesting that compartmentalisation within the condensed phase is indeed permissive to compartmentalised RNA catalysts.

Seeing as the ribozyme benefits from enhanced catalytic rate and supressed circularisation activity in the coacervate compartment, the next step was to assess whether such a change in average substrate length brough about by ribozyme ligation was enough to, in turn, alter the physical properties of the host coacervates, effectively achieving a rudimentary coupling of compartment phenotype to genotype. To this end, complex coacervates were prepared with inactive or active ribozyme by combining 1 mM total monomer RNA of equimolar ribozyme and substrate (10% Cy5-tagged) with 0.75 mM $(Lys)_{19-72}$ or 3 mM $(Lys)_{5-24}$ in passivated environments and were observed over the course of 24 hours by confocal microscopy. Adding all reagents together immediately allowed ligation to occur in the presence of $(Lys)_n$ peptides such that they would benefit from the rate enhancements, potentially amplifying any phenotypic incongruities that might emerge. The coacervates partitioned the Cy5-tagged S1 substrate for all conditions tested during the 24-hour duration of the experiment **(Figure 33a-b, e-f)**. Although all displayed a round morphology indicative of surface tension and consequently, substantial hydration, different behaviours could be attributed to the different populations of active and inactive coacervates, which was especially notable for those formed with 0.75 mM $(Lys)_{19-72}$. Over 24 hours, $(Lys)_{19-72}$ coacervates with the active ribozyme maintained a stable area and similarly, the number of

particles per 100 μm^2 remained nearly constant **(Figure 33c-d; grey bars)**. Their persistence as individual droplets suggested that they did not grow through passive coarsening or coalescence. In contrast, inactive ribozyme-containing coacervates showed a decrease in the number of particles per 100 μm^2 and a concomitant increase in coacervate area; specifically, an approximate 35% increase in droplet surface area was associated with a 40% decrease in the average number of particles, suggesting that individual coacervates were merging with one another to form larger ones over time **(Figure 33c-d: black bars)**. In conditions of incubation more than or equal to 2 hours, the inactive system registered a final droplet area approximately double that of the active system. Since half of the polyanionic coacervate component i.e., RNA S1 substrate, is expected to elongate with time in the active system, the dissimilar phenotypes observed between active and inactive coacervates formed with $(Lys)_{19\text{-}72}$ peptides can be attributed to substrate tandem ligation by the active ribozyme. Thus, it appears possible to modulate compartment phenotype based on the action of an active compartment genotype under the conditions tested.

The behaviours of coacervates formed from the shorter $(Lys)_{5\text{-}24}$ peptides were slightly different than those formed from the longer $(Lys)_{19\text{-}72}$ peptides. Fluorescence microscopy-based monitoring revealed no difference between $(Lys)_{5\text{-}24}$ coacervates with active or inactive systems in a passivated environment **(Figure 33g-h)**. In fact, coacervates comprised of $(Lys)_{5\text{-}24}$ peptides with either active or inactive ribozyme systems displayed a small increase in coacervate area and decrease in population density (particles per 100 μm^2), suggesting a reduced compartment stability and that, unlike the active $(Lys)_{19\text{-}72}$ coacervates, coalescence was not completely inhibited for the active system but rather slightly supressed, allowing some coacervate growth over time via fusion events and perhaps passive coarsening. Interestingly, irregularly shaped coacervates formed from the shorter $(Lys)_{5\text{-}24}$ and active RNA system were initially observed, which eventually relaxed to a spherical morphology after 24 hours **(Figure 33e)**. This phenomenon lends further support to the argument that the physical properties of coacervates are being altered over time, likely in response to ribozyme catalysis. In short, we observed a resistance towards formation of larger droplets, suggestive of fusion events among coacervates, when longer poly-lysine and active RNA were present. In these particular conditions, coacervates retained their initial morphology throughout the experiment. The loss of resistance towards coacervate merging events for shorter lysine peptides suggested that there is correlation not just between the compartment phenotype and length of the RNA component, but also length of poly-lysine component. The macromolecular characteristics of both components were

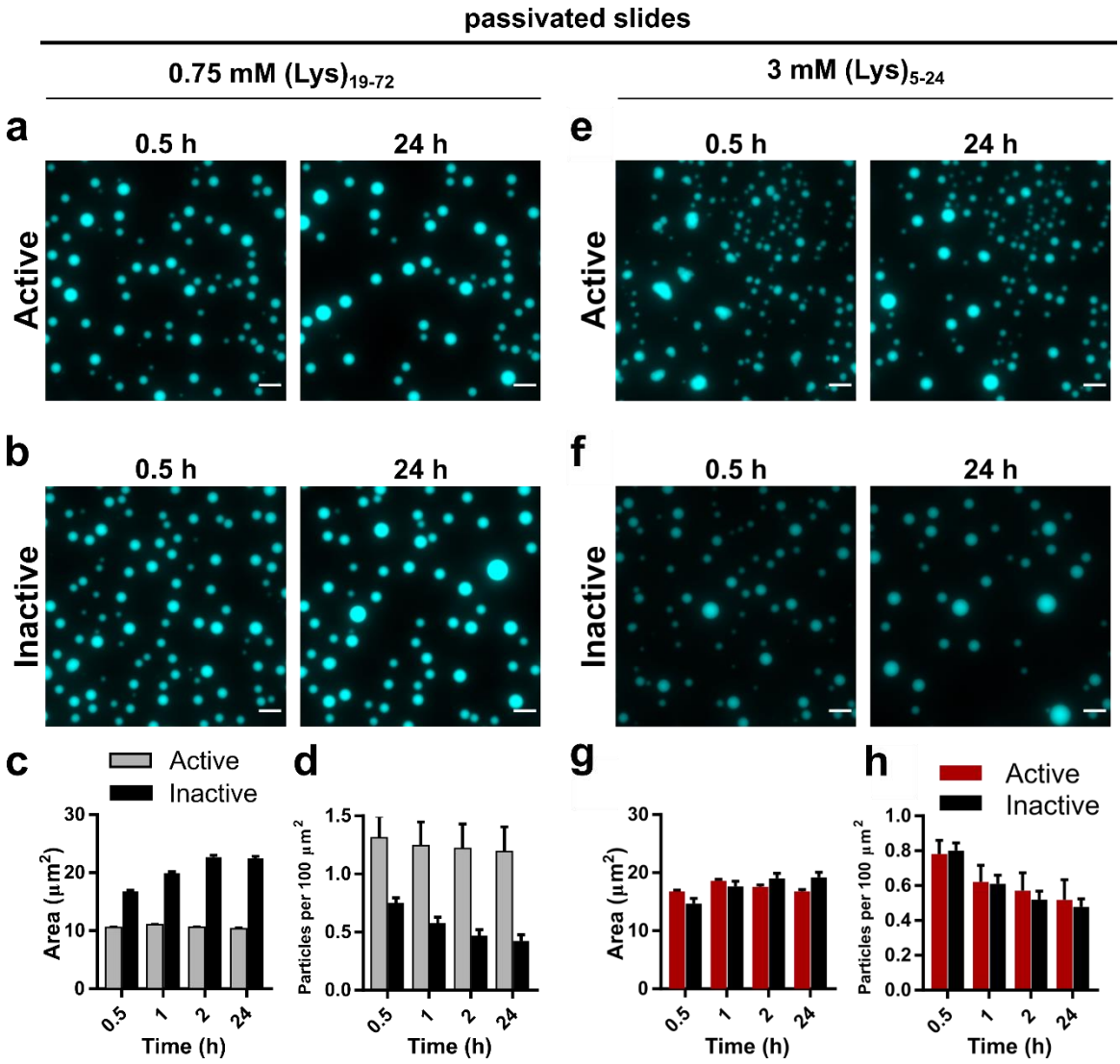


Figure 33: (Data was acquired by Kristian Le Vay and reproduced from (235)***) Screening of droplet morphology over time on passivated surfaces with active or inactive RNA and (Lys)***$_{19\text{-}72}$*** or (Lys)***$_{5\text{-}24}$***. a-b)*** *Representative widefield fluorescence images of 0.75:1 (Lys)*$_{19\text{-}72}$*:RNA coacervates with* ***a)*** *active or* ***b)*** *inactive ribozyme and either 10% Cy5-tagged S1 substrate on passivated slides.* ***c-d)*** *Bar plot showing both the change in* ***c)*** *droplet area over time and* ***d)*** *particles per area (100 µm*^{2}*) over time.* ***e-f)*** *Representative widefield fluorescence images of 3:1 (Lys)*$_{5\text{-}24}$*:RNA coacervates with* ***e)*** *active ribozyme or* ***f)*** *inactive ribozyme and 10% Cy5-tagged S1 substrate.* ***g-h)*** *Bar plot showing both the change in* ***g)*** *droplet area over time and* ***h)*** *particles per area (100 µm*^{2}*) over time. All data are presented as mean ± standard deviation (n=9). Scale bars = 10µm.*

shown to play a role in determining the overall compartment phenotype, confirming the hypothesis that changes in either poly-ionic components that effectively weaken or strengthen the interactions between the polymers could, to varying degrees, dictate the overall morphology and physical properties of the comprising coacervates.

To further the investigation into the potential phenotypic effects on the compartment emerging from ribozyme catalysis, the interactions that a microcompartment would

have with its surrounding environment were examined. While surface passivation is typically the standard method to image liquid-liquid phase separation *in situ*, a chemical system in nature is in reality inextricably linked to its environment. This served as motivation to monitor coacervate behaviour in unpassivated environments, in hopes of extracting additional data that could informative about coacervate properties, the interactions with its surrounding, and possibly design rules for the generation of RNA-peptide coacervates with predictable characteristics and behaviour. For these reasons, the above experiment was repeated on an unpassivated polystyrene surface, which exhibits intermediate hydrophobic properties and are designed for optimal non-polar protein measurements in solution (Greiner μClear, medium binding). In this unpassivated environment, coacervates with the active RNA system behaved similarly compared to the passivated environment. Coacervates formed from $(Lys)_{19\text{-}72}$ or $(Lys)_{5\text{-}24}$ peptides and active ribozyme retained their spherical morphology and maintained the same area over 24 hours, despite being surrounded by an environment with which it could potentially interact **(Figure 34a, c-d, e, g-h)**. Coacervates composed of $(Lys)_{19\text{-}72}$ and the inactive ribozyme system grew by coalescence and coarsening as in the passivated environment, but also demonstrated strong surface wetting capabilities **(Figure 34b-d)**. In contrast to the passivated environment, coacervates composed of $(Lys)_{5\text{-}24}$ and inactive RNA now also displayed growth and fusion events rather than remaining relatively static **(Figure 34f-h)**. In fact, in all cases employing inactive ribozyme coacervates, the gradual wetting and spreading of coacervates was observed, which eventually produced a condensed phase film on the surface, inferred from the gradual loss of fluorescent hotspots signifying homogeneity in the focal plane. The results obtained here underpin the importance of having a catalytically active system contained within the compartment, as only those coacervates, irrespective of whether short (5-24 residues) or long (19-72 residues) lysine polymers were used, did not merge into a single large phase but rather persisted as distinct particles throughout the experiments. This also differs from what was observed in a passivated environment, where inactive coacervates with short poly-lysine also resisted growth and coalescence as in the active system **(Figure 33e-h)**. Therefore, a logical deduction from the results presented is that RNA-peptide coacervates hosting catalytically active species that directly affect coacervate component length, even those formed from relatively short peptides, can counteract the thermodynamic-entropic forces that push the system towards the formation of a single homogeneous condensed phase rather allowing them to persist as individual droplet compartments.

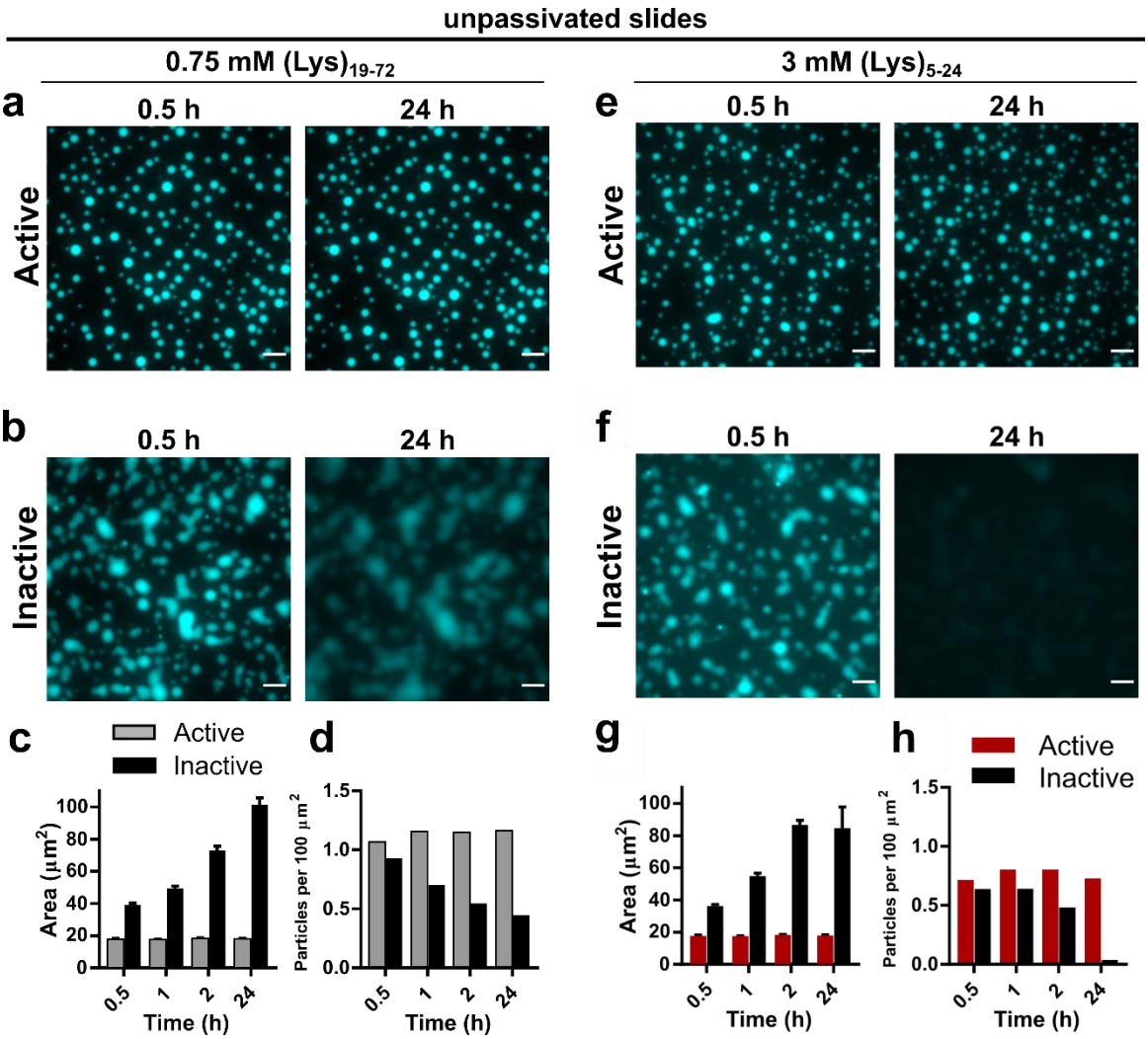


***Figure 34: (Data was acquired by Kristian Le Vay and reproduced from** (235)) **Screening of droplet morphology over time on unpassivated surfaces with active or inactive RNA and (Lys)$_{19-72}$ or (Lys)$_{5-24}$. a-b)** Representative widefield fluorescence images of 0.75:1 (Lys)$_{19-72}$:RNA coacervates with **a)** active or **b)** inactive ribozyme and either 10% Cy5-tagged S1 substrate on unpassivated slides. **c-d)** Bar plot showing both the change in **c)** droplet area over time and **d)** particles per area (100 µm^2) over time. **e-f)** Representative widefield fluorescence images of 3:1 (Lys)$_{5-24}$:RNA coacervates with **e)** active ribozyme or **f)** inactive ribozyme and 10% Cy5-tagged S1 substrate on unpassivated slides. **g-h)** Bar plot showing both the change in **g)** droplet area over time and **h)** particles per area (100 µm^2) over time. Data from **(c, g)** are presented as mean ± standard deviation (n=9), data from (d, h) are from a single measurement. Scale bars = 10µm.*

To better understand the cause of the above differences in behaviour spurred by shifting of the average RNA length of the system from short to long, $(Lys)_n$ peptides were added to solutions of equimolar substrate and ribozyme RNA that were allowed to react beforehand and were subsequently monitored by fluorescence microscopy. These pre-reacted RNA mixtures could potentially be informative of the final physical state of coacervates with active systems, since RNA catalysis did not occur simultaneously with coacervation but

rather in advance, such that the already elongated products subsequently form the compartments. Mixtures containing the active ribozyme and therefore, pre-concatenated RNA initially formed non-spherical, adhesive, gel-like condensates with both peptides after 1 hour of incubation, which relaxed to form spherical droplets over the course of 24 hours at 30 °C, whereas coacervates with inactive ribozyme yielded spherical, non-adhesive condensates from the onset of $(Lys)_n$ addition **(Figure 35a-b)**. These observations suggested that large morphological differences between active and inactive systems are not observed when peptide and RNA are combined from the start because the average RNA length is initially identical. As the reaction proceeded in the coacervates with active systems, a transition to a more viscous or gel-like state may occur while maintaining the initial spherical morphology. This effect hinted at the different timescales for coacervation and ligation, indicating that the spontaneous coalescence of phase-separated coacervates upon $(Lys)_n$ addition occurs faster than the rate of ribozyme-mediated RNA concatenation.

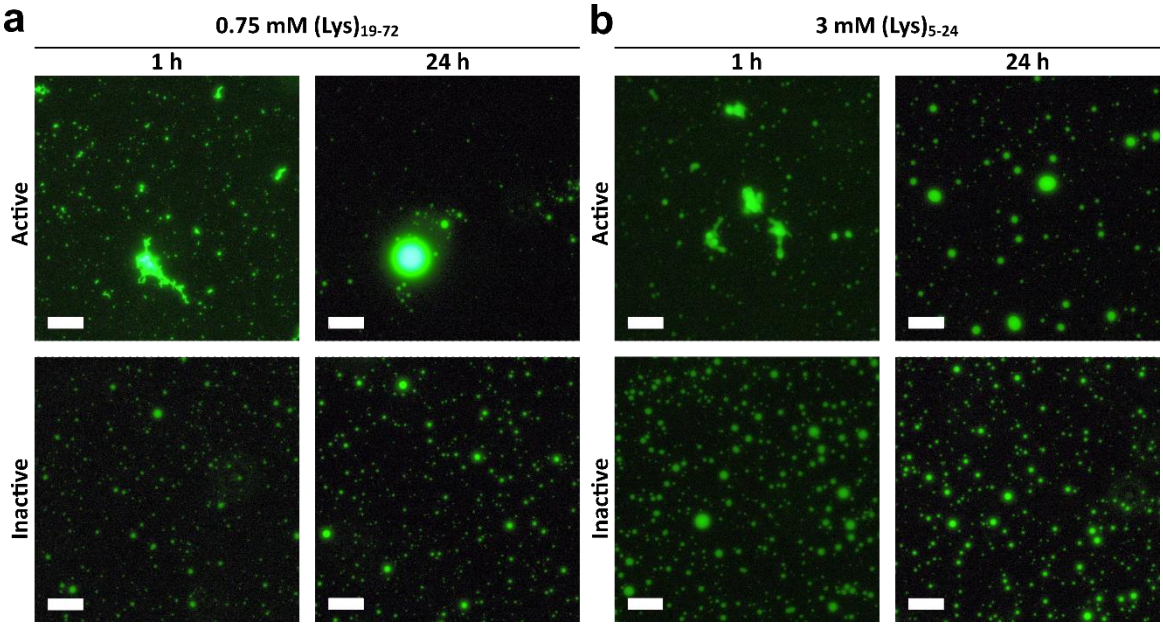


Figure 35: Monitoring coacervates formed from pre-reacted active or inactive RNA with either 0.75 mM $(Lys)_{19-72}$ or 3 mM $(Lys)_{5-24}$. *Equimolar concentrations (10.5 µM each) of ribozyme and substrate (10% FAM-tagged) were combined to achieve a 1 mM total RNA charge, then allowed to react for 2 hours at 30 °C in solution, before the addition of either* ***a)*** *0.75 mM $(Lys)_{19-72}$ or* ***b)*** *3 mM $(Lys)_{5-24}$ and further incubating at 30 °C for 24 hours to monitor coacervates composed of pre-reacted RNA. Images were acquired by fluorescence microscopy 1 hour and 24 hours after the addition of $(Lys)_n$ peptides. Scale bars = 20 µm.*

Given the effect of RNA concatenation on the coacervate droplets, the effect of ribozyme concatenation activity on the interactions between different populations of droplets was explored. Using the previously established conditions and concentration ratios,

namely, 1 mM total monomer concentration of RNA with equimolar ribozyme and substrate were combined with either sub-stoichiometric $(Lys)_{19-72}$ or excess $(Lys)_{5-24}$ peptides (0.75 or 3 mM monomer concentration, respectively), two populations of coacervates produced under the same conditions but containing either Cy5- or FAM-tagged S1 substrate (10 % total substrate concentration) were combined and monitored by confocal microscopy for mixing and passive content exchange over the course of 24 hours. Mixing within individual coacervates and among populations of coacervates was quantified by the calculation of a Pearson correlation coefficient (PCC), which measures the correlation of pixel intensities between the two fluorescence channels, or in other words, the co-localisation of two pixels from different fluorescent channels (*238*). After image processing, two coefficients were calculated by Kris Le Vay (former postdoctoral associate in our group): $PCC_{droplet}$, which measures the degree to which fluorophores mix within single merged droplets revealing diffusion dynamics within individual coacervates, and PCC_{pop}, which is calculated on a population level and measures the degree of mixing between the two populations informing on the degree of coalescence. Positive values indicate spatial colocalization of fluorophores, while negative values indicate spatial separation of fluorophores and therefore, the presence of either distinct sub-compartments within individual coacervates for $PCC_{droplet}$ or distinct populations of coacervates for PCC_{pop}.

At the start of the experiment, discrete Cy5- or FAM-tagged coacervate populations with spherical morphologies were visible, as well as partly fused and non-homogenously mixed coacervates containing both fluorophores that are physically separated, evidencing coalescence and fusion events as the mechanism of growth in the system **(Figure 36a-b, g-h)**. For coacervates with 0.75 mM $(Lys)_{19-72}$, $PCC_{droplet}$ values were relatively low after 30 minutes, confirming that that the initially formed droplets were in fact distinct **(Figure 36c)**. The value gradually increases over time tending towards unity after 24 hours, indicating that the differently tagged substrates were able to equilibrate within each coacervate via diffusion. PCC_{pop} values were negative after 30 minutes for both active and inactive systems with $(Lys)_{19-72}$, corroborating the previous claim that discrete populations of coacervates are initially formed **(Figure 36d)**. Nonetheless, the trend changed, highlighting different behaviours of active and inactive systems as the PCC_{pop} values for the active system persisted in the negative range whereas that of the inactive system became largely positive after 24 hours. Moreover, the coacervate area and number remained constant for active coacervates, whereas inactive coacervates showed an increase in area and decrease in number as previously reported **(Figure 36e-f)**. Taken together, the data provided convincing

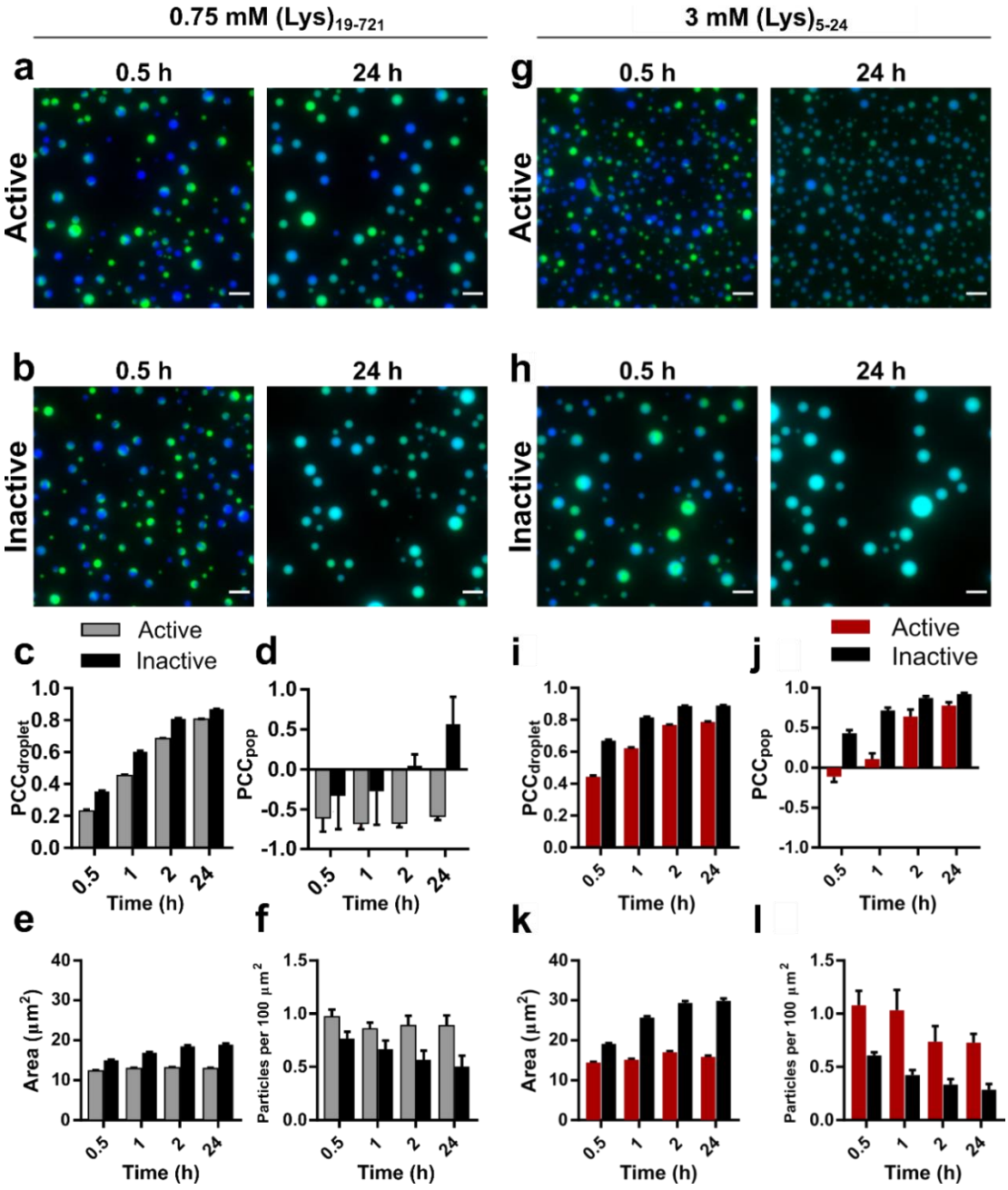


Figure 36: (Data was acquired by Kristian Le Vay and reproduced from (235)***) Mixing and passive content exchange among coacervate populations comprised of active or inactive RNA systems with either 0.75 mM (Lys)$_{19-72}$ or 3 mM (Lys)$_{5-24}$. a-b)*** *Representative widefield fluorescence images of two mixed populations of 0.75:1 mM (Lys)$_{19-72}$:RNA coacervates with **a)** active or **b)** inactive ribozyme and either 10% Cy5- or FAM-tagged S1 substrate. **c-d)** Bar plots showing the change in Pearson Correlation Coefficient (PCC) for **c)** individual droplets or **d)** the whole population. **e-f)** Bar plot showing both the change in **e)** droplet area over time and **f)** particles per area (100 µm^2) over time. **g-h)** Representative widefield fluorescence images of two mixed populations of 3:1 mM (Lys)$_{5-24}$:RNA coacervates with **g)** active and **h)** inactive ribozyme and either 10% Cy5- or FAM-tagged S1 substrate. **i-j)** Bar plots showing the change in Pearson Correlation Coefficient (PCC) for **i)** individual droplets or **j)** the whole population. **k-l)** Bar plot showing both the change in **k)** droplet area over time and **l)** particles per area (100 µm^2) over time. All data are presented as mean ± standard deviation. Scale bars = 10 µm. Blue colour = Cy5; green colour = FAM.*

evidence that the inactive coacervate system did indeed grow by coalescence and that the fused coacervates all equilibrated leading to a homogeneous population of coacervates after 24 hours. The same could not be said about active droplets, which at the end of experiment, still partially retained their individual identities as coacervates containing either FAM- or Cy5-tagged substrates.

A similar trend was noted for coacervates prepared with $(Lys)_{5-24}$, where final $PCC_{droplet}$ values reached the same values as with the longer $(Lys)_{19-72}$ **(Figure 36i)**. However, initial $PCC_{droplet}$ values were higher, resulting in a smaller change in $PCC_{droplet}$ values and a faster equilibration of fluorophores within coacervates, suggesting that the $(Lys)_{5-24}$ coacervates were more liquid-like, possessed a lower viscosity and higher diffusivity than coacervates formed from $(Lys)_{19-72}$ for both conditions with active or inactive ribozyme. As for PCC_{pop} values, a difference in behaviour between coacervates with active or inactive ribozyme was also observed, however, but was less pronounced than those with the longer $(Lys)_{19-72}$ peptides, likely due to the difference in condensed phase properties imparted by the longer length peptides **(Figure 36j)**. The change in both coacervate area and number over time approximately coincided with that observed in the previous experiments, particularly that active coacervates tended to remain relatively static throughout the experiment, while the inactive coacervates exhibited changes in area and number corresponding the fusion events **(Figure 36k-l)**. Therefore, with shorter peptides, the presence of active ribozyme slowed the rate of coalescence and content exchange rather than inhibiting it completely. These data further support the hypothesis that active systems impart unique characteristics to the compartments hosting them, as opposed to inactive systems. Specifically, a striking difference in their growth behaviour was observed likely due to the suppression of coalescence and perhaps, passive growth.

Although coacervates retained their spherical morphology over the course of the experiment, the inhibition of growth and wetting might also be caused by the transition of the condensed phase to a more viscous, gel-like state as a result of the elongation of substrate RNA over time. Thus, it seems plausible that coacervates formed from $(Lys)_n$ and active RNA initially were liquid-like but transitioned to a more gel-like phase as catalysis proceeded, thereby enabling the formation of spherical but gel-like condensates. This claim is supported by the data shown involving the production of coacervates with pre-reacted RNA **(Figure 35)**, although more data is required to draw conclusive inferences regarding the exact mechanism of modulation between ribozyme activity and compartment phenotype.

3.2.6. Summary and conclusions

In this section, the chief motivation was to design a ribozyme system capable of repeated RNA ligation thereby significantly increasing the average RNA length in the system, in effort to explore the effects of such a reaction on the physical properties of complex coacervate microcompartments. A well-characterised RNA ligase (R3C ligase) was re-engineered to ligate the same substrate together, and was shown to produce ligated RNA strands well over 500 nucleotides starting from 31 nucleotide substrates and included circular products as well. Next, ribozyme activity was assayed while compartmentalised in complex coacervates formed with lysine oligopeptides, revealing a strong rate enhancement effect for ribozyme catalysis as well as the inhibition of circular product formation. Reciprocally, ribozyme ligation activity endowed the coacervate compartments with altered material properties including the inhibition of droplet growth, surface wetting and inter-compartment content exchange. In conclusion, the data presented demonstrated how ribozyme sequence information could be connected to the physico-chemical characteristics of the coacervates, effectively linking genotype and phenotype in membrane-less compartments composed of RNA and peptides.

In this system, the phenotypic differences originated from the ribozyme-driven concatenation reaction, which increases the length of the anionic coacervate component. The sequence of the ribozyme in effect comprised the genotype of our protocellular system, where different variants (e.g., active or inactive sequences) produced different protocellular phenotypes. Other than genotypic variations (mutating sequences), the realisation of a phenotype-genotype linkage is also essential for true open-ended evolution, otherwise phenotypes with obvious fitness advantages would have no way of being propagated to future generations (*239*). The potential of artificially coupling genotype-phenotype for the directed evolution / selection of biomolecules was initially demonstrated by Tawfik and Griffiths already in 1998 (*161*), where inverse micelles produced via a microfluidic platform behaved as microcompartments effectively linking the phenotype of an enzyme (activity) to its genotype (DNA). Recently, the development of technologies that permit the sequencing of genetic components from individual coacervates was achieved (*240*). Then, for instance, the combination of this system with single-coacervate sequencing technologies will undoubtedly progress the knowledge of phase-separated systems and elucidate genotypic characteristics that directly affect individual fitness in a varying population of coacervates, potentially allowing selection experiments with multiple selective pressures.

Being a fundamental principle for protocellular Darwinian evolution, the linking of genotype and phenotype in an *in vitro* membrane-less model system also opens the door for more exploratory inquiries into the possible co-evolution of nucleic acids and proteins. As adduced in the introduction, many examples of symbiotic relationships, both commensal and mutualistic, between RNA and peptides have been empirically demonstrated. In this work, the mutualistic relationship lies between the genetic component of a compartment, which could potentially allow for the *in vitro* selection and evolution with two selective pressures directed at each, such as faster ligation for the ribozyme and maintenance of identity for compartment. An important limitation to acknowledge in this work is that to achieve such a rise in average substrate length, half of the compartment-comprising RNA was ribozyme RNA, which is not truly plausible in an evolutionary setting where active or inactive coacervates would be competing. In principle, this could be addressed by simply lowering the ribozyme concentration and selecting for faster catalysts, however, that might come at the cost of catalytic rate of increase in the average substrate length.

The peptide sequence is also of interest, as different sequences likely have varying effects on ribozyme catalysis and provide further opportunities for improvement. Similarly, different chemistries may exert a phenotypic effect on coacervates as well, particularly one that could drive the synthesis of longer peptides. RNA can catalyse both amide bond formation as well as ester bond formation, and thus could potentially elongate peptides or depsipeptides, which are mixed amino and hydroxy acid polymers (*241*, *242*).

In conclusion, this work represented perhaps the first experimental model of how a ribozyme can alter the phenotype of its compartment, and will enable the design of selection experiments based on populations of protocells with varying RNA genotypes, rather than simply on RNA function, while varying the degrees of survival with respect to environmental pressures such as temperature, ionic composition or strength. The current system employed for characterisation of RNA activity-responsive coacervates cannot be adopted for directed evolution experiments of the hosted RNA since it is not self-replicating, and thus must rely on the experimenter to replicate the sequence and introduce variability. Future work might explore the system in combination with RNA ligase ribozymes that undergo self-replication with the introduction of variation by mutation.

3.3. Far-from-equilibrium settings for RNA replication

3.3.1. Background

As we have seen, the RNA World hypothesis theorises a transitional state during the early evolution of life when the genetic and metabolic functions of a protocell was entirely orchestrated by RNA polymers. Conversely, replication of nucleic acid in extant cells is driven exclusively by protein catalysts. Natural selection has, over billions of years, evolved proteins that specifically solve the inherent obstacles that are faced during the replication of nucleic acids, particularly, the general (sequence-independent) templating mechanism of polymerases and duplex unwinding capabilities of helicases. For these reasons among others, the realisation of experimental systems for general templated RNA-dependent RNA replication *in vitro* (by an RNA replicase, for instance) has been severely limited. To achieve complete cycles of replication and allow downstream catalysis of functional RNA sequences that were previously copied, templates must be recycled and newly synthesised strands have to adopt their tertiary conformation required for catalysis **(Figure 37)**. Thus, a means to separate RNA duplexes resulting from templated replication becomes a pre-requisite for the sequence replication as well as the survival and evolution of RNA-based biological entities.

The melting of RNA duplexes *in situ* has been hindered by the following obstacles. First, RNA duplexes are inherently more stable than both DNA duplexes and their hybrid counterparts, presumably due to helix conformation as DNA or hybrid duplexes adopt B-form helices while RNA duplexes adopt an A-form helix, base pairing and base stacking as well as hydration level and counterion binding (*243*). Second, the metal ion requirements of certain ribozymes as well as of non-enzymatic templated primer extension reactions also greatly enhances duplex stability. Third, while traditional temperature-induced denaturation functions flawlessly for DNA, RNA polymers are significantly more labile with regards to elevated temperatures, alkaline pH and divalent metal ions due to the presence of 2'-hydroxyl groups, which promote extensive RNA degradation in solution (*244*). Finally, assuming strand separation can be achieved, the problem of strand reannealing must also be addressed, since the hybridisation of longer duplexes is favoured over the binding of shorter oligomers or nucleotides to the template (*173*). Thus, there seems to be trade-off between the conditions that favour RNA catalysis / stability and those that favour strand separation.

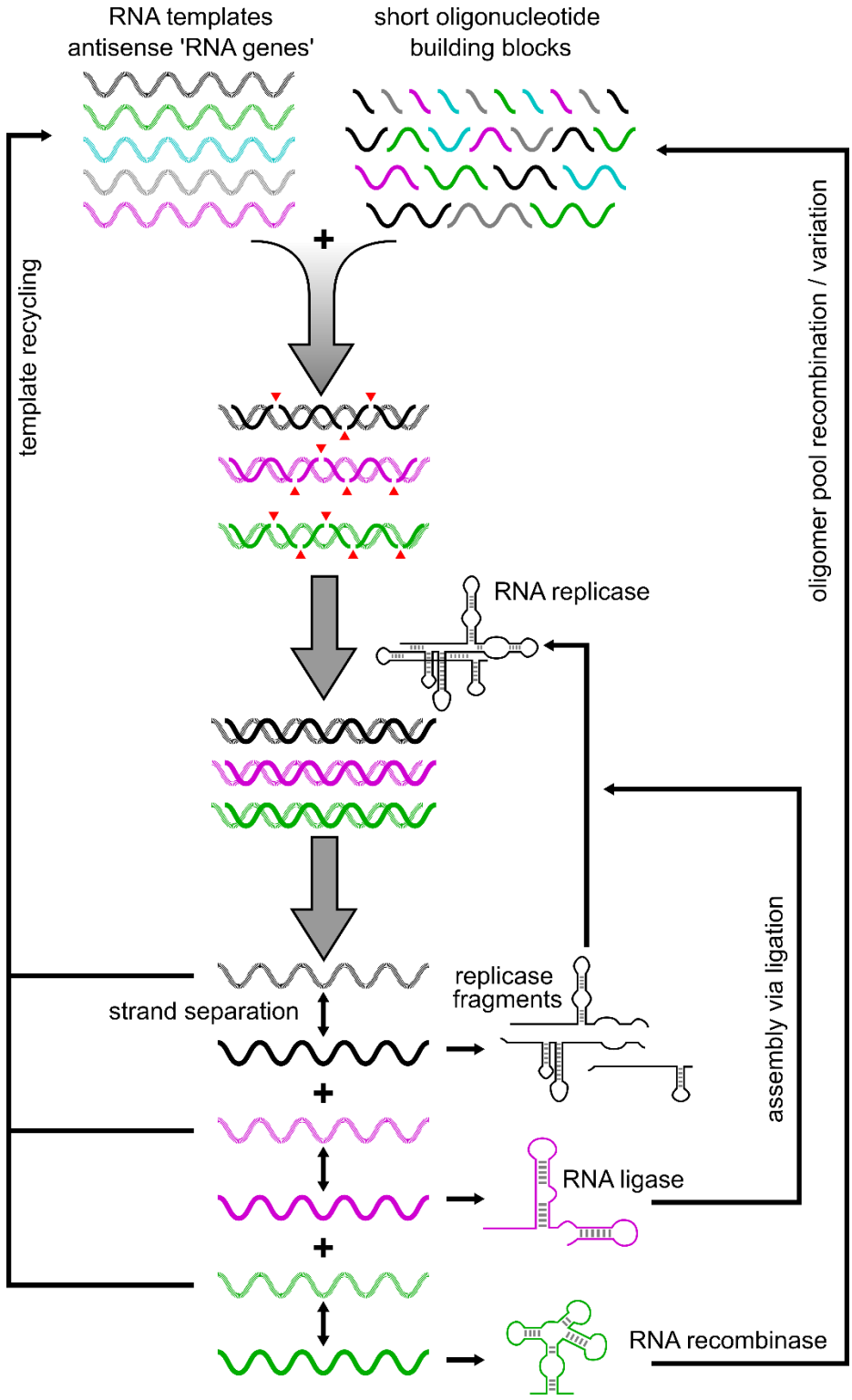


Figure 37: Schematic of RNA-based genetics in an RNA World. *The emergence of an RNA replicase would allow general template dependent replication of RNA genes. The templates would direct the hybridisation of oligomers found in the pool by complementarity, which would subsequently be synthesised by the RNA replicase. The resulting RNA duplex must be separated before the copied sequences can perform the desired function and the templates recycled to direct the synthesis of more ribozymes. The strand-separation and reannealing of duplex RNA has proved to be the bottleneck for achieving such RNA-base replication schemes.*

Due to the factors mentioned above, thermal denaturation of RNA duplexes has been limited to short strands that are less vulnerable to hydrolysis with increasing pH, metal ion concentration and temperature (*101*). Various stratagems have been concocted in efforts to circumvent the strand separation / reannealing problem. For instance, non-enzymatic template-dependent RNA polymerisation was shown to copy through double-stranded regions of RNA with the help of invader RNA strands that promote toe-hold mediated strand displacement and allow polymerisation through dsRNA (*63*). However, this mechanism necessitates short invader strands with sequences identical to the template, of which the source remains unclear. Another approach involves the use of nucleic acids with chimeric backbones and wobble bases, such that the effective melting temperature decreases and strand separation becomes attainable without sacrificing much in terms of stability (*245*, *246*). The disadvantage here is that chimeric nucleic acid backbones have yet to be shown to possess catalytic activity, and the ligation chemistry required for heterogeneous backbone ligation is different from that employed for homogeneous backbone ligation. To prevent strand reannealing, some have resorted to templated RNA ligation in viscous solvents (includes glycerol as a buffer component) generated periodically by wet-dry cycles of water evaporation and condensation, such that temperature induced melting is not followed by strand reannealing but rather by monomer binding, a consequence of the reduced diffusive coefficient of long strands as opposed to short polymers in viscous solvents (*247*, *248*). Finally, RNA synthesis from a circular genome has been discussed as a possible solution to the strand separation problem (*249*). Empirical support has come from the synthesis of an active minimal hammerhead ribozyme by the triplet polymerase from a circular RNA template (*250*). Nonetheless, this and other works typically rely on post-synthesis purification of the *de novo* RNA to observe any subsequent catalytic activity, precisely due to the presence of the template in solution and absence of single stranded RNA species. Therefore, despite the many attempts at RNA-mediated RNA replication, a scenario that clearly demonstrates the general synthesis of RNAs in a continuous one-pot system including complete cycles of RNA replication is still lacking.

Underexplored and potential solutions to this conundrum are geochemically plausible far-from-equilibrium systems, such as porous rock systems exposed to CO_2-rich atmospheres comprising heated air-water interfaces (AWI-system) that can be reproducibly replicated in the laboratory by microfabrication and have been shown to exhibit periodic changes in magnesium ion concentration and pH level, as discussed in the introduction(*214*). A working hypothesis was that the low pH would enable the transient melting of

RNA duplexes while the co-accumulation of RNA and magnesium ions would promote folding and catalysis at the air-water interface, thereby evading strand inhibition and allowing catalysis. To this end, a ribozyme system capable of general templated ligation for the synthesis of RNA sequences from shorter oligomer pools was characterised under different synthesis setups. Then, once the system was established and thoroughly explored, the feasibility of AWI-systems to host such an RNA catalyst was assessed together with our collaborators, and an attempt to couple *de novo* RNA synthesis of a minimal hammerhead ribozyme with strand separation and cleavage of its cognate substrate in the AWI-system was undertaken. The results obtained demonstrate the versatility of templated RNA ligases to perform sequence-independent templated ligation and the ability of AWI-systems to promote RNA replication in a one-pot reaction (*251*).

3.3.2. Templated RNA ligation of functional RNAs by *sunY*

A literature search revealed the existence of a group I self-splicing intron that was re-engineered to become a double-stranded RNA ligase that requires a 5'-guanosine as a leaving group. The intron is located in the *sunY* gene of bacteriophage T4, which had unknown function at the time, hence its name *sunY* (split gene, unknown function, why), but follow-up studies demonstrated that the *sunY* gene encodes for the SunY protein that functions as an anaerobic ribonucleotide reductase (*nrdD*), an essential enzyme for nucleotide metabolism (*252, 253*). A modified version of the *sunY* group I intron was generated by Doudna and Szostak via deletion of the 5'- and 3'-splice junctions (stems P1, P9.1 and P9.2), the distal non-essential stem (P2), the open reading frame (ORF) encoding a homing endonuclease and finally, including three point substitution mutations from the analogous thymidylate synthase (*td*) gene group I intron that exhibited higher self-splicing activity *in vitro* (*79*). The resulting ribozyme is 182 nucleotides long and adopts a complex secondary structure **(Figure 38a)**.

In this format, the *sunY* ribozyme can catalyse phospho-ester exchange reactions between two oligomers on an external template with relatively simple sequence requirements, namely, a G•U Wobble base at the ligation junction for optimal activity and possibly, 3'-5' regiospecificity **(Figure 38b)** (*74, 254*). As a general rule, all efficient splicing events of group I introns occur directly downstream of a G residue, owing to the self-splicing chemistry, meaning that any G residue could potentially act as a recognition site for splicing to occur, given enough sequence complementarity between substrate oligomer and template. The reaction carried out resembles the exon ligation step (the second step of self-

splicing), in which the 3'-hydroxyl of the upstream oligonucleotide performs an in-line attack on the phosphate between the terminal 5'-guanosine and the second residue of the downstream oligomer resulting in ligation of the two oligomers and release of a guanosine **(Figure 38c)**. Despite the versatility of this ribozyme in regards to the types of cascade reactions that can be designed, it was soon abandoned by its discoverers probably because the templated RNA assembly reactions persistently led to the formation of dead-end duplexes, effectively inhibiting downstream applications.

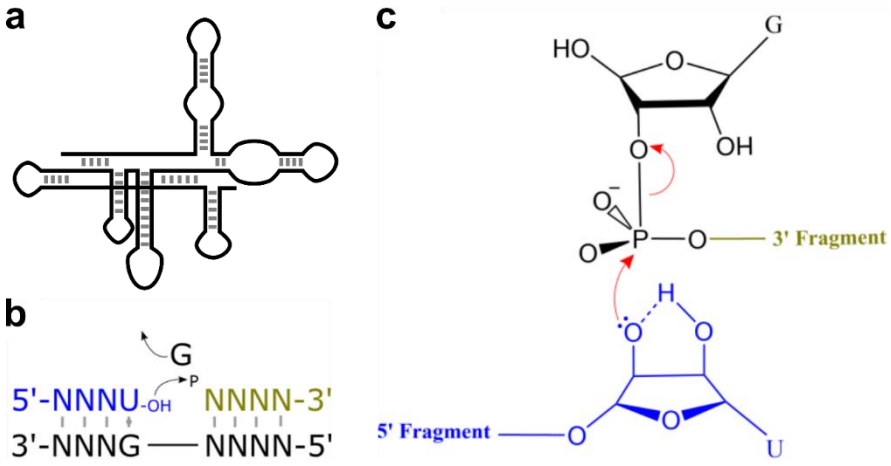


Figure 38: Structure and function of the sunY ribozyme. a) *Schematic of the secondary structure of the group I intron-derived sunY ribozyme.* ***b)*** *Sequence-optimisation of the ligation junction for catalytic rate of the sunY -catalysed ligation.* ***c)*** *Mechanism of ligation promoted by the ribozyme leading to ligation of 5'- and 3'-substrates and release of guanosine, analogous to the exon ligation step during splicing.*

Initially, the *sunY* ribozyme was tested for activity and further characterised in vitro to gain insight into the conditions that are optimal as well as to explore the sequence space for substrate design. To this end, an aptamer assembly reaction was designed in which three oligomers are ligated on a template to form a functional fluorogenic aptamer **(Figure 39a)**. In this way, sequence specific ligation could be ascertained, since after separation of product and template by denaturing PAGE and gel staining with the aptamer ligand solution, only correct sequences with properly folded tertiary structures will bind the ligand, serving as an indirect measure of fidelity. The Mango III and minimal Broccoli aptamers were selected for their short sequence and strong fluorescence enhancement upon ligand binding (*255*, *256*). Each aptamer was interrupted at two critical residues required to adopt a functional tertiary conformation such that the oligomer fragments would not associate non-covalently and bind the ligand even in the absence of ligation, producing unwanted fluorescence and false positives. The template sequence was designed as the

complement except for the ligation junctions where G•U wobble bases introduced for enhanced rate and the ends were appended with additional nucleotides to allow for different migration distances in PAGE **(Figure 39b-c)**.

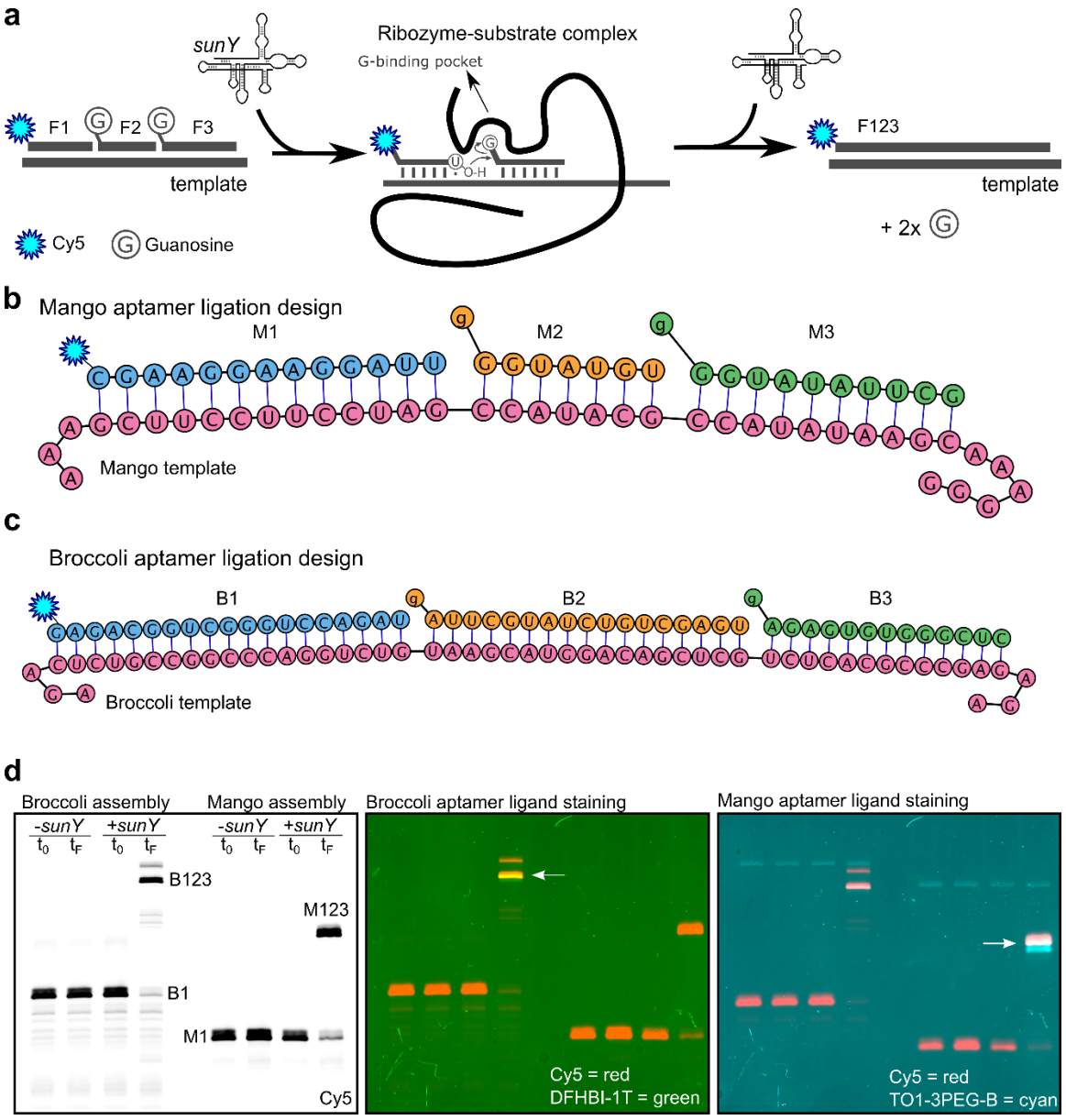


Figure 39: The sunY ribozyme ligates functional aptamers by templated ligation of shorter oligomers. a) *Schematic showing the assembly design for ligation by sunY ribozyme: upon template binding, the first fragment (Cy5-tagged) is ligated to the other 2 fragments (carrying a 5'-guanosine) leading to full-length aptamer assembly. The ribozyme coordinates the ligation site in the active site by binding to the terminal 5'-guanosine in the G-binding pocket.* ***b)*** *Design of mango aptamer assembly.* ***c)*** *Design of broccoli aptamer assembly.* ***d)*** *20% PAGE analysis of broccoli and mango assembly by the sunY ribozyme. Reactions were in 30 mM Tris-HCl pH 7.5, 100 mM KCl, 20 mM $MgCl_2$ and 10% v/v PEG 8000 with 1 µM sunY, 10 µM Cy5-M1/B1, 20 µM of each M2/M3 or B2/B3 and 10 µM of template for 3.5 hours at 45 °C. Left gel shows Cy5 channel, centre gel shows Cy5 (red) and broccoli ligand DFHBI-1T (green) overlay, right gel shows Cy5 (red) and mango ligand TO1-3PEG-Biotin (cyan) overlay. Note that the Mango aptamer ligand can also be bound by the broccoli aptamer, although to a lesser degree.*

The systems were first tested under multiple turnover conditions with 10-fold excess substrate for 3.5 hours at 45 °C. Satisfyingly, the PAGE analysis revealed that almost all substrate was being converted into product in the presence of *sunY* ribozyme **(Figure 39d, Cy5 channel)**. After staining with the respective aptamer ligand solution, the ligated product was indeed functional and could fold and bind its cognate ligand in a polyacrylamide matrix, confirming the sequence of the ligated aptamer **(Figure 39d, white arrows)**. Of the two assembly reactions tested, the Mango III aptamer system was selected for further investigation as the products of this ligation reaction yielded fewer unspecific ligation products, visible in the Broccoli ligation assembly as an extraneous band above the main ligation products. This result agrees with what can be predicted from their sequences, since Broccoli aptamer is has more similar sequence repeats and thus is more prone to non-specific ligation events.

Seeing as the incubation time in the previous experiment was excessive, the timescales at which the *sunY* ribozyme can perform the ligation assembly reaction of the Mango III aptamer was examined. As such, a kinetic experiment was performed with 5-fold excess substrate and template over ribozyme and sampled the reaction over the course of one hour at 45 °C. The PAGE analysis revealed that product formation takes place rather rapidly and almost reaches completion during the 60-minute incubation **(Figure 40a)**. Global fitting of a two-step irreversible reaction with incomplete product conversion (see Equations (1), (2) and (3) in Methods section) yielded two rate constants, k_1 for conversion of M1 to M12 and k_2 for the conversion of M12 to M123, with best fit values of $k_1 = 0.139$ min^{-1} and estimated 95% confidence interval of [0.135-0.142], while best fit values of $k_2 = 0.268$ min^{-1} and an estimated 95% confidence interval of [0.257-0.280] **(Figure 40b)**. The kinetics of M123 product formation were characterised by a rapid initial increase followed by a slow, gradual increase towards a plateau, which are typical of first-order reactions that are dependent on substrate concentration. At the endpoint, final M123 yields plateau at approximately 78%, with the remaining substrate being divided between unligated substrate (M1, ~13%) and the intermediate (M12, ~9%), suggesting that substrate inhibition and formation of unproductive complexes are hindering stoichiometric conversion of substrate to product. Together, these data demonstrate that the *sunY* ribozyme possesses the capabilities for multiple turn-over ligation and a catalytic rate sufficient for the assembly ligations of RNA sequences at short timescales.

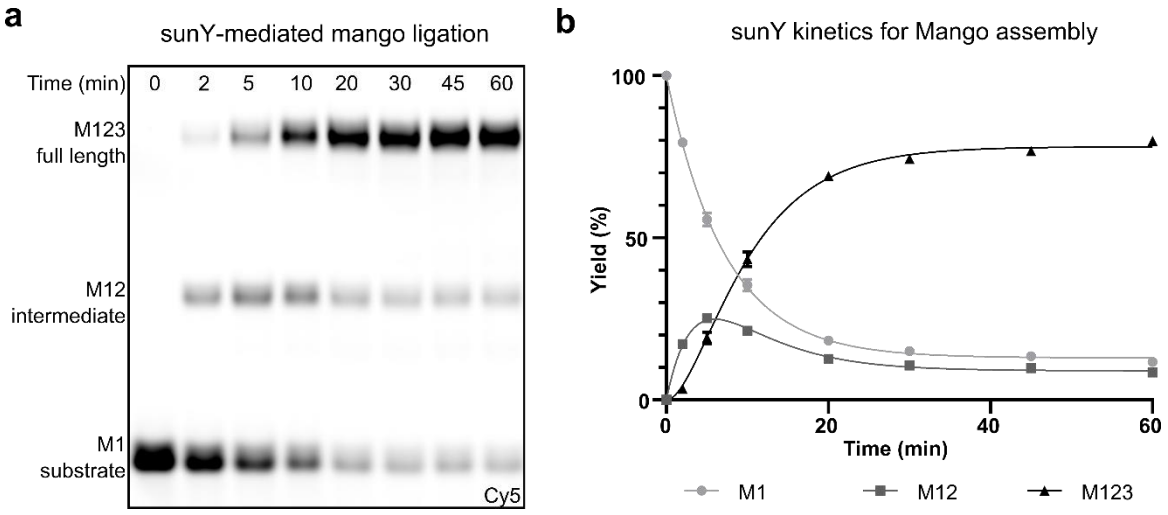


Figure 40: Kinetic analysis of mango assembly by sunY. a) *A 20% PAGE showing the mango ligation reaction performed with 1 µM sunY, 5 µM of each substrate and template, in 30 mM Tris-HCl pH 7.5, 100 mM KCl and 100 mM $MgCl_2$ at 45 °C. Samples were quenched at the specified timepoints and subsequently resolved.* ***b)*** *Scatter plot showing band intensity estimations from* ***a)****. Data are presented as mean ± standard deviation (n=3) and are fit globally with a two-step irreversible reaction (see Equations (1), (2) and (3) in Methods).*

3.3.3. Simultaneous ligation on both substrate strands

Having demonstrated templated ligation and estimated the reaction rate, a different sequence design assembly was subsequently explored, particularly one that is not strictly dependent on the presence of a template RNA. To this end, a reaction scheme in which all the RNA oligomers template each other in a form of splinted ligation was set up **(Figure 41a)**. All fragments initially present serve as both substrate and template so as to eliminate the need for a full-length template that is not directly involved in a phosphorester exchange reaction, and effectively combining the functions in the system. By interrupting a quasi-random dsRNA at two points on one strand and a single point on the complementary strand and further optimising sequence for minimal substrate inhibition while maximising ligation activity, a 24-mer dsRNA product could therefore be ligated from five substrate oligomer fragments by the *sunY* ribozyme **(Figure 41b)**. Sequence optimisation was manually performed by first comparing free energy (ΔG) predictions such that heterodimers (hybridisation between top and bottom strands; e.g., F1+F5, F3+F4, etc.) have higher binding energy (lower ΔG) than homodimers, and in a second step optimising for minimal inhibition (highest ΔG) between oligomer fragments from the same strand (hybridisations between F1+F2, F1+F3, F4+F5). The fragments F1 and F5 were fluorescently tagged at the 5'- and 3'-end with FAM and TAMRA, respectively, for visualisation by PAGE.

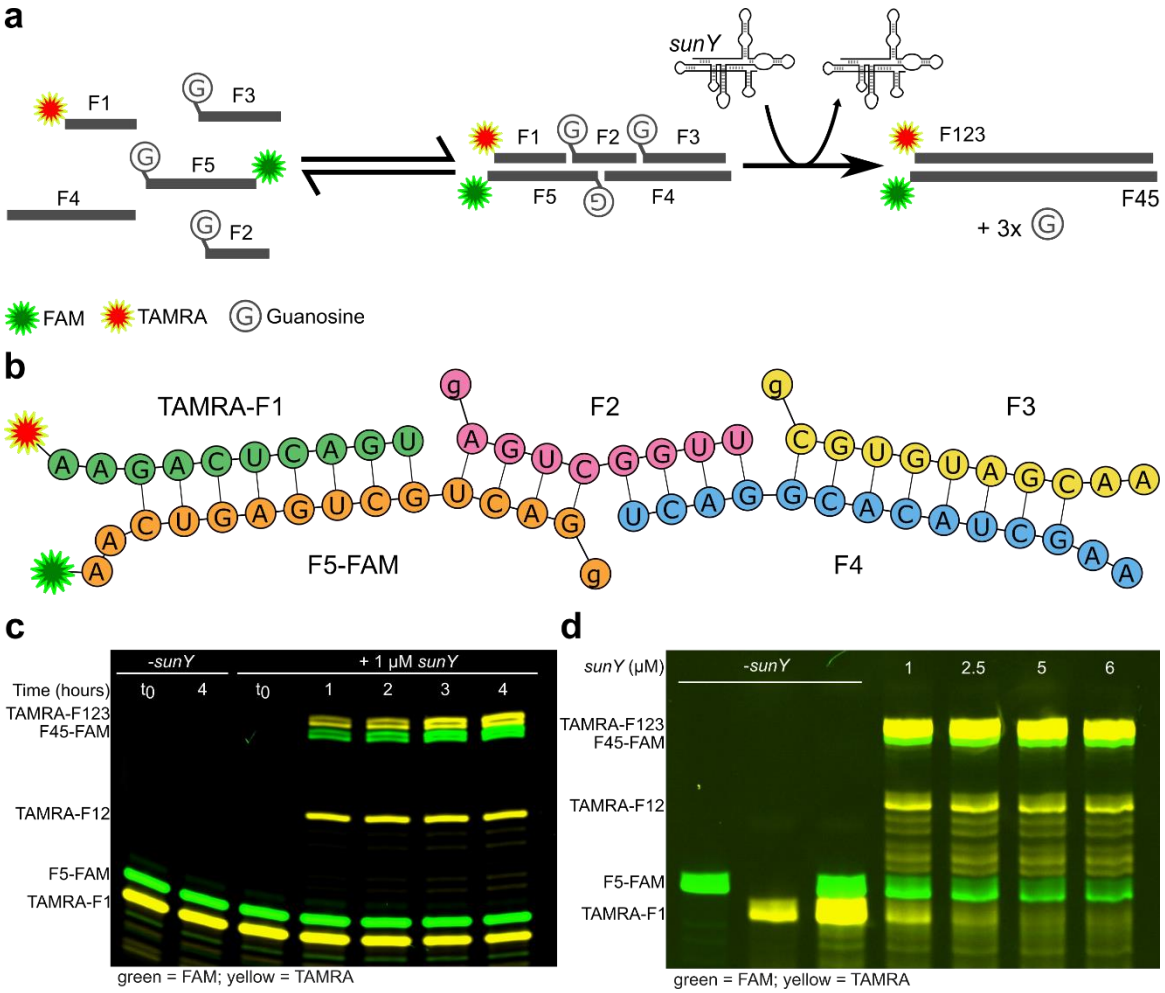


Figure 41: The sunY ribozyme ligates short splints into full length sequences. a) *Schematic of experimental design for sunY ligation using short splints exclusively.* ***b)*** *Sequence design for splinted ligation. Fragment 1 (F1) has a 5'-TAMRA fluorophore whereas fragment 5 opposite has a 3'-FAM fluorophore.* ***c)*** *A 20% PAGE showing the reaction products of splinted ligation with 10 µM of each substrate (F1-F5) and 2.5 µM sunY in 30 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$ and 10% v/v PEG 8000 at 45 °C. Negative control lacks ribozyme (-sunY; lanes 1-2). Samples were taken and quenched at the specified timepoints before resolving.* ***d)*** *A 20% PAGE showing the effect of increasing sunY concentration on reaction yields. The reactions contained 5 µM of each substrate (F1-F5), 30 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$ and 100 mM KCl, with varying concentration of sunY ribozyme. Negative controls lack ribozyme (-sunY; lanes 1-3). Reactions were quenched and resolved after 20 hours of incubation at 45 °C.*

This system was expected to perform worse than the aptamer assemblies previously employed, since the hybridisation strengths of the duplexes formed with a splint are significantly lower than that of duplexes formed with a full-length template. With that in mind, a time dependent analysis of splinted ligation was performed with 4-fold excess substrates over ribozyme for 4 hours at 45 °C **(Figure 41c)**. Product formation was detected for both strands, indicating that ligation was taking place independently of the location of the junction. Omitting *sunY* from the reaction prevented product formation as anticipated.

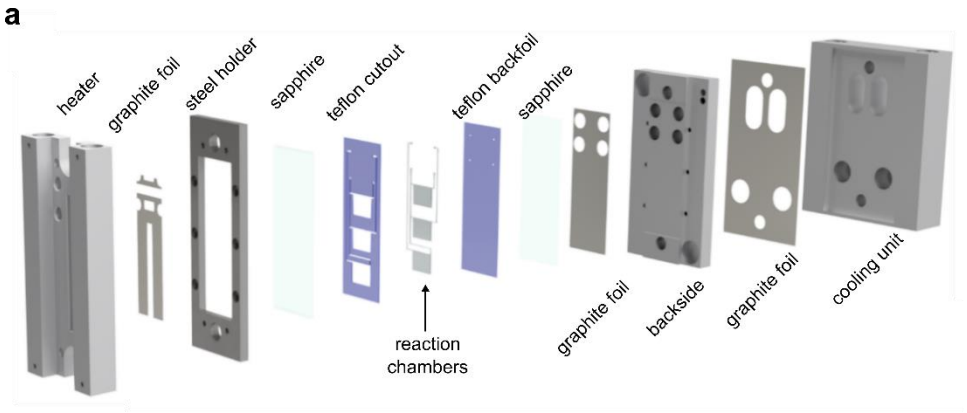
Despite substantial ligation activity after 4 hours of incubation, a significant amount of substrate remained unreacted. In efforts to enhance yields, the reaction was repeated with altered parameters, namely, an increased concentration of Mg^{2+} ions and *sunY* ribozyme, a reduced concentration of substrates and a longer incubation time **(Figure 41d)**. A higher yield than previously was obtained, although it was accompanied by a significant amount of RNA degradation, seen as banding patterns below the main substrate and product bands, likely a result of the increased incubation time in higher salt concentrations. When the concentration of *sunY* was 2.5 µM, the reaction showed enhanced yields but was still under a sub-stoichiometric regime of ribozyme to substrate. Consequently, this ribozyme concentration was selected for implementation in subsequent experiments. The results obtained here confirmed the versatility of the *sunY* ribozyme as a ligase system for the templated synthesis of RNA from shorter fragments. Despite its modularity, designing ribozyme cascades or synthesis networks remained limited *in vitro* due to strand inhibition. In an attempt to find a solution to this conundrum, far-from-equilibrium environments that could promote product release for downstream activity and template recycling were adopted and tested for compatibility with RNA catalysts.

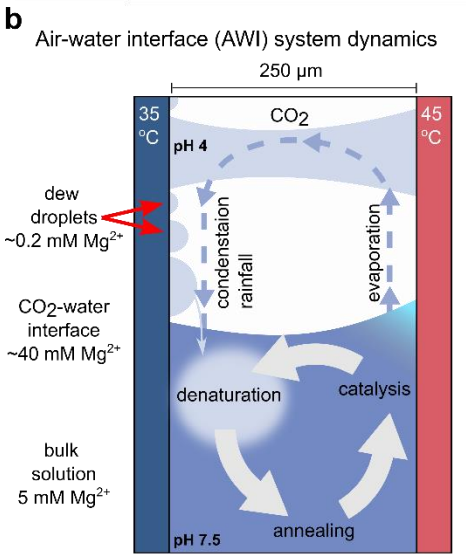
3.3.4. RNA ligation in a far-from-equilibrium setting

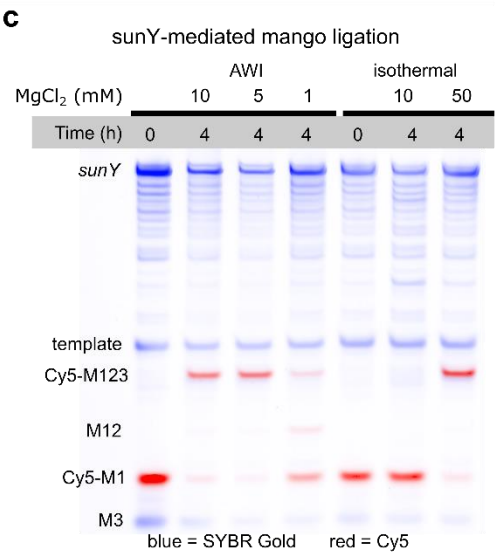
Such an out-of-equilibrium setting has been previously explored and thoroughly discussed, particularly heated air-water interfaces (AWI-systems) that can be reproducibly microfabricated by sandwiching the different components to create the reaction chamber between a heater and cooler unit **(Figure 42a)** (*214–218*). In CO_2-based AWI-systems, a temperature gradient across the compartment generates the dynamics that are pre-requisite for strand separation **(Figure 42b)**. Briefly, the gradient leads to water evaporation on the warm side and re-condensation on the cooler side. The condensing droplets can grow by coalescence until they become sufficiently large to precipitate back into solution. If the droplets are too small and fail to grow, they can eventually dry out. At times, large droplets can form water bridges between the warm and cool side, which ultimately fall back into solution. The level of bulk buffer solution therefore constantly rises and falls, drying material and subsequently dissolving it. Importantly, falling dew droplets are depleted of ions and slightly acidic as a result of the CO_2 in the air. The system was hypothesised to be suitable for RNA mediated ligation under sub-optimal magnesium chloride concentrations required for in vitro catalysis due to the ability of the AWI-system to concentrate nucleic acids and ions at the air-water interface. To this end, the magnesium dependence of the

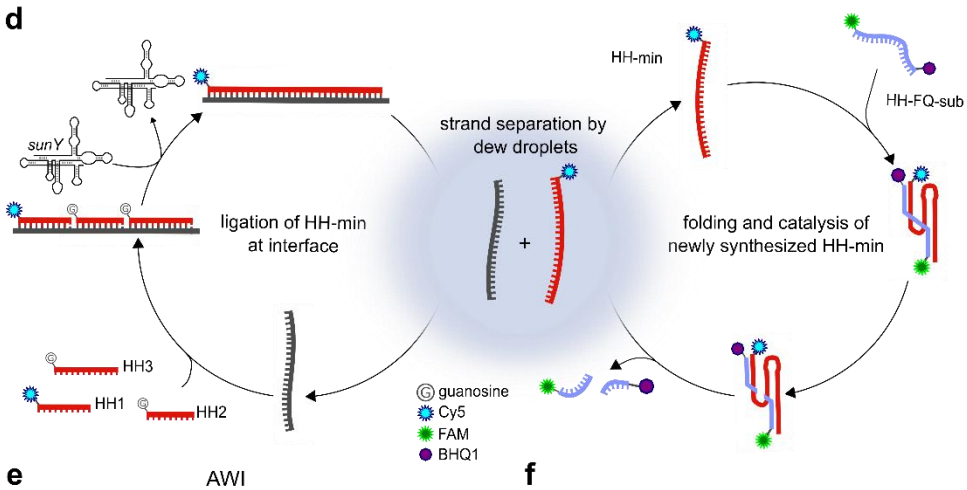
sunY ribozyme was initially explored by our collaborators in Munich with the previously characterised Mango III assembly system. The ribozyme required a 10-fold lower bulk magnesium concentration compared to the isothermal control at 45 °C to achieve similar ligation yields of full-length product, Cy5-M123 **(Figure 42c)**. After 4 hours of incubation at 5 mM $MgCl_2$ bulk concentration in the AWI-system, 89% ligation yield was obtained, while 50 mM were required at constant temperature to obtain 86% product yields. Surprisingly, product formation could be detected even as low as 1 mM bulk concentration of $MgCl_2$ in the AWI-system. The data confirmed that the system is not only suitable for ribozyme activity, but also rescues activity at bulk magnesium chloride concentrations considerably lower than those required under isothermal conditions, likely due to the local up-concentration of solutes at warm side of the air-water interface achieved in the AWI-system.

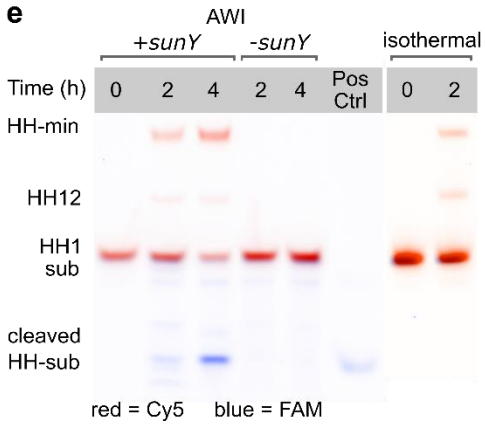
The positive results motivated further investigation of RNA catalysis in the AWI-system, particularly because most substrate was converted to full-length product. It was postulated that the release of template and ligation product driven by precipitating dew droplets could allow folding of the nascent strand into functionally active conformation, similar to what occurs during transcription in extant cells. Therefore, a HH-min assembly system was designed for *sunY*-mediated ligation and tested in the AWI-system **(Figure 42d)**. To generate the three substrate oligomers to be ligated, HH-min structure was interrupted at the stem loop and the catalytic site (CUGA bulge) to ensure inactivity of individual fragments even through non-covalent association. In this system, HH-min is synthesised by *sunY* from three RNA substrates (HH1-3) via templated ligation. Ligation assays with 2- or 4-fold excess substrates over template revealed the formation of full-length HH-min under isothermal conditions with 50 mM $MgCl_2$, whereas similar yields were obtained in the AWI-system at 10-fold lower bulk magnesium chloride (5 mM), corroborating the pervious hypothesis regarding the up-concentration of solutes achieved by such a dynamic far-from-equilibrium environment **(Figure 42e)**. While isothermal conditions showed stoichiometric conversion of substrate in relation to template (single-turnover), excess HH-min product over template could be detected in the AWI-system, suggesting that far-from-equilibrium conditions enabled template recycling whereas all template:product complexes remained tightly bound under isothermal conditions. If the HH-min synthesised in solution was indeed being released from the template to allow for multi-turnover ligation, then the solution should contain free full-length HH-min substrate. To confirm template release and probe the activity of *de novo* synthesised HH-min, the experiment was repeated but included a small amount of the cognate HH-min substrate carrying a 5'-FAM and 3'-BHQ1


a
heater
graphite foil
steel holder
sapphire
teflon cutout
teflon backfoil
sapphire
graphite foil
backside
graphite foil
cooling unit
reaction chambers


b
Air-water interface (AWI) system dynamics
250 µm
35 °C
45 °C
pH 4
pH 7.5
CO2
dew droplets ~0.2 mM Mg2+
CO2-water interface ~40 mM Mg2+
bulk solution 5 mM Mg2+
condenstaion
rainfall
evaporation
denaturation
catalysis
annealing


c
sunY-mediated mango ligation
AWI
isothermal
MgCl2 (mM) 10 5 1 10 50
Time (h) 0 4 4 4 0 4 4
sunY
template
Cy5-M123
M12
Cy5-M1
M3
blue = SYBR Gold
red = Cy5


d
sunY
ligation of HH-min at interface
strand separation by dew droplets
HH-min
HH-FQ-sub
folding and catalysis of newly synthesized HH-min
HH3
HH1
HH2
guanosine
Cy5
FAM
BHQ1


e
AWI
+sunY
-sunY
isothermal
Time (h) 0 2 4 2 4 Pos Ctrl 0 2
HH-min
HH12
HH1 sub
cleaved HH-sub
red = Cy5
blue = FAM

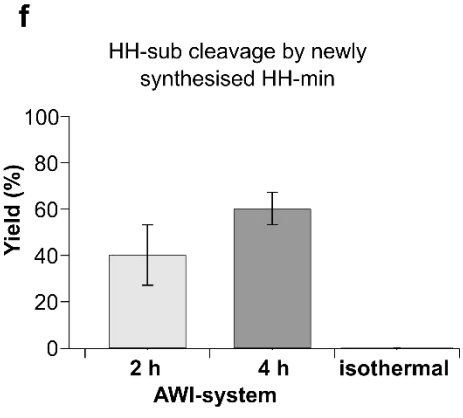

f
HH-sub cleavage by newly synthesised HH-min
Yield (%)
100
80
60
40
20
0
2 h
4 h
isothermal
AWI-system

***Figure 42: (Figure was produced by Annalena Salditt and reproduced from** (251))**: Microfabricated heated rock pores support ribozyme catalysis. a)** Scheme showing the components that are sandwiched to create the reaction chambers containing the heated air-water interface. **b)** Schematic showing the dynamics occurring inside the reaction chamber as a result of temperature gradient and CO_2-buffer interface. **c)** A 15% PAGE showing the reaction products of the sunY-mediated mango aptamer ligation (see Figure 37) in solution (bulk; isothermal) and in the AWI-system at various magnesium concentrations. The reactions contained 20 µM of each substrate, 5 µM template and 2.5 µM sunY in 30 mM Tris-HCl pH 7.5, 100 mM KCl and the specified $MgCl_2$ concentrations. Isothermal reactions were carried out at 45 °C. **d)** Reaction scheme for the sunY-mediated templated ligation of Hammerhead ribozyme (HH-min) from three oligomers (HH1-3) at the interface in the AWI-system, followed by denaturation of newly synthesised HH-min:template complex by precipitating dew droplets and finally folding, binding, and cleavage of its cognate substrate (HH-FQ-sub). **e)** A 15% PAGE showing the products of HH-min assembly by sunY and HH-FQ-sub cleavage. The reactions contained 10 or 12 µM of substrates (H1-3) and 5 or 3 µM of template, 2.5 µM sunY, 0.4 µM HH-FQ-sub in 30 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$ and 100 mM KCl. Isothermal controls were identical except for 50 mM $MgCl_2$ and constant 45 °C incubation. Positive control shows cleaved HH-FQ-sub in solution. **f)** Bar graph showing yields from PAGE analysis of cleaved HH-FQ-sub after 2 and 4 hours of incubation in the AWI-system. Data are presented as mean ± standard error of mean (n=10 for 2 h; n=4 for 4 h).*

tags (HH-FQ-sub), in hopes of detecting catalytic substrate cleavage. PAGE analysis delightfully revealed that HH-min formation was accompanied by HH-FQ-sub cleavage strictly under far-from-equilibrium conditions in a time dependent manner, lagging behind HH-min synthesis, with $40 \pm 13\%$ and $60 \pm 7\%$ cleavage yields after 2 and 4 hours, respectively **(Figure 42e-f)**. Unsurprisingly, omitting *sunY* from the reaction mix resulted in no observable HH-min assembly or subsequent substrate cleavage. The data obtained validated the primary hypothesis and demonstrated the compatibility of a far-from-equilibrium environment to host general template-directed RNA synthesis followed by product release, catalysis and template recycling.

3.3.5. Summary and conclusions

In this section, periodic changes in salt concentration and pH driven by water cycles in a closed environment were explored as a possible solution to the problem of strand separation and reannealing during RNA replication. First, a general templated RNA ligase ribozyme derived from group I intron (*sunY; or nrdD*) was characterised under various conditions and with a number of different substrate systems, demonstrating the versatility of the *sunY* ribozyme at performing user-defined ligations. Next, it was shown together with our collaboration partners at the LMU Munich (Braun Lab) that not only does the ribozyme retain ligase ability in the AWI-system, but actually also performs better under sub-optimal bulk salt concentration as a result of the concentration effect at the warm side

of the reaction chamber. Finally, strand release after ligation was also demonstrated by co-incubating the cognate substrate of a synthesised catalytic RNA, and that fidelity was high enough to enable function in the *de novo* synthesised RNA.

AWI-systems can be adapted to the requirements of other ribozyme systems by altering temperatures or chamber geometries. Of particular interest are RNA polymerase ribozymes that catalyse templated primer extension reactions of equal sequence length to themselves. As previously mentioned, the problem of separating *de novo* RNA from its template is exacerbated by salts in solution. It remains to be seen how other complex ribozyme systems, such as these RNA polymerase ribozymes, might also benefit from the salt and pH oscillations to enhance the yields by the positive feedback of template release and by reduced metal ion- and pH-mediated degradation of RNAs.

The data presented further strengthen the hypothesis that dynamic, far-from-equilibrium environments were cradles for the emergence of higher order biomolecular complexity. These systems are dependent on geothermal activity and could have been quite common during the Hadean eon. Nonetheless, it seems unlikely that conditions identical to the laboratory model would be found exhibiting dynamics at the same timescales, since geothermal sources uncommonly generate a temperature gradient of approximately 10 °C across 250 µm. Such temperature gradients do exist in nature although at much larger distances, which would have not necessarily inhibited dew droplet dynamics and thermo-convective currents, but perhaps rather just slowed down the cycle, such that solutions still undergo periodic changes in salt concentration and pH albeit at timescales that range in hours or days instead of seconds or minutes. As these systems were shown to host a variety of other prebiotically relevant functions, future work might build on these results and perhaps explore coupling different ribozyme reactions together with the aim of generating an RNA-based genetic and metabolic networks within a prebiotically plausible protocellular environment.

4. Closing Remarks

The work presented here primarily dealt with the design and characterisation of catalytic RNA systems in experimental settings that are of high relevance to the emergence of life and evolution of early protocells. The first section dealt with genetic information transfer predating the advent of coded protein synthesis. The second section dealt with genotype-phenotype coupling in a protocell model that lacks a physical membrane. The last section dealt with the problem of strand separation and reannealing in an RNA-mediated replication systems. Although these topics might initially seem somewhat unrelated, they all share recurring themes that are central to many of the hypotheses proposed for the emergence and evolution of primordial life.

First, all systems explored comprise in one way or another a compartment that contains the ribozyme system, be it a liposome, a complex coacervate or simply a thermoconvective current that concentrate biomolecules in a defined physical space. For any replicating genetic system, the inevitable emergence of parasitic sequences that are able to selfishly promote their own replication can quickly lead to a systemic collapse by depleting building blocks and cellular resources that would otherwise be utilised for protocell maintenance and survival (*163*). Thus, a major role for any compartment in a protocellular system must be provision of adequate protection from parasitic takeover. Unlike in bulk solution, the emergence of a parasite in a population of protocells will not doom the population as whole, rather only that compartment in which it emerged. However, in our population of model GUVs, freezing and thawing transiently enabled content exchange among the populations. Similarly, coacervates lack a physical membrane and potentially could also be invaded by parasitic sequences that arise. Several studies have demonstrated that compartmentalisation provided a means to prevent parasitic takeover as physical boundaries helped to protect replicators from the excessive dissemination of parasitic sequences throughout the protocell population (*257*, *258*). Surprisingly, even transient compartmentalisation allows replicator survival in the presence of parasites if there is selection pressure for compartments containing active replicators (*162*). Under additional selective pressures, both parts can coexist exhibiting oscillatory dynamics, since the parasitic sequence is dependent on the replicase protein for proliferation and must therefore allow replication of the replicase RNA transcript (*166*). These oscillatory dynamics were in turn shown to drive the diversification of the replicator population, aiding to prevent the collapse of the host replicating system (*259*). Thus, despite their incompleteness as a traditional compartment due

to transient periods of unfiltered exchange with the surrounding environment, it seems conceivable that compartments similar to those employed in this work would nonetheless prevent the unabated proliferation of parasitic sequences and allow for the persistence of RNA replicators within the population.

Several examples of catalytic RNAs have been presented and explained how each system functions differently than the next. Notwithstanding the different chemistries, a recurrent theme that emerges is the ability and ease with which RNA secondary structures incorporate non-Watson-Crick base pairing, specifically, Wobble base pairs between guanine and uracil. It has long been recognised that this base pair, which is evolutionarily conserved in many extant tRNAs and group I introns (*260*, *261*), affects the duplex geometry of the RNA, analogous to how proline residues cause kinks in amino acid polymers (*224*, *262*). This occurs because the glycosidic angles (χ-angle) for both residues in G•U wobble base pairs differ from those that occur in Watson-Crick base pairs. The shifted angles cause differences in the electrostatic gradient in the major groove, since the exocyclic amino group of cytosine in G**-**C base pairs is replaced with a carbonyl group of uracil in G•U base pairs, enhancing the negative electrostatic potential and providing a binding pocket to which divalent metal ions can bind with higher affinity than major grooves with traditional Watson-Crick base pairing. Wobble bases were shown to not only affect RNA secondary and tertiary structure, but also rate of catalysis. In fact, the *sunY* ribozyme enjoys approximately a 3-fold enhancement in ligation rate if the nucleophilic hydroxyl group belongs to a uracil rather than a cytosine base (*74*). In natural group I introns, the site of cleavage contains a G•U base pair that is almost universally conserved, as introns are able to recognize RNA structure and not sequence (*263*). Moreover, the presence of G•U base pair at the flanking site has been shown to promote canonical 3'-5' phosphodiester bond formation, as opposed to G**-**C or A**-**U base pairs where 2'-5' phosphodiester bond formation competes with canonical linkages (*254*). Ribozyme ligases could therefore be engineered to some extent into a specific regioselective and kinetically tuned catalyst by introducing a Wobble base pair or two at the ligation junction and its flanking base pair.

Today, a large number of natural ribozymes have been discovered and a variety of synthetic ribozymes have been evolved *in vitro*, and riboswitches have been found to be relatively common for regulating gene expression at the mRNA level *in vivo* (*264*, *265*). Notwithstanding the widespread occurrence of functional RNA elements, only few allosteric ribozymes have so far been found in nature, which are typically composed of an aptamer in close communication with a group I self-splicing intron. In contrast, allosteric

protein regulators are widespread in all domains of life, suggesting that the division of labour between riboswitches and ribozymes exercised by extant cells might have once been united in ligand-responsive catalytically -active allosteric ribozymes that could achieve responsive control over a wide range of metabolic reactions. These ideas lend support to the co-evolution of biomolecules where the overall fitness of an evolving system assessed by natural selection is not biased towards a single biopolymer but rather on the level of the whole interacting system.

We have seen that origin of life research is highly interdisciplinary, bridging the gaps among geology, chemistry and physics. The reason behind this phenomenon is that the abiogenesis requires that physical phenomena give rise to chemical systems, which interact with their environment to ultimately give rise to biological life as we know it. Knowing that life is inextricably linked to the environment in which it lives, we attempted to address traditional problems in origin of life research with a multidisciplinary approach by taking into consideration potential environmental factors that might benefit RNA catalysis for experimental design. At this present point in time, we can still appreciate how traditional, purely geological, chemical or biological investigations have contributed to the maturation of the field of origin of life. However, while these studies have contributed with crucial insights into the early stages of the Earth and of life on a young planet, an interdisciplinary approach must be adopted to truly tease out, from empirical observations, a scenario that describes the coming together of chemistry to locally and continuously defy the second law of thermodynamics. Jacques Monod ended his essay ‘Chance and Necessity’ by concluding that life is product of chance and was by no means directed to a finite end, our destiny not written anywhere and that it is up to us to choose what to make of it. In the wake of all this exciting new research that has come to light and more that will undoubtably follow, he might have wanted to revise his statement.

5. Materials and Methods

5.1. Equipment

The equipment used during the study are the following: Proflex PCR systems from Applied Biosystems by Life Technologies, Nanodrop One^{C}, X-cell SureLock Mini-Cell Electrophoresis (10 x 10 cm), Owl Easycast B1 mini gel electrophoresis system, Megafuge 8R centrifuge, and Sorvall Lynx 6000 ultracentrifuge with F14-6x250y rotor, PowerEase Touch 120W (Invitrogen) from Thermo Fisher Scientific. Benchtop Centrifuge 5425 and 5425 R (cooled), CryoCube (-80 °C freezer), Concentrator plus, dry block Thermomixer C from Eppendorf. Sprout minifuge centrifuge and EasyPhor PAGE maxi WAVE electrophoresis (20 x 20 cm) from Biozym. Analog vortex mixer, Scout and PR series scales, Orbital Shaker from Ohaus. Sapphire Biomolecular Imager from Azure Biosystems. CLARIOstar plus microplate reader from BMG labtech. Standard incubator model B 28 from Binder. Sonorex RK 31 (water bath sonicator) and and Sonopuls HD 4100 with Sonotrode TS106 from Bandelin. UV light source UVLS-26 EL series from Analytic Jena. pH/mV meter FiveEasy from Mettler Toledo. Magentic stirrer / heater from Heidolph. Rocking platform from VWR. Blue light scanner BIO-1000F from Microtek. Incubator-shaker I26 from New Brunswick Scientific. Power supply EV3330 from Consort. Upright fluorescence mircoscope Axiolab 5 KMAT from Zeiss mounted with 505 nm LED module, dry objectives A-Plan 20x/0,45 Ph2 and 40x/0,65 Ph2 and Axiocam 208 color microscope camera. Leica Thunder inverted widefield microscope equipped with a sCMOS camera Leica DFC9000 GTC using a 63x / NA 1.47 oil immersion objective.

5.2. Materials

1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) was purchased from Avanti Polar. Mineral oil (neat), dimethyl sulfoxide (DMSO), Rotiphorese 40% (19:1 acrylamide:bis-acrylamide) ready-to-use solution, N, N, N′, N′-Tetramethylethylendiamin (TEMED), chloroform, 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), *N*-(2-hydroxyethyl)piperazine-*N*′-(3-propanesulfonic acid) (EPPS), ethylenediaminetetraacetic acid (EDTA), sodium acetate (NaOAc), Triton X-100, bromophenol blue sodium salt, xylene cyanol sodium salt, ammonium persulfate (APS), standard agarose, isopropyl-β-D-1-thiogalactopyranosid (IPTG), carbenicillin, glycerine (99%), calcium chloride ($CaCl_2$), magnesium sulfate ($MgSO_4$) heptahydrate, micro-granulated yeast extract, imidazole and urea were all purchased from Carl Roth. Formamide, 5-carboxy-tetramethylrhodamine *N*-

succinimidyl ester (TAMRA-NHS-ester), polyethylene glycol 8000 (PEG8000), absolute ethanol, 2-propanol (isopropanol), D-glucose, boric acid, glacial acetic acid, sodium tetraborate hexahydrate (NaB), sodium hexametaphosphate $(NaPO_3)_6$, sodium chloride (NaCl), Sigmacote, glutamic acid, Tween 20, imidazole, low viscosity silicone oil (20 cSt), Kolliphor-P188, potassium chloride (KCl), Nickel-NTA affinity resin and peptone from Glycine max (soybean) were purchased from Sigma. Bacto Agar was purchased from Becton, Dickinson and Company (BD). Magnesium chloride ($MgCl_2$) hexahydrate and tris(hydroxymethyl)aminomethane (Tris-HCl) were purchased from Th.Geyer. D-sucrose, dithiothreitol (DTT), DNase I (RNase-free), spermidine and RNA-grade glycogen, SYBR Gold and SYBR Safe were purchased from Thermo Fisher. Gamma-aminohexyl-GTP and NTPs were purchased from Jena Bioscience. Sodium hydroxide (NaOH) pellets, potassium hydroxide (KOH) pellets and 37% hydrochloric acid solution were purchased from VWR. Pre-coated TLC plates and Nucleospin Plasmid EasyPure miniprep kit were purchased from Machery-Nagel. All DNA oligonucleotides were synthesised by IDT. RNA was either transcribed in-house or purchased from IDT when mentioned. All DNA oligomers were ordered from IDT with standard desalting and used without further purification. Cellulose acetate filter spin columns (Spin-X) were bought from Costar. Ultrapure Milli-Q water (Millipore) was used for all experiments in this study. Fluorescently labelled oligomers were ordered with HPLC purification, whereas unlabelled oligomers were desalted and used without further purification. All RNA oligomers were dissolved in RNase-free water and stored at -80 ºC. All peptides were purchased from Sigma-Aldrich and used without further purification (poly-L-lysine hydrobromide (4–15 kDa, 19-72 residues, monomer: 209 g/mol), poly-L-lysine hydrobromide (1-5 kDa, 5-24 residues, monomer: 209 g/mol)). RNA oligomer length standards were purchased from various suppliers and used according to manufactures instructions (low range ssRNA ladder and low molecular weight dsDNA ladder from NEB, 10/60 and 20/100 ssDNA oligo standards from IDT). Q5 High Fidelity DNA polymerase 2X Master Mix, *E. coli* inorganic pyrophosphatase (IPP), T4 RNA ligase 2 (T4Rnl2), Monarch DNA cleanup kit (5 µg), Monarch RNA cleanup kit (10 µg and 50 µg), *E. coli* BL21 DE3 competent cells and *E. coli* TOP10 competent cells were purchased from NEB. GoTaq G2 Hot-Start green Master Mix was purchased from Promega. Vivaspin 20 polyethersulfone (PES) 30 kDa MWCO centrifugal filters were purchased from Vivascience AG. Amicon UF 3 kDa MWCO regenerated cellulose spin filters were purchased from Merck.

5.3. Media and Buffers

- **10X Tris-HCl-Borate-EDTA (TBE):** 0.89 M Tris base, 0.89 M boric acid, 20 mM EDTA pH 8.0
- **50X Tris-HCl-Acetate-EDTA (TAE):** 2 M Tris base, 1 M acetic acid, 50 mM EDTA pH 8.0
- **20% urea-PAGE:** 48% w/v urea (8 M), 1X TBE, 20% 19:1 acrylamide:bisacrylamide
- **0% urea-PAGE:** 48% w/v urea (8 M), 1X TBE
- **LB Media:** 1% w/v peptone-tryptone, 0.5% w/v yeast extract, 100 mM NaCl
- **LB plates:** LB media with 1.5% w/v bactoagar, supplemented with 100 µg/mL carbenicillin
- **SOC media:** 2% w/v peptone-tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5 mM KCl, 5 mM $MgCl_2$, 5 mM $MgSO_4$, 20 mM glucose
- **Protein purification buffers:**
    - Binding / lysis buffer: 30 mM HEPES pH 7.5, 150 mM glutamic acid, 0.05% v/v tween 20, 10 mM imidazole
    - High-salt wash buffer: 30 mM HEPES pH 7.5, 150 mM glutamic acid, 0.05% v/v tween 20, 10 mM imidazole, 1 M NaCl
    - Elution buffer: 30 mM HEPES pH 7.5, 150 mM glutamic acid, 0.05% v/v tween 20, 100 mM imidazole
    - Dialysis / storage buffer: 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 5 mM freshly prepared DTT, 0.1% v/v Triton X-100, 50% v/v glycerin

5.4. General Methods

Purification of recombinant T7 RNA polymerase. His-tagged T7 RNA polymerase expression plasmid (pBH161, P_{lac}-inducible T7 RNAP expression cassette in a pUC19 vector) was a generous gift from Leopold Ulrich at the Protein Production Core Facility of the Max Planck Institute for Biochemistry in Martinsried. For transformation, 50 ng of plasmid (miniprep purity) was added to 50 µL of competent BL21 DE3 *E. coli* cells, mixed and incubated for 30 minutes on ice. The cells and plasmid were then heat shocked for 1 minute at 42 °C, then directly cooled on ice for 5 minutes. The heat-shocked cells were diluted in 1 mL of pre-warmed (37 °C) SOC media and recovered by growing for 1 hour at 37 °C and

900 rpm on a thermoblock. After recovery, cells were pelleted by spinning 2 minutes at 12,000 g, 900 µL of supernatant was discarded, the pellet was resuspended and 100 µL of cells was plates on pre-warmed LB agar plate supplemented with 100 µg/mL of carbenicillin and left to grow overnight in a static incubator at 37 °C. The next day, single colonies were picked and used to inoculate 5 mL of LB media supplemented with carbenicillin and allowed to grow overnight in a shaker incubator at 30 °C, 700 rpm. The following day, 5 mL of culture was used to inoculate 2 L of LB supplemented with carbenicillin and left to grow at 37 °C, 220 rpm for ~4 hours until the OD_{600} reached 1.0 - 1.2. Upon reaching the desired OD_{600}, T7 RNA polymerase expression was induced by the addition of IPTG to a final concentration of 0.1 mM, and the cells were grown for another 4 hours. After induction and growth, cells were pelleted at 6000 g for 45 mins at 4 °C, the supernatant was discarded and the pellet was frozen in liquid nitrogen and stored at -80 °C overnight. The next day, the pelleted cells were thawed and weighed (approximately 12.4 g for a 2 L culture) and resuspended in three times their weight of protein purification binding / lysis buffer (approximately 38 mL). After thorough resuspension, the cells were transferred to a 50 mL conical tube and sonicated on ice using the Bandelin Sonopuls with the TS 106 probe and the following settings: 60% amplitude, pulsing 1 second on, 2 seconds off for 20 minutes. This was repeated three times until the suspension turned less cloudy and lighter in colour. After sonification, the lysate was clarified by spinning in an ultracentrifuge at 45,000 g for 1 hour at 4 °C. Meanwhile, a column was packed with 20 mL of Ni-NTA slurry (5 mL of resin / Litre of culture for a total column volume (CV) of 10 mL), the resin was allowed to settle, the ethanol was eluted and the resin was rinsed twice with 5 CV (50 mL) of de-ionized water, then equilibrated twice with 5 CV (50 mL) of binding / lysis buffer. The clarified lysate was loaded onto the column and allowed to flow through twice to ascertain resin saturation with His-tagged T7 RNA polymerase and prevent non-specific protein binding. After binding, the resin was subjected to high salt wash steps to elute non-specific protein binders with 5 CV (50 mL) of high-salt wash buffer followed by 5 CV (50 mL) of binding buffer repeated twice. Finally, the protein of interest was eluted thrice in 2.5 CV (25 mL x 3 for a total of 75 mL elution volume). The samples were pooled together, split into 4 equal volumes (~20 mL each) and concentrated on 30 kDa molecular weight cut-off (MWCO) in a swing bucket rotor spinning at 3260 g at 4 °C until the total volume was reduced to <1 mL. The concentrated protein solution was dialysed overnight at 4 °C against 2 L of dialysis / storage buffer using 2 slide-a-lyzer mini dialysis units with

3 kDa MWCO. The concentration of protein was determined on Nanodrop using absorbance at 280 nm and setting 1 ABS = 1 mg/mL, then calculating the molar concentration by applying Beer's law using the extinction coefficient of T7 RNA polymerase (ε = 138,380 M^{-1} cm^{-1}) and a 1 cm pathlength. The protein stocks were aliquoted, frozen in liquid nitrogen and stored at -80 °C until use. Purity was determined by Nanodrop and SDS-PAGE analysis as well as functional analyses. The purified T7 RNA polymerase was used in all subsequent run-off i*n vitro* transcriptions for RNA synthesis.

Preparation of DNA templates. Different methods were implemented to generate the DNA templates that were used for run-off *in vitro* transcription (IVT). For short RNAs less than 70 nucleotides, partially double-stranded DNA templates containing an upstream double-stranded T7 promoter and downstream sequence of interest with a leading GpGpG sequence to increase transcription yields were prepared by heating equimolar concentrations (final concentration of 50 µM) T7 promoter sequence and reverse complement of T7 promoter with the sequence of interest at 95 °C for 2 minutes and then quickly cooling the solution on ice. The resulting partially double-stranded DNA was used directly as input template for IVT. For RNAs between 70-120 nucleotides, double-stranded DNA template was generated by annealing two complementary oligos at the T7 promoter region and extending the promoter with the Taq G2 hot start master mix (Promega). For a 50 µL reaction volume, 25 µL of water containing 1 µM of each oligonucleotide was mixed with 25 µL GoTaq G2 hot start master mix and incubated in a thermocycler with the following settings: 2 minutes at 98 °C, 2 cycles comprising denaturation for 5 seconds at 98 °C, annealing at 54 °C for 15 seconds and extension at 72 °C for 2 minutes, followed by a final extension at 72 °C for 2 minutes. The resulting fill-in reaction was purified on silica columns using the Monarch DNA & PCR clean-up kit (NEB), eluted in 15 µL ultrapure water and used immediately or stored at -80 °C. For RNAs longer than 120 nucleotides, the sequence was ordered as a gBlock (IDT) along with primers to amplify it by PCR. Primers were designed such that the optimal annealing temperature coincided with the extension temperature for the polymerase (72 °C). The DNA templates were PCR amplified in a 50 µL reaction volume starting from 10 ng gBlock and 1 µM of each primer with the Q5 high fidelity master mix (NEB) with the following settings: initial denaturation at 98 °C for 30 seconds, 20 cycles of denaturation at 98 °C for 7 seconds and annealing/extension for 50 seconds at 72 °C, followed by a final extension step of 2 minutes at 72 °C. The amplicon was recovered with the Monarch DNA & PCR clean-up kit (NEB), eluted in 15 µL ultrapure water and used

immediately or stored at -80 °C. Quantification (ng/µL) and sample purity (A260/A280 and A260/A230) was determined on the Nanodrop. Products were also routinely checked by agarose gel electrophoresis as described below.

Template screen by agarose gel electrophoresis. Standard agarose and 1X TAE were used to prepare 50 mL of 2-3% agarose gel solution by melting the agarose in a cotton sealed Erlenmeyer flask on high power in the microwave. The solution must be kept under control as it should not come to a boil. Once the agarose was completely dissolved, the solution was cooled to ~65 °C before 3-5 µL of SYBR Safe was mixed in and it was slowly poured into a 10x10 cm cast and 14-well comb. For sample preparation, 5 µL of PCR reaction products were combined with 1 µL of 6X DNA Loading Dye. Taq Fill-in reactions were loaded directly after the reaction (6 µL). An appropriate molecular weight marker (~200 ng of Low Molecular Weight Ladder from NEB) was used to estimate the size of amplicon. The gels were electrophoresed at 80-100 V for 60-80 minutes and imaged directly on the Blue Light Scanner BIO-1000F from Microtek.

Preparation of RNA. RNA was either ordered from IDT or prepared by run-off *in vitro* transcription (IVT) with the T7 RNA polymerase purified from a recombinant source as detailed above. The DNA templates were prepared as detailed above. For partially double-stranded DNA templates prepared by annealing, the final concentration of template was 1 µM, whereas DNA templates prepared by fill-in reaction or PCR amplification were used in whole after column clean-up. Typical reactions volumes were 100 µL, however for large scale production of RNAs the reaction was scaled up (between 500 – 4,000 µL transcriptions). The reaction conditions were: DNA template as mentioned, 30 mM Tris-HCl pH 7.8, 30 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 5 mM of each NTP, 1 U/mL *E. coli* inorganic pyrophosphatase, 0.5 µM T7 RNA polymerase. The reaction proceeded for 2-6 hours at 37 °C, after which the reaction was supplemented with 5 mM $CaCl_2$ and 0.1 volumes of DNase I (Thermo) and further incubated for 30-60 minutes at 37 °C. For transcription volumes greater than 100 µL, the sample was first concentrated by spinning at 15,000 x g, 4 °C in Amicon ultrafiltration columns with 3 kDa molecular weight cut-off regenerated cellulose filters (Merck). Following concentration, the RNA was purified with the Monarch RNA clean-up kit (NEB) following manufacturer's instructions, eluted in the 50-100 µL of ultrapure water. Samples with volumes for 100 µL were directly cleaned up

with the Monarch RNA clean-up kit without concentration and eluted in 30-50 µL of ultrapure water. The amount of RNA recovered was determined on the Nanodrop. Note that transcription reactions should be assembled at room temperature to avoid precipitation of reagents (especially NTPs), and it is recommended against incubating the reaction overnight for increased yields as the concentration of free magnesium increases as the reaction proceeds potentially leading to RNA hydrolysis.

Preparative urea polyacrylamide gel electrophoresis (PAGE). RNAs produced by run-off IVT were purified by 8-15% PAGE, depending on the transcript length. The 20% and 0% acrylamide gel solutions were combined in appropriate ratios to obtain the final desired polyacrylamide percentage, then polymerisation was initiated by the addition of 0.001 volumes of TEMED and 0.01 volumes of 10% APS (for 70 mL gel solution, 70 µL TEMED and 700 µL 10% APS). Gels were cast with 2 mm spacers and large-well combs (5 or 10 wells per 20 cm wide gel) to maximise RNA load per well, which typically was approximately 7-8 µg of RNA per mm^2 of well surface area for optimal resolution. The cast gels were allowed to polymerise at room temperature for more than one hour. After polymerisation, the gel was mounted in the running chamber that was filled with 1X TBE running buffer, the comb was carefully removed and the wells were flushed with 1X TBE using a needle and syringe to clear any gel debris. The gel was then pre-run at constant wattage that results in an approximate 45 mA current for 1 hour to warm the gel and flush out excess TEMED and APS. A suitable amount of RNA was diluted in two volumes of formamide containing xylene cyanol dye, which runs slower than bromophenol blue and hence is a better measure of resolution since the gel can be run for longer times. Care should be taken such that the xylene cyanol and transcript do not have the same electrophoretic mobility in the percentage of acrylamide cast. The RNA in loading buffer was heat denatured at 85 °C for 3-5 minutes before loading the sample into the wells. The gel was subsequently run at the same power settings as the pre-run, until the xylene cyanol dye reached 1-3 cm from the bottom of the gel. After electrophoresis, the gel was removed from the plates, wrapped in plastic foil. In a dark room, the gel was placed on a pre-coated TLC plate and very briefly illuminated with 254 nm UV light. The main product band was marked, excised from the gel with a clean scalpel, transferred to a 2 mL or 5 mL tube depending on the size of the gel slice and crushed to a fine paste using the plunger of a syringe. The crushed gel debris was weighed and a suitable volume of 0.3 M sodium acetate pH 5.2 gel elution buffer was added. Typically, that volume was chosen to be 2 times the weight of the gel slice i.e.,

800 µL of elution buffer for a 400 mg gel slice. The tube was moved to a rotator and the RNA was left to elute overnight at 4 °C. The next day, the gel slurry was filtered through a Spin-X cellulose acetate 0.45 µm filter to remove all gel debris. The recovered eluate was supplemented with 20 ng of RNA-grade glycogen to aid precipitation and the RNA was precipitated with 1.2 volumes of 2-propanol and cooling to -20 °C for an hour. The precipitated RNA was pelleted by spinning the tube for 1 hour at 21,000 g at 4 °C. The supernatant was thoroughly discarded and the pellet was washed with 1 mL of cold 80% ethanol and spun again for 20 minutes at the same speed and temperature. The supernatant was discarded after centrifugation, the pellet was dried under vacuum for 5-10 minutes and suspended in 30-50 µL ultrapure water. Absorbance and purity were determined on the Nanodrop, while the molar concentration of RNA was calculated using the absorbance at 260 nm and the specific extinction coefficient of the oligonucleotide estimated using the online tool OligoCalc (*266*). Purified RNAs were aliquoted and stored at -80 °C until use.

Analytical urea PAGE. Analysis of RNA catalysis was routinely performed by electrophoretic mobility shift assays by denaturing PAGE. The percentage of acrylamide was chosen based on the length of the reactants and products of the reaction. The 20% and 0% acrylamide gel solutions were combined in appropriate ratios to obtain the final desired polyacrylamide percentage, then polymerisation was initiated by the addition of 0.001 volumes of TEMED and 0.01 volumes of 10% APS (for 35 mL gel solution, 35 µL TEMED and 350 µL 10% APS). The gels were cast with 1 mm spacers with different number of wells depending on the number of samples (24, 30, 34, or 48 wells per 20 cm wide gel). The cast gels were allowed to polymerise at room temperature for more than one hour or sometimes even overnight. After polymerisation, the gel was mounted in the running chamber that was filled with 1X TBE running buffer, the comb was carefully removed and the wells were flushed with 1X TBE using a needle and syringe to clear any gel debris. The gel was then pre-run at constant wattage that results in an approximate 45 mA current for 45 minutes to warm the gel and flush out excess TEMED and APS. Meanwhile, samples were prepared in RNA loading buffer containing appropriate amounts of EDTA in formamide to properly quench the reaction with traces of bromophenol blue (typically, the final concentration of EDTA was in excess of that of magnesium chloride, whereas formamide was at least 70% v/v). Before loading, RNA samples were always heat denatured by incubating 3-5 minutes at 85 °C and cooling them quickly on ice. The wells were flushed im-

mediately before sample loading to remove leached urea from the wells. 1-3.5 µL of denatured sample was loaded per well, depending on the number of teeth. Smaller volumes were always preferred if the amount of RNA allowed it, such that the bands after electrophoresis were crisp and clear. The gel was run at constant 20-25 W until the bromophenol blue reached 2-3 cm from the bottom of the gel. When specified, the gel was stained with SYBR Gold. Briefly, 2 µL of SYBR Gold was diluted in approximately 200 mL of 1X TBE, the gel was transferred to the staining solution and allowed to incubate in the dark for 5-10 minutes, after which it was de-stained twice for 5 minutes in ultrapure water. If the gel required no staining as was the case for substrates with fluorescent tags, it was simply allowed to incubate for 5 minutes in ultrapure water. The gel was subsequently imaged on the Azure Sapphire RGB laser scanner with the respective excitations (488 nm for fluorescein-based chromophores (6-FAM), 520 nm for SYBR Gold and Cyanine 3 (Cy3), 658 nm for Cyanine 5 (Cy5)).

Quantitative analysis of analytical PAGE. When specified, gel images were subject to quantitative analysis using the built-in AzureSpot software of the laser scanner**.** Briefly, lane length and diameter were manually selected on the gel image were manually selected to include all bands in the lane. Then, background was subtracted using either the rubber band method or rolling ball method with diameter of 1000 or using the rubber band method when more appropriate. Finally, bands were manually designated and the output was an automatically calculated band percentage for each band, given by the intensity of the band over the intensity of the sum of the bands. Unless otherwise stated, band percentages were used to plot reaction yields.

Giant unilamellar vesicle (GUV) components and protocol. All GUVs were composed of 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC) as the phase transition temperature of this lipid (-2 °C) enables reliable transient membrane permeabilization and content exchange during freezing and thawing episodes. Encapsulation was realised via an inverted emulsion transfer-based method. The emulsion transfer method relies on a density gradient between inner and outer phase achieved by equimolar solutions of internal sucrose and external glucose. Briefly, reverse micelles were generated by rubbing a tube containing the inner phase and lipids in the carrier oil over a rack. The reverse micelles (inverted emulsion) were then transferred to a lipid in oil solution layered over outer phase, and centrifuged at low speeds to generate vesicles by passing the reverse micelles through the lipid-outer

phase interface thus acquiring the second membrane leaflet. All GUVs produced were used on the same day. GUVs encapsulating RNA were stored on ice until use.

5.4. Specific Methods

5.4.1. Methods of Section 3.1

Hammerhead assay in solution. Reactions were carried out in volumes of 50 µL. Substrate cleavage was monitored in a reaction containing 5 µM hammerhead ribozyme, 2.5 µM Cy5-tagged substrate in hammerhead buffer that consisted of 20 mM Tris-HCl pH 8.3, 900 mM sucrose, and 4 mM MgCl2. The reaction was set up on ice and started by the addition of MgCl2 and incubating at 25 °C. Samples were taken at specific time points by quenching 5 µL of reaction with 20µL of hammerhead loading buffer containing 10 mM EDTA, 0.01% bromophenol blue, 98% formamide and placed on ice. After all samples were collected, they were denatured at 85 °C for 5 minutes and cooled quickly on ice before denaturing PAGE analysis (20%). Gel imaging and analysis were performed as described in the General Methods. Data was fitted to an exponential plateau function.

Hammerhead assay in GUVs. GUVs encapsulating hammerhead ribozymes were prepared as described above using hammerhead buffer (900 mM sucrose, 20 mM Tris-HCl pH 8.3, 4 mM $MgCl_2$) and 5 µM of either HH-min. Substrate GUVs were also prepared the same way except that it included 2.5 µM Cy5-tagged substrate RNA instead of ribozyme. The outer phase was the hammerhead buffer but with 900 mM of glucose rather than sucrose, which is necessary for the density gradient. The two GUV populations were united and washed once with 700 µL of buffered outer phase followed by spinning at 2000 g for 5 min. The volume was reduced to ~25 µL and split into two samples of 10 µL each. One sample was a control held at room temperature while the other was subjected to a cycle of freezing and thawing. Both samples were subsequently incubated at 37 °C for 1 h before the GUVs were washed, the RNA was recovered from the GUVs and analysed by 20% denaturing PAGE, as described in General Methods.

Initial R3C ligase characterization in solution. Activity was assayed in 50 mM EPPS buffer pH 8.5 supplemented with 20 mM magnesium chloride, 900 mM sucrose at 42 °C. The substrate concentrations 25 were equal in both control and reaction: 0.5 µM Cy5-A, 4.5 µM Hyper-A, and 12 µM B, whereas the ribozyme was added at a final concentration of 0.5 µM in the reaction only. Samples were collected at the specified time points by

quenching 1 µL in 9 µL R3C loading buffer (25 mM EDTA in formamide). Quenched samples were then heat denatured at 85 °C for 5 minutes and RNAs separated by denaturing PAGE (12%) and analysed as described in General Methods.

Serial transfer of R3C autocatalytic ligase in bulk solution. Typical 10 µL reaction volumes were prepared, the first containing 10 µM Cy5-A with 10 µM of B and 1 µM ribozyme in 20 mM MgCl2 and 50 mM EPPS pH 8.5. The second containing 9 µM Hyper-A, 1 µM Cy5-A and 10 µM B without any ribozyme in the same buffer, and lastly a positive control with 9 µM Hyper-A, 1 µM Cy5-A, 10 µM B and 1 µM F1 ribozyme. The reactions were incubated for 30 minutes at 42 °C, after which 1 µL of the reaction was transferred to 9 µL of fresh substrate mix and the incubation was repeated. This was done several times to achieve 1x106 dilution of the starting reactants. Samples were taken at the initial and final timepoints of a generation by quenching 1 µL of reaction with 9 µL R3C RNA loading buffer (25 mM EDTA in formamide). Samples were heat denatured at 85 °C for 5 minutes before denaturing PAGE analysis (12%).

R3C autocatalytic ligase assay in bulk solution. To determine the activity of the R3C ligase ribozyme in vitro, only the Cy5-tagged inactive A substrate was used as a reporter for ligation yields. 10 µL reactions were prepared on ice, with 5 µM Cy5-A, 7.5 µM B, and varying concentrations of either ribozyme F1 or the unmodified substrate Hyper-A. The R3C buffer consisted of 50 mM EPPS pH 8.5, 900 mM sucrose, and 20 mM MgCl2. The reaction was started by the addition of MgCl2 and incubating it at 42 °C. Samples were consistently taken at specific time points by quenching 1 µL of the reaction with 9 µL of R3C RNA loading buffer (containing 25 mM EDTA, 95% formamide and 0.01% bromophenol blue) and placed on ice. After all samples were collected, they were denatured at 85 °C for 5 minutes and immediately placed on ice, before denaturing urea-PAGE analysis (12%). Gel imaging and analysis were performed as described in General Methods.

FT-dependent activity of encapsulated R3C ligase. The different GUV populations were prepared independently as described above. Briefly, ribozyme GUVs encapsulated 1 or 2 µM of ribozyme F1 in R3C buffer (50 mM EPPS pH 8.5, 20 mM MgCl2, 900 mM sucrose). Substrate A GUVs contained a mixture of 10 µM Hyper-A and 2.5 µM Cy5-A in R3C buffer and substrate B GUVs encapsulated 15 µM of B in R3C buffer. Empty GUVs were used as a negative control and contained only R3C buffer with no RNA. The outer

phase also contained the R3C buffer but with 900 mM glucose instead of sucrose. The GUV populations were combined in a 1:1:1 ratio to obtain a starting sample with both substrate vesicles and either ribozyme GUVs or empty GUVs as the control. The GUVs were washed once with ~700 µL buffered glucose outer phase and centrifuged at 2000 g for 5 min. The supernatant was discarded, and the pellet was resuspended in ~15 µL buffered glucose outer phase and 10 µL were transferred to PCR tubes and subjected to different numbers of FT cycles (none, one, two or three successive cycles) followed by a 1-h incubation at 42 °C. Samples were subsequently washed twice in 500 µL unbuffered glucose and prepared for 12% denaturing urea PAGE as described in General Methods.

Autocatalytic R3C ligase in bulk solution under freeze-thaw conditions. In a total reaction volume of 10 µL, the different components (9 µM Hyper-A; 1 µM Cy5-A; 15 µM B) were combined at room temperature either with 1 µM F1 (grey bars) or without any F1 ribozyme seed (Neg Ctrl, black bars) in 1X R3C buffer (50 mM EPPS; 20 mM MgCl2) and subjected to several freeze-thaw cycles (one FT cycle comprises 10 minutes at -80 °C and 10 minutes at room temperature). After each cycle, 1 µL of sample was quenched with 9 µL RNA loading buffer (10 mM EDTA, 98% formamide, 0.05% bromophenol blue) for analysis. When mentioned, the reactions were further incubated for 60 minutes at 42 °C and sampled in the same way. The samples analysed by denaturing urea-PAGE (12%) after heat denaturation at 85 °C for 5 minutes, see details in General Methods.

Activity of autocatalytic R3C ligase in the presence of GUVs. Empty GUVs were prepared according to the methods section with R3C buffer (20 mM $MgCl_2$ and 50 mM EPPS pH 8.5). A 10 µL reaction volume containing R3C buffer and either 0.5 µM Cy5-A and 1 µM B only (Neg Ctrl) or the same substrate concentrations with 1 µM F1 ribozyme. The reactions were prepared on ice, then increasing volume fractions of GUVs were added to the reaction (0 v/v = 10 µL of 1X buffer was added; 0.25 v/v = 5 µL 1x buffer and 5 µL GUV preparation was added; 0.5 v/v = 10 µL GUV preparation was added). The reactions were then incubated for 1 hour at 42 °C and stopped with 30 µL GUV loading buffer. After heat denaturation at 85 °C for 5 minutes, samples were analysed by denaturing urea-PAGE (12%) and imaged as described in General Methods.

Serial dilution of autocatalytic R3C ligase encapsulated in GUVs. The GUV populations were prepared as described above with the same RNA concentrations and buffer conditions. A population of substrate vesicles (separate GUV populations containing either 10 μM Hyper-A and 2.5 μM Cy5-A or 15 μM of B in R3C buffer) was prepared to be used as a fresh feedstock. The reaction was started by combining 10 μL of GUVs containing unlabelled F1 with 20 μL containing equal amounts of both substrate GUVs (initial mixing ratio 1:1:1), followed by a wash step and centrifugation at 2000 g for 5 min. After discarding 15 μL of the supernatant, settled GUVs were resuspended, and a 5 μL sample was taken and washed in 500 μL of 900 mM unbuffered glucose and subsequently resuspended in 20 μL of the GUV loading buffer (G0 PAGE sample). A second 5 μL sample for PAGE analysis was taken from the remaining 10 μL GUV mixture after three consecutive FT-cycles followed by incubation at 42 °C (G1). In all further generations, 5 μL of the GUVs sample from the previous generation was mixed 1:1 with 5 μL of feedstock substrate GUV containing equal amounts of substrate A and B (see above). These new generations were centrifuged at 2000 g for 5 min and subjected to 3 FT cycles followed by incubation at 42 °C after which 5 μL were removed for PAGE analysis. After all generations were sampled, they were heat denatured and loaded on a 12% denaturing urea-PAGE and analysed as described in General Methods. Data points were fitted to a single exponential using GraphPad Prism for descriptive purposes.

5.4.2. Methods of Section 3.2

Temperature and Ratio Screen of laddering ribozyme. Reactions were carried out with varying ribozyme to substrate ratios, maintaining a total negative charge of 1 mM. Reactions with component ratios of 1:1, 1:2, and 1:4 contained ribozyme:substrate concentrations 10.5:10.5 μM, 8:16 μM, and 5:20 μM respectively. Buffer conditions were 50 mM Tris-HCl, pH 8.6, and 10 mM $MgCl_2$. The reactions were incubated for 2 hours at 30 °C, 37 °C and 45 °C, then quenched by the addition of 9 volumes of RNA loading buffer (10 mM EDTA, 0.05 % bromophenol blue, 95 % formamide). They were resolved on an 8% urea PAGE and stained and imaged as described in General Methods.

Poly-L-lysine titration PAGE. Ribozyme assays were carried out with a 10 μL total reaction volume and the following components: 50 mM Tris-HCl pH 8.6, 10.5 μM ribozyme, 10.5 μM substrate, 10 mM $MgCl_2$ and varying concentrations of poly-L-lysine. Ratios of

RNA to $(Lys)_n$ were based on a fixed charge concentration of the RNA (1 mM total monomer charge), calculated by multiplying molar concentration with length and summing. Positive charge concentrations were calculated based on the lysine hydrobromide monomer repeat molecular weight (209 g/mol) and mass of poly-L-lysine, considering a single positive charge for each residue. Reactions were set up at room temperature by first adding the all the components except the RNA to allow equilibration of poly-L-lysine in the buffer. The reaction was then started by adding a mixture of ribozyme and substrate and incubated in a thermocycler at 30 °C for 2 hours. Reactions were stopped by adding 1 volume of 5 M NaCl, 1 volume of 1.25 M hexametaphosphate (HMP) and 12 volumes of RNA loading buffer containing 10 mM EDTA, 0.05 % bromophenol blue, 95 % formamide. The resulting samples were briefly vortexed, denatured for 5 minutes at 85 °C, cooled quickly on ice and centrifuged for 5 minutes at 2000 x g (Color Sprout Plus, Biozym). PAGE analysis proceeded as described in General Methods.

OD measurements. Measurements were performed using a NanoDrop One^C (Thermo Fischer Scientific) by measuring absorbance at 500 nm as a proxy for coacervate formation. The RNAs were pre-reacted at 2X concentration (21 μM each) in 1X buffer (50 mM Tris-HCl pH 8.6, 10 mM $MgCl_2$) for 3 hours at 30 °C. Varying concentrations of poly-L-lysine in 1X buffer were then added to the reacted RNA. For every poly-L-lysine concentration tested, 2 μL of pre-reacted RNA was added to 2 μL of poly-L-lysine and mixed by pipetting 10 times. The resulting solution was incubated for 5 minutes, after which absorbance was measured. At least 3 biological replicates were measured for each concentration. From these data, we selected specific peptide:RNA ratios for both peptides for further experimentation (0.75:1 (Lys)19-72):RNA and 3:1 (Lys)5-24:RNA). The criterion for this selection was the formation of liquid coacervate droplets (determined by turbidity measurements and microscopy) at a peptide:RNA ratio that did not supress ribozyme activity (determined by PAGE analysis of lysine titration). Both concentrations occur shortly before the respective turbidity maxima in the peptide titration turbidity curve.

RNase R digest. The reaction products from 10 μL reactions containing 1:1 ribozyme:substrate with either no poly-L-lysine, 0.75 mM $(Lys)_{19\text{-}72}$ or 3 mM $(Lys)_{5\text{-}24}$ and incubated for 3 hours at 45 °C were recovered using a 50 μg Monarch RNA clean-up kit (NEB) and eluted in 12 μL ultrapure water. RNase R (Applied Biological Materials, Richmond, Canada) was used to digest a portion of the purified reaction products. The digest reaction

mixture (20 μL) contained 1 μg of RNA, 1X RNase R buffer and 30 U of RNase R, and was incubated for 3.5 hours at 37 °C. The digested RNA was recovered with the 10 μg Monarch RNA clean-up kit (NEB), eluted in 6 μL of ultrapure water and mixed with 6 μL of RNA loading buffer. 1.5 μL of the undigested RNA and 5 μL of the digested RNA were resolved on an 8 % PAGE, stained and imaged as described in General Methods.

R3C ladder ribozyme (v2F) kinetics. Time-dependent assays were performed using individual aliquots for each time point to mitigate the effects of coacervate adhesion to the PCR tube walls over time. The reaction mixture contained: 50 mM Tris-HCl pH 8.6, 10 mM $MgCl_2$, 1 mM RNA charge concentration (9.5 μM substrate, 1 μM Cy5-tagged substrate, 10.5 μM ribozyme), and either 3 mM $(Lys)_{5-24}$ or 0.75 mM $(Lys)_{19-72}$ charge concentration. Aliquoted reactions were quenched by the addition of 2 μL 5 M NaCl, 2 μL 1.25 M HMP, and 34 μL RNA loading buffer (10 mM ETDA, 0.05 % bromophenol blue, 95 % formamide) at designated time points (t = 0, 2, 5, 10, 20, 30, 60, 120, 180, 240 minutes). Sample preparation and PAGE were performed as described in General Methods. The average substrate length and relative band percentages were calculated by summing the products' respective relative abundance multiplied by length. The data were fitted in GraphPad Prism using both a first order ('one-phase association') and second order ('two-phase association') kinetic model. First and second order kinetics were discriminated between using the extra sum-of-squares F test for nested models. The simpler first order model was rejected when $P < 0.05$.

RNA concentration screen. Reactions were carried out in the standard buffer system (50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$) with different concentrations of total RNA but maintaining the same 1:1 ratio of ribozyme to substrate. The concentration of each RNA was either 5.25 μM designated as 0.5x, 10.5 μM designated as 1x, or 21 μM designated as 2x. All reactions had 10 % Cy5-tagged substrate included. The reactions were set up on ice followed by a 10-minute incubation period at 30 °C. The reactions were quenched with different volumes of RNA loading buffer such that the final concentration of labelled species is the same in the quenched sample buffer (half the volume of leading buffer for the 0.5x and twice the volume of loading buffer for the 2x). The 1x reaction was quenched with 4 volumes of RNA loading buffer (10 mM EDTA, 98% formamide, 0.01% bromophenol blue). All samples were heat denatured for 5 minutes at 85 °C and subsequently resolved on an 8% urea PAGE and imaged as described in the General Methods.

PEG 8000 and magnesium chloride screens. Reactions were carries out in the standard buffer system (50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$), and were supplemented with either PEG 8000 alone, MgCl2 alone, or both in combination. All reactions comprised a 1:1 ribozyme to substrate ratio and final positive charge concentration of 1 mM (10.5 μM final concentration of each RNA with 10 % Cy5-tagged substrate). The reactions were set up on ice, started by adding the RNAs and incubated at 30 °C for 10 minutes. After incubation, the reactions containing either PEG 8000 or $MgCl_2$ were quenched with 20 volumes of RNA loading buffer (10 mM EDTA, 98 % formamide, 0.01% bromophenol blue). For the conditions that combine PEG 8000 and $MgCl_2$, a positive control with added $(Lys)_n$ was performed. Consequently, these reactions were quenched by the addition of 1 volume of 5 M NaCl, followed by 1 volume of 1.25 M hexametaphosphate and 18 volumes of RNA loading buffer. All samples were heat denatured for 5 minutes at 85 °C and subsequently resolved on an 8% urea PAGE and imaged as described in General Methods.

Activity assay for inactive ribozyme dv2F. The reactions were carried out in 10 μL volumes with standard buffer (50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$). Ribozyme (either active or inactive) and substrate concentrations were equimolar at 10.5 μM each for a total RNA monomer concentration of 1 mM, except in the controls (Sub only & No Sub) where either one or the other was added. The reactions were either in solution, supplemented with 0.75 mM $(Lys)_{19\text{-}72}$ or 3 mM $(Lys)_{5\text{-}24}$ calculated in charge concentration. They were combined on ice, incubated for 2 hours at 30 °C before being quenched by the addition of 1 volume of 5 M NaCl, followed by 1 volume of 1.25 M hexametaphosphate and 18 volumes of RNA loading buffer. All samples were heat denatured for 5 minutes at 85 °C and subsequently resolved on an 8% urea PAGE and imaged as described in General Methods.

Synthesis of FQ-B Substrate. TAMRA-NHS-Ester was bought from Sigma. γ-aminohexylGTP was bought from Jena Bioscience. A 5'-phosphorylated pentamer with a 3'-end quencher (BHQ2) was ordered from IDT. The substrate was first in vitro transcribed with the modified initiator nucleotide. The reaction differed from the IVT described above by having 6 mM ATP, UTP and CTP, 2 mM GTP, and 4 mM gamma-aminohexyl-GTP. The resulting RNA was column purified (Monarch RNA Clean-up Kit) and reacted with the TAMRA-NHS-ester according to manufacturer's instructions. Briefly, around 60 ODs of

the 5'-aminohexyl RNA was dried under vacuum and resuspended in 600 µL sodium tetraborate pH 8.4 then combined with 100 µL of TAMRA-NHS-ester in DMSO at 10 µg/µL of fluorophore and incubated for 4 hours in a dry block at 25 °C, 600 rpm. The RNA was precipitated in 1.2 volumes of isopropanol, washed once with 80% ethanol, resuspended in ultrapure water and quantified on a Nanodrop as described in General Methods. Subsequently, the fluorophore-tagged RNA was ligated to the 3' quencher-tagged pentamer with T4 RNA ligase 2 (NEB) and a DNA splint complementary to the 15 nucleotides from the 3'-end. The reaction setup (200 µL – 1 mL) was as follows: 3 µM of TAMRA-B, 3.5 µM Penta-BHQ2 (Supplementary Table 1), 3.5 µM DNA splint, 8% DMSO, 1X T4 RNA ligase 2 buffer, 9 mM $MgCl_2$, 100 µM ATP, 50 U T4 RNA ligase 2, 15% PEG8000. The nucleic acids were mixed with water and DMSO, heated to 85 °C and placed on ice for annealing. The remaining reagents were added in the order listed. The reaction was incubated at 25 °C for 1 hour, then at 16 °C for 16 hours. After isopropanol precipitation, the pellet was suspended in formamide and gel purified on a denaturing 20% PAGE gel. Recovery and quantification of RNA was as described in General Methods.

Quenching efficiency and single turnover kinetics for FQ-B. To determine the quenching efficiency, fluorescence measurements for 1 µM of either TAMRA-B or TAMRA-B-BHQ2 (FQ-B) in R3C buffer containing 20 mM $MgCl_2$, 50 mM EPPS pH 8.5, 900 mM sucrose were acquired in a microplate reader by exciting at 545 nm and detecting emission at 590 nm. The quenching efficiency was estimated using the inverse of the ratio of FQ-B fluorescence over TAMRA-B fluorescence multiplied by 100. For single turnover kinetics, reactions were assembled on ice then transferred to a microplate and fluorescence (Ex / Em = 545 / 590 nm) was measured over time every 120 seconds in a microplate reader preheated to 37 °C. The reaction contained 2 µM FQ-B with 5 µM of each rt-F and rt-short-A in a buffer composed of 20 mM $MgCl_2$, 50 mM EPPS pH 8.5, 900 mM sucrose, whereas the control contained only the fluorescent substrate in buffer but no ribozyme.

TIRF microscopy of coacervates and FQ-B system. This experiment was performed with the help of Suchet Nanda from the Dehmelt laboratory at TU Dortmund University, faculty of CCB. TIRF microscopy was performed on an Olympus IX-81 microscope, equipped with a TIRF-MITICO motorized TIRF illumination combiner, an Apo TIRF 60×1.45 NA oil immersion objective and a ZDC autofocus device. Experiments employed spectral emis-

sion ranges in yellow (TAMRA), therefore, a dichroic mirror (ZT514) was used in combination with an emission filter (HC 542/27), the 514 nm line of a 400 mW Argon ion laser as well as wide-field illumination via the MT20 light source (Olympus). Detection was performed with an EMCCD camera (C9100-13) at medium gain without binning. The microscope was equipped with a temperature-controlled incubation chamber which was pre-warmed to 30 °C. Time-lapse microscopy was carried out with image acquisition every 2 minutes for 240 minutes.

Fluorescence microscopy of pre-reacted RNA. Reactions containing 2x RNA concentrations (21 µM of each ribozyme and substrate, of which 10% substrate was FAM-tagged) for a total RNA monomer concentration of 2 mM were prepared in standard buffer containing 50 mM Tris-HCl pH 8.6 and 10 mM $MgCl_2$ and allowed to react for 2 hours at 30 °C. Similarly, 2x $(Lys)_n$ concentrations (1.5 mM $(Lys)_{19\text{-}72}$ and 6 mM $(Lys)_{5\text{-}24}$) were prepared in standard buffer. After incubation, 5 µL of reacted RNA was combined with 5 µL of one of the prepared $(Lys)_n$ solutions in an PCR tube, mixed by pipetting 10 times, transferred to a glass slide previously passivated with Sigmacote and rinsed with water, covered with a 0.1 mm passivated cover slip, and sealed with transparent nail polish. Following 5 minutes at room temperature to allow the seal to dry, the slides were transferred to an incubator set at 30 °C. Slides were imaged at room temperature after 1 hour and 24 hours on the fluorescence microscope AxioLab 5 with camera Axiocam 208 Color and a 40x / 0.65 NA objective by exciting at 505 nm. Images were captured with the Zeiss Labscope v3.2 image acquisition software. Images were contrast adjusted in ImageJ and used without further manipulation.

Widefield microscopy (from Le Vay *et al.* *(232)*). Microscopy was performed on a Leica Thunder inverted widefield microscope equipped with a sCMOS camera Leica DFC9000 GTC using a 63x / NA 1.47 objective. Fluorescence channels were Ex / Em = 484 - 496 / 507 - 543 nm for FAM, and Ex / Em = 629 - 645 / 669 - 741 nm for Cy5. The sample stage was pre-warmed to 30 °C. Samples were loaded into clear bottomed 384 well plates (Greiner µclear, medium binding). Coacervate droplets were formed by combining a solution of RNA with a solution containing poly(L-lysine), buffer and magnesium chloride and mixing using a pipette in a PCR tube. The mixture was incubated on ice for five minutes, then 5 µL was loaded into the well plate. Individual wells were sealed with a drop of silicon

oil to prevent evaporation. After loading, plates were immediately incubated at 30 °C. Imaging was performed as soon as the coacervate suspension had settled, typically first at 30 minutes after mixing. A 3 x 3 image grid, centred at the middle of the well, was captured for each sample at various time points. In most cases, multi-well plates were passivated using Pluronic F-68 to prevent droplet wetting and adhesion by incubating the solution in the well for 30 minutes followed by washing with buffer.

Image processing and analysis (from Le Vay *et al.* *(232)*). Following data collection grid tiles were inspected, and out of focus images were discarded. Only images that remained in focus for all timepoints in a given sample were carried forward for analysis. Image masks were produced by segmentation in the brightfield channel using the Cellpose algorithm (*267*) (cyto2 model, average cytoplasm diameter = 30 - 100 pixels, flow threshold = 0.4, cell probability threshold = 0). A rolling ball background was subtracted in all fluorescence channels (radius = 100 pixels). Masks and images were then used to measure the number of particles per image, particle area and volume corrected fluorescence intensity. Both inter- and intra- particle Pearson correlation coefficients were calculated for samples containing two populations of droplets with orthogonal labels. For figure preparation, time course images were aligned using HyperStackReg (*268*) (v.5.6, translation transformation) and displayed without background subtraction.

5.4.3. Methods of Section 3.3

***sunY*-mediated Broccoli and Mango III aptamer assembly.** Reaction of 10 µL volumes containing 30 mM Tris-HCl pH 7.5, 100 mM KCl, 20 mM $MgCl_2$ and 10% v/v PEG 8000 were combined with 10 µM of Cy5-tagged M1 or B1 fragments, 10 µM of template and 20 µM of fragment M2 and M3 or B2 and B3 were prepared on ice. The reactions were initiated by adding sunY ribozyme to a final concentration of 1 µM and incubating at 45 °C. For control reactions, water was added instead of sunY ribozyme. After 3.5 hours of incubation, the reactions were quenched with 14 volumes (140 µL) of RNA loading buffer comprised of 10 mM EDTA in formamide and trace bromophenol blue. They were subsequently heat denatured at 85°C for 5 minutes before resolving on a 20% PAGE as described in General Methods.

Kinetics of Mango aptamer assembly by *sunY*. Reactions of 20 µL were prepared on ice containing 30 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$ and 100 mM KCl with 5 µM of each

Cy5-M1, M2, M3 and template. Reactions were initiated by the addition of sunY ribozyme to a final concentration of 1 µM and incubating them at 45 °C. Timepoints were taken that correspond to minutes 0, 2, 5, 10, 20, 30, 45 and 60 by quenching 1 µL of the reaction in 14 µL RNA loading buffer containing 10 mM EDTA in formamide and trace amounts of bromophenol blue. Samples were heat denatured at 85 °C for 5 minutes before resolving on a 20% PAGE as described in General Methods. The data produced was fitted with an irreversible two-step reaction equation which also accounts for incomplete conversion of substrate to final product:

$$Y_{M1} = \left(M1_{max} - M1_{left}\right) * \ e^{-k_1 * X} + M1_{left} \ (1)$$

$$Y_{M12} = \frac{k_1 * \left(M1_{max} - M1_{left}\right)}{k_2 - k_1} - M12_{left}\ ) * (e^{-k_1 * X} - e^{-k_2 * X}) + M12_{left} * (1 - \ e^{-k_1 * X}) \ (2)$$

$$Y_{M123} = \left(M1_{max} - M1_{left} - M12_{left}\right) * [1 + \frac{1}{k_1 - k_2}(k_2 * e^{-k_1 * X} - k_1 * e^{-k_2 * X})] \ (3)$$

where k_1 is the rate of conversion of M1 to M12, k_2 is the rate of conversion of M12 to M123, while $M1_{left}$ and $M12_{left}$ are the unreacted substrate and intermediate, respectively, and M1 max is the starting amount of substrate. The parameters were set as shared and constrained accordingly. Equations (1), (2) and (3) describe the best fit for RNA species M1, M12, and M123, respectively.

Time-dependent characterisation of splinted RNA synthesis. Reactions of 10 µL were prepared on ice containing 10 µM of each substrate (TAMRA-F1, F2, F3, F4 and F5-FAM) in 30 mM Tris-HCl pH 7.5, 50 mM $MgCl_2$ and 10% v/v PEG8000. Reactions were initiated with the addition of 2.5 µM *sunY* and incubation at 45 °C. Samples were taken at the specified timepoints (0, 1, 2, 3 and 4 hours) by quenching 1 µL of sample with 14 µL of RNA loading buffer containing 10 mM EDTA in formamide and trace amounts of bromophenol blue. Samples were heat denatured at 85 °C for 5 minutes before resolving on a 20% PAGE as described in General Methods.

Effect of sunY concentration on splinted ligation. Reactions of 10 µL volumes were prepared on ice with 5 µM of each substrate (TAMRA-F1, F2, F3, F4 and F5-FAM) in 30 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$ and 100 mM KCl. Reactions were initialised by

the addition of sunY to the specified concentrations to achieve different ratios of ribozyme:substrates, namely, 1:5, 1:2, 1:1 and 6:5. Reactions were incubated at 45 °C overnight (~20 hours) and subsequently quenched with 14 volumes of RNA loading buffer containing 10 mM EDTA in formamide and trace amounts of bromophenol blue, before heat denaturation for 5 minutes at 85 °C and resolving the products on a 20% PAGE as described in General Methods.

Preparation of AWI chambers and sample injection (from Salditt *et al.*(*251*)) AWI-system chambers were built according to the schematic shown and based on designs previously published in Matreux *et al.*(*213*) . The desired shape was cut out from a 250 µm thick Teflon foil (HolscotEurope) using a GRAPHTEC CE6000-40 Plus cutter. The resulting Teflon slide defined the shape of the reaction chambers., designed to have three reaction compartments. The compartments are all 11 mm wide and 8 mm high, corresponding to a maximum capacity of 22 µL for a 250 µm foil thickness. A second foil with 100 µm thickness was cut out to serve as hydrophobic backside. Both Teflon slides were sandwiched between two sapphire plates and fixed to a metal block with a steel frame. A 25 µm thickness graphite (PANASONIC) foil was sandwiched between both elements to increase heat conductivity between the cold sapphire and the metal surface. All parts of the sandwich had a width of 22 mm and a height of 60 mm. Prior to assembly, all components were cleaned with ethanol and ultrapure water and from dust particles. A heater was screwed onto the front side, and 200 µm thick graphite foils (PANASONIC) were sandwiched between the each of the heating and cooling units and the reaction chamber to improve heat conductivity throughout the system. The temperature gradient was maintained by cooling the backside with a refrigerated water bath (CF41 and 300 F, Julabo) and heating the front with an electrical heater regulated by an Arduino microcontroller board. For each experiment, a reaction volume of 15 µL was injected with a Hamilton syringe through a Teflon tubing (Techlab), which were fixed by PTFE fittings (Techlab). The inlets were then sealed with soft silicon and PTFE fittings. All compartments were connected to the same CO_2 inflow to ensure equivalent partial pressure conditions. Finally, to monitor the cool water bath temperature and set the temperature of the heater accordingly, a temperature sensor was screwed into one of the holes in the cool backside along with one of the fittings.

Synthesis of Mango aptamer in AWI-system (from Salditt *et al.*(*251*)). RNA ligation reactions were carried out with final concentrations of 20 µM for all ligation substrates,

5 μM of template, 30 mM Tris-HCl pH 7.5, 100 mM KCl, and varying bulk magnesium concentrations as specified. Ligation was initiated by adding *sunY* ribozyme to a final concentration of 2.5 μM. The first substrate (M1) contained a Cy5 label at the 5'-end. A reaction volume of 12 μL was prepared for the isothermal control and 17 μL were prepared for the experiments under AWI conditions. Upon *sunY* addition, 2 μL of each sample was quenched with 2× gel loading buffer (90% formamide, 5% glycerol, 50 mM EDTA, 0.015% Orange G). The remaining 10 μL of the isothermal controls were incubated at 45 °C in T100 thermal cycler (Bio-Rad), whereas the remaining 15 μL for the far-from-equilibrium conditions was injected into the fully assembled AWI-system through a Teflon tubing (Techlab) using a Hamilton syringe. The reaction was left to proceed for 4 hours before sample recovery and quenching. All samples were heat denatured before resolving on a 15% PAGE.

Synthesis of HH-min and subsequent HH-FQ-sub cleavage (from Salditt *et al.*(*251*)). Reactions with a final concentration of 12 μM and 3 μM of all ligation substrates (HH1-3) and template, respectively. The first RNA ligation substrate (HH1) was Cy5-tagged at the 5'-end. The buffer for AWI sample contained 30 mM Tris-HCl·pH 7.5, 100 mM KCl and 5 mM $MgCl_2$. Isothermal buffer was identical except that it contained 50 mM $MgCl_2$. Reactions were initialised by the addition of sunY to a final concentration of 2.5 μM. To verify the activity of the in situ synthesized HH-min hammerhead ribozyme, 0.4 μM HH-FQ-sub was added. which has a 5'-FAM tag and a 3'-BHQ1 quencher. An initial sample (T_0) was obtained by quenching 2 μL of the reaction with 18 μL of an appropriate RNA loading buffer. Isothermal controls were incubated at a constant 45 °C. Negative controls in the absence of *sunY* were also performed. The positive control was cleaved HH-FQ-sub in solution. All samples were heat denatured before resolving on a 15% PAGE.

6. Acknowledgements

I first want to thank Hannes for giving me the opportunity to work for him for these almost five years now. Who knows where I would have ended up if it was not for him, probably back home living with my parents in a country plagued by corruption and greed… and for that I am truly grateful. The next person to thank is Kris, who not only trained me to work with RNA at the start, but also supported me with ideas and discussions throughout. I would also like to thank everyone in the group, past and present members for the nice discussions, fun times, laughs, disappointments and all. I've also to thank other labs with whom we collaborated, in particular the Schwille lab at the MPI for Biochemistry in Martinsried, the Boekhoven lab at the TUM in Garching, and finally the Braun lab at the LMU in Munich. A special thanks goes to Carsten Donau who recently (or so it seems to me) graduated and is to become a father soon. We had some fun times back in Munich.

From outside the scientific field, I have to thank my parents for providing me with an education without the crippling debt. My accomplishments are just as much theirs. For that I will be forever grateful. Last but definitely not least, I want to thank my wife Claudia for supporting throughout this long journey and following me from Spain to Munich to Dortmund, soon to Boston and hopefully wherever I end up after that. I love you.

7. References

1. C. de Duve, *Am Sci*. **83**, 428–437 (1995).
2. Z. Tóth, *J Chem Educ*. **73**, 539 (1996).
3. A. Strecker, *Liebigs Ann Chem*. **75**, 27–45 (1850).
4. C. Zimmer, *Life's Edge: the search for what it means to be alive* (Picador, 2021).
5. A. Butlerov, *C R Acad Sci*, 145–147 (1861).
6. R. Breslow, *Tetrahedron Lett*. **1**, 22–26 (1959).
7. P. Berche, *Clin Microbiol Infect*. **18**, 1–6 (2012).
8. J. Peretó, J. L. Bada, A. Lazcano, *Orig Life Evol Biosph*. **39**, 395–406 (2009).
9. P. L. Luisi, *The Emergence of Life: From Chemical Origins to Synthetic Biology* (Cambridge University Press, Second Edi., 2016).
10. A. Lazcano, *J Mol Evol*. **83**, 214–222 (2016).
11. S. L. Miller, J. W. Schopf, A. Lazcano, *J Mol Evol*. **44**, 351–353 (1997).
12. S. Tirard, *Orig Life Evol Biosph*. **40**, 215–220 (2010).
13. S. Tirard, *J Genet*. **96**, 735–739 (2017).
14. I. Fry, *Endeavour*. **30**, 24–28 (2006).
15. J. D. Watson, F. H. C. Crick, *Nature*. **171**, 737–738 (1953).
16. S. Zamenhof, G. Brawerman, E. Chargaff, *Biochim Biophys Acta*. **9**, 402–405 (1952).
17. S. L. Miller, *Science*. **117**, 528–529 (1953).
18. S. L. Miller, *J Am Chem Soc*. **77**, 2351–2361 (1955).
19. H. C. Urey, *Proc Natl Acad Sci*. **38**, 351–363 (1952).
20. J. Oró, *Biochem Biophys Res Commun*. **2**, 407–412 (1960).
21. J. Oró, *Nature*. **191**, 1193–1194 (1961).
22. S. W. Fox, K. Harada, *Science*. **133**, 1923–1924 (1961).
23. G. M. Tener, H. G. Khorana, R. Markham, E. H. Pol, *J Am Chem Soc*. **80**, 6223–6230 (1958).
24. G. Schramm, H. Grötsch, W. Pollmann, *Angew Chem Int Ed*. **1**, 1–7 (1962).
25. M. Meselson, F. W. Stahl, *Proc Natl Acad Sci*. **44**, 671–682 (1958).
26. F. Crick, *Nature*. **227**, 561–563 (1970).
27. I. Haruna, K. Nozu, Y. Ohtaka, S. Spiegelman, *Proc Natl Acad Sci*. **50**, 905–911 (1963).
28. S. Spiegelman, I. Haruna, I. B. Holland, G. Beaudreau, D. Mills, *Proc Natl Acad Sci*. **54**, 919–927 (1965).
29. D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc Natl Acad Sci*. **58**, 217–224 (1967).
30. R. Levisohn, S. Spiegelman, *Proc Natl Acad Sci*. **63**, 805–811 (1969).
31. A. Rich, M. Kasha, B. Pullman, Eds., On the problems of evolution and biochemical information transfer. *Horizons in Biochemistry* (1962), pp. 103–126.
32. F. H. C. Crick, *J Mol Biol*. **38**, 367–379 (1968).
33. L. E. Orgel, *J Mol Biol*. **38**, 381–393 (1968).
34. C. R. Woese, *The genetic code: the molecular basis for genetic expression* (Harper & Row, 1967).
35. C. R. Woese, *Proc Natl Acad Sci*. **99**, 8742–8747 (2002).
36. B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat Commun*. **7**, 11328 (2016).
37. D. M. Fialho *et al.*, *Org Biomol Chem*. **16**, 1263–1271 (2018).
38. A. Lelièvre-Büttner, T. Schnarr, M. Debiais, M. Smietana, S. Müller, *Eur J Chem*. **29** (2023).
39. O. Leslie E., *Crit Rev Biochem Mol Biol*. **39**, 99–123 (2004).
40. R. A. Sanchez, J. P. Ferris, L. E. Orgel, *J Mol Biol*. **30**, 223–53 (1967).

41. L. Frick, J. P. Mac Neela, R. Wolfenden, *Bioorg Chem*. **15**, 100–108 (1987).
42. W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J Mol Biol*. **67**, 25–33 (1972).
43. R. Krishnamurthy, *Nat Commun*. **9**, 5175 (2018).
44. T. P. Roche *et al.*, *Eur J Chem*. **29** (2023).
45. K. N. Drew, J. Zajicek, G. Bondo, B. Bose, A. S. Serianni, *Carbohydr Res*. **307**, 199–209 (1998).
46. M. W. Powner, B. Gerland, J. D. Sutherland, *Nature*. **459**, 239–242 (2009).
47. J. Xu *et al.*, *Nat Chem*. **9**, 303–309 (2017).
48. B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat Chem*. **7**, 301–307 (2015).
49. K. E. Koch, thesis, ETH Zurich, Zurich (1992).
50. S. Becker *et al.*, *Science*. **352**, 833–836 (2016).
51. M. Dizdaroglu, G. Kirkali, P. Jaruga, *Free Radic Biol Med*. **45**, 1610–1621 (2008).
52. H. Okamura *et al.*, *Chem Commun*. **55**, 1939–1942 (2019).
53. S. Becker *et al.*, *Science*. **366**, 76–82 (2019).
54. H. Sawai, *J Mol Evol*. **17**, 48–51 (1981).
55. W. Huang, J. P. Ferris, *J Am Chem Soc*. **128**, 8914–8919 (2006).
56. J. P. Ferris, *Philos T R Soc B*. **361**, 1777–1786 (2006).
57. J. P. Ferris, G. Ertem, *Science*. **257**, 1387–1389 (1992).
58. C. A. Jerome, H.-J. Kim, S. J. Mojzsis, S. A. Benner, E. Biondi, *Astrobiology*. **22**, 629–636 (2022).
59. P.-A. Monnard, A. Kanavarioti, D. W. Deamer, *J Am Chem Soc*. **125**, 13734–13740 (2003).
60. T. Hassenkam, D. Deamer, *Sci Rep*. **12**, 10098 (2022).
61. P. Chamanian, P. G. Higgs, *PLoS Comput Biol*. **18**, e1010458 (2022).
62. M. Todisco *et al.*, *ACS Nano*. **12**, 9750–9762 (2018).
63. L. Zhou *et al.*, *Elife*. **8** (2019).
64. S. J. Zhang, D. Duzdevich, J. W. Szostak, *J Am Chem Soc*. **142**, 14810–14813 (2020).
65. S. J. Zhang, D. Duzdevich, D. Ding, J. W. Szostak, *Proc Natl Acad Sci*. **119** (2022).
66. L. Zhou, D. K. O'Flaherty, J. W. Szostak, *Angew Chem Int Ed*. **59**, 15682–15687 (2020).
67. L. Zhou, D. K. O'Flaherty, J. W. Szostak, *J Am Chem Soc*. **142**, 15961–15965 (2020).
68. S. C. Kim, D. K. O'Flaherty, C. Giurgiu, L. Zhou, J. W. Szostak, *J Am Chem Soc*. **143**, 3267–3279 (2021).
69. C. Tuerk, L. Gold, *Science*. **249**, 505–510 (1990).
70. A. D. Ellington, J. W. Szostak, *Nature*. **346**, 818–822 (1990).
71. S. Gwiazda, K. Salomon, B. Appel, S. Müller, *Biochimie*. **94**, 1457–1463 (2012).
72. N. Paul, G. F. Joyce, *Proc Natl Acad Sci*. **99**, 12733–12740 (2002).
73. E. J. Hayden, N. Lehman, *Chem Biol*. **13**, 909–918 (2006).
74. J. A. Doudna, N. Usman, J. W. Szostak, *Biochemistry*. **32**, 2111–2115 (1993).
75. A. Wochner, J. Attwater, A. Coulson, P. Holliger, *Science*. **332**, 209–212 (2011).
76. J. A. Doudna, J. W. Szostak, *Mol Cell Biol*. **9**, 5480–5483 (1989).
77. B. I. Kankia, L. A. Marky, *J Phys Chem B*. **103**, 8759–8767 (1999).
78. A. Mariani, C. Bonfio, C. M. Johnson, J. D. Sutherland, *Biochemistry*. **57**, 6382–6386 (2018).
79. J. A. Doudna, S. Couture, J. W. Szostak, *Science*. **251**, 1605–1608 (1991).
80. R. Green, J. W. Szostak, *Science*. **258**, 1910–1915 (1992).
81. G. F. Dolan, U. F. Müller, *RNA*. **20**, 202–213 (2014).
82. W. E. Draper, E. J. Hayden, N. Lehman, *Nucleic Acids Res*. **36**, 520–531 (2007).
83. T. S. Jayathilaka, N. Lehman, *ChemBioChem*. **19**, 217–220 (2018).

84. M. J. Fedor, *J Mol Biol*. **297**, 269–291 (2000).
85. H. Mutschler, A. Wochner, P. Holliger, *Nat Chem*. **7**, 502–508 (2015).
86. K. Le Vay, E. Y. Song, B. Ghosh, T.-Y. D. Tang, H. Mutschler, *Angew Chem Int Ed*. **60**, 26096–26104 (2021).
87. R. Hieronymus, J. Zhu, S. Müller, *Nucleic Acids Res*. **50**, 368–377 (2022).
88. J. Zhu, R. Hieronymus, S. Müller, *ChemBioChem*. **24** (2023).
89. M. P. Robertson, J. R. Hesselberth, A. D. Ellington, *RNA*. **7**, S1355838201002199 (2001).
90. M. P. Robertson, W. G. Scott, *Science*. **315**, 1549–1553 (2007).
91. M. P. Robertson, G. F. Joyce, *Chem Biol*. **21**, 238–245 (2014).
92. Y. Ikawa, K. Tsuda, S. Matsumura, T. Inoue, *Proc Natl Acad Sci*. **101**, 13750–13755 (2004).
93. J. Ishikawa *et al.*, *Arch Biochem Biophys*. **490**, 163–170 (2009).
94. E. H. Ekland, D. P. Bartel, *Nucleic Acids Res*. **23**, 3231–3238 (1995).
95. E. H. Ekland, J. W. Szostak, D. P. Bartel, *Science*. **269**, 364–370 (1995).
96. N. H. Bergman, W. K. Johnston, D. P. Bartel, *Biochemistry*. **39**, 3115–3123 (2000).
97. I.-M. Johansson, C. Forsman, *Eur J Biochem*. **224**, 901–907 (1994).
98. S. Lindskog, *Pharmacol Ther*. **74**, 1–20 (1997).
99. W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, D. P. Bartel, *Science*. **292**, 1319–1325 (2001).
100. J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni, P. Holliger, *Elife*. **7** (2018).
101. D. P. Horning, G. F. Joyce, *Proc Natl Acad Sci*. **113**, 9786–9791 (2016).
102. K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce, *Proc Natl Acad Sci*. **117**, 2906–2913 (2020).
103. X. Portillo, Y.-T. Huang, R. R. Breaker, D. P. Horning, G. F. Joyce, *Elife*. **10** (2021).
104. R. Cojocaru, P. J. Unrau, *Science*. **371**, 1225–1232 (2021).
105. A. Eschenmoser, *Tetrahedron*. **63**, 12821–12844 (2007).
106. E. J. Javaux, *Nature*. **572**, 451–460 (2019).
107. H. J. Hofmann, K. Grey, A. H. Hickman, R. I. Thorpe, *Geol Soc Am Bull*. **111**, 1256–1262 (1999).
108. N. Noffke, D. Christian, D. Wacey, R. M. Hazen, *Astrobiology*. **13**, 1103–1124 (2013).
109. T. Djokic, M. J. van Kranendonk, K. A. Campbell, M. R. Walter, C. R. Ward, *Nat Commun*. **8**, 15263 (2017).
110. K. Hickman-Lewis, B. Gourcerol, F. Westall, D. Manzini, B. Cavalazzi, *Precambrian Res*. **342**, 105689 (2020).
111. M. Schidlowski, *Precambrian Res*. **106**, 117–134 (2001).
112. E. A. Bell, P. Boehnke, T. M. Harrison, W. L. Mao, *Proc Natl Acad Sci*. **112**, 14518–14521 (2015).
113. T. Hassenkam, M. P. Andersson, K. N. Dalby, D. M. A. Mackenzie, M. T. Rosing, *Nature*. **548**, 78–81 (2017).
114. T. Tashiro *et al.*, *Nature*. **549**, 516–518 (2017).
115. M. Brasier, O. Green, J. Lindsay, A. Steele, *Orig Life Evol Biosph*. **34**, 257–269 (2004).
116. M. Brasier, N. McLoughlin, O. Green, D. Wacey, *Philos T R Soc B*. **361**, 887–902 (2006).
117. M. D. Hopkins, T. M. Harrison, C. E. Manning, *Earth Planet Sci Lett*. **298**, 367–376 (2010).
118. S. A. Wilde, J. W. Valley, W. H. Peck, C. M. Graham, *Nature*. **409**, 175–178 (2001).
119. R. Brasser, S. J. Mojzsis, S. C. Werner, S. Matsumura, S. Ida, *Earth Planet Sci Lett*. **455**, 85–93 (2016).
120. J. W. Valley *et al.*, *Nat Geosci*. **7**, 219–223 (2014).
121. D. Trail *et al.*, *Proc Natl Acad Sci*. **115**, 10287–10292 (2018).
122. N. H. Sleep, K. J. Zahnle, R. E. Lupu, *Philos T R Soc A*. **372**, 20130172 (2014).

123. N. H. Sleep, K. Zahnle, P. S. Neuhoff, *Proc Natl Acad Sci*. **98**, 3666–3672 (2001).
124. N. T. Arndt, E. G. Nisbet, **40**, 521–549 (2012).
125. B. Damer, D. Deamer, *Astrobiology*. **20**, 429–452 (2020).
126. M. J. van Kranendonk *et al.*, *Astrobiology*. **21**, 39–59 (2021).
127. K. A. Campbell *et al.*, *Earth Sci Rev*. **148**, 44–64 (2015).
128. F. Westall *et al.*, *Astrobiology*. **18**, 259–293 (2018).
129. D. Milshteyn, B. Damer, J. Havig, D. Deamer, *Life*. **8**, 11 (2018).
130. S. Becker *et al.*, *Nat Commun*. **9**, 163 (2018).
131. S. Nader *et al.*, *ACS Earth Space Chem*. **7**, 252–259 (2023).
132. K. Von Damm, *Annu Rev Earth Planet Sci*. **18**, 173–204 (1990).
133. W. Martin, J. Baross, D. Kelley, M. J. Russell, *Nat Rev Microbiol*. **6**, 805–814 (2008).
134. M. J. Russell, *Acta Biotheor*. **55**, 133–179 (2007).
135. M. J. Russell, R. M. Daniel, A. J. Hall, *Terra Nova*. **5**, 343–347 (1993).
136. S. Duval *et al.*, *Interface Focus*. **9** (2019).
137. F. Westall, A. Brack, A. G. Fairén, M. D. Schulte, *Front Astron Space Sci*. **9** (2023).
138. A. Hofmann, C. Harris, *Chem Geol*. **257**, 221–239 (2008).
139. H. Follmann, C. Brownson, *Naturwissenschaften*. **96**, 1265–1292 (2009).
140. A. Kanavarioti, P.-A. Monnard, D. W. Deamer, *Astrobiology*. **1**, 271–281 (2001).
141. M. Gull, *Challenges*. **5**, 193–212 (2014).
142. S. Roy, N. V. Bapat, J. Derr, S. Rajamani, S. Sengupta, *J Theor Biol*. **506**, 110446 (2020).
143. K. Zahnle, L. Schaefer, B. Fegley, *Cold Spring Harb Perspect Biol*. **2**, a004895–a004895 (2010).
144. J. N. Bahcall, M. H. Pinsonneault, S. Basu, *Astrophys J*. **555**, 990–1012 (2001).
145. G. Feulner, *Rev Geophys*. **50** (2012).
146. S. Kadoya, J. Krissansen-Totton, D. C. Catling, *Geochem Geophys Geosys*. **21** (2020).
147. F. J. Millero, *Geochim Cosmochim Acta*. **59**, 661–677 (1995).
148. A. A. Lacis, G. A. Schmidt, D. Rind, R. A. Ruedy, *Science*. **330**, 356–359 (2010).
149. S. A. Benner *et al.*, *ChemSystemsChem*. **2**, e1900035 (2020).
150. H. Trinks, W. Schröder, C. K. Biebricher, *Orig Life Evol Biosph*. **35**, 429–445 (2005).
151. V. L. Bronshteyn, A. A. Chernov, *J Cryst Growth*. **112**, 129–145 (1991).
152. C. Menor-Salván, D. M. Ruiz-Bermejo, M. I. Guzmán, S. Osuna-Esteban, S. Veintemillas-Verdaguer, *Eur J Chem*. **15**, 4411–4418 (2009).
153. J. Attwater, A. Wochner, V. B. Pinheiro, A. Coulson, P. Holliger, *Nat Commun*. **1**, 76 (2010).
154. F. J. Ciesla, S. A. Sandford, *Science*. **336**, 452–454 (2012).
155. R. Lohrmann, L. E. Orgel, *J Mol Evol*. **9**, 323–328 (1977).
156. P. B. Rimmer *et al.*, *Astrobiology*. **21**, 1099–1120 (2021).
157. G. Tsuji, S. Fujii, T. Sunami, T. Yomo, *Proc Natl Acad Sci*. **113**, 590–595 (2016).
158. T. Litschel *et al.*, *New J Phys*. **20**, 55008 (2018).
159. M. Eigen, *Naturwissenschaften*. **58**, 465–523 (1971).
160. M. Eigen, R. Winkler, *Steps towards life: a perspective on evolution* (Oxford University Press, 1992).
161. D. S. Tawfik, A. D. Griffiths, *Nat Biotechnol*. **16**, 652–656 (1998).
162. S. Matsumura *et al.*, *Science*. **354**, 1293–1296 (2016).
163. G. F. Joyce, J. W. Szostak, *Cold Spring Harb Perspect Biol*. **10**, a034801 (2018).
164. A. Akoopie, J. T. Arriola, D. Magde, U. F. Müller, *Sci Adv*. **7** (2021).
165. F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc Natl Acad Sci*. **98**, 4552–4557 (2001).

166. Y. Bansho, T. Furubayashi, N. Ichihashi, T. Yomo, *Proc Natl Acad Sci*. **113**, 4045–4050 (2016).
167. S. Mansy, *Int J Mol Sci*. **10**, 835–843 (2009).
168. D. W. Nooner, J. Oro, Chapter 12. *Hydrocarbon Synthesis from Carbon Monoxide and Hydrogen*. **178** (1979), pp. 159–171.
169. G. Proskurowski *et al.*, *Science*. **319**, 604–607 (2008).
170. A. Wang, J. W. Szostak, *Emerg Top Life Sci*. **3**, 537–542 (2019).
171. K. P. Adamala, A. E. Engelhart, J. W. Szostak, *Nat Commun*. **7**, 11041 (2016).
172. S. S. Mansy *et al.*, *Nature*. **454**, 122–125 (2008).
173. K. Adamala, J. W. Szostak, *Science*. **342**, 1098–1100 (2013).
174. F. Anella, C. Danelon, *Life*. **4**, 929–943 (2014).
175. F. Anella, C. Danelon, *Life*. **7**, 17 (2017).
176. S. Sarkar, S. Dagar, A. Verma, S. Rajamani, *Sci Rep*. **10**, 4483 (2020).
177. Ö. D. Toparlak, M. Karki, V. E. Ortuno, R. Krishnamurthy, S. S. Mansy, *Small*. **16**, 1903381 (2020).
178. L. Jin, N. P. Kamat, S. Jena, J. W. Szostak, *Small*. **14**, 1704077 (2018).
179. M. E. Glasner, N. H. Bergman, D. P. Bartel, *Biochemistry*. **41**, 8103–8112 (2002).
180. M. Fiore *et al.*, *Astrobiology*. **22**, 598–627 (2022).
181. D. Fayolle *et al.*, *Sci Rep*. **7**, 18106 (2017).
182. C. Bonfio *et al.*, *J Am Chem Soc*. **141**, 3934–3939 (2019).
183. L. Liu *et al.*, *Nat Chem*. **12**, 1029–1034 (2020).
184. D. S. Kelley *et al.*, *Nature*. **412**, 145–149 (2001).
185. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature*. **405**, 299–304 (2000).
186. S. Sarkar, S. Dagar, S. Rajamani, *ChemSystemsChem*. **3**, e2100014 (2021).
187. M. P. Joshi, A. A. Sawant, S. Rajamani, *Chem Sci*. **12**, 2970–2978 (2021).
188. A. P. Costa, X. Xu, D. J. Burgess, *Pharm Res*. **31**, 97–103 (2014).
189. B. Ghosh, R. Bose, T.-Y. D. Tang, *Curr Opin Colloid Interface Sci*. **52**, 101415 (2021).
190. L.-W. Chang *et al.*, *Nat Commun*. **8**, 1273 (2017).
191. M. Abbas, W. P. Lipiński, K. K. Nakashima, W. T. S. Huck, E. Spruijt, *Nat Chem*. **13**, 1046–1054 (2021).
192. L. Li *et al.*, *Macromolecules*. **51**, 2988–2995 (2018).
193. S. Choi, M. O. Meyer, P. C. Bevilacqua, C. D. Keating, *Nat Chem*, 1–8 (2022).
194. E. A. Frankel, P. C. Bevilacqua, C. D. Keating, *Langmuir*. **32**, 2041–2049 (2016).
195. T.-Y. Dora. Tang, D. van Swaay, A. deMello, J. L. Ross Anderson, S. Mann, *Chemical Communications*. **51**, 11429–11432 (2015).
196. M. Abbas, W. P. Lipiński, J. Wang, E. Spruijt, *Chem Soc Rev*. **50**, 3690–3705 (2021).
197. M. Frenkel-Pinter *et al.*, *Nat Commun*. **11**, 3137 (2020).
198. S. Tagami, J. Attwater, P. Holliger, *Nat Chem*. **9**, 325–332 (2017).
199. M. Seal *et al.*, *J Am Chem Soc*. **144**, 14150–14160 (2022).
200. B. Drobot *et al.*, *Nat Commun*. **9**, 3643 (2018).
201. R. R. Poudyal *et al.*, *Nat Commun*. **10**, 490 (2019).
202. J. M. Iglesias-Artola *et al.*, *Nat Chem*. **14**, 407–416 (2022).
203. J. D. Bernal, *Proc Phys S B*. **62**, 597–618 (1949).
204. H. James Cleaves II *et al.*, *Chem Soc Rev*. **41**, 5502 (2012).
205. A. Ricardo, M. A. Carrigan, A. N. Olcott, S. A. Benner, *Science*. **303**, 196–196 (2004).
206. G. Costanzo, R. Saladino, C. Crestini, F. Ciciriello, E. Di Mauro, *J Biol Chem*. **282**, 16729–16735 (2007).
207. J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature*. **381**, 59–61 (1996).

208. D. Gibbs, R. Lohrmann, L. E. Orgel, *J Mol Evol*. **15**, 347–354 (1980).
209. R. Mizuuchi *et al.*, *Chem Commun*. **55**, 2090–2093 (2019).
210. E. Biondi, S. Branciamore, M.-C. Maurel, E. Gallori, *BMC Evol Biol*. **7**, S2 (2007).
211. C. B. Mast, S. Schink, U. Gerland, D. Braun, *Proc Natl Acad Sci*. **110**, 8030–8035 (2013).
212. M. Kreysing, L. Keil, S. Lanzmich, D. Braun, *Nat Chem*. **7**, 203–208 (2015).
213. T. Matreux *et al.*, *Nat Chem*. **13**, 1038–1045 (2021).
214. A. Ianeselli, C. B. Mast, D. Braun, *Angew Chem*. **131**, 13289–13294 (2019).
215. A. Ianeselli *et al.*, *Nat Phys*. **18**, 579–585 (2022).
216. M. Morasch *et al.*, *Nat Chem*. **11**, 779–788 (2019).
217. A. V. Dass *et al.*, *ChemSystemsChem*. **5** (2023).
218. A. Ianeselli *et al.*, *Nat Chem*. **14**, 32–39 (2022).
219. E. Salibi, B. Peter, P. Schwille, H. Mutschler, *Nat Commun*. **14**, 1222 (2023).
220. A. Moga, N. Yandrapalli, R. Dimova, T. Robinson, *ChemBioChem*. **20**, 2674–2682 (2019).
221. A. Carbonell, R. Flores, S. Gago, *Nucleic Acids Res*. **39**, 2432–2444 (2011).
222. K. R. Birikh, P. A. Heaton, F. Eckstein, *Eur J Biochem*. **245**, 1–16 (1997).
223. R. Desai, D. Kilburn, H.-T. Lee, S. A. Woodson, *J Biol Chem*. **289**, 2972–2977 (2014).
224. G. Varani, W. H. McClain, *EMBO Rep*. **1**, 18–23 (2000).
225. D. H. Mathews *et al.*, *Proc Natl Acad Sci*. **101**, 7287–7292 (2004).
226. S. H. Bernhart *et al.*, *Algorithms Mol Biol*. **1**, 3 (2006).
227. R. Lorenz *et al.*, *Algorithms Mol Biol*. **6**, 26 (2011).
228. T. Czerniak, J. P. Saenz, *Proc Natl Acad Sci*. **119** (2022).
229. N. Yandrapalli, T. Robinson, *Lab Chip*. **19**, 626–633 (2019).
230. Avanti Polar, Phase Transition Temperatures for Glycerophospholipids. *https://avantili-pids.com/tech-support/physical-properties/phase-transition-temps*.
231. R. Rubio-Sánchez *et al.*, *J Am Chem Soc*. **143**, 16589–16598 (2021).
232. J. A. Doudna, J. R. Lorsch, *Nat Struct Mol Biol*. **12**, 395–402 (2005).
233. E. Spruijt, A. H. Westphal, J. W. Borst, M. A. Cohen Stuart, J. van der Gucht, *Macromolecules*. **43**, 6476–6484 (2010).
234. A. E. Neitzel *et al.*, *Macromolecules*. **54**, 6878–6890 (2021).
235. K. K. Le Vay, E. Salibi, B. Ghosh, T. D. Tang, H. Mutschler, *Elife*. **12** (2023).
236. E. Kizilay, A. B. Kayitmazer, P. L. Dubin, *Adv Colloid Interface Sci*. **167**, 24–37 (2011).
237. C. Jeancolas *et al.*, *Chem Commun*. **57**, 7517–7520 (2021).
238. K. W. Dunn, M. M. Kamocka, J. H. McDonald, *Am J Physiol Cell Physiol*. **300**, C723–C742 (2011).
239. J. W. Szostak, D. P. Bartel, P. L. Luisi, *Nature*. **409**, 387–390 (2001).
240. D. Wollny *et al.*, *Nat Commun*. **13**, 2626 (2022).
241. A. Jenne, M. Famulok, *Chem Biol*. **5**, 23–34 (1998).
242. F. Müller *et al.*, *Nature*. **605**, 279–284 (2022).
243. B. I. Kankia, L. A. Marky, *J Phys Chem B*. **103**, 8759–8767 (1999).
244. K. Le Vay, E. Salibi, E. Y. Song, H. Mutschler, *Chem Asian J*. **15**, 214–230 (2020).
245. J. V Gavette, M. Stoop, N. V Hud, R. Krishnamurthy, *Angew Chem Int Ed*. **55**, 13204–13209 (2016).
246. S. Bhowmik, R. Krishnamurthy, *Nat Chem*. **11**, 1009–1018 (2019).
247. C. He, I. Gállego, B. Laughlin, M. A. Grover, N. V. Hud, *Nat Chem*. **9**, 318–324 (2017).
248. C. He, A. Lozoya-Colinas, I. Gállego, M. A. Grover, N. V Hud, *Nucleic Acids Res*. **47**, 6569–6577 (2019).
249. A. S. Tupper, P. G. Higgs, *J Theor Biol*. **527**, 110822 (2021).
250. E. L. Kristoffersen, M. Burman, A. Noy, P. Holliger, *Elife*. **11** (2022).

251. A. Salditt *et al.*, *Nat Commun*. **14**, 1495 (2023).
252. F. Michel *et al.*, *Genes Dev*. **6**, 1373–1385 (1992).
253. P. Young, M. Ohman, M. Q. Xu, D. A. Shub, B. M. Sjöberg, *J Biol Chem*. **269**, 20229–20232 (1994).
254. J. N. Pitt, A. R. Ferré-D'Amaré, *J Am Chem Soc*. **131**, 3532–3540 (2009).
255. G. S. Filonov, C. W. Kam, W. Song, S. R. Jaffrey, *Chem Biol*. **22**, 649–660 (2015).
256. A. Autour *et al.*, *Nat Commun*. **9**, 656 (2018).
257. E. Szathmáry, J. M. Smith, *Nature*. **374**, 227–232 (1995).
258. C. Jeancolas, C. Malaterre, P. Nghe, *iScience*. **23**, 101756 (2020).
259. T. Furubayashi *et al.*, *Elife*. **9**, 1–15 (2020).
260. F. Juhling *et al.*, *Nucleic Acids Res*. **37**, D159–D162 (2009).
261. A. M. Pyle *et al.*, *Biochemistry*. **33**, 13856–13863 (1994).
262. M. Ruff *et al.*, *Science*. **252**, 1682–1689 (1991).
263. J. A. Doudna, B. P. Cormack, J. W. Szostak, *Proc Natl Acad Sci*. **86**, 7402–7406 (1989).
264. R. R. Breaker, *ACS Chem Biol*. **15**, 2020–2030 (2020).
265. S. S. S. Panchapakesan, R. R. Breaker, *Nat Chem Biol*. **17**, 375–382 (2021).
266. W. A. Kibbe, *Nucleic Acids Res*. **35**, W43–W46 (2007).
267. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, *Nat Methods*. **18**, 100–106 (2021).
268. V. Sharma, ImageJ Plugin HyperStackReg V5.6 (v5.6). Zenodo (2018).

8. Appendix

8.1. Table of RNA sequences:

Key: bold letters are mutations, TP = triphosphate, P = phosphate, lowercase (g) is leaving group for sunY substrates, IVT = *in vitro* transcription, IDT = Integrated DNA Technologies (solid state synthesis).

Name	Length	5' end	Sequence (5' to 3')	3' end	Source
HH-min	39	-	GGGAGCUGAACUGAUGAGUCCGUGAGGACGAAAGGCACA	-	IVT
HH-mut	39	-	GGGAGCUGAACU**A**AUGAGUCCGUGAGGAC**A**AAAGGCACA	-	IVT
HH-FQ-Sub	12	FAM	UGCCUCUUCAGC	BHQ1	IDT
HH-Sub	27	Cy5	UUCACCUCAGAACUGUGCCUCUUCAGC	-	IDT
R3C ligase F1	66	-	GGAAGUUGUUAUCACUUGUUACGUAAGUAACACUUUG-GAUGGGUUGAAGUAUGAGACCGCAACUUA	-	IVT
Hyper-A	52	-	GGAAGUUGUUAUCACUUGUUACGUAAGUAACACUUUG-GAUGGGUUGAAGUAU	-	IDT
Cy5-A	52	Cy5	GGAAGUUGUUAUCACUUGUUACGUAAGUAACACU-UUG**A**AUGGGUUGAAGUAU	-	IDT
B	14	TP	GAGACCGCAACUUA	-	IVT
rt-F	66	-	GGAAGUUGUGUCUUAAUGUUACGUAAGUAACAGAUGGAAU-UGGUUGAAGUAUGAGACCGCAACUUA	-	IVT
rt-A-short	10	-	GGUCUCAUAU	-	IDT
FQ-B	19	TAMRA-TP	GAGACCGCAACUUA|CACAU	BHQ2	IVT
Penta-BHQ2	5	P	CACAU	BHQ2	IDT
v2F	65	-	GGAAGUUGUUAUCACUUGUUACGUAAGUAACACUUUG-GAUGGGUUGAAGUAUGAUAACGGCAGUU	-	IDT/ IVT
dv2F	65	-	GGAAGUUGUUAUCAC**C**UGUUACGU**CC**GUAACA**GCG**UG-GAUGGGUU**UUU**GUAU**UU**UAACGGCAGUU	-	IDT/IVT
S1	31	TP	GAGACCGCAACUUAACUGCUGUUGGAAGUAU	-	IVT
Cy5-S1	31	Cy5	GAGACCGCAACUUAACUGCUGUUGGAAGUAU	-	IDT
FAM-S1	31	FAM	GAGACCGCAACUUAACUGCUGUUGGAAGUAU	-	IDT
sunY	182	-	GGGAAAAUCUGCCUAAACGGGGAAACACUCACUGAGUCAAUCCCGUG-CUAAAUCAG-CAGUAGCUGUAAAUGCCUAACGACUAUCCCUGAUGAAUGUAAGGGAGU AGGGUCAAGCGACCCGAAACGGCAGACAACUCUAAGAGUUGAAGAU-AUAGUCUGAACUGCAUGGUGACAUGCAGGAUC	-	IVT
M1	13	Cy5	CGAAGGAAGGAUU	-	IDT
M2	8	-	gGGUAUGU	-	IDT
M3	11	-	gGGUAUAUUCG	-	IDT
Mango Template	39	-	GGGAAACGAAUAUACCGCAUACCGAUCCUUCCUUCGAAA	-	IVT
B1	19	Cy5	GAGACGGUCGGGUCCAGAU	-	IDT
B2	18	-	gAUUCGUAUCUGUCGAGU	-	IDT
B3	14	-	gAGAGUGUGGGCUC	-	IDT

Broccoli Template	58	-	GGGAGAGAGCCCGCACUCUGCUCGACAGGUACGAAUGUCUG-GACCCGGCCGUCUCAGA	-	IVT
Mango	44	-	GGGCGCGCGAAGGAAGGAUUGGUAUGUGGUAUAUUCGCGCGCCC	-	IVT
F30-Broccoli	108	-	GGGUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAU-AUUCGUAUCUGUCGAGUAGA-GUGUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGAUCAUUCAUGGCAA	-	IVT
TAMRA-F1	10	TAMRA	AAGACUCAGU	-	IDT
F2	9	-	gAGUCGGUU	-	IDT
F3	11	-	gCGUGUAGCAA	-	IDT
F4	14	-	AAGCUACACGGACU	-	IDT
F5-FAM	15	-	gGACUGCUGAGUCAA	FAM	IDT
HH1	11	Cy5	GGAGCUGAACU	-	IDT
HH2	12	-	gGAUGAGUCCGU	-	IDT
HH3	17	-	gGAGGACGAAAGGCACA	-	IDT
HH Template	47	-	GGGAAAUGUGCCUUUCGUCCUCGCGGACUCAUCGGUUCAGCUCCAAA	-	IVT

For sequences produced by IVT, the DNA templates were designed as follows: the sequence was first converted to DNA (by substituting Uracil with Thymine) and the T7 promoter sequence (GAAATTAATACGACTCACTATA) was appended to the 5'-end. Next, the reverse complement of the resulting sequence was generated and used as the DNA template for IVT.

8.2. Uncropped PAGE images

Figure 19b: serial transfers in solution with Cy5-A only as substrate and F1 seed.

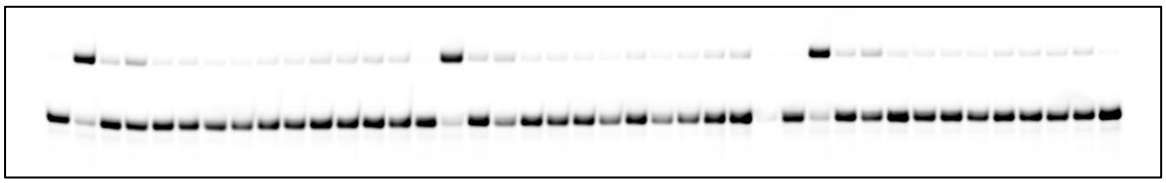


Figure 19c: serial transfers in solution with Hyper-A substrate and no F1 seed.

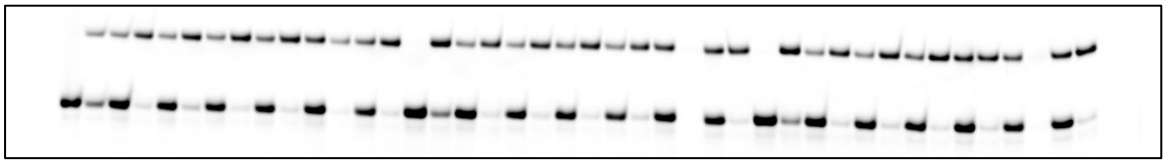


Figure 19d: serial transfers in solution with Hyper-A and F1 seed.

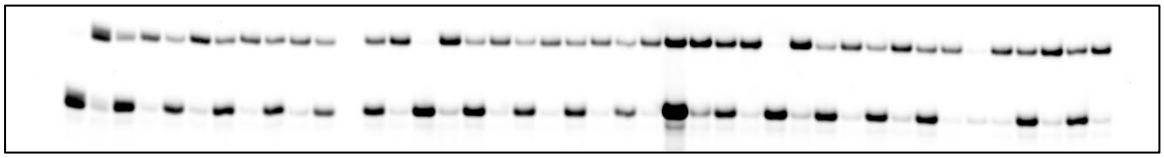


Figure 20: Ligation activity at 0.1 μM and 0.25 μM F1 seed.

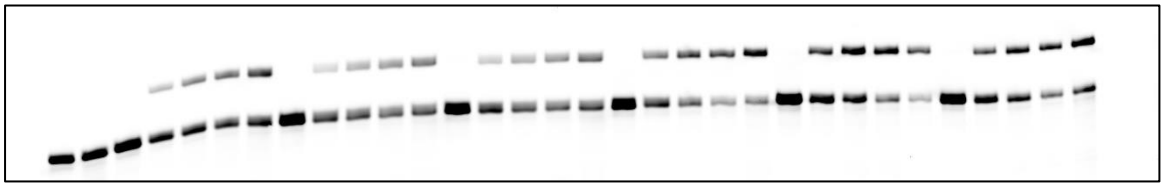


Figure 20: Ligation activity at 0.5 μM and 1 μM F1 seed.

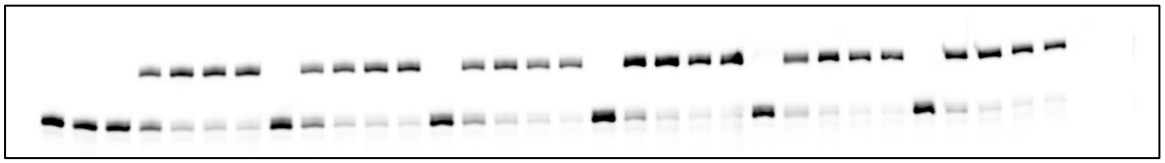


Figure 20: Ligation activity at 0.5 μM and 1 μM Hyper-A seed.

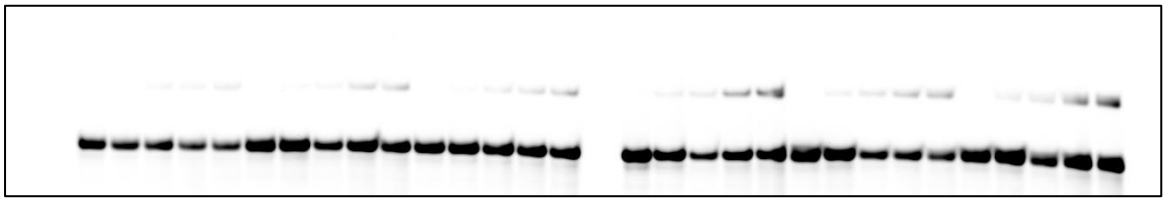


Figure 20: Ligation activity at 1.5 μM, 2 μM and 2.5 μM Hyper-A seed.

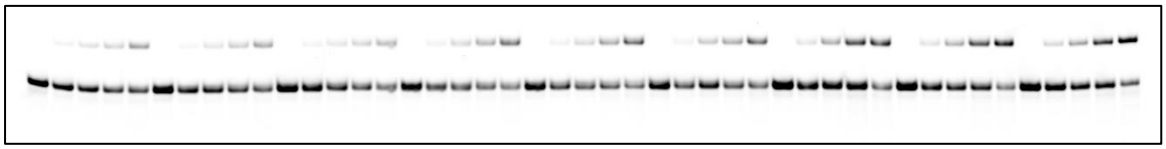


Figure 21e: R3C F1 autocatalysis in GUVs after 0, 1 or 2 freeze-thaw cycles.

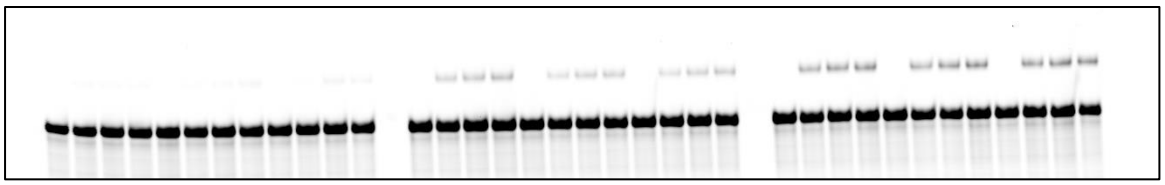


Figure 22b: Freeze-thawing R3C solutions.

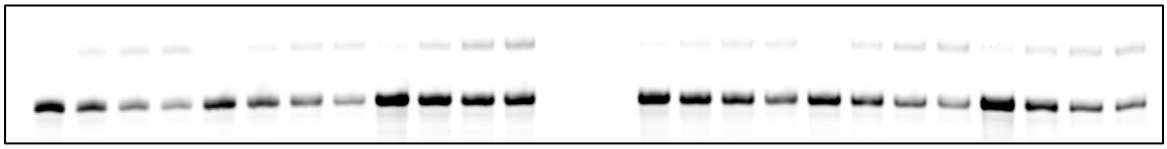


Figure 22b: Freeze-thawing R3C solutions with subsequent incubation at 42 °C.

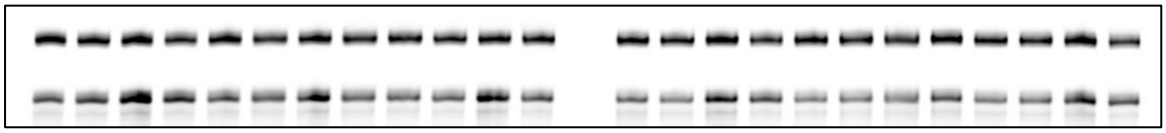


Figure 22d: R3C autocatalysis in solution in presence of GUVs.

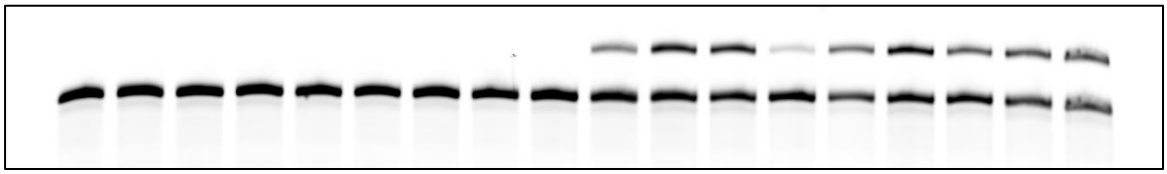


Figure 23: serial transfer of R3C GUVs.

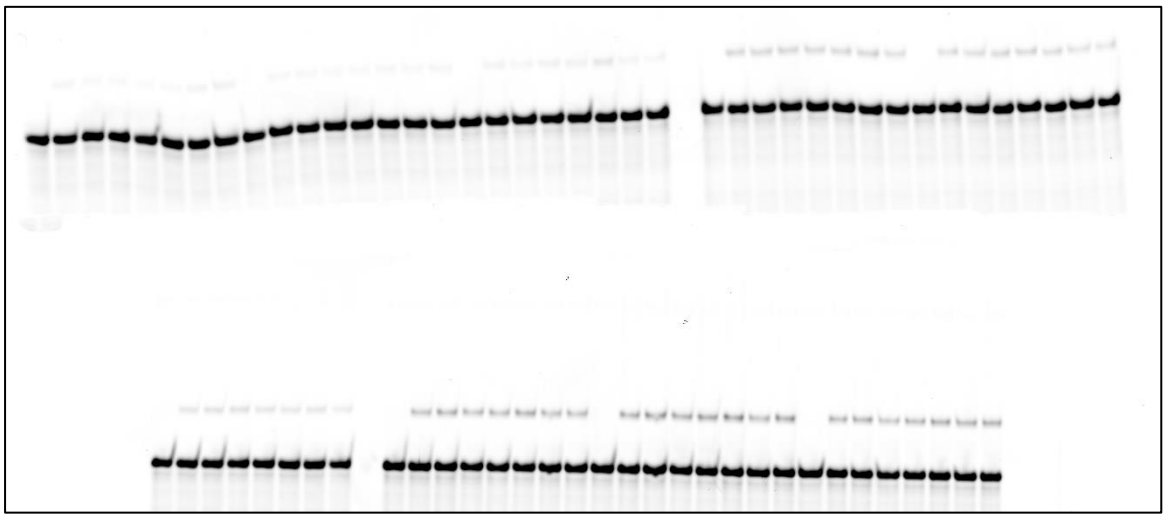


Figure 29: Laddering ribozyme kinetics in solution.

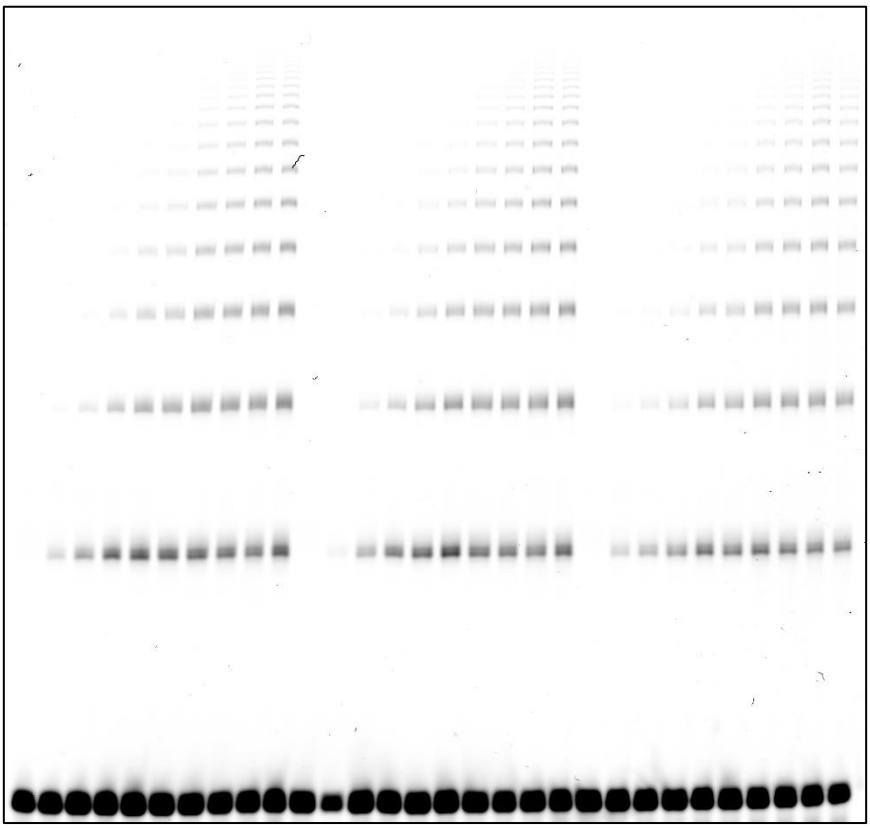


Figure 29: Laddering ribozyme kinetics with 3 mM $(Lys)_{5\text{-}24}$.

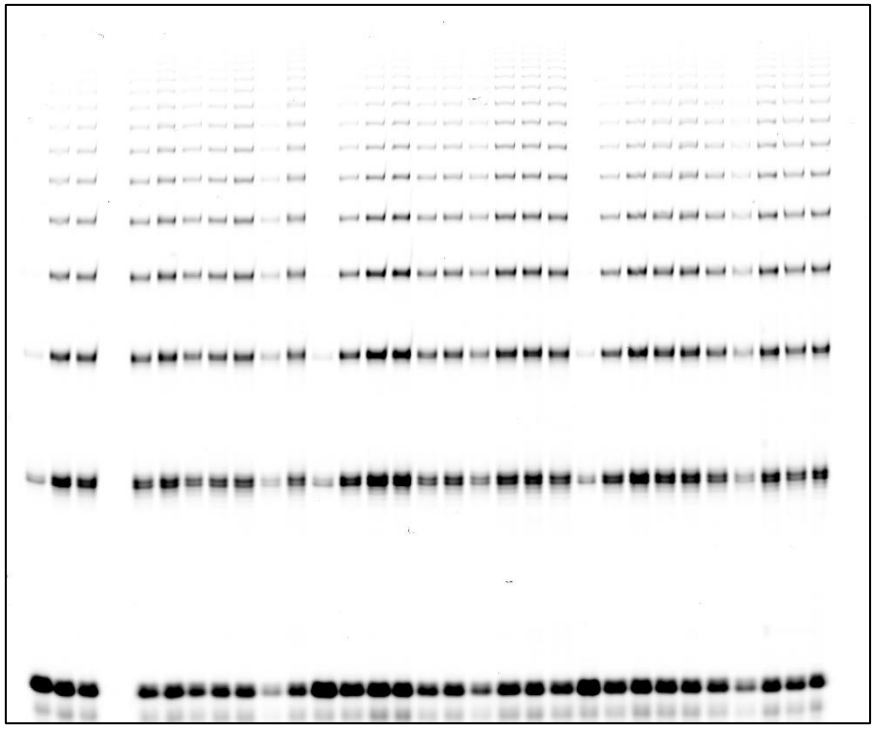


Figure 29: Laddering ribozyme kinetics with 0.75 mM $(Lys)_{19\text{-}72}$.

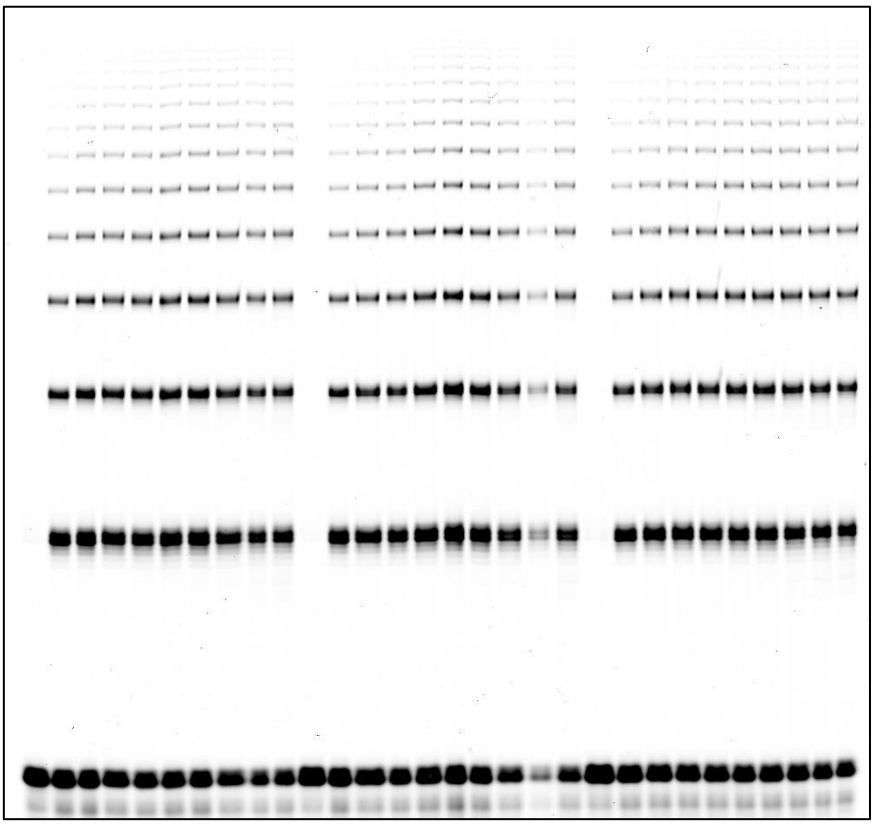


Figure 30b: R3C ladder activity in increasing magnesium chloride concentrations.

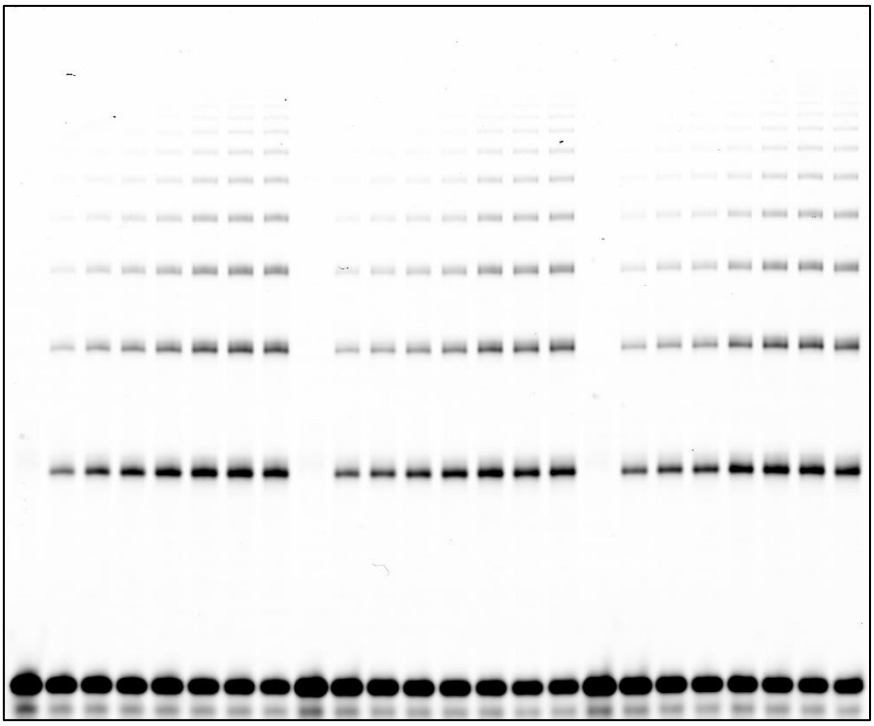


Figure 30c: R3C ladder activity in increasing PEG8000.

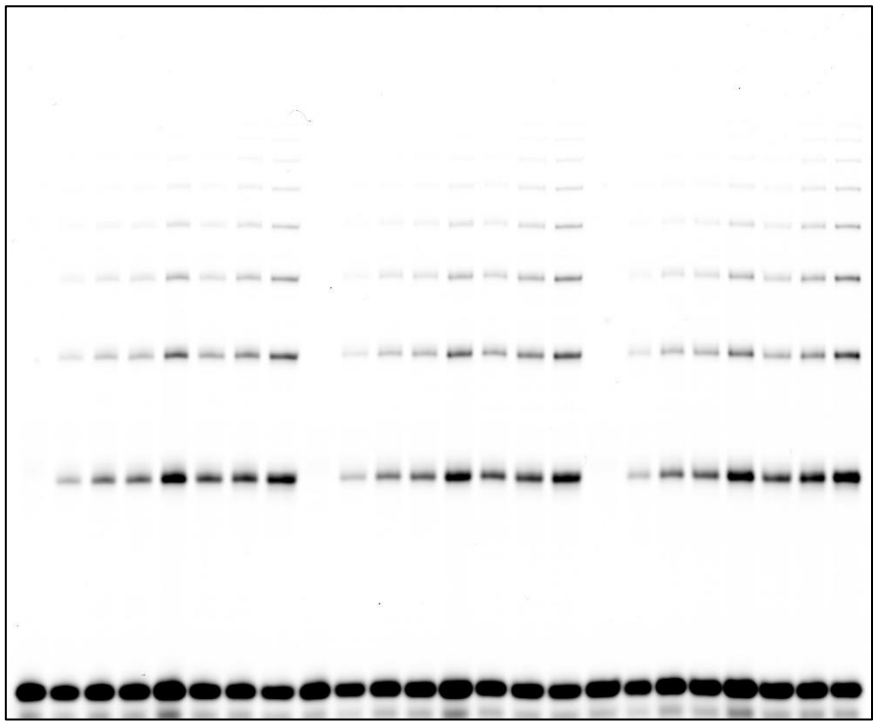


Figure 30d: R3C ladder activity in increasing magnesium chloride and PEG8000.

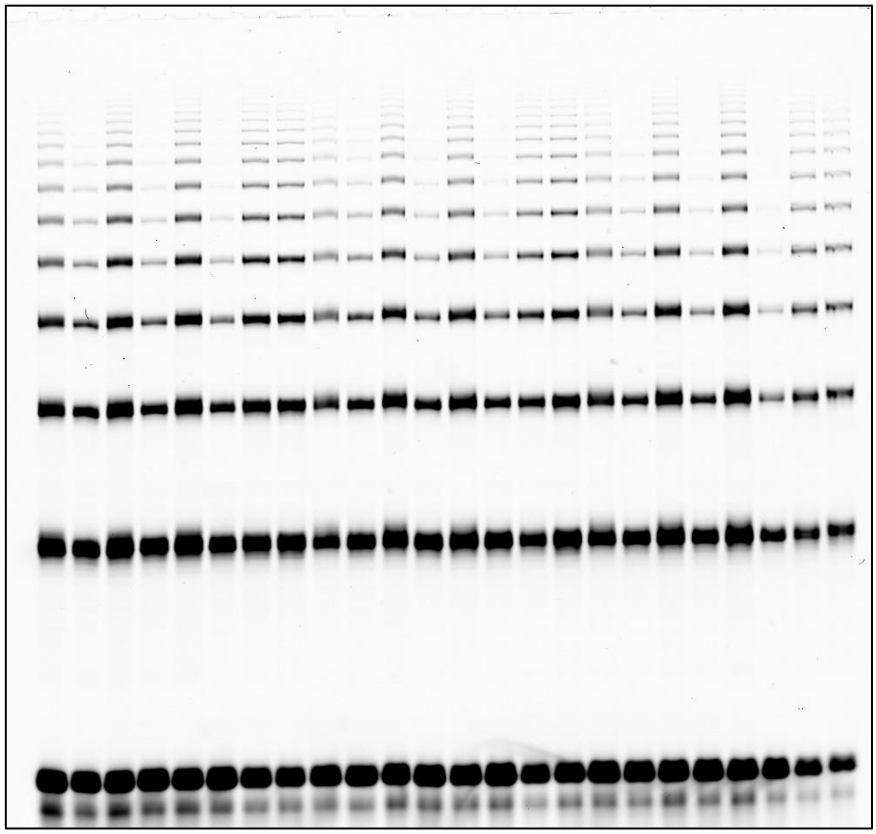


Figure 40: *sunY* kinetics for Mango III assembly.