

Functional capabilities of short deoxyribonucleic acids beyond storing information



Dissertation der Fakultät Chemie & Chemische Biologie zur Erlangung des Doktorgrades der
Naturwissenschaften (Dr. rer. nat.) der Technischen Universität Dortmund

Corbin Machatzke

Technische Universität Dortmund 2026

Diese Dissertation wurde unter der Leitung von Prof. Dr. Hannes Mutschler (Technische Universität Dortmund, Professur für Biomimetische Chemie) angefertigt.

Erstgutachter:	Prof. Dr. Hannes Mutschler
Zweitgutachter:	Prof. Dr. Guido Clever
Tag der Abgabe:	01.06.2026

Table of Contents

1	Acknowledgments	1
2	Summary	2
2.1	English.....	2
2.2	Deutsch.....	2
3	List of Publications.....	4
4	Introduction	5
4.1	Synthetic cells	5
4.1.1	The minimal cell.....	5
4.1.2	Building protocells from the bottom up	7
4.2	Functional ssDNA	10
4.2.1	DNAzymes	11
4.2.2	Aptamers	12
4.2.3	The Lettuce light-up aptamer	13
4.2.4	Retrons are a possible expression vector of intracellular ssDNA	15
4.2.5	The 17E DNAzyme	18
4.3	Complex coacervates as protocells.....	19
4.3.1	Static complex coacervates	19
4.3.2	Active complex coacervates	22
5	Goals.....	25
6	Summary of the individual publications	26
6.1	Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2.....	26
6.2	Regulating Nucleic Acid Catalysis Using Active Droplets.....	27
6.3	DNA affects the phenotype of fuel-dependent coacervate droplets.....	29
6.4	Protocol for the recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets	30
7	Discussion	32
7.1	ssDNA as a biotechnological tool	32

7.2	Towards catalysis, genotypes, and evolution in synthetic protocells.....	36
7.2.1	The influence of the active droplet environment on their DNA-cargo	36
7.2.2	Cargo influence on active droplets.....	38
7.3	Conclusion.....	40
8	References	41
9	Publications	60
9.1	Publication 1.....	60
9.1.1	Personal contribution.....	60
9.1.2	Main Text	61
9.2	Publication 2.....	74
9.2.1	Personal contribution.....	74
9.2.2	Main Text	75
9.3	Publication 3.....	82
9.3.1	Personal Contribution.....	82
9.3.2	Main Text	83
9.4	Publication 4.....	97
9.4.1	Personal Contribution.....	97
9.4.2	Main Text	98
10	Affidavit	138

Acknowledgments

1 Acknowledgments

When I was accepted into this position four years ago, I did not truly believe I could finish this journey. Surrounded by so many brilliant minds, I felt slightly inadequate and out of place among my colleagues. Even now, after submitting this thesis, I still do not fully understand how I managed to do it.

The answer is that I did not do it alone. There were many people who supported me along the way, and this is the place to honor them:

Prof. Dr. Hannes Mutschler is the best supervisor anyone could ask for. He provided me with ample guidance and advice throughout my PhD. I still vividly remember walking back to the lab after my first TAC report, talking to him, and feeling truly valued and accomplished.

I would also like to thank Prof. Dr. Guido Clever for taking the time to evaluate my work.

In the lab, I did not just meet great colleagues; instead, I found some wonderful friends. Listing all the memories I enjoyed from the past four years could fill another 100 pages, so I will try to keep it short:

Wonderful shenanigans involving the Sewer Rats; the breakfast club; exploring new restaurants in Dortmund; and discovering a new money sink in paper crack. Thank you to everyone who joined in, you made these years truly special!

Outside of the Lab, many more people deserve my gratitude:

I thank my parents, whose continuous support made me the person I am today. Thank you to Sebastian and Dennis for listening to me rant about work and for the time we spent together. Thank you to Dominik and Daniel for the countless hours of distraction in stressful times.

I also need to thank the cutest cat in the world, Gnokki. Although she does not know it, she helped me through the hardest phase of writing this thesis with a picture a day from my advent calendar.

Lastly, I want to thank the most important person in my life. Without you I would not even have strived to achieve this, nor been able to push through the bad times. I love you Ernie.

2 Summary

2.1 English

This thesis elucidates the functional capabilities of deoxyribonucleic acids, with emphasis on synthetic biology and origin-of-life research. In the first part, a method is presented for the intracellular expression of ssDNA using a retron-based expression platform. The results demonstrate that the bacterial retron Eco2 is capable of expressing a truncated light-up aptamer, and that purified ssDNA retains aptamer functionality. This represents, to the best of the author's knowledge, the first instance of functional ssDNA aptamers being expressed intracellularly.

In the second part of this thesis the impact of dissipative, non-equilibrium compartments on the activity of a catalytic DNA (DNAzyme) is investigated. To this end, a carbodiimide-fuel-dependent complex coacervate system is employed to characterize which sequence elements determine partitioning of the DNAzyme into droplets. It is then demonstrated that DNAzyme activity can be inhibited by fuel addition and restored after fuel consumption, with the cycle being repeatable multiple times.

In the third part, the impact of short, single-stranded DNA oligonucleotides (ssDNA) on the same coacervate system is examined. In order to do this, a randomized 30 nucleotides (nt) long ssDNA library was combined with the droplet system and the sequence composition of strongly and weakly-partitioning sequences were compared. The base composition and sequence of the DNA have been shown to strongly modulate the lifetime of the droplet, with G-rich termini preventing and A-rich termini accelerating dissolution. Collectively, these findings establish a phenotype–genotype correlation that may guide the development of synthetic protocell models.

The last part of this thesis proposes a sequencing method that has been optimized for the isolation of short ssDNA oligos from synthetic cells. Following library preparation and sequencing, data analysis modules extract maximal information from the sequencing data. The objective of this work is to standardize the data sets used in the bottom-up origin of life and synthetic life fields where genotypes are typically short nucleic acids.

Collectively, these publications contribute to the advancement of knowledge in the field of functional ssDNA, both from a synthetic life research perspective and as a potential tool for future biochemical applications.

Summary

2.2 Deutsch

Diese Dissertation untersucht die funktionellen Eigenschaften von Desoxyribonukleinsäuren (DNA) mit besonderem Schwerpunkt auf der synthetischen Biologie und der Erforschung des Ursprungs des Lebens. Im ersten Teil wird eine Methode zur intrazellulären Expression von ssDNA unter Verwendung einer Retron-basierten Expressionsplattform präsentiert. Die Ergebnisse zeigen, dass Retron Eco2 in der Lage ist, einen verkürzten Lettuce-Light-up-Aptamer zu exprimieren und dass die aufgereinigte ssDNA ihre Funktionalität beibehält. Dies stellt, soweit dem Verfasser bekannt ist, den ersten Nachweis funktionaler, intrazellulär synthetisierter ssDNA Aptamere dar.

Im zweiten Teil wird der Einfluss von Kompartimenten auf eine katalytisch aktive DNA (DNAzym) untersucht. Dafür wird ein transientes, Carbodiimid-getriebenes Koazervatsystem verwendet, um herauszufinden, welche Sequenzeigenschaften des DNA-Moleküls die Aufnahme in Tröpfchen bestimmt. Es konnte gezeigt werden, dass die DNAzym-Aktivität durch Zugabe des Carbodiimids gehemmt und nach dessen Verbrauch wiederhergestellt wird.

Im dritten Teil dieser Arbeit wird der Einfluss kurzer, einzelsträngiger DNA (ssDNA) auf dasselbe Koazervatsystem analysiert. Hierfür wurde eine Bibliothek aus 30 nt langer ssDNA mit den Koazervaten gemischt und die partitionierenden mit den nicht-partitionierenden Sequenzen verglichen. Es zeigte sich, dass Basenzusammensetzung und Sequenz der ssDNA Moleküle die Lebensdauer der Tröpfchen stark modulieren: G-reiche Termini verhindern deren Auflösung, während A-reiche Termini die Lebensdauer reduzieren. Zusammengefasst etablieren diese Ergebnisse eine Phänotyp–Genotyp-Korrelation, die als Grundlage für die Entwicklung synthetischer Protozell-Modelle dienen kann.

Der letzte Teil dieser Arbeit stellt eine Sequenzierungsmethode vor, die für die Isolierung kurzer ssDNA-Oligonukleotide aus synthetischen Zellen optimiert wurde. Zusätzlich zur Sequenzierung werden Datenanalyse-Module, zugeschnitten auf ssDNA in synthetischen Zellen, präsentiert. Ziel dieser Arbeit war es, Datensätze, die im „Bottom-up“-Ansatz der Lebensentstehungs- und synthetischen Lebensforschung verwendet werden, zu standardisieren.

Zusammenfassend tragen diese Publikationen wesentlich zum Fortschritt des Wissens über funktionelle einzelsträngige DNA bei; sowohl aus Perspektive der synthetischen Lebensforschung als auch hinsichtlich ihrer potenziellen Anwendung in zukünftigen biochemischen Systemen.

3 List of Publications

(Co-)First authored publications:

1. Vibhute, M.A.*; **Machatzke, C.***; Bigler, K.; Krümpel, S.; Summerer, D.; Mutschler, H. Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2. *eLife* **2026**, 13:RP99554.
2. Holtmannspötter, A.L.*; **Machatzke, C.***; Begemann, C.; Salibi, E.; Donau, C.; Späth, F.; Boekhoven, J.[†]; Mutschler, H.[†] Regulating nucleic acid catalysis using active droplets. *Angewandte Chemie International Edition* **2024**, 63(47), e202412534.
3. **Machatzke, C.***; Holtmannspötter, A.L.*; Mutschler, H.[†]; Boekhoven, J.[†]. DNA affects the phenotype of fuel-dependent coacervate droplets. *Nature Communications* **2026**, 17, 2953.
4. Holtmannspötter, A.L.*; **Machatzke, C.***; Boekhoven, J.[†]; Mutschler, H.[†] Recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets. *STAR protocols* **2026**, 7, 1, 104293.

In all above-mentioned papers, the two first authors contributed equally.

4 Introduction

4.1 Synthetic cells

“There was Eru, the One, who in Arda is called Ilúvatar; and he made first the Ainur, the Holy Ones, that were offspring of his thought, and they were with him before aught else was made.” These constitute the opening words of the *Ainulindalë*, the first chapter of J. R. R. Tolkien’s *Silmarillion*.¹ It is one of a plethora of works of fiction trying to answer a fundamental question: How is life created?

The cradle of creation is an unknown mystery, filled by Tolkien and others with their own speculative interpretations of its historical occurrence. Throughout the years, the scientific community has engaged in similar practices: Etching out a possible path from lifeless rock to living organism. The term “primordial soup” may be found in fantasy literature, but it actually originates from the first scientific evidence for the generation of amino acids under early-Earth-like conditions. This evidence was provided by Stanley L. Miller in 1953, who demonstrated the creation of the building blocks of life without life itself.² The (Urey-)Miller experiment provided evidence for theories that had been formulated much earlier in the 20th century, most prominently articulated by Oparin and Haldane.^{3–5} Their theories postulated the phototrophic synthesis of organic material – amino acids, sugars and similar – without the aid of organisms, followed by their dissolution in the ocean, yielding the primordial “soup”. Despite the years of iteration and refinement that the theory has undergone, it remains the foundation of contemporary research in the field of origin of life research.

4.1.1 The minimal cell

In an attempt to elucidate the issue from the other end, scientists also endeavor to track extant organisms back towards their origins (Figure 1). This theory is rooted in the concept of the Last Universal Common Ancestor (LUCA), which postulates that all extant life on Earth is traced back to a single organism that existed over three billion years ago.^{6,7} LUCA is estimated to have possessed a substantial level of complexity, with a gene count ranging from 500 to 600 genes, and in some estimations, exceeding 1,500 genes.^{8,9} Irrespective of the precise gene count, the requirement for an organism to sustain processes such as replication, transcription, translation, lipid bilayer maintenance, membrane transport, and energy generation is a significant departure from prebiotic chemistry. In order to bridge this knowledge gap, scientists are attempting to

Introduction

construct model systems that represent intermediate stages between these two extremes. These model systems are referred to as protocells or synthetic cells.

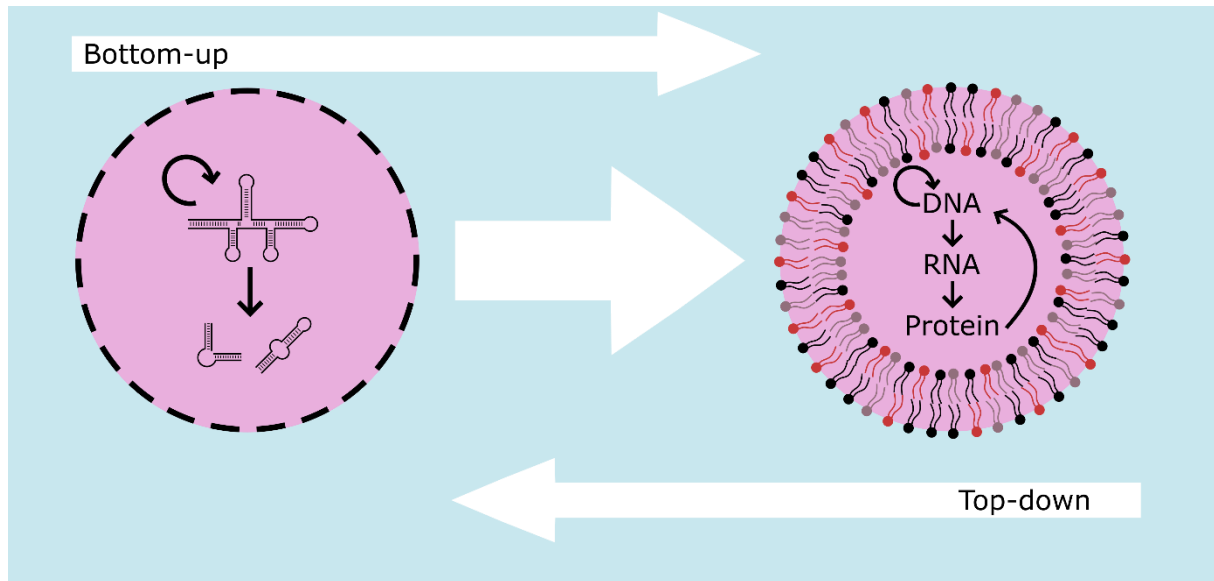


Figure 1: The evolution from primitive, prebiotically plausible protocells into something resembling LUCA remains unresolved. The top-down approach tries to illuminate the problem starting from a modern cell and simplifying it by removing non-essential genes. The bottom-up approach starts from building protocells from prebiotically relevant components under relevant conditions and increases complexity from there.

One approach towards the creation of synthetic cells or protocells involves beginning from a living cell and starting to remove components: genetic circuits, non-essential pathways or regulatory functions. This method is referred to as the top-down approach, whereby the initial steps involve the closure of the aforementioned gap from the final ("top") state. Synthetic cells of this kind are frequently designated "minimal cells", as they retain the characteristics of cells, yet contain a mere fraction of the original components, all of which are necessary for survival.¹⁰ *Mycoplasma mycoides* was selected as a starting point due to its status as the simplest cell capable of autonomous growth.^{11–13} Based on the genome of *M. mycoides*, a synthetic genome containing 1.08 mega base pairs was synthesized and implanted into *M. capricolum* recipient cells.¹³ The newly transplanted cells were identified as *M. mycoides*, thereby demonstrating the feasibility of constructing synthetic genomes and their transplantation into recipient cells.

This initial step demonstrated that it is possible to transplant a synthetic genome into a recipient cell and thereby induce the expression of its traits. Subsequently, the next step was the reduction of the genome size. In the following six-year period, the Venter Lab succeeded in reducing the synthetic genome, termed JCVI-syn1, from 1079 to 531 kilo base pairs.¹² The challenge in this

Introduction

simplification process is the identification of essential and non-essential genes: Merely comparing the genomes of different (but related) organisms is insufficient, since both genomes may contain different genes for the same essential function, but neither of the two genes would be identified as essential. Therefore, the process of optimization was executed through an iterative procedure, whereby a synthetic genome of progressively smaller size was designed and evaluated.¹²

The final genome, termed JCVI-syn3, contains a total of 473 genes, removing 428 from JCVI-syn1.0. This strain mainly contains genes related to transcription and translation, essential metabolism, and membranes, with a total of 79 genes of unknown, but essential, function. In subsequent works, the researchers performed slight modifications of the gene composition, for instance they reincorporated 19 previously removed genes into JCVI-syn3 to enhance its doubling time, or removing individual genes of unknown function deemed non-essential.^{14–16} So far, JCVI-syn3 has been employed to address questions concerning the precursors of cellular life, including the composition of lipid membranes, cell division, metabolism, and adaptive evolution.^{17,18} It is noteworthy that *M. mycoides* is not the only organism utilized for minimal cells. For example, *Buchnera* spp. is another organism employed for the study of minimal cells, although it exclusively appears as a symbiont of aphids.^{19,20}

4.1.2 Building protocells from the bottom up

Whilst these works of simplifying current cells are undoubtedly beneficial, it is also imperative to bridge the gap between the present and the primordial soup from the other side. Mirroring the top-down nomenclature, this approach is called bottom-up: The protocell is constructed through the assembly of its individual components.

In order to construct these systems, it is necessary to pose the following question: what was the phenotype of an organism comprising an extremely limited number of genes? The process of DNA being converted into RNA, and subsequently into protein, is a complex series of events involving the concerted action of over 100 genes, both RNA and protein, in the production and maintenance of this process. Conversely, a system constructed from the bottom up will contain a smaller number of components, and therefore be subject to greater limitations.

Naturally, if the Central Dogma of Molecular Biology cannot apply to early-Earth protocells, we must speculate how life might have looked like prior to this point. The most prominent theory in this field is the RNA world hypothesis, which was developed by Woese, Crick and

Introduction

Orgel in the 1960s and 1970s.^{21–23} This theory postulates, that RNA functioned as both the carrier of genetic information and the catalyst that drove cellular reactions. It is important to note that this does not exclude the presence of other biomaterials, such as peptides or fatty acids, which may have a supportive role in early protocells.²⁴

This theory gained traction when the first catalytical RNAs were discovered within living organisms. In 1982, Cech discovered that in the case of introns, while supported by protein cofactors, the RNA itself performs the catalytic function.²⁵ Approximately a year later, Altman published a paper regarding Ribonuclease P, also demonstrating that the catalytic moiety of the enzyme is in fact the RNA, not the protein.²⁶ The discovery of catalytically active RNAs supported the theories from the late 1960s with experimental evidence, leading to the first coining of the term “RNA-world” by Gilbert in 1986.²⁷

The concept of the RNA world remains somewhat ambiguous, with different scientists offering slightly divergent interpretations. However, all theories share two important assumptions: Firstly, RNA fulfils the function of genetic carrier, and replication is performed using Watson-Crick-mediated copying. Secondly, RNA has been identified as the catalyst of these early cells, and the involvement of genetically encoded proteins in catalysis has been ruled out.²⁸ The discovery that the ribosome is a ribozyme has been interpreted by some scientists as almost definitive proof, showing that RNA is used to construct proteins in one of the most ancient molecular machineries present and highly conserved in all clades of life.^{28–30}

All protocells can be characterized by their implementation of compartmentalization, which serves to effectively distinguish the internal environment of the system from the external environment (Figure 2). In addition to protecting from the outside, compartments have been shown to modulate the catalytic activity of ribozymes. For instance, the Hammerhead ribozyme exhibited a rate of activity that was more than fivefold faster in the presence of coacervates comprising a synthetic polymer and a peptide, a result attributed to the up-concentration of biomolecules within the coacervates.³¹ A similar outcome was observed with multiple ligating ribozymes in the presence of complex coacervates.^{32–34}

In other cases, the use of compartments allows for the selection of specific sequences from a sequence pool. The selection of specific, often longer or more complex sequences is of paramount importance for the selection of functional ribozymes out of a pool of short RNAs. One type of compartment that has the capacity to perform this function are air-water interfaces,

Introduction

which have been engineered to mimic the structure of volcanic rock pores.³⁵ It has been demonstrated that these rock pores possess the capacity to enrich prebiotic molecules, thereby forming protocells that are encased within the solid rock and exhibiting specific selection of longer fragments.^{36,37} In a subsequent publication, it was further demonstrated that these compartments have the ability to harbor active ribozymes, thus facilitating catalysis and even solving the strand separation problem.³⁸

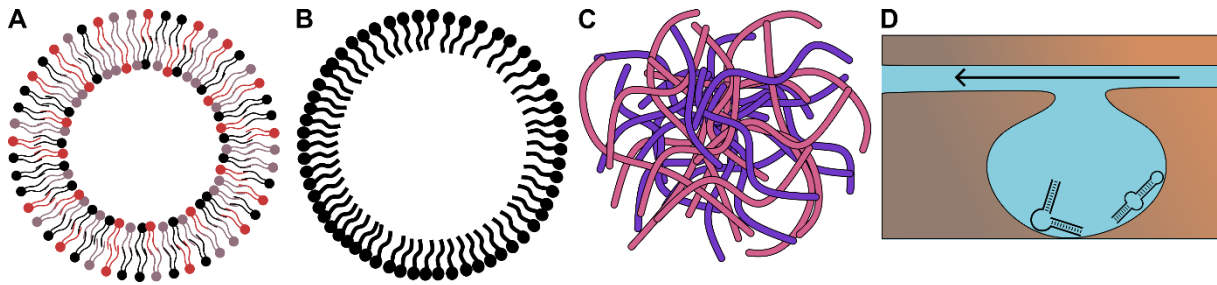


Figure 2: Different types of compartments for prebiotically plausible protocells. A) Lipid vesicles created from (phospho-)lipids. B) Micelles and other simple assemblies created from fatty acids. C) Complex coacervates based on two different polymers. D) A rock pore, acting as a physical compartment.

In addition to the utilization of protocells as a means to emulate the RNA world hypothesis, another application of protocells involves the generation of synthetic life that is not associated with the origin of life on Earth.³⁹ To achieve this objective, it is helpful to shift our focus from the fulfilment of individual hallmarks, such as compartmentalization and metabolism, to the encoding of the fundamental prerequisites of life: self-sustaining, self-replicating, mutating and open-endedly evolving through selection.⁴⁰

In order to achieve this objective, cell-free expression systems (CFES) can be encapsulated and employed as protocell models.⁴¹ CFES utilize the transcription-translation machinery of entire organisms to produce proteins from a DNA template, and can be entirely prepared from purified components or from unpurified cell extract.⁴² CFES have been previously combined, both with phospholipid vesicles,⁴³ or with complex coacervates.⁴⁴ This combination enables the expression of a gene to be coupled to its own genetic material, thus facilitating evolutionary experiments. In order to demonstrate this principle, a multitude of genes coding for RNAs or proteins, either catalytic or of other functions, have been subjected to evolutionary experiments in these systems.^{45–50} While none of them are self-sustaining or fully self-replicating, meaning they cannot produce an exact copy of the entire system, they demonstrate the open-ended evolution capabilities that are prerequisites to life.

Introduction

Self-replicating and self-sustained protocell models have also been previously postulated. These systems typically prioritize on the vesicle encasing the protocell, facilitating modes of growth by, for example, producing the necessary molecules for the membrane.^{51–53} Division, however, remains a more challenging subject: In the bacterium *Escherichia coli*, a complex series of events occurs within the cell to generate concentration gradients. This process is initiated by multiple proteins, including MinC, MinD, and MinE, which work in conjunction with FtsZ to form a ring around the membrane. Subsequently, FtsA and ZipA constrict this ring, leading to the division of the cell.^{54–59} Kohyama et al. have demonstrated a method for reconstituting this entire machinery in liposomes, resulting in controlled and even division of protocells.⁶⁰ They also produced some of the required proteins via a CFES in the protocell. However, this protocell is unable to self-sustain, as the CFES cannot be regenerated.

Some extant organisms also use less complex methods of division, like budding.⁶¹ An approach to this was presented by Zhu and Szostak: A multilamellar vesicle consisting of fatty acids is fed via fatty acid micelles, leading to a linear growth of the original vesicle. Subsequently, moderate shear force proved adequate to disrupt these tubular vesicles, resulting in the formation of multiple daughter cells.⁶² This system does not necessitate a sophisticated machinery, trading the loss of precise control over the size and composition of the daughter cells for simplifying implementation.

4.2 Functional ssDNA

While RNA is the sole catalytically active nucleic acid found in living cells so far, there is no compelling reason for scientists to restrict their methods to the same constraints when attempting to create life from the bottom up. RNA exhibits a limited degree of chemical diversity, with a mere four bases comprising amines, ketones and alcohols as functional groups available for chemical diversity. In addition, the anionic backbone necessitates elevated salt concentrations for adequate folding, further constricting the available design parameters.⁶³

To expand the design space, the consideration of other nucleic acids beyond RNA is important. DNA is the only other nucleic acid that exists in vivo, and therefore the obvious choice. The most significant distinction between RNA and DNA pertains to the 2' hydroxyl group, which removes a functional group from the backbone while simultaneously enhancing temperature- and pH-stability.⁶⁴ This enables ssDNA to function in environments that would be hostile to an RNA molecule, such as in the presence of strong condensing agents like carbodiimides.⁶⁵

Introduction

Additionally, DNA has the capacity to modify the characteristics of specific artificial cells, for instance facilitating a higher fluidity in RNA/peptide complex coacervates, showing unique behaviors that can be utilized in building artificial cells.⁶⁶ Apart from RNA and DNA, scientists have designed a plethora of analogs of nucleic acids not found *in vivo*. This group of synthetic nucleic acid alternatives is referred to as xeno nucleic acids (XNA).^{67,68}

4.2.1 DNAzymes

Given the absence of naturally occurring DNAzymes, all existing ones have been selected or evolved *in vitro*.^{69–72} The methods, initially developed for the selection of functional RNA such as ribozymes and aptamers, are applicable to functional DNA with the added benefit of circumventing reverse transcription prior to amplification each round. The first DNAzyme selected through this method is the RNA cleaving Gr-5 DNAzyme, published in 1994.⁷³ The complete library utilized for the selection process contained just over 1,000 sequences, and the activity exhibited was minimal with a ~100-fold increase in comparison to the uncatalyzed reaction. The specific activity selected for is the most prevalent activity exhibited by all DNAzymes: the cleavage of a specific RNA substrate via transesterification (Figure 3).

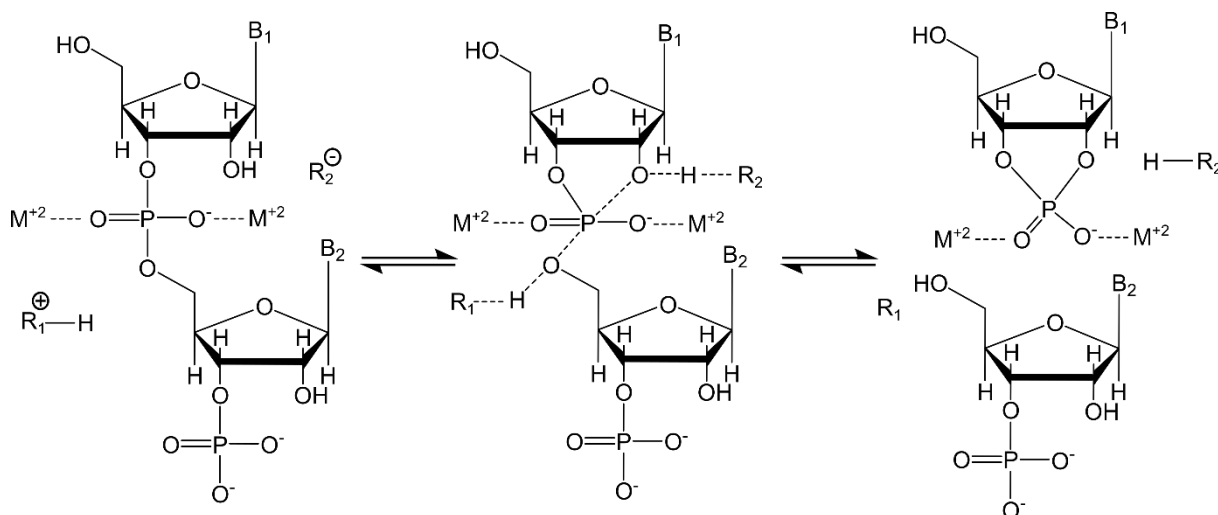


Figure 3: General mode of catalysis for RNA cleaving DNAzymes. R₁ and R₂ are part of the DNAzyme, while M⁺² represents a divalent metal ion, the most common cofactor. The DNAzyme performs multiple functions during catalysis: Stabilization of the phosphate, arranging geometry of the in-line attack, supporting the 2' alcohol by inducing a negative charge and stabilizing the 5' leaving group.

A multitude of RNA-cleaving DNAzymes have been selected, all exhibiting specificity towards various cofactors or conditions. The general mode of catalysis is consistent across all of them: The 2'-OH of the 5'-terminal ribose of the substrate RNA performs an in-line nucleophilic attack at the phosphate, with the 3'-terminal RNA fragment acting as the leaving group. This

Introduction

reaction yields two cleaved RNA fragments: A 5'-fragment with a 2'-3'-cyclic phosphate and a 3'-fragment with a free 5'-alcohol group.⁷⁴ In most cases, this transesterification reaction requires a metal ion as a cofactor, with divalent metal ions exhibiting the highest prevalence.⁷⁵

Intriguingly, RNA-cleaving DNazymes can be utilized as a biosensor if a readout is coupled to the underlying cleavage reaction. One common methodology entails the incorporation of a quencher and a fluorophore on opposite ends of an RNA substrate. Upon DNzyme catalyzed cleavage, it will induce an increase in fluorescence as fluorophore and quencher disassociate after RNA cleavage.⁷⁶ A typical class of target molecules to sense with DNazymes are metal ions. The employment of a DNzyme specific to a single metal ion can be utilized to sense the presence of that specific cofactor. The development of biosensors specific to Pb^{2+} ,^{77,78} Cu^{2+} ,⁷⁹ UO_2^{2+} ,^{80,81} and Hg^{2+} ,^{82,83} has been documented. In addition, a considerably number of DNazymes that are specific to other ions, such as Zn^{2+} ,^{84,85} Ca^{2+} ,⁸⁶ Cd^{2+} ,⁸⁷ Cr^{+3} ,⁸⁸ Ln^{+3} ,^{89,90} and Ag^{+} ⁹¹ have been identified.^{92,93} These DNazymes have the potential to serve as further biosensors with relevance to public health and food safety.

In addition to RNA-cleaving DNazymes, other catalytic activities have been documented. The repertoire of DNazymes encompasses a variety of functions, including DNA cleavage,^{94,95} RNA ligation,^{96,97} (de)phosphorylation,^{98,99} peroxidase activity,¹⁰⁰ and depurination,¹⁰¹ among others. This extensive array of activities gives rise to numerous potential protocell designs incorporating DNazymes. It is particularly intriguing to examine the elongation of other nucleic acids, for example via ligation. This would enable the development of synthetic genomes that are exclusively dependent on nucleic acid-based catalysts.

4.2.2 Aptamers

Another class of functional nucleic acids is aptamers. Aptamers are the nucleic acid analog to antibody proteins. These molecules assume a unique three-dimensional fold, enabling highly specific binding of a ligand, ranging from small molecules to proteins.¹⁰² Selection is also performed via in vitro selection, like SELEX, in a similar design to DNzyme discovery: A library of potential aptamers is presented with the ligand of interest, tethered to a solid phase. Subsequent to an incubation step, all non-binders are washed away. Binders are amplified via (error-prone) PCR and the cycle is repeated until strong binders are identified.¹⁰³ The backbone chemistries of these binders can be based on RNA, DNA, or XNA. Since 2010, there has been

Introduction

an increase in the number of publications of DNA-based aptamers due to their higher stability compared to RNA and the ease of amplification compared to XNA.¹⁰⁴

A frequently utilized class of aptamers used in research settings are fluorescent light-up aptamers (FLAPs). FLAPs bind a small molecule ligand that is almost non-fluorescent when free in solution but exhibits a significant increase in fluorescence when bound to the aptamer.¹⁰⁵ This allows for a rapid fluorescent readout for a wide range of experiments, for example as a readout for RNA transcription¹⁰⁶ or as a fluorescent marker in cells.¹⁰⁷

Aptamers of other classes often target therapeutically relevant targets for two reasons. First, they can be used to visualize these targets in diagnostics. Second, they can be used to modulate them in some way, either directly by binding to them or indirectly via recruitment of other drugs or immune system components.¹⁰⁴ Currently, aptamers specific to treating eye disorders,¹⁰⁸ thrombosis,¹⁰⁹ and cancer,¹¹⁰ among others,¹¹¹ are approved or in clinical trials as therapeutic treatments. The most significant advantage of aptamers over antibodies as therapeutic agents is primarily logistics. This is due to the fact that aptamers can be entirely produced in vitro and are more stable than antibodies.^{104,112}

4.2.3 The Lettuce light-up aptamer

The Lettuce DNA light-up aptamer utilized in the first publication listed in this thesis was also selected in vitro using the SELEX protocol.^{72,113,114} Lettuce is a DNA mimicry of the green fluorescent protein (GFP), akin to the RNA aptamers Spinach and Broccoli.^{107,115,116} All of them bind the same ligand, 3,5-Difluoro-4-hydroxybenzylidene imidazolinone (DFHBI-1T), which is nearly non-fluorescent in solution but exhibits a pronounced (>1.000 fold) increase in fluorescence upon binding the RNA aptamers, while the increase in fluorescence is more modest upon binding Lettuce (~30 fold).¹¹³

The Lettuce FLAP is composed of two stems, named P1 and P2.^{113,117} Truncation of the P1 stem results in the generation of various shortened variants, such as 4-Lettuce, 8-Lettuce, or 11-Lettuce (Figure 4). The P2 stem may be shortened by redesigning the FLAP, such that its 5' and 3' terminus are positioned within the P2 stem, resulting in the deletion of the entire P1 stem. This process yields the shortest possible variant of the Lettuce aptamer, termed “Bibb Lettuce”.¹¹⁷ The main non-base paired region of the Lettuce aptamer forms a complex binding pocket for its ligand involving multiple non-canonical base pairs and two G-quartets forming an antiparallel G-quadruplex. This G-quadruplex, composed of four surrounding antiparallel

Introduction

DNA strands, a single Cytosine from P2, and a monovalent cation, serves as the coordinating agent for the ligand, DFHBI-1T. This results in the formation of a tight bond to the aptamer, thereby inducing fluorescence enhancement.¹¹⁷

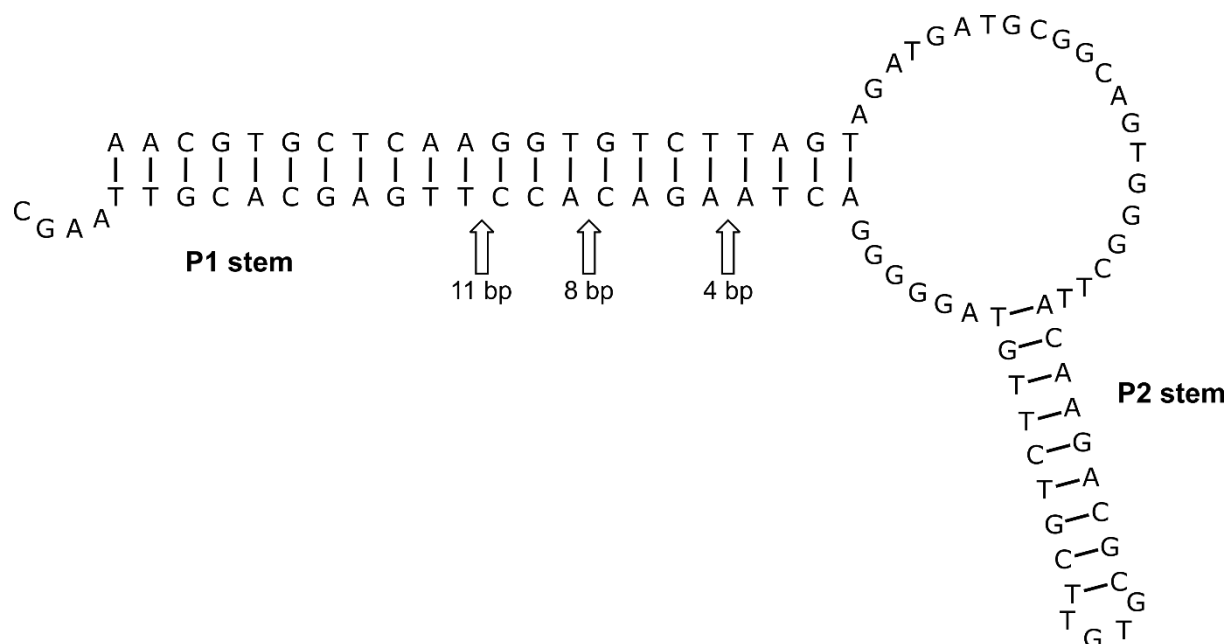


Figure 4: Structure of the Lettuce light-up aptamer. The P1 stem stabilizes the active fold generated by a G-quadruplex forming in the open ring area and can be truncated as indicated by the arrows to generate the shortened variants 4-, 8- and 11-Lettuce. The P2 stem is partly relevant for fluorescence but can be shortened as well. This minimal variant with both P1 and P2 shortened is named Bibb Lettuce.

The selection of Spinach2 and Broccoli was specifically driven to function *in vivo* for live cell imaging. These constructs have been extensively utilized to image various target classes, predominantly by co-expression with the RNA of interest.^{114,118–120} RNA aptamers can be selected for *in vivo*. Therefore, the RNA of interest is synthesized within cells by the cell's intrinsic transcription machinery. The use of membrane-permeable small-molecule ligands, such as DFHBI-1T, allows for direct delivery of the ligand by supplementing the growth medium. This facilitates a direct protocol of transforming an *in vitro* preselected DNA library encoding for RNA aptamer variants on a plasmid into cells, expressing the different RNA aptamers in cells, sorting the cells with FACS, and finally selecting the brightest variants for sequencing and characterization.

This selection cycle cannot be replicated for DNA aptamers, since cells do not innately possess the machinery to express ssDNA, and free ssDNA in cells generally results in a stress response linked to DNA damage or viral infection.^{121–124} Therefore, a new method of expressing

Introduction

functional ssDNA intracellularly is needed in order to evolve ssDNA light-up aptamers that properly work in cells.

4.2.4 Retrons are a possible expression vector of intracellular ssDNA

Retrons were initially identified in the 1980s as short, multi-copy, single-stranded DNA sequences (msDNA). The initial finding was made in *Myxococcus xanthus*, a gram-negative bacterium derived from a soil sample.¹²⁵ Subsequent investigations revealed that analogous msDNAs were present in additional species, suggesting that these msDNAs are not uncommon but rather prevalent among bacteria.^{126,127} Following this discovery, structural analysis of these msDNAs was conducted, revealing their composition to be not only ssDNA, but instead a hybrid of ssDNA and RNA, with a covalent bond between the 2'-alcohol of a conserved G residue and the 5'-terminus of the ssDNA.¹²⁸ Following this discovery, a gene cassette from *E. coli* found to encode for retron Eco1 was cloned into *E. coli* K12 strain and sequenced, which revealed multiple components within this gene cassette: A non-coding RNA sequence containing the RNA covalently bound to the msDNA and the reverse complement of the msDNA and an open reading frame coding for two proteins: One exhibits notable parallels with retroviral reverse transcriptases (RT), while the other displays striking similarities to ribonuclease H.¹²⁹

The biological function of retons remained elusive for a long time, until, their role as an anti-phage defense system in bacteria was discovered 36 years following their initial discovery.¹³⁰ Bacterial organisms harbor defense islands, which are regions within the respective organism that contain numerous genes that are unrelated to each other, yet are always associated with anti-phage defense.^{131,132} Retrons are situated within these defense islands, which led to the identification of their biological function as toxin-antitoxin systems.¹³⁰

Toxin-antitoxin systems are comprised of a toxin capable of disrupting a vital cellular process and an anti-toxin that prevents the toxins' action, either through direct inhibition or indirectly by inhibiting translation or reversing the toxic effect.^{133–135} Once a specific trigger condition is met, the anti-toxin ceases its inhibitory function, and the toxin becomes active. Toxin-antitoxin systems have been classified into three distinct categories: The first one is post-segregationally killing, in which the loss of a critical gene leads to cell death.¹³⁶ The second category is abortive infection, which occurs when bacteria are infected with phages, leading to abortive cell death to halt viral replication and growth.¹³⁷ The final category is persister formation, a protective

Introduction

dormant state that cells can enter to survive antibiotic treatment.^{133,138} Retrons specifically fall into the category of abortive infection, as they encode a toxin that will lead to cell death upon phage infection.

A comprehensive investigation involving metagenomic database searches has led to the identification of over a thousand distinct retron genes.¹³⁹ A notable observation from this analysis is the striking similarity in the structural organization of these genes. The gene cassette always commences with a non-coding RNA (ncRNA), called the *msr* (multicopy single-stranded RNA), followed by the *msd* (multicopy single-stranded DNA). The *msd* then serves as the template for the RT, which produces the msDNA. Following the *msd* starts the open reading frame, first encoding the retron specific RT, and finally encoding the toxin, which is referred to as effector protein.¹⁴⁰ The prediction of the specific msDNA sequence from metagenomic data is challenging; consequently, each retron must be investigated individually to identify its specific sequence.¹⁴⁰ Furthermore, the identification of effector proteins is also a tedious process due to their diverse range of activities.^{139,141–144}

Following intracellular transcription, the RT is translated and immediately binds its own ncRNA, recognizing the *msr* and *msd* regions for reverse transcription.^{145–147} A strongly conserved guanosine located near the 5'-terminus of the *msr* is recognized, and its 2' OH is used as a priming site for the msDNA, reverse transcribing the entire *msd*.¹⁴⁷ RNase H, an RNase important to be present in the host organism, recognizes specifically RNA-DNA-hybrid duplexes and digests the RNA, leaving the newly reverse transcribed msDNA intact.^{147–150} The finished RNA-DNA covalently linked hybrid structure binds to both the reverse transcriptase and the effector protein, creating an observatory complex that inhibits the effector protein until a phage infection triggers the release of the effector protein and activation of the toxin (Figure 5).^{143,144,151}

An important application of retons at present is the intracellular production of ssDNA. This has particularly garnered interest in the context of CRISPR/Cas applications, wherein the donor template DNA can be expressed intracellularly in conjunction with the entire Cas ribonucleotide complex, as opposed to the conventional method of manual transfection into cells.^{152,153} This approach is particularly compelling given the challenges associated with transfection of ssDNA, including its low efficiency and the variability in outcomes.^{154,155} Retrons offer potential solutions to these issues by facilitating high-throughput screening of a broad spectrum of mutations due to their high editing efficiency.¹⁵⁶

Introduction

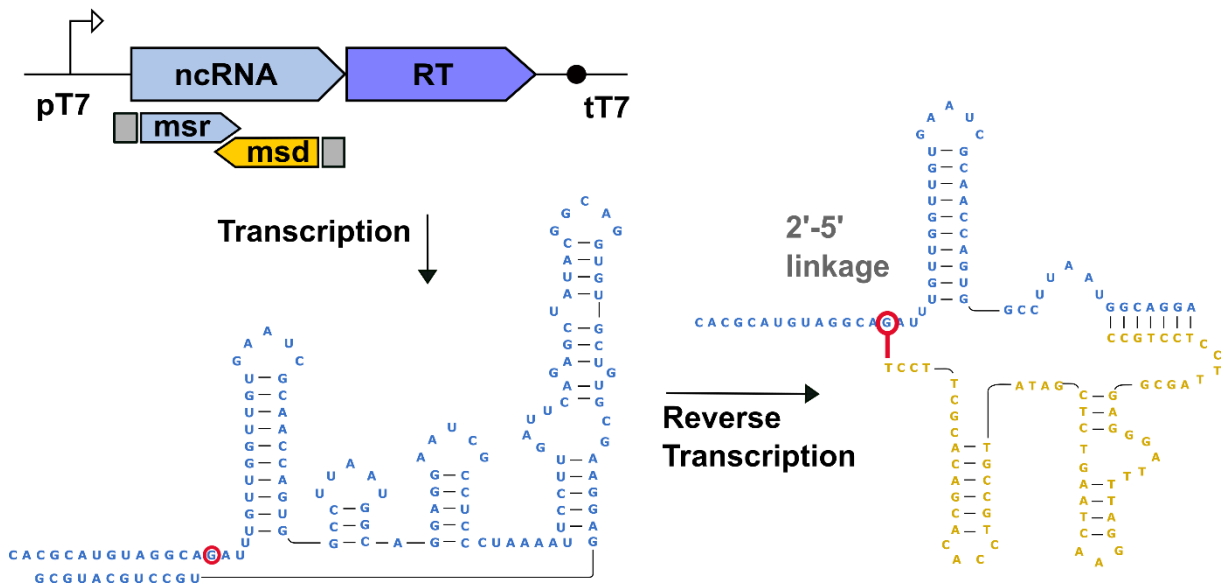


Figure 5: Design of retron with the structure of retron Eco2 as an example. The gene cassette encodes for the ncRNA, followed by the reverse transcriptase and the effector protein, which are fused into one protein in the case of Eco2. After transcription and translation of the protein, the RT binds the ncRNA and starts reverse transcribes half of it into ssDNA, with the template RNA being removed by RNase H. The 5' of the ssDNA is thereby linked to the 2'-OH of a highly conserved guanine residue on the ncRNA.

The expression of Cas related ssDNA is only one application of retrons as a ssDNA expression platform. Furthermore, the expression of functional ssDNA within the organism also enables the evolution of ssDNA, resulting in sequences that demonstrate superior functionality in vivo compared to those selected in vitro. The 10-23 DNAzyme has been expressed in *E. coli* using retron Eco1, targeting the mRNA of the cyan fluorescence protein. This allows for a straightforward readout of DNAzyme activity due to the direct correlation of protein concentration and fluorescence.¹⁵⁷ The next logical step would be to improve upon the activity of the 10-23 DNAzyme using this system.

From the extensive list of retrons, retron Eco2 (Ec67 in alternative nomenclature) was used for intracellular ssDNA display in Publication 1 of this thesis.¹¹⁴ Isolated from *E. coli*, retron Eco2 is among the earliest retron systems identified, with its initial documentation in 1989.¹⁵⁸ Retron Eco2 possesses two notable characteristics that distinguish it from numerous other retrons: First, the effector protein is fused with the reverse transcriptase, resulting in the expression of both as a single protein. Secondly, retron Eco2 possesses its own RNase H activity, thereby eliminating its reliance on the host to provide this essential protein in sufficient quantities for induced overexpression.¹⁴⁷ The msDNA of retron Eco2 is 67 nucleotides in length, with the final seven nucleotides at the 3' end hybridizing with the corresponding seven nucleotides at the 3' end of the *msr*. The 5'-end of the msDNA is covalently linked to the 2'-OH of G15 of the

Introduction

msr.¹²⁹ The effector protein of retron Eco2, fused with the C-terminus of the reverse transcriptase, contains a topoisomerase-primase-like domain that can degrade tRNAs leading to inhibition of translation.^{139,143}

4.2.5 The 17E DNAzyme

The DNAzyme employed in the second publication of this thesis is the 17E DNAzyme (Figure 6).⁶⁵ The 33 nucleotide long 17E DNAzyme was originally selected in 2003 from a pool of $\sim 10^{14}$ 40mer sequences.⁸⁴ The DNAzyme was selected in two phases: A first phase consisting of 12 rounds of selection, followed by sequencing and a subsequent second selection over 6 rounds using an optimized library with 20% sequence degeneracy compared to the most abundant sequence of the first 12 rounds.⁸⁴ While the 17E DNAzyme was selected as a Zn(II)-dependent catalyst,⁸⁴ subsequent analysis demonstrated higher activity in the presence of Pb(II), low activity in the presence of Mn(II) and Co(II), and minimal activity with further divalent metal ions.^{159,160} Intriguingly, there is almost no activity observable in the presence of Mg(II), which is likely the most prevalent cofactor for RNA-cleaving DNAzymes.¹⁵⁹

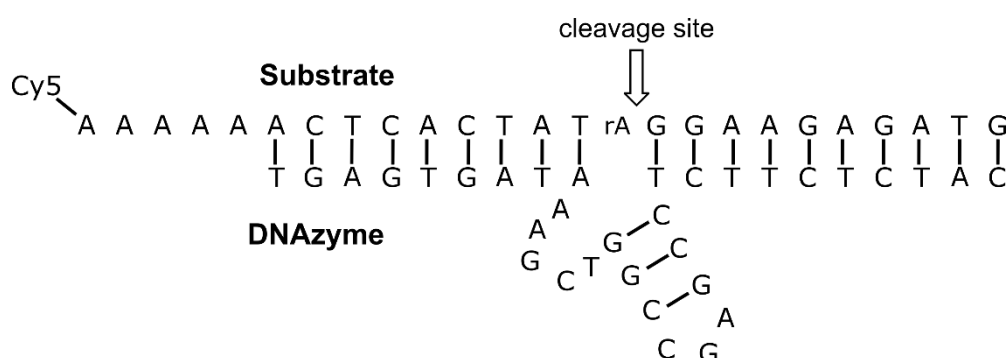


Figure 6: Structure of the 17E DNAzyme and substrate as utilized in the subsequent project. The active site of the DNAzyme is in the bulge and short stem loop, complexing a Zn^{2+} ion. The cleavage site is right next to the rA, utilizing the 2'-OH of the RNA moiety as a nucleophile for the in-line attack.

The 17E DNAzyme utilizes a distinct two-step mechanism: In the initial step, a hydrophilic attack is initiated by the 2'-OH of the RNA on the adjacent phosphate group, resulting in the formation of a 2'-3' cyclic phosphate and the release of the 5'-fragment.¹⁵⁹ Subsequent to this, the cyclic phosphate is hydrolyzed, resulting in the formation of either a 2'- or 3'-monophosphate group on the 5'-fragment of the substrate. This mechanism elucidates the pivotal role of the single RNA base at the cleavage site in catalysis, while the remainder of the substrate can be substituted with DNA moieties. In the case of all divalent ions, the expected mode of action is a Lewis acid catalysis, though specifically for the Pb(II) a general acid

Introduction

catalysis cannot be ruled out.¹⁶¹ This finding is consistent with the mode of catalysis for other closely related RNA-cleaving DNAs.^{162–164}

Despite being selected at a pH of 7, and the catalytic mechanism involving a single deprotonation, the 17E DNase exhibits a moderate activity at acidic pH values.^{159,160} This distinctive property of the 17E DNase is a primary factor that led to its selection as the genotype cargo for the complex coacervates, given that the fueling reaction involving 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC) necessitates an acidic pH of ~5.3.^{165,166} Additionally, the utilization of an all-DNA catalyst and majority-DNA substrate offers enhanced stability against the reactivity of the fuel, which is a significant benefit. The 2'-alcohol group has the capacity to function as a nucleophile subsequent to activation of a carboxylic acid moiety by EDC, resulting in irreversible linkage of RNA and precursor peptide. In the event that this reaction occurs within the active site or binding arms of a nucleic acid catalyst, it is to be expected that this will likely result in inactivation by steric inhibition of the catalytically active fold.

4.3 Complex coacervates as protocells

4.3.1 Static complex coacervates

Complex coacervates are a form of liquid-liquid phase separation, bearing resemblance to an oil droplet suspended in an aqueous solution. However, in contrast to phase separation occurring due to the use of different solvents, such as water and oil, the two phases of complex coacervates are both aqueous phases: A polymer poor phase, making up the bulk of the solution, and a polymer rich phase, containing a majority of the polymers that create the complex coacervates.^{167–170} Complex coacervates are distinguished from simple coacervates by their propensity to contain at least two different polymers. In contrast, simple coacervates consist of a single polymer that carries all the moieties needed to form coacervates.¹⁷¹

Introduction

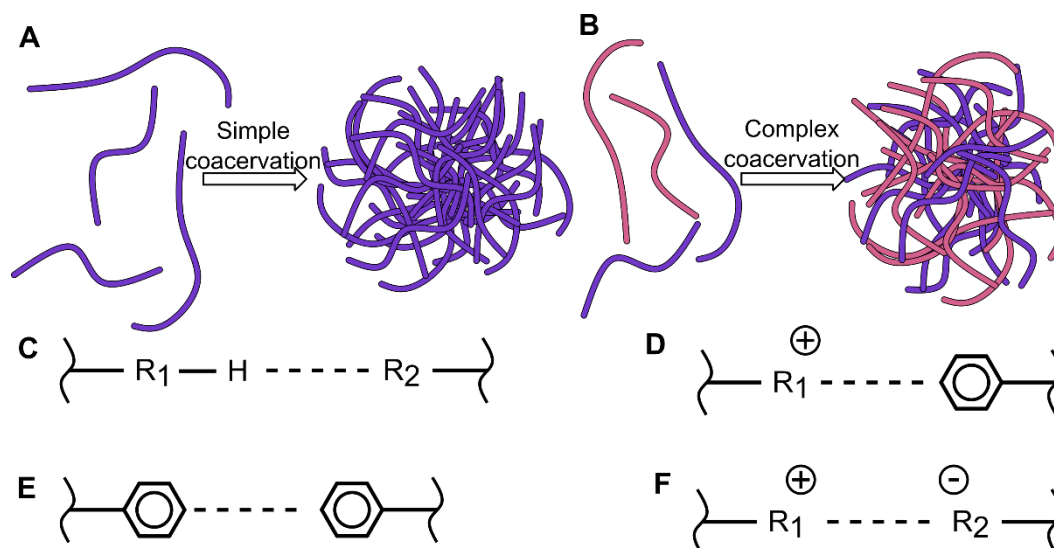


Figure 7: Some of the most common compositions of coacervates. A) Simple coacervation occurs when a single polymer species phase separates. B) Complex coacervation occurs when at least two different polymer species phase separate. C) Hydrogen bonds, D) cation- π interaction, E) π - π interaction and F) charge-charge interaction are some of the most common forces driving the phase separation. Most polymers are oppositely charged, but a mix of multiple attractive forces can be utilized in the formation of complex coacervates, for example by employing a heteromeric peptide with both charged moieties and aromatic residues.

While complex coacervates can be generated via a diverse set of attractive forces, including hydrogen bonds, cation- π , or π - π interactions, the strongest and most common interaction is the attraction between a positively and a negatively charged polymer (Figure 7).^{171–176} In this case, the positive charges of a poly-cation and the negative charges of a poly-anion will attract each other and form the polymer-dense phase, which generally has gel-like properties, compared to the very liquid polymer-poor phase. The process of droplet formation is contingent upon the interaction of charged moieties; consequently, a substantial alteration in the ionic strength of the solution, for example through the excessive addition of salts, will lead to the dissolution of droplets, as the charge-charge interactions will be perturbed by the increased salt concentrations.^{169,177}

Complex coacervates present an attractive alternative to vesicles for the compartmentalization of bottom-up protocell models.¹⁷⁸ While vesicles are the more obvious choice as chassis for synthetic cells, mainly since they are the predominantly used compartment by all forms of extant life, complex coacervates offer two distinct advantages. Firstly, complex coacervates are not contingent upon any form of cross-membrane transport for the purpose of facilitating the movement of nutrients into the cell or the expulsion of waste from the cell. The liquid-liquid phase interface allows for diffusion of various water-soluble molecules in and out of the

Introduction

protocell. The concentration of a given species within the protocell is thus contingent upon the partitioning coefficient of that species. In an ideal scenario, a nutrient would possess a high partitioning coefficient, while waste would exhibit a lower partitioning coefficient. This form of passive transport vastly reduces the requirements for a protocell system, since neither components nor energy for transport is needed.

Secondly, complex coacervates are chemically less complex than most vesicle protocell systems. This property makes them much more robust and easier to adopt to various physicochemical environments. For instance, prebiotically relevant fatty acid vesicles exhibit a low tolerance for even moderate concentrations of divalent ions, a condition that is paramount to the activity of nearly all nucleic acid catalysts.^{179,180} Additionally, encapsulation efficiency constitutes a significant limiting factor during the preparation of vesicle-based protocells, wherein a substantial quantity of material is required to encapsulate sufficient amounts of cargo.^{181–183} Furthermore, the incorporation of more complex cell-like behavior, such as cross-membrane transport or division, is considerably more challenging for vesicles compared to complex coacervates, which can exhibit behavior analogous to budding in the case of certain fueled complex coacervates.^{165,184,185}

The compounds assembling to complex coacervates encompass a broad spectrum of structural classes, ranging from small molecules to polymers with a molecular weight of hundreds of thousands of Dalton.¹⁶⁸ Coacervates can comprise molecules found in cellular environments or entirely synthetic polymers that lack any connection to biological systems. For instance, the synthetic polymer polydiallyldimethylammonium (PDDA) is an ideal candidate for a polycation due to the high robustness of the formed coacervates.^{186–189} In addition, poly(allylamines) can also be utilized as robust alternatives, though they have been less extensively studied.^{31,190,191}

Another prevalent class of polycations includes oligopeptides, with polylysine and polyarginine being a prominent example. These polycations are notable for their ease of production and the robust generation of complex coacervates they exhibit under most conditions, a property attributable to their positive charge.^{31,32,34,44,187,189,191–194} In addition to these homopolymers, another prevalent class of oligopeptides are heteropolymers with specific sequences. These heteropolymers often mix positively charged lysine and arginine with uncharged amino acids like glycine, which serves to space the charges out, or with aromatic amino acids like phenylalanine, which also benefits from π - π - or cation- π -interactions.^{165,194,195}

Introduction

Regarding the poly-anions, a notable class commonly employed to generate complex coacervates are polysaccharides. A plethora of different sugars, like carboxymethyl amylose, carboxymethyl dextran and dextran sulphate fall within this category.^{44,187,189,196,197} A significant advantage of the sugar backbone is the abundance of hydroxy groups available for modification and the high solubility in water even before the incorporation of charged modifications to the polymer.¹⁹¹

Nucleic acids constitute another common class of poly-anion for complex coacervates. Specifically, RNA has been identified as the common carrier anion in the generation of complex coacervates. The homopolymer polyuridylic acid (polyU) can be synthesized by the protein PNPase from the diphosphate monomers UDP, representing a rapid and efficient method for producing high quantities of polyU RNA.¹⁹⁸ This polyU can then be utilized to generate complex coacervates, exhibiting the advantage of approximating real-life examples of liquid-liquid phase separation, such as p-bodies or stress granules.¹⁹⁹

In contrast to the utilization of homopolymeric RNA, the employment of specific RNA sequences has also been demonstrated to facilitate the formation of complex coacervates. Thus, functional RNA can be utilized to create complex coacervates, with recent studies showing varied behaviors of the functional RNA, depending on the polycation.¹⁹² For instance, the incorporation of a ribozyme sequence can result in the creation of a protocell, in which the catalysis by this ribozyme is amplified through phase separation.^{32,34,195}

Lastly, there are smaller, non-polymeric molecules, such as ATP and other triphosphate nucleic acid monomer derivatives.¹⁹³ A notable benefit of these coacervates is their resemblance to living systems, as ATP is present in abundance in living cells, and if the protocell already composes ATP, a fuel for most biological reactions is readily available. Furthermore, the ability to produce ATP in situ enables the generation of active coacervates through the implementation of an esterification/hydrolysis reaction cycle.²⁰⁰

4.3.2 Active complex coacervates

Active molecular assemblies constitute a subgroup of molecular assemblies. In these active systems, the components necessary for supramolecular assembly are not invariably present in solution; rather, they are produced in situ.²⁰¹ Frequently, these assemblies are driven by a fuel cycle. In such a cycle, a precursor or monomer block required for the assembly is activated by the fuel, thereby enabling the assembly and consuming the fuel.^{202–205} If the activation of the

Introduction

building block is reversible without regeneration of the fuel, it can return to the precursor state again, consuming the fuel as a result. This process culminates in a transient steady state of assembly for as long as fuel is present, followed by a collapse and return to the dissolved state after the complete consumption of the fuel.

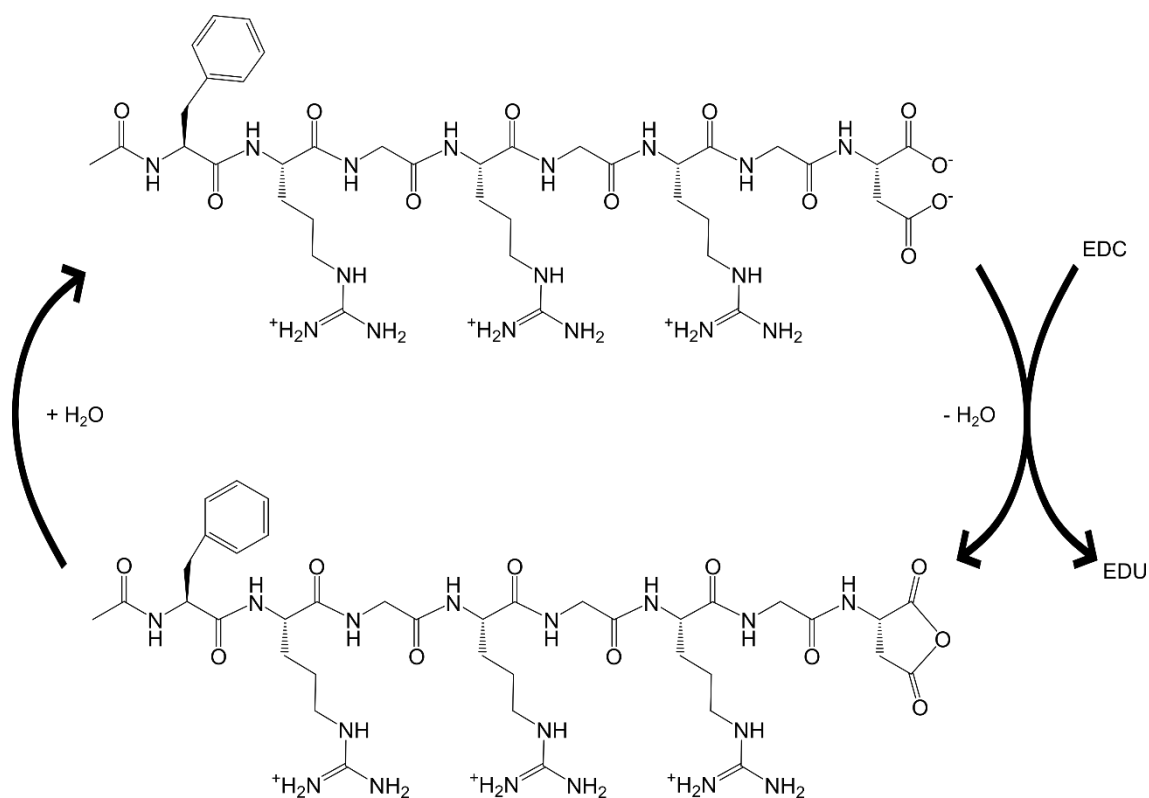


Figure 8: The fueling system utilized for the active complex coacervates in this study. A precursor peptide (top) gets activated by fuel (EDC) producing the waste product EDU and the activated peptide. The activation reaction is a condensation reaction, masking two negative charges of the carboxylic acid moieties at the C-terminus and the last residue. This anhydride can be hydrolyzed by H_2O , once again yielding the precursor peptide. The lifetime of the activated peptide is ~ 1 minute at pH 5.3 and 25 $^{\circ}C$.⁶⁵

In the system employed in the publications two, three and four of this thesis, a precursor peptide is supplied in conjunction with a poly-anion and a fuel (Figure 8).^{65,206,207} This precursor peptide, with the sequence Ac-FRGRGRGD, contains the essential characteristics of a positively charged polymer necessary for complex coacervation. The structure contains three positive charges in the arginine sidechains, a glycine spacer between the charges, and phenylalanine for the purpose of facilitating π - π -interactions. In contrast to the aforementioned design principles, the C-terminus of the peptide contains an aspartate residue with a negatively charged carboxylic group and a secondary negative charge on the C-terminus itself. This results in a net charge of +1 for this peptide. It has been determined that this is insufficient to form complex coacervates with the peptide and the supplied poly-anion polyU.^{165,208} A carbodiimide

Introduction

fuel, EDC, has been shown to catalyze the condensation of the two carboxylic acids at the C-terminus, resulting in the activation of the peptide, increasing the total charge to +3.^{165,209,210} Water can hydrolyze the reaction, resulting in the formation of the precursor peptide once more. The formation of complex coacervates is contingent upon the presence of an activated peptide, whereas their dissolution is initiated once all peptide has been deactivated within the solution. This fuel cycle, when initiated by 25 mM EDC in presence of sufficient precursor to form complex coacervates, takes about 15 minutes.¹⁶⁵

This system exhibits behavior that is analogous to that of living cells. The presence of fuel is a prerequisite for the growth and survival of the protocells. In the absence of fuel, the activated peptide undergoes hydrolysis. Following the complete removal of the activated peptide from the solution, the protocells undergo a terminal phase, akin to how life cannot persist without a continuous supply of energy. In addition to its capacity to emulate life-like behavior, the timescales of droplet longevity and the activity of the 17E DNAzyme align. Due to this, the modulation of the enzyme's activity can be achieved through the creation of droplets, the subsequent partitioning of the DNAzyme into the polymer-rich phase, followed by fuel consumption, resulting in droplet dissolution and sequestering of DNAzyme in solution, thereby restoring its catalytic activity.⁶⁵

5 Goals

During my doctoral studies, I explored functional roles of DNA beyond its canonical functions in genetic information storage and replication. In particular, I focused on how molecular context shapes the behavior of functional single-stranded DNA in intracellular and protocellular systems. My work showed that these DNA species are strongly influenced by their surrounding environment and can, in turn, modulate that environment.

Aim 1: Establish a retron-based intracellular expression platform for functional ssDNA.

The first part of this thesis focusses on the intracellular expression of the Lettuce light-up aptamer highlighting the importance of the environment on ssDNA folding and activity. Intracellularly selected RNA-based light-up aptamers are adapted to their environment and exhibit strong fluorescence in their host organism, showing that the environment in which functional ssDNA aptamers have been selected, is fundamental for their activity.^{115,119} My goals were to establish retrons as a versatile ssDNA expression platform and lay the foundation for in vivo selection of functional ssDNA instead of RNA.

Aim 2: Modulate the activity of a DNzyme using transient complex coacervates.

Another case highlighting the importance of the direct environment for functional nucleic acids is presented in the second publication, in which we aimed at modulating the activity of the 17E DNzyme by formation and dissolution of transient complex coacervates. In this scenario, the addition of carbodiimide fuel drastically changes the physicochemical properties of the DNzyme's reaction environment turning a dilute, aqueous phase into a very dense, gel-like polymer-rich phase of complex coacervates. The goal of this work was to modulate the activity of the DNzyme using this system, either strongly enhancing, or completely inhibiting its activity.

Aim 3: Investigate if ssDNA can influence the phenotype of a transient complex coacervate system.

The goal of this work was to explore whether ssDNA can also influence its own environment, producing morphological changes in metabolically active complex coacervates. This would allow a bidirectional relationship between environment and ssDNA: The environment regulating the behavior of the nucleic acid, while the nucleic acids shape the environment around them. Such a feedback mechanism would allow for Directed Evolution to take place.

Summary of the individual publications

Aim 4: Establish an orientation-specific, sequence-agnostic library preparation, sequencing and data analysis protocol in the context of artificial cells.

To analyze ssDNA samples isolated from transient droplet environment, we aimed to develop a custom sequencing workflow based on library preparation methods originally designed for ancient, damaged, genomic DNA.²¹¹ Unlike dsDNA or RNA, short, ssDNA with no fixed terminal sequences to attach adapters to is a very difficult target to sequence and no suitable commercial kits or established protocols are available for this application. The overarching goal was to use this workflow to identify sequences that preferentially partition into transient complex coacervate droplets and subsequently characterize how these sequences affect droplet morphology and lifetime.

6 Summary of the individual publications

6.1 Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2

Vibhute, M.A.*; **Machatzke, C.***; Bigler, K.; Krümpel, S.; Summerer, D.; Mutschler, H.⁺: Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2. *eLife* **2026**, 13:RP99554

The potential of ssDNA as a tool in molecular biology is impressive due to its low production cost and ease to handle compared to RNA or proteins. A significant challenge are the intracellular delivery and stability of ssDNAs. One potential solution to this problem would be the expression of ssDNA directly in the cells, a process which can be achieved by retron systems.

To this end, the retron Eco2 was utilized in **Publication 1** to express a DNA light-up aptamer in *E. coli* intracellularly. First, in vitro experiments were conducted to optimize the incorporation site and specific cargo design used for this project. In silico folding prediction of the RT-DNA was performed using ViennaRNA. Four different sites in the RT-DNA that were predicted to remain unpaired (V1-V4) were selected as incorporation sites. Oligonucleotides containing different length variants (a P1 stem length of 4, 8, 11 or 20 nucleotides, as well as the even smaller variant Bibb Lettuce) of the Lettuce light-up aptamer were incorporated in each of these positions.

Summary of the individual publications

The fluorescence of these oligonucleotides was tested using in gel-staining with DFHBI-1T, as well as bulk measurements to calculate K_D values for these constructs and their ligand. All constructs with at least an 8 nt P1 stem exhibited K_D values that were not significantly different from the positive control (full-length Lettuce aptamer). In contrast, for the variant with a 4 nt long P1 stem, no fluorescence was detected when Lettuce was incorporated into positions V1, V2 or V3. For the V4 variant with a 4 nt P1 stem, the K_D value was 10-fold higher compared to the other functional V4 DNA constructs with longer P1 stems. In gel staining confirmed, that the 8 nt long P1 stem at incorporation site V4 is the shortest, fully fluorescent construct.

The intracellular expression of retron Eco2 with cargo was confirmed by isolating the ssDNA from the cells after expression and analyzing it on urea-PAGE. This demonstrated that the expression of all length variants is, in principle, feasible, though expression levels decreased ~75% when any cargo was introduced into the RT-DNA. Consequently, the expression levels were inadequate for direct in-cell fluorescence detection due to a strong autofluorescence background. Instead, the expression of functional Lettuce aptamer was verified by isolating the RT-DNA, concentrating the sample, and analyzing the recovered material on a gel. Using this approach, the successful intracellular expression of functional Lettuce aptamer could be demonstrated.

To the best of our knowledge, this is the first instance of intracellular expression of a functional ssDNA aptamer using an endogenous gene expression platform. Whilst the current expression levels using the parental Eco2 retron are inadequate for the intracellular detection of fluorogenic cargo, this platform has the potential to provide a solution for the direct expression of functional, non-coding ssDNA in cells such as other aptamers or catalytic DNazymes.

6.2 Regulating Nucleic Acid Catalysis Using Active Droplets

Holtmannspötter, A.L.*; **Machatzke, C.***; Begemann, C.; Salibi, E.; Donau, C.; Späth, F.; Boekhoven, J.[†]; Mutschler, H.[†]: Regulating nucleic acid catalysis using active droplets. *Angewandte Chemie International Edition* **2024**, 63(47), e202412534

The regulation of chemical reactions is of fundamental importance to all forms of life and is frequently achieved by means of compartmentalization of components or catalysts. In order to achieve this in a protocell system, **Publication 2**'s goal was to enable regulation of 17E DNzyme activity by the formation and decay of fueled complex coacervates.⁶⁵

Summary of the individual publications

To this end, a variant of the 17E DNzyme that encompassed two A₁₀ partitioning arms was designed. This modification was expected to enable hybridization with the polyU RNA used as the major polyanion in coacervate formation. Indeed, the modified termini enhanced the partitioning of the DNzyme by more than 300-fold. This ensured that over 99% of the catalyst was retained within the complex coacervates. There was no discernible difference in the effectiveness of substrate cleavage between the A₁₀ tagged and untagged DNzyme variants, indicating that these sequence alterations did not impair the activity of the DNzyme.

Upon addition of fuel, a pronounced difference in activity was observed between the two versions of the 17E DNzyme. While the untagged DNzyme maintained a comparable level of activity to the unfueled conditions, the tagged DNzyme became nearly completely inactive. Subsequently, the underlying cause of this phenomenon was investigated. Utilizing differential scanning calorimetry (DSC) and circular dichroism (CD), it was shown that the DNzyme underwent complete unfolding in the presence of the activated peptide. Given that activated peptide predominantly resides within the complex coacervates, it can be concluded that at a minimum 99% of the A₁₀ tagged DNzyme became unfolded and therefore inactive in the presence of droplets.

This behavior was then used to establish temporal controlled activation-inhibition cycles of the DNzyme using the half-life of the carbodiimide fuel as pacemaker. By the addition of batches of fuel, the DNzyme activity was suspended for approximately 15 minutes before it resumed substrate cleavage at its baseline rate. These cycles could be repeated multiple times within the same solution. In addition, extended but reversible pausing of DNzyme activity could be achieved by continuous fuel addition to the sample.

In conclusion, the dynamic active behavior of the complex coacervates / DNzymes observed in this study mimicked, to some degree, the reversible control of enzymatic activity via transient sequestration into membrane-free compartments. As such, it demonstrates the possibility of a bottom-up route towards the formation of synthetic cells that are based on a completely synthetic, non-biological chemistry.

6.3 DNA affects the phenotype of fuel-dependent coacervate droplets

Machatzke, C.*; Holtmannspötter, A.L.*; Mutschler, H.; Boekhoven, J.⁺: DNA affects the phenotype of fuel-dependent coacervate droplets. *Nature Communications* **2026**, 17, 2953.

Publication 3 explores the impact of nucleic acids on the complex coacervate systems into which they partition.²⁰⁶ To this end, the effects of 30mer ssDNAs with different sequence designs on the droplet properties was tested.

First, sequence-related features associated with enhanced partitioning into complex coacervates were investigated. To this end a N₃₀ ssDNA library was incubated with transient droplets after which the polymer-rich and polymer-poor phases were separated, and the contained DNA sequenced and analyzed. The results were remarkable: although only a modest increase in G content (13%) was observed in sequences recovered from the droplet phase compared with the supernatant phase (26.5% vs. 29%), this enrichment was not evenly distributed across the sequence pool. Instead, the strongest increase in G-content was observed at the first and last positions of the sequences as well as at neighboring positions where G-content increased to 32%, whereas it remained at about 25% in the aqueous polymer-poor phase. This effect diminished toward the center of the sequences, where the relative increase in G prevalence was only 6% (26.5% vs. 25%). These findings suggest that the presence of G-rich termini may promote partitioning into the polymer-rich phase of the droplets.

The sequences obtained by deep-sequencing were furthermore analyzed for motif enrichment. The results corroborated the aforementioned findings: an enrichment of repetitive G and A stretches was observed in sequences recovered from the polymer-rich phase of the droplets. These repeats exhibit a tendency to localize at the termini of sequences, with a length of five to eight nucleotides. The samples isolated from the droplet phase contained on average twice as many motifs with five to seven consecutive A or G runs compared to DNA isolated from the polymer-poor phase. Furthermore, the average Gibbs energy of folding was predicted to be lower for sequences from the droplet sample compared to the supernatant sample, suggesting that unstructured sequences preferentially reside inside droplets.

Next, the effect of tailormade A-rich and G-rich sequences on the droplet phenotype was investigated, with specific focus on droplet lifetime and diffusivity of the polyanionic species inside the complex coacervates using FRAP experiments. The findings were substantial: While unbiased sequences (N₃₀) or A-rich libraries had no significant effect on the droplet life cycle

Summary of the individual publications

or its material properties, the addition of an A₃₀ homopolymer nearly halved the lifetime of fueled coacervates while also markedly reducing the propensity of droplets to coalesce by rendering them more rigid through their hybridization of the poly(U) polyanionic component of the coacervates. A further detailed characterization of the system revealed that only seven terminal As are sufficient to achieve a substantial reduction in the droplet's lifespan. This observation underscores the pivotal role of the terminal nucleotides in this process.

In contrast, G-rich motifs were shown to enhance the lifetime of droplets. An G₈N₁₄G₈ oligonucleotide kinetically trapped the droplet lifecycle, thereby impeding complete dissolution after EDC hydrolysis, even after extended periods of time (>1 hour). This finding indicated that the polyG termini triggered the irreversible formation of quasi-static droplets that were incapable of complete dissolution, thereby explaining the high content of these sequences in the sequencing sample.

In conclusion, the findings of this publication offer generalizable design principles for achieving high partitioning of nucleic acid cargo into static and dissipative droplets formed from poly(U) and polycationic peptides. The main requirements are the presence of G- or A-repeats at the termini, a design, which is likely compatible with a broad range of functional nucleic acids, given the tolerance by most of them towards terminal modifications: The selection of the A₁₀ partitioning arms modification for **Publication 2** was made on the basis of rational design, and no substantial discrepancy in activity was observed between the tagged and untagged DNAzyme, confirming this approach.

Combining the findings from **Publications 2** and **3**, it could be shown that ssDNA cargo can modulate the behavior of active complex coacervates and vice versa. This relationship establishes a linkage between a primitive “genotype” (DNA sequence) and a phenotype (droplet properties), thereby paving the way for potential evolution experiments within these systems.

6.4 Protocol for the recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets

Holtmannspötter, A.L.*; **Machatzke, C.***; Boekhoven, J.; Mutschler, H.⁺: Recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets. *STAR protocols* **2026**, 7(1), 104293

Summary of the individual publications

The majority of commercially available methods for sequencing are focused on either RNA or long and double-stranded DNA. This renders the handling of single-stranded DNA a complex undertaking, owing to the absence of necessary protocols, equipment, and kits. Due to the error rate of simple replicators, the initial genotypes employed in nucleic acid-based synthetic cells will most likely be short nucleic acids.²¹² Therefore, the goal of **Publication 4** was to establish a workflow that enables the recovery of short, ssDNA from active (but also static) coacervate droplets and to enable the preparation of orientation specific libraries for next-generation sequencing without introducing major sequence biases.²⁰⁷ In addition, the publication describes a bioinformatic workflow for sequence analysis and motif discovery.

As a first step, it was necessary to establish a protocol of robust separation of the polymer-rich and polymer-poor phases in the short time windows provided by the lifetime of EDC-fueled coacervate droplets. Next, the ssDNA had to be recovered from both phases separately. This objective was accomplished by implementing a time-sensitive spin down protocol, followed by the removal of the poly(U) RNA via RNases. After this, a column-based purification of the ssDNA from the residual droplet material was conducted.

Following the recovery of ssDNA, an orientation-specific, sequence-agnostic ligation of sequencing adaptors to the ssDNA was performed. The protocol was derived from a library preparation methodology, which employs a one-pot reaction that uses four different oligos to ligate a ssDNA of interest.²¹¹ The protocol was adapted to better suit short ssDNA input sequences, i.e. to improve adapter ligation yields as the desired applications of these two methods differ greatly.

After sequencing, computational downstream sequence processing and analysis based on custom scripts and the integration of available tools are described in detail. Typical sequence characteristics are, for example, nucleotide content and position, predicted minimal free energies of folding including the most abundant consensus structures as well as the enrichment of sequence motifs.

The method published in **Publication 4** has been employed to study the sequences obtained in **Publication 3**. The implementation of this approach is expected to promote methodological standardization across laboratories in the large field of nucleic acid-driven coacervation, thereby enabling more robust comparisons between datasets.

7 Discussion

7.1 ssDNA as a biotechnological tool

The first part of this thesis aimed to demonstrate the expression of a DNA-based aptamer in a living organism for the first time. A platform for robust expression of functional ssDNA inside cells may be of considerable interest for biotechnology, diagnostics and therapeutics, as it expands the range of biomolecules to monitor and modulate cellular activity.

The expression of RT-DNA with the Eco2-retron used in this work showed a considerable amount of tolerance towards sequence modifications, and several constructs that are significantly longer than the original wildtype *msd* could be expressed successfully. However, expression levels dropped ~75% when cargo was incorporated into the retron scaffold. This effect is rarely reported on, since retrons containing a ssDNA cargo are primarily used for genome editing, where only editing efficiency is quantified, not RT-DNA expression.^{213,214} In the few cases where RT-DNA levels have been quantified, a significant reduction in RT-DNA abundance across different retron systems and cargos was reported.^{153,157} Next to lower expression levels, folding of functional nucleic acids without the help of proteins is challenging since intracellular concentrations of free Mg^{2+} are low compared to optimal in vitro conditions (reported values range between 0.25 and 1 mM).^{215,216} While the total Mg^{2+} concentration is higher, most of it is fully or partially chelated by various biomolecules, with ATP being the most prominent one.²¹⁵ As such, with Mg^{2+} being a mandatory cofactor for Lettuce folding and fluorogenic ligand binding, it is uncertain if these conditions are sufficient to induce an intracellular fluorescence response upon Lettuce expression.¹¹³

Due to these limitations, a direct intracellular detection of Lettuce aptamer fluorescence remained elusive at this point. From a starting volume of 1 liter of *E. coli* culture, 2.6 pmol of Eco2-encoded 8 Lettuce DNA were successfully isolated through the implementation of an optimized phenol-based extraction protocol. While merely serving as a lower bound for intracellular concentration, since presumable a large fraction of DNA was lost during isolation, 2.6 pmol of DNA from 1 liter of culture at an optical density (OD) of 2 (assuming 8×10^8 cells/mL per OD) would correspond to a copy number of approximately 1, which would be too low for the detection of intracellularly expressed Lettuce aptamer by DHFBI-1T fluorescence.

Discussion

I propose that this low copy number is a direct result of the toxin domain encoded in the reverse transcriptase gene of Eco2. Toxin-Antitoxin systems typically reside in a latent, tightly regulated state, with the capability to rapidly activate the toxin activity once the antitoxin's neutralizing effect is reduced due to degradation or e.g. a conformational change.²¹⁷ In the case of retrons, every copy of the RT-DNA is covalently bound to the corresponding RNA, thereby functioning as the anti-toxin to the toxin.^{130,151,218} This means that with every expressed copy of RT-DNA, stoichiometric amounts of the corresponding toxin are also expressed. Thus, changing the structure of the antitoxin, for example, by encoding cargo in the RT-DNA, might reduce its efficacy of silencing and thereby increase toxicity. The survival benefit of poorly RT-DNA expressing cells might therefore be the reason for lower expression levels of retron Eco2 when any cargo is encoded in the RT-DNA, regardless of length or position, since strongly expressing cells might die early or grow poorly after induction of Eco2 expression.

In the case of Eco2 the toxin domain is fused to the RT enzyme, making it challenging to remove the toxin.²¹⁹ In preliminary work, I explored the potential of mutating the active site of the toxin to hinder its activity. The toxin in retron Eco2 is a TOPRIM domain, containing the canonical DxD motif for catalysis. I targeted the catalytically important residue E374A to inhibit the TOPRIM activity, as suggested by previous publications.^{143,219} A notable observation was a return to a no-cargo baseline cell growth regardless of cargo, while in the case of active toxin, some cargo designs significantly reduced the OD of cultures they were expressed in (Figure 9A). This finding indicates a strong negative effect of the presence of the wildtype toxin domain on growth, which can be mitigated by mutating the catalytic site of the toxin. The potential implications of this phenomenon, specifically whether it will result in elevated copy numbers of cargo within cells, are yet to be elucidated and will be the focus of subsequent research endeavors.

Once retron Eco2 or a related retron has been engineered to enable reliable, high-copy-number expression of modified RT-DNA, it can be used as an expression platform for intracellular evolution of DNA-aptamers. For example, I compared the fluorescence of Lettuce and the RNA light up aptamer Broccoli, both in the presence of the same amount of DFHBI-1T (Figure 9B). A significant disparity in fluorescence intensity was observed, with an observed fluorescence difference of approximately 75-fold between the two samples. If ssDNA-based light-up aptamers are to be used in intracellular applications, it is paramount to address this issue of low fluorescence compared to their RNA counterparts.

Discussion

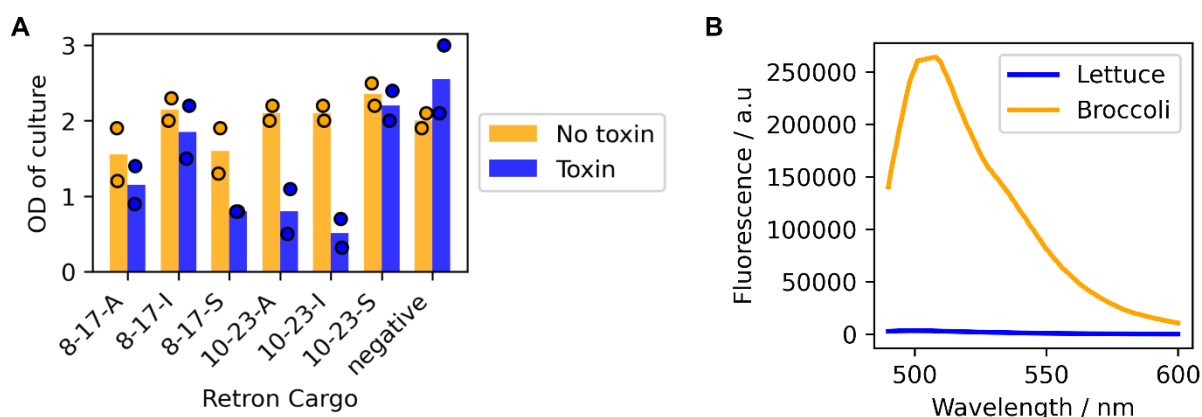


Figure 9: Preliminary data showing the effect of mutating the toxin active site on cell growth. A) OD measurements of cells expressing different ssDNA cargos in the retron cassette either in presence of the active (blue) or mutated (orange) toxin. Mutating the toxin restores growth rates equal to the negative control (no cargo). $n = 2$. B) Fluorescence emission spectra of the Lettuce aptamer and the Broccoli aptamer under the same conditions (40 mM HEPES pH 7.4, 100 mM KCl, 1 mM $MgCl_2$, 1 μ M DFHBI-1T, 10 μ M aptamer). Fluorescence of Broccoli is ~ 75 -fold higher than Lettuce.

To this end, a (semi-)continuous evolution of the Lettuce aptamer, akin to the selection of Broccoli, could be employed.¹¹⁵ For Broccoli, 4-6 rounds of SELEX were performed on a random RNA library to identify DFHBI binders. Alternatively, in the case of Lettuce, a library based on the known aptamer sequence could be used as a starting point. Following SELEX, the Broccoli library was reverse transcribed, cloned into a vector, and transformed into bacteria. In the case of Lettuce, a suitable expression system would be necessary. I presented one such system in this work, using the retron Eco2 scaffold. The cells carrying the best aptamer clones of this round could be isolated using fluorescence assisted cell sorting (FACS) and either be screened directly or used for further rounds of selection involving e.g. the introduction of additional variability into the collapsed library e.g. by using error-prone PCR or *E. coli* mutator strains.^{220,221}

While the selection of a superior ssDNA light-up aptamer would constitute a proof of principle, a logical subsequent step would be the selection of RNA-cleaving DNazymes. RNA-cleaving nucleic acid catalysts hold particular interest as potential therapeutic agents, as they can be easily programmed to target any mRNA, thus allowing the knockdown of many different targets.^{222–224} Methods established for the selection of ribozymes could be adopted for the purpose of optimizing DNazymes as well. For example, during the intracellular selection of a hammerhead ribozyme (HHz), the mRNA of a short toxic peptide was used as the ribozyme target. Using this strategy it was ensured that only cells carrying catalytically active ribozyme variants would propagate.²²⁵ To date, the expression of a DNzyme in a retron scaffold has

Discussion

only been demonstrated once.¹⁵⁷ Here, the target was the mRNA of cyan fluorescing protein (CFP), resulting in a direct readout (reduction of CFP fluorescence) when the mRNA was cleaved. Analogous to the HHz selection, this system could be used for DNAzyme evolution by combining a fluorescence readout with cell sorting, thus linking the DNAzyme activity to a high-throughput selection protocol required for directed evolution.

Achieving this would enable targeting therapeutically relevant gene products such as currently “undruggable proteins” that lack target sites for small molecule therapeutics.^{226–230} A prominent member of this class of challenging drug targets are, for example, the oncogenic mutations of the Ras protein family.²³¹ K-Ras, a major regulator of cell-growth and division can be permanently activated by single mutations located in the P-loop of the GTP binding site such as, for example G12.²³² While such Kras mutations are notoriously difficult to target with small molecules inhibitors, allele-specific therapeutics that target hyperactive variants are in great demand, and one strategy could be the down-regulation of their expression. Consequently, RNA-cleaving nucleic acid enzymes that target specific Kras variants are a major focus of research. For example, a recent proof-of-concept approach based on a transfected RNA-cleaving XNAzyme was shown to specifically target mutants carrying the mutated K-Ras G12V oncogene in mammalian cells, resulting in oncogenic cells to return to a healthy phenotype.^{233,234}

The choice to utilize XNAzymes instead of DNAzymes in the aforementioned study was primarily determined by considerations of stability. ssDNA is recognized as stress factor (i.e. DNA damage or viral infection) in cells and therefore degraded swiftly once identified.²³⁵ This means that transfections with DNAzymes typically result in very short half-lives of the nucleic acid therapeutic, and high dosages would be required to achieve a therapeutic effect.²³⁶ Retrons would reduce this problem two-fold: First, due to potential continuous production of the ssDNA therapeutic in the target cell, the effective dosage can be a lot lower. Secondly, the stability of retons in vivo is exceptionally high due to the protection of the termini by the associated retron binding protein. In Publication 1, even after four hours, we could not detect any degradation of the retron DNA in *E. coli* cells. As such, this exceptional stability of retons might confer an advantage that both RNA and XNA lack.

Discussion

7.2 Towards catalysis, genotypes, and evolution in synthetic protocells

In Publications 2-4 I made valuable contributions to the development of synthetic protocellular model systems that approximate several hallmarks of life in a simplified manner. Specifically, a chemically fueled (“metabolically active”) complex coacervate system was employed to simulate the feeding and starvation cycles of protocells. By using DNA as a cargo molecule, I aimed to introduce both information carriers but also biocatalysts in this fascinating model system. Publication 2 focused on modulating the activity of a DNAzyme using the fueled droplet system, while Publications 3 and 4 focused on the sequence features that govern DNA partitioning into the complex coacervate system and the impact of specific DNA sequences on the morphology and lifetime of the droplets.

7.2.1 The influence of the active droplet environment on their DNA-cargo

In my work with the 17E DNAzyme in the transient droplet system, I successfully proved that this catalyst can be temporally and, to some degree spatially, controlled by the formation and decay of transient complex coacervates. This is in contrast to the overwhelming majority of previous work on nucleic acid catalysis in the presence of complex coacervates, which focusses on enhancing their catalytic activity in static droplets.^{171,197,237–239}

Moving forward, the optimal subsequent steps would be the introduction of a catalytic activity other than RNA cleavage. RNA cleavage, by design, is a destructive process, which inherently does not pair well with the general goal of designing an artificial cell capable of Darwinian evolution. Instead, ligation or polymerization of nucleic acids are the most compelling candidates.

Both polymerization and ligation are constructive processes, meaning they yield longer, new or more complex products than their substrates. Due to longer polyanions generally generating more stable droplets in this transient system, producing longer nucleic acid fragments is beneficial for droplet stability and therefore the desired phenotype for Darwinian evolution if lifetime under fuel-starvation is chosen as a measure for “droplet fitness”.²⁰⁸ Thus, a system could be envisioned, in which the nucleic acid catalysts primary function is ligation of other nucleic acids, thereby enhancing the stability of the droplet in which it is located. Such a process would prolong the lifespan of the droplet, ultimately benefiting the nucleic acid cargo itself, if for example, the coacervate environment would shield it from a hypothetical nucleolytic

Discussion

process. In such a scenario, the mutual benefits would drive both, the selection for more active DNazymes as well as of droplets with greater longevity under starvation conditions.

A DNzyme candidate for this kind of mutualistic system could be the 9DB1 DNzyme, which catalyzes the ligation between 5'-triphosphate-RNA and a 3'-OH RNA substrates or the 5'-AppRNA ligase DNzyme, which ligates RNA molecules using 5'-adenylated RNA as activated substrates.^{240,241} Neither of these DNzymes has been tested as cargo in a complex coacervate system so far and are therefore attractive targets for future studies. Contrary, a number of RNA-ligating ribozymes have been previously tested in the presence of static complex coacervate systems.^{32,195,242,243} From these, a particularly interesting candidate is the R3C ribozyme.²⁴⁴ A variant of it, named R3C ladder ribozyme, catalyzes the formation of long RNA concatemers from simple RNA repeats.^{33,34} This specific variant of the R3C ribozyme was tested in the presence of poly-Lysine, forming complex coacervates with the peptide and the RNA. A system similar to this, but employing fueled complex coacervates, could potentially enable Darwinian evolution experiments. This is especially true since changes in droplet phenotype were already observed in the presence of the R3C ladder ribozyme.³⁴

Currently, the limiting factor of using ribozymes in the fueled complex coacervates presented in publications 2-4 is the highly reactive EDC, which leads to undesired side reactions between nucleic acids and peptides.^{245,246} These side reactions are likely to impair any protein- or RNA biocatalyst introduced to the system. In order to mitigate the impact of this phenomenon, it will therefore be necessary to explore alternative peptide activation chemistries using more specific or less reactive fuel molecules.

In the field of supramolecular assemblies, a variety of different fueling strategies have been suggested. Most of them are equally incompatible with nucleic acids as the EDC chemistry presented in these publications, like other condensing agents or acid/base catalysts.^{247,248} However, some forms of activation chemistry borrow or are inspired by natural processes, and thus are compatible with biomolecules. For example, fueling cycles using ATP have been used to assemble polymers, vesicles and complex coacervates before, in one case even to regulate cytochrome c activity.²⁴⁹⁻²⁵¹ Similarly, active droplets were formed using transient non-enzymatic phosphorylation reactions using monoamidophosphate.²⁵²

Alternatively, enzymatic activity can be used to modulate the charges on a peptide to form complex coacervates with RNA, like a kinase/phosphatase combination.²⁵³ Either of these

Discussion

approaches would be better suited to regulate the assembly of complex coacervates, if the goal of the assembly is DNA catalysis. Most interestingly, the phosphorylation/dephosphorylation reaction could be catalyzed by ribozymes, allowing the biomimetic integration of enzymatically driven lifecycles into the complex coacervate cargo itself.^{254,255}

7.2.2 Cargo influence on active droplets

After identifying how active droplets can regulate the activity of their cargo, I investigated how different DNA cargos can influence the droplet phenotype. The goal was to identify nucleic acid sequence features and motifs that have a specific effect on droplet lifetime and morphology. In order to investigate the impact of ssDNA on the fueled complex coacervate droplets, a ssDNA recovery and sequencing protocol was developed first. This protocol is described in detail in Publication 4. I used this protocol to analyze the content of the complex coacervate droplets that were generated for Publication 3.

Although the content of prebiotically plausible cells is of great interest to the origin of life community, little research exists on quantifying the nucleotide content of artificial cells. In 2022, a single investigation of RNA content in individual complex coacervates was performed using a modified RNA sequencing protocol.¹⁸⁹ This previous work could rely on the extensive research in the field of single-cell RNA sequencing for sample purification, library preparation and downstream analysis.^{256,257}

Biologically relevant sequencing samples are usually either RNA or dsDNA, due to their cellular prevalence. ssDNA, on the other hand, is much rarer in a cellular context, resulting in an underrepresentation of methods for deep sequencing of short ssDNAs. One subfield that works with samples containing short ssDNAs is the field of ancient / damaged DNA sequencing that often has to analyze samples of mixed double stranded and single stranded identities.^{258,259} The methods employed for these samples denature the genomic DNA into long ssDNAs before sequencing adaptors bound to magnetic beads are ligated to them with the help of T4 DNA ligase.^{258,260} The method I developed and presented in Publication 4 is based on a similar method, in which the ligation procedure is a one-pot, single-step reaction that does not require the use of magnetic beads.²¹¹ To the best of my knowledge, the method developed in Publication 4 is the first sequencing method solely targeted at short, ssDNA sequences. As demonstrated in Publications 3 and 4, these sequences can be analyzed with a rich portfolio of bioinformatic

Discussion

tools and may help to standardize analysis workflows methods in the field of synthetic cell and origin of life research.

Based on our sequencing results of the droplet and supernatant DNA content, I formulated the following hypothesis: The free, unfolded termini of a ssDNA molecule interact with the droplet material either via Watson-Crick base pairing (poly(A) with poly(U)) or via hydrogen bonds and salt bridges (poly(G) with Arg).²⁶¹ Considering the prevalence of poly(G) motifs, I hypothesize that the latter interaction is more significant and results in more stable droplets than the former.

A notable limitation of the current approach was the inability to amplify recovered DNA for subsequent rounds of selection. Due to the experimental design, it was impossible to enrich sequences by polymerase chain reactions (as typically employed in directed evolution experiments) while maintaining the complete randomness of the sequences. A process involving amplification and reselection may result in a sequence or a pool of sequences that offer a more specific benefit to the droplets, as opposed to the more general modes of action that have been identified.

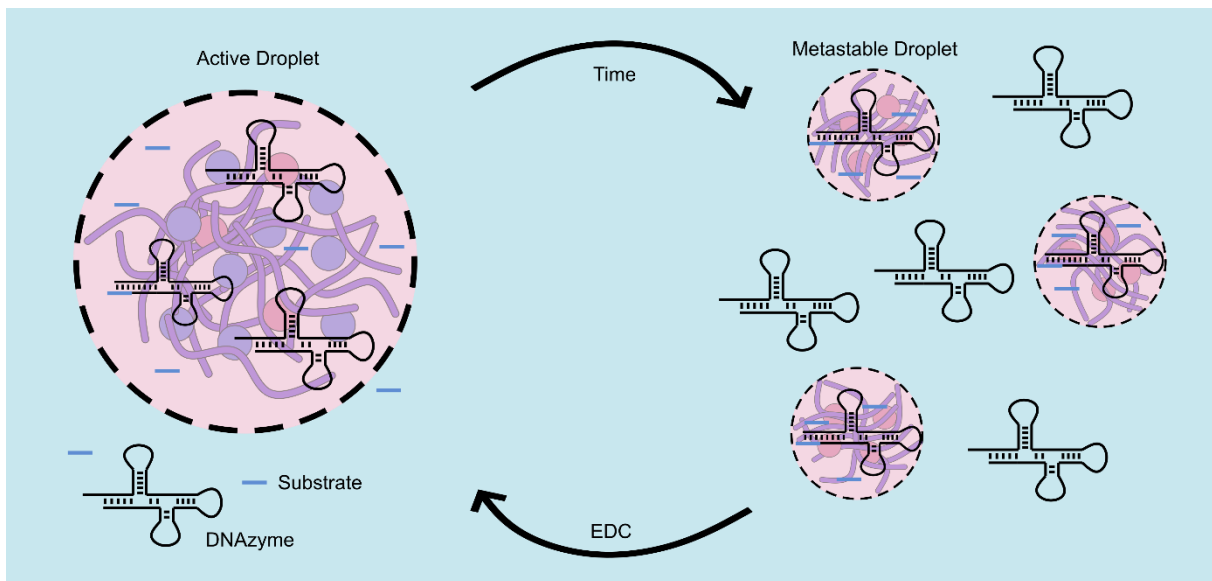


Figure 10: Design for DNA ligating DNAzyme evolution using active complex coacervates. A library of potential DNAzymes and a G-rich Substrate is combined with the droplet system. Some of the DNAzymes will be capable of ligating the Substrate to themselves, adding a G-rich tail. A fueling cycle will result in active droplets, followed by dissolution of droplets. Metastable droplets, containing the successfully ligated DNAzyme, will remain due to the interaction of the G-rich substrate and the droplet material. These metastable droplets can operate as seeds for a new fueling round, continuously enriching active DNAzymes. Harvesting of the metastable droplets will yield the active species for amplification, identification and characterization.

Discussion

To this end, I have conducted some preliminary experiments and designed a library with fixed primers. Given the inevitable terminal nature of the primers, this approach eliminates the biases that were identified in the aforementioned research, revealing that the termini play a pivotal role in inducing changes to the droplet phenotype. MeRPy-PCR, a special PCR variant was also employed, in which only the desired strand is pulled down after the PCR by binding one primer to a solid-phase anchor.²⁶² The results of this work are only preliminary but will be explored further to potentially find further, more specific modes of action to influence the droplet phenotype.

In the long term, this platform could be used to evolve DNazymes. For example, a DNA-ligating DNzyme capable of self-ligation to a substrate with G-rich termini could take advantage of the fact that metastable droplets are preferentially formed by $G_8N_{14}G_8$ sequences to create seeds for droplets in a new fuel cycle (Figure 10). More efficient ligase DNazymes would create these seeds more quickly and enrich themselves in the surviving metastable droplets over multiple fueling rounds. PCR-based amplification, using solid-phase anchored primers as described above, could then amplify the specific DNazymes based on the ligated, G-rich termini. This process would continuously select for the optimal DNzyme DNA ligase.

7.3 Conclusion

In this thesis, I explored the capabilities of functional deoxynucleic acids in two very different contexts – living bacterial organisms and purely synthetic compartments. However, when viewed in a broader context, both systems aim to offer platforms for the future evolution of functional deoxynucleic acids, connecting these projects in their core concepts.

In the first part of my thesis, I characterized a retron-based platform for the intracellular expression of ssDNA. In the future, this platform could be used for intracellular DNA-apatamer selection such as, for example, improved GFP-mimicking DNA light-up aptamers that are more compatible with cytosolic environments than current variants. While I could demonstrate intracellular expression of light up aptamers as well as potential routes towards further improvements of the platform, the conduction of selection experiments was not possible within the time frame of this thesis.

In the second part of this thesis, I explored the feasibility of using highly simplified chemical systems consisting of synthetic ssDNA, peptides, and condensing agents to develop model

References

protocells with a distinct genotype-phenotype linkage thereby providing a conceptual foundation for future directed evolution experiments. After establishing the model system presented above, and providing valuable insight into designs and constraints that govern evolvability in fuel-driven complex coacervate protocells, due to time constraints, Darwinian evolution could not yet be realized. Currently, the main limitation of the system is the condensing fuel used to drive complex coacervation being partly incompatible with nucleic acids. This reduces the potential design space to build an evolving system by, for example, excluding ribozymes. Addressing this challenge will likely require a revised activation strategy using alternative chemistries, like an ATP driven fuel cycle, to allow for robust directed evolution in synthetic protocells.

Nonetheless, this work is a big leap towards evolution in synthetic protocells, and while I might not have etched the entire path from lifeless rock to living organism, I did start chiseling towards this common goal.

8 References

- (1) Tolkien, J. *The Silmarillion*; HarperCollins, 1999.
- (2) Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, 117 (3046), 528–529. DOI: 10.1126/science.117.3046.528.
- (3) Fry, I. The origins of research into the origins of life. *Endeavour* **2006**, 30 (1), 24–28. DOI: 10.1016/j.endeavour.2005.12.002.
- (4) Miller, S. L.; Schopf, J. W.; Lazcano, A. Oparin's ``Origin of Life": Sixty Years Later. *Journal of molecular evolution* **1997**, 44 (4), 351–353. DOI: 10.1007/PL00006153.
- (5) Tirard, S. J. B. S. Haldane and the origin of life. *Journal of genetics* **2017**, 96 (5), 735–739. DOI: 10.1007/s12041-017-0831-6.
- (6) Fox, G. E.; Stackebrandt, E.; *et al.* The phylogeny of prokaryotes. *Science* **1980**, 209 (4455), 457–463. DOI: 10.1126/science.6771870.
- (7) Woese, C. The universal ancestor. *Proceedings of the National Academy of Sciences of the United States of America* **1998**, 95 (12), 6854–6859. DOI: 10.1073/pnas.95.12.6854.
- (8) Koonin, E. V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nature reviews. Microbiology* **2003**, 1 (2), 127–136. DOI: 10.1038/nrmicro751.
- (9) Ouzounis, C. A.; Kunin, V.; Darzentas, N.; Goldovsky, L. A minimal estimate for the gene content of the last universal common ancestor--exobiology from a terrestrial perspective. *Research in microbiology* **2006**, 157 (1), 57–68. DOI: 10.1016/j.resmic.2005.06.015.

References

- (10) Luisi, P. L.; Ferri, F.; Stano, P. Approaches to semi-synthetic minimal cells: a review. *Die Naturwissenschaften* **2006**, 93 (1), 1–13. DOI: 10.1007/s00114-005-0056-z.
- (11) HJ Morowitz. The completeness of molecular biology. *Israel journal of medical sciences* **1984**.
- (12) Hutchison, C. A.; Chuang, R.-Y.; *et al.* Design and synthesis of a minimal bacterial genome. *Science (New York, N.Y.)* **2016**, 351 (6280), aad6253. DOI: 10.1126/science.aad6253.
- (13) Gibson, D. G.; Glass, J. I.; *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (New York, N.Y.)* **2010**, 329 (5987), 52–56. DOI: 10.1126/science.1190719.
- (14) Moger-Reischer, R. Z.; Glass, J. I.; *et al.* Evolution of a minimal cell. *Nature* **2023**, 620 (7972), 122–127. DOI: 10.1038/s41586-023-06288-x.
- (15) Glass, J. I. *Synthetic bacterium JCVI-Syn3.0 strain 6d, complete genome*. <http://www.ncbi.nlm.nih.gov/nuccore/CP016816.2>.
- (16) Safronova, N.; Junghans, L.; Saenz, J. P. Temperature change elicits lipidome adaptation in the simple organisms *Mycoplasma mycoides* and JCVI-syn3B. *Cell Reports* **2024**, 43 (7), 114435. DOI: 10.1016/j.celrep.2024.114435.
- (17) Justice, I.; Kiesel, P.; *et al.* A tuneable minimal cell membrane reveals that two lipid species suffice for life. *Nat Commun* **2024**, 15 (1), 9679. DOI: 10.1038/s41467-024-53975-y.
- (18) Seidel, Z. P.; Kumar, R.; *et al.* Meeting Proceedings from 4th Minimal Cell Workshop: Exploring JCVI Minimal Cell Fundamental Insights and Integrative Applications. *ACS synthetic biology* **2025**, 14 (6), 1905–1911. DOI: 10.1021/acssynbio.5c00159.
- (19) Gil, R.; Sabater-Muñoz, B.; *et al.* Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99 (7), 4454–4458. DOI: 10.1073/pnas.062067299.
- (20) Moran, N. A.; Mira, A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome biology* **2001**, 2 (12), RESEARCH0054. DOI: 10.1186/gb-2001-2-12-research0054.
- (21) Crick, F. H. The origin of the genetic code. *Journal of molecular biology* **1968**, 38 (3), 367–379. DOI: 10.1016/0022-2836(68)90392-6.
- (22) Orgel, L. E. Evolution of the genetic apparatus. *Journal of molecular biology* **1968**, 38 (3), 381–393. DOI: 10.1016/0022-2836(68)90393-8.
- (23) Woese, C. R.; Dugre, D. H.; Saxinger, W. C.; Dugre, S. A. The molecular basis for the genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **1966**, 55 (4), 966–974. DOI: 10.1073/pnas.55.4.966.
- (24) Woese, C. R. On the evolution of cells. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, 99 (13), 8742–8747. DOI: 10.1073/pnas.132266999.
- (25) Kruger, K.; Grabowski, P. J.; *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **1982**, 31 (1), 147–157. DOI: 10.1016/0092-8674(82)90414-7.

References

- (26) Guerrier-Takada, C.; Gardiner, K.; *et al.* The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **1983**, 35 (3 Pt 2), 849–857. DOI: 10.1016/0092-8674(83)90117-4.
- (27) Gilbert, W. Origin of life: The RNA world. *Nature* **1986**, 319 (6055), 618. DOI: 10.1038/319618a0.
- (28) Robertson, M. P.; Joyce, G. F. The origins of the RNA world. *Cold Spring Harb Perspect Biol* **2012**, 4 (5), a003608. DOI: 10.1101/cshperspect.a003608.
- (29) Steitz, T. A.; Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in Biochemical Sciences* **2003**, 28 (8), 411–418. DOI: 10.1016/S0968-0004(03)00169-5.
- (30) Joyce, G. F. *The RNA World: Life Before DNA and Protein*. <https://ntrs.nasa.gov/citations/19980211165>.
- (31) Poudyal, R. R.; Guth-Metzler, R. M.; *et al.* Template-directed RNA polymerization and enhanced ribozyme catalysis inside membraneless compartments formed by coacervates. *Nat Commun* **2019**, 10 (1), 490. DOI: 10.1038/s41467-019-08353-4.
- (32) Le Vay, K.; Song, E. Y.; *et al.* Enhanced Ribozyme-Catalyzed Recombination and Oligonucleotide Assembly in Peptide-RNA Condensates. *Angewandte Chemie International Edition* **2021**, 60 (50), 26096–26104. DOI: 10.1002/anie.202109267.
- (33) Iglesias-Artola, J. M.; Drobot, B.; *et al.* Charge-density reduction promotes ribozyme activity in RNA-peptide coacervates via RNA fluidization and magnesium partitioning. *Nat. Chem.* **2022**, 14 (4), 407–416. DOI: 10.1038/s41557-022-00890-8.
- (34) Le Vay, K. K.; Salibi, E.; *et al.* Ribozyme activity modulates the physical properties of RNA-peptide coacervates. *eLife* **2023**, 12. DOI: 10.7554/eLife.83543.
- (35) Ianeselli, A.; Salditt, A.; *et al.* Physical non-equilibria for prebiotic nucleic acid chemistry. *Nat Rev Phys* **2023**, 5 (3), 185–195. DOI: 10.1038/s42254-022-00550-3.
- (36) Morasch, M.; Liu, J.; *et al.* Heated gas bubbles enrich, crystallize, dry, phosphorylate and encapsulate prebiotic molecules. *Nature Chem* **2019**, 11 (9), 779–788. DOI: 10.1038/s41557-019-0299-5.
- (37) Kreysing, M.; Keil, L.; Lanzmich, S.; Braun, D. Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nature Chem* **2015**, 7 (3), 203–208. DOI: 10.1038/nchem.2155.
- (38) Salditt, A.; Karr, L.; *et al.* Ribozyme-mediated RNA synthesis and replication in a model Hadean microenvironment. *Nat Commun* **2023**, 14 (1), 1495. DOI: 10.1038/s41467-023-37206-4.
- (39) Schwill, P.; Spatz, J.; *et al.* MaxSynBio: Avenues Towards Creating Cells from the Bottom Up. *Angewandte Chemie International Edition* **2018**, 57 (41), 13382–13392. DOI: 10.1002/anie.201802288.
- (40) Kriebisch, C. M.; Bantysh, O.; *et al.* A roadmap toward the synthesis of life. *Chem* **2025**, 11 (3), 102399. DOI: 10.1016/j.chempr.2024.102399.

References

- (41) Gonzales, D. T.; Suraritdechachai, S.; Tang, T.-Y. D. Compartmentalized Cell-Free Expression Systems for Building Synthetic Cells. *Advances in biochemical engineering/biotechnology* **2023**, *186*, 77–101. DOI: 10.1007/10_2023_221.
- (42) Hunt, A. C.; Rasor, B. J.; *et al.* Cell-Free Gene Expression: Methods and Applications. *Chemical reviews* **2025**, *125* (1), 91–149. DOI: 10.1021/acs.chemrev.4c00116.
- (43) Noireaux, V.; Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2004**, *101* (51), 17669–17674. DOI: 10.1073/pnas.0408236101.
- (44) Dora Tang, T.-Y.; van Swaay, D.; *et al.* In vitro gene expression within membrane-free coacervate protocells. *Chem. Commun.* **2015**, *51* (57), 11429–11432. DOI: 10.1039/C5CC04220H.
- (45) Abil, Z.; Restrepo Sierra, A. M.; *et al.* Darwinian Evolution of Self-Replicating DNA in a Synthetic Protocell. *Nat Commun* **2024**, *15* (1), 9091. DOI: 10.1038/s41467-024-53226-0.
- (46) Fujii, S.; Matsuura, T.; *et al.* In vitro evolution of α -hemolysin using a liposome display. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (42), 16796–16801. DOI: 10.1073/pnas.1314585110.
- (47) Ichihashi, N.; Usui, K.; *et al.* Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat Commun* **2013**, *4*, 2494. DOI: 10.1038/ncomms3494.
- (48) Holstein, J. M.; Gylstorff, C.; Hollfelder, F. Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. *ACS synthetic biology* **2021**, *10* (2), 252–257. DOI: 10.1021/acssynbio.0c00538.
- (49) Tawfik, D. S.; Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* **1998**, *16* (7), 652–656. DOI: 10.1038/nbt0798-652.
- (50) Ghadessy, F. J.; Ong, J. L.; Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98* (8), 4552–4557. DOI: 10.1073/pnas.071052198.
- (51) Murtas, G. Internal lipid synthesis and vesicle growth as a step toward self-reproduction of the minimal cell. *Systems and synthetic biology* **2010**, *4* (2), 85–93. DOI: 10.1007/s11693-009-9048-1.
- (52) Ji, P.; Harjung, A.; *et al.* Photochemical synthesis of natural lipids in artificial and living cells. *Nat Commun* **2025**, *16* (1), 5068. DOI: 10.1038/s41467-025-60358-4.
- (53) Bhattacharya, A.; Brea, R. J.; Niederholtmeyer, H.; Devaraj, N. K. A minimal biochemical route towards de novo formation of synthetic phospholipid membranes. *Nat Commun* **2019**, *10* (1), 300. DOI: 10.1038/s41467-018-08174-x.
- (54) Bi, E. F.; Lutkenhaus, J. FtsZ ring structure associated with division in *Escherichia coli*. *Nature* **1991**, *354* (6349), 161–164. DOI: 10.1038/354161a0.

References

- (55) Hu, Z.; Lutkenhaus, J. Topological regulation of cell division in *Escherichia coli* involves rapid pole to pole oscillation of the division inhibitor MinC under the control of MinD and MinE. *Molecular microbiology* **1999**, *34* (1), 82–90. DOI: 10.1046/j.1365-2958.1999.01575.x.
- (56) McQuillen, R.; Xiao, J. Insights into the Structure, Function, and Dynamics of the Bacterial Cytokinetic FtsZ-Ring. *Annual review of biophysics* **2020**, *49*, 309–341. DOI: 10.1146/annurev-biophys-121219-081703.
- (57) Raskin, D. M.; Boer, P. A. de. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96* (9), 4971–4976. DOI: 10.1073/pnas.96.9.4971.
- (58) Rothfield, L.; Justice, S.; García-Lara, J. Bacterial cell division. *Annual review of genetics* **1999**, *33*, 423–448. DOI: 10.1146/annurev.genet.33.1.423.
- (59) Scholey, J. M.; Brust-Mascher, I.; Mogilner, A. Cell division. *Nature* **2003**, *422* (6933), 746–752. DOI: 10.1038/nature01599.
- (60) Kohyama, S.; Merino-Salomón, A.; Schwille, P. In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nat Commun* **2022**, *13* (1), 6098. DOI: 10.1038/s41467-022-33679-x.
- (61) Angert, E. R. Alternatives to binary fission in bacteria. *Nature reviews. Microbiology* **2005**, *3* (3), 214–224. DOI: 10.1038/nrmicro1096.
- (62) Zhu, T. F.; Szostak, J. W. Coupled growth and division of model protocell membranes. *Journal of the American Chemical Society* **2009**, *131* (15), 5705–5713. DOI: 10.1021/ja900919c.
- (63) Ward, W. L.; Plakos, K.; DeRose, V. J. Nucleic acid catalysis: metals, nucleobases, and other cofactors. *Chemical reviews* **2014**, *114* (8), 4318–4342. DOI: 10.1021/cr400476k.
- (64) Lönnberg, H. Cleavage of RNA phosphodiester bonds by small molecular entities: a mechanistic insight. *Organic & biomolecular chemistry* **2011**, *9* (6), 1687–1703. DOI: 10.1039/C0OB00486C.
- (65) Holtmannspötter, A.-L.; Machatzke, C.; *et al.* Regulating Nucleic Acid Catalysis Using Active Droplets. *Angewandte Chemie International Edition* **2024**, *63* (47), e202412534. DOI: 10.1002/anie.202412534.
- (66) Nakashima, K. K.; Mihoubi, F. Z.; *et al.* Differential stability and dynamics of DNA-based and RNA-based coacervates affect non-enzymatic RNA chemistry. *Nat Commun* **2025**, *16* (1), 9296. DOI: 10.1038/s41467-025-64335-9.
- (67) Chaput, J. C.; Egli, M.; Herdewijn, P. The XNA alphabet. *Nucleic acids research* **2025**, *53* (13). DOI: 10.1093/nar/gkaf635.
- (68) Malik, T. N.; Chaput, J. C. XNA enzymes by evolution and design. *Current Research in Chemical Biology* **2021**, *1*, 100012. DOI: 10.1016/j.crchbi.2021.100012.
- (69) Ellington, A. D.; Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **1990**, *346* (6287), 818–822. DOI: 10.1038/346818a0.

References

- (70) Joyce, G. F. Amplification, mutation and selection of catalytic RNA. *Gene* **1989**, 82 (1), 83–87. DOI: 10.1016/0378-1119(89)90033-4.
- (71) Klug, S. J.; Famulok, M. All you wanted to know about SELEX. *Molecular biology reports* **1994**, 20 (2), 97–107. DOI: 10.1007/BF00996358.
- (72) Tuerk, C.; Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)* **1990**, 249 (4968), 505–510. DOI: 10.1126/science.2200121.
- (73) Breaker, R. R.; Joyce, G. F. A DNA enzyme that cleaves RNA. *Chemistry & Biology* **1994**, 1 (4), 223–229. DOI: 10.1016/1074-5521(94)90014-0.
- (74) Liu, M.; Chang, D.; Li, Y. Discovery and Biosensing Applications of Diverse RNA-Cleaving DNAzymes. *Accounts of chemical research* **2017**, 50 (9), 2273–2283. DOI: 10.1021/acs.accounts.7b00262#Abstract.
- (75) Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends in Biochemical Sciences* **2016**, 41 (7), 595–609. DOI: 10.1016/j.tibs.2016.04.010.
- (76) Gong, L.; Zhao, Z.; *et al.* DNAzyme-based biosensors and nanodevices. *Chem. Commun.* **2015**, 51 (6), 979–995. DOI: 10.1039/C4CC06855F.
- (77) Ji, R.; Niu, W.; *et al.* Target-inspired Pb²⁺-dependent DNAzyme for ultrasensitive electrochemical sensor based on MoS₂-AuPt nanocomposites and hemin/G-quadruplex DNAzyme as signal amplifier. *Biosensors & bioelectronics* **2019**, 144, 111560. DOI: 10.1016/j.bios.2019.111560.
- (78) Wu, H.; Wang, S.; *et al.* A label-free lead(II) ion sensor based on surface plasmon resonance and DNAzyme-gold nanoparticle conjugates. *Analytical and bioanalytical chemistry* **2020**, 412 (27), 7525–7533. DOI: 10.1007/s00216-020-02887-z.
- (79) Xu, W.; Tian, J.; *et al.* A rapid and visual turn-off sensor for detecting copper (II) ion based on DNAzyme coupled with HCR-based HRP concatemers. *Scientific reports* **2017**, 7, 43362. DOI: 10.1038/srep43362.
- (80) Yun, W.; Jiang, J.; *et al.* Ultrasensitive electrochemical detection of UO₂²⁺ based on DNAzyme and isothermal enzyme-free amplification. *RSC Adv.* **2016**, 6 (5), 3960–3966. DOI: 10.1039/C5RA22773A.
- (81) Liu, J.; Brown, A. K.; *et al.* A catalytic beacon sensor for uranium with parts-per-trillion sensitivity and millionfold selectivity. *Proceedings of the National Academy of Sciences of the United States of America* **2007**, 104 (7), 2056–2061. DOI: 10.1073/pnas.0607875104.
- (82) Yun, W.; Zhong, H.; *et al.* Simple, one-step and amplified Hg²⁺ detection strategy based on DNAzyme motor. *Sensors and Actuators B: Chemical* **2018**, 277, 456–461. DOI: 10.1016/j.snb.2018.09.050.
- (83) Ravikumar, A.; Panneerselvam, P.; Radhakrishnan, K. Fluorometric determination of lead(II) and mercury(II) based on their interaction with a complex formed between graphene oxide and a DNAzyme. *Mikrochimica acta* **2017**, 185 (1), 2. DOI: 10.1007/s00604-017-2585-5.

References

- (84) Li, J.; Zheng, W.; Kwon, A. H.; Lu, Y. In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme. *Nucleic acids research* **2000**, 28 (2), 481–488. DOI: 10.1093/nar/28.2.481.
- (85) Santoro, S. W.; Joyce, G. F.; *et al.* RNA cleavage by a DNA enzyme with extended chemical functionality. *J. Am. Chem. Soc.* **2000**, 122 (11), 2433–2439. DOI: 10.1021/ja993688s.
- (86) Zhou, W.; Zhang, Y.; Ding, J.; Liu, J. In Vitro Selection in Serum: RNA-Cleaving DNazymes for Measuring Ca²⁺ and Mg²⁺. *ACS Sens.* **2016**, 1 (5), 600–606. DOI: 10.1021/acssensors.5b00306.
- (87) Huang, P.-J. J.; Liu, J. Rational evolution of Cd²⁺-specific DNazymes with phosphorothioate modified cleavage junction and Cd²⁺ sensing. *Nucleic acids research* **2015**, 43 (12), 6125–6133. DOI: 10.1093/nar/gkv519.
- (88) Zhou, W.; Vazin, M.; *et al.* In Vitro Selection of Chromium-Dependent DNazymes for Sensing Chromium(III) and Chromium(VI). *Chemistry (Weinheim an der Bergstrasse, Germany)* **2016**, 22 (28), 9835–9840. DOI: 10.1002/chem.201601426.
- (89) Huang, P.-J. J.; Lin, J.; *et al.* Ultrasensitive DNzyme beacon for lanthanides and metal speciation. *Analytical chemistry* **2014**, 86 (3), 1816–1821. DOI: 10.1021/ac403762s.
- (90) Huang, P.-J. J.; Vazin, M.; Liu, J. In vitro selection of a new lanthanide-dependent DNzyme for ratiometric sensing lanthanides. *Analytical chemistry* **2014**, 86 (19), 9993–9999. DOI: 10.1021/ac5029962.
- (91) Saran, R.; Liu, J. A Silver DNzyme. *Analytical chemistry* **2016**, 88 (7), 4014–4020. DOI: 10.1021/acs.analchem.6b00327.
- (92) Khan, S.; Burciu, B.; *et al.* DNzyme-Based Biosensors: Immobilization Strategies, Applications, and Future Prospective. *ACS nano* **2021**, 15 (9), 13943–13969. DOI: 10.1021/acsnano.1c04327.
- (93) McGhee, C. E.; Loh, K. Y.; Lu, Y. DNzyme sensors for detection of metal ions in the environment and imaging them in living cells. *Current opinion in biotechnology* **2017**, 45, 191–201. DOI: 10.1016/j.copbio.2017.03.002.
- (94) Wang, S.; Liu, C.; *et al.* The Triple Roles of Glutathione for a DNA-Cleaving DNzyme and Development of a Fluorescent Glutathione/Cu²⁺-Dependent DNzyme Sensor for Detection of Cu²⁺ in Drinking Water. *ACS Sens.* **2017**, 2 (3), 364–370. DOI: 10.1021/acssensors.6b00667.
- (95) Chandra, M.; Sachdeva, A.; Silverman, S. K. DNA-catalyzed sequence-specific hydrolysis of DNA. *Nature chemical biology* **2009**, 5 (10), 718–720. DOI: 10.1038/nchembio.201.
- (96) Büttner, L.; Seikowski, J.; *et al.* Synthesis of spin-labeled riboswitch RNAs using convertible nucleosides and DNA-catalyzed RNA ligation. *Bioorganic & medicinal chemistry* **2013**, 21 (20), 6171–6180. DOI: 10.1016/j.bmc.2013.04.007.
- (97) Flynn-Charlebois, A.; Wang, Y.; *et al.* Deoxyribozymes with 2'-5' RNA ligase activity. *J. Am. Chem. Soc.* **2003**, 125 (9), 2444–2454. DOI: 10.1021/ja028774y.

References

- (98) Chandrasekar, J.; Silverman, S. K. Catalytic DNA with phosphatase activity. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (14), 5315–5320. DOI: 10.1073/pnas.1221946110.
- (99) Walsh, S. M.; Sachdeva, A.; Silverman, S. K. DNA catalysts with tyrosine kinase activity. *J. Am. Chem. Soc.* **2013**, *135* (40), 14928–14931. DOI: 10.1021/ja407586u.
- (100) Kosman, J.; Juskowiak, B. Peroxidase-mimicking DNAzymes for biosensing applications: a review. *Analytica chimica acta* **2011**, *707* (1-2), 7–17. DOI: 10.1016/j.aca.2011.08.050.
- (101) Höbartner, C.; Pradeepkumar, P. I.; Silverman, S. K. Site-selective depurination by a periodate-dependent deoxyribozyme. *Chemical communications (Cambridge, England)* **2007** (22), 2255–2257. DOI: 10.1039/b704507g.
- (102) Jayasena, S. D. Aptamers: An Emerging Class of Molecules That Rival Antibodies in Diagnostics. *Clinical Chemistry* **1999**, *45* (9), 1628–1650. DOI: 10.1093/clinchem/45.9.1628.
- (103) Qian, S.; Chang, D.; He, S.; Li, Y. Aptamers from random sequence space: Accomplishments, gaps and future considerations. *Analytica chimica acta* **2022**, *1196*, 339511. DOI: 10.1016/j.aca.2022.339511.
- (104) Dunn, M. R.; Jimenez, R. M.; Chaput, J. C. Analysis of aptamer discovery and technology. *Nat Rev Chem* **2017**, *1* (10). DOI: 10.1038/s41570-017-0076.
- (105) Ouellet, J. RNA Fluorescence with Light-Up Aptamers. *Frontiers in chemistry* **2016**, *4*, 29. DOI: 10.3389/fchem.2016.00029.
- (106) Wagner, A.; Mutschler, H. Design of Novel Synthetic RNA Replicons Based on Emesvirus zinderi. *ACS synthetic biology* **2024**, *13* (6), 1773–1780. DOI: 10.1021/acssynbio.4c00097.
- (107) Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science (New York, N.Y.)* **2011**, *333* (6042), 642–646. DOI: 10.1126/science.1207339.
- (108) Ruckman, J.; Green, L. S.; *et al.* 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *Journal of Biological Chemistry* **1998**, *273* (32), 20556–20567. DOI: 10.1074/jbc.273.32.20556.
- (109) Nimjee, S. M.; Keys, J. R.; *et al.* A novel antidote-controlled anticoagulant reduces thrombin generation and inflammation and improves cardiac function in cardiopulmonary bypass surgery. *Molecular therapy : the journal of the American Society of Gene Therapy* **2006**, *14* (3), 408–415. DOI: 10.1016/j.ymthe.2006.04.006.
- (110) Soundararajan, S.; Chen, W.; *et al.* The nucleolin targeting aptamer AS1411 destabilizes Bcl-2 messenger RNA in human breast cancer cells. *Cancer research* **2008**, *68* (7), 2358–2365. DOI: 10.1158/0008-5472.CAN-07-5723.
- (111) Cesarini, V.; Appleton, S. L.; Franciscis, V. de; Catalucci, D. The recent blooming of therapeutic aptamers. *Molecular aspects of medicine* **2025**, *102*, 101350. DOI: 10.1016/j.mam.2025.101350.

References

- (112) Keefe, A. D.; Pai, S.; Ellington, A. Aptamers as therapeutics. *Nature reviews. Drug discovery* **2010**, 9 (7), 537–550. DOI: 10.1038/nrd3141.
- (113) VarnBuhler, B. S.; Moon, J.; *et al.* Detection of SARS-CoV-2 RNA Using a DNA Aptamer Mimic of Green Fluorescent Protein. *ACS chemical biology* **2022**, 17 (4), 840–853. DOI: 10.1021/acschembio.1c00893.
- (114) Vibhute, M. A.; Machatzke, C.; *et al.* Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2. *eLife* **2026**, 13. DOI: 10.7554/eLife.99554.3.
- (115) Filonov, G. S.; Moon, J. D.; Svensen, N.; Jaffrey, S. R. Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **2014**, 136 (46), 16299–16308. DOI: 10.1021/ja508478x.
- (116) Strack, R. L.; Disney, M. D.; Jaffrey, S. R. A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nature methods* **2013**, 10 (12), 1219–1224. DOI: 10.1038/nmeth.2701.
- (117) Passalacqua, L. F. M.; Banco, M. T.; *et al.* Intricate 3D architecture of a DNA mimic of GFP. *Nature* **2023**, 618 (7967), 1078–1084. DOI: 10.1038/s41586-023-06229-8.
- (118) Spille, J.-H.; Kubitscheck, U. Labelling and imaging of single endogenous messenger RNA particles in vivo. *Journal of cell science* **2015**, 128 (20), 3695–3706. DOI: 10.1242/jcs.166728.
- (119) Strack, R. L.; Jaffrey, S. R. Live-cell imaging of mammalian RNAs with Spinach2. *Methods in enzymology* **2015**, 550, 129–146. DOI: 10.1016/bs.mie.2014.10.044.
- (120) Zou, J.; Huang, X.; *et al.* Selection of Intracellularly Functional RNA Mimics of Green Fluorescent Protein Using Fluorescence-Activated Cell Sorting. *Journal of molecular evolution* **2015**, 81 (5-6), 172–178. DOI: 10.1007/s00239-015-9718-4.
- (121) Sassanfar, M.; Roberts, J. W. Nature of the SOS-inducing signal in Escherichia coli. The involvement of DNA replication. *Journal of molecular biology* **1990**, 212 (1), 79–96. DOI: 10.1016/0022-2836(90)90306-7.
- (122) Zou, L.; Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science (New York, N.Y.)* **2003**, 300 (5625), 1542–1548. DOI: 10.1126/science.1083430.
- (123) Sass, T. H.; Lovett, S. T. The DNA damage response of Escherichia coli, revisited: Differential gene expression after replication inhibition. *Proceedings of the National Academy of Sciences of the United States of America* **2024**, 121 (27), e2407832121. DOI: 10.1073/pnas.2407832121.
- (124) Zou, Y.; Liu, Y.; Wu, X.; Shell, S. M. Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses. *Journal of cellular physiology* **2006**, 208 (2), 267–273. DOI: 10.1002/jcp.20622.
- (125) Yee, T.; Furuichi, T.; Inouye, S.; Inouye, M. Multicopy single-stranded DNA isolated from a gram-negative bacterium, Myxococcus xanthus. *Cell* **1984**, 38 (1), 203–209. DOI: 10.1016/0092-8674(84)90541-5.

References

- (126) Dhundale, A. R.; Furuichi, T.; Inouye, S.; Inouye, M. Distribution of multicopy single-stranded DNA among myxobacteria and related species. *Journal of bacteriology* **1985**, *164* (2), 914–917. DOI: 10.1128/jb.164.2.914-917.1985.
- (127) Furuichi, T.; Dhundale, A.; Inouye, M.; Inouye, S. Branched RNA covalently linked to the 5' end of a single-stranded DNA in *Stigmatella aurantiaca*: structure of msDNA. *Cell* **1987**, *48* (1), 47–53. DOI: 10.1016/0092-8674(87)90354-0.
- (128) Dhundale, A.; Lampson, B.; *et al.* Structure of msDNA from *Myxococcus xanthus*: evidence for a long, self-annealing RNA precursor for the covalently linked, branched RNA. *Cell* **1987**, *51* (6), 1105–1112. DOI: 10.1016/0092-8674(87)90596-4.
- (129) Lampson, B. C.; Sun, J.; *et al.* Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science (New York, N.Y.)* **1989**, *243* (4894 Pt 1), 1033–1038. DOI: 10.1126/science.2466332.
- (130) Millman, A.; Bernheim, A.; *et al.* Bacterial Retrons Function In Anti-Phage Defense. *Cell* **2020**, *183* (6), 1551-1561.e12. DOI: 10.1016/j.cell.2020.09.065.
- (131) Doron, S.; Melamed, S.; *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (New York, N.Y.)* **2018**, *359* (6379). DOI: 10.1126/science.aar4120.
- (132) Makarova, K. S.; Wolf, Y. I.; Snir, S.; Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology* **2011**, *193* (21), 6039–6056. DOI: 10.1128/jb.05535-11.
- (133) Harms, A.; Brodersen, D. E.; Mitarai, N.; Gerdes, K. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Molecular cell* **2018**, *70* (5), 768–784. DOI: 10.1016/j.molcel.2018.01.003.
- (134) Page, R.; Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature chemical biology* **2016**, *12* (4), 208–214. DOI: 10.1038/nchembio.2044.
- (135) Unterholzner, S. J.; Poppenberger, B.; Rozhon, W. Toxin-antitoxin systems: Biology, identification, and application. *Mobile genetic elements* **2013**, *3* (5), e26219. DOI: 10.4161/mge.26219.
- (136) Gerdes, K.; Rasmussen, P. B.; Molin, S. Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences of the United States of America* **1986**, *83* (10), 3116–3120. DOI: 10.1073/pnas.83.10.3116.
- (137) Dy, R. L.; Richter, C.; Salmond, G. P. C.; Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual review of virology* **2014**, *1* (1), 307–331. DOI: 10.1146/annurev-virology-031413-085500.
- (138) Harms, A.; Maisonneuve, E.; Gerdes, K. Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science (New York, N.Y.)* **2016**, *354* (6318). DOI: 10.1126/science.aaf4268.

References

- (139) Mestre, M. R.; González-Delgado, A.; *et al.* Systematic prediction of genes functionally associated with bacterial retrons and classification of the encoded tripartite systems. *Nucleic acids research* **2020**, *48* (22), 12632–12647. DOI: 10.1093/nar/gkaa1149.
- (140) Khan, A. G.; Rojas-Montero, M.; *et al.* An experimental census of retrons for DNA production and genome editing. *Nature biotechnology* **2025**, *43* (6), 914–922. DOI: 10.1038/s41587-024-02384-z.
- (141) Nakamura, K. L.; Zhang, K.; *et al.* New retron systems from environmental bacteria identify triggers of anti-phage defense and expand tools for genome editing. *PLoS biology* **2025**, *23* (10), e3003042. DOI: 10.1371/journal.pbio.3003042.
- (142) Rychlik, I.; Sebkova, A.; Gregorova, D.; Karpiskova, R. Low-molecular-weight plasmid of *Salmonella enterica* serovar Enteritidis codes for retron reverse transcriptase and influences phage resistance. *Journal of bacteriology* **2001**, *183* (9), 2852–2858. DOI: 10.1128/jb.183.9.2852-2858.2001.
- (143) Wang, Y.; Wang, C.; *et al.* Structural basis of the RNA-mediated Retron-Eco2 oligomerization. *Cell discovery* **2025**, *11* (1), 73. DOI: 10.1038/s41421-025-00823-y.
- (144) Wang, Y.; Guan, Z.; *et al.* Cryo-EM structures of *Escherichia coli* Ec86 retron complexes reveal architecture and defence mechanism. *Nature microbiology* **2022**, *7* (9), 1480–1489. DOI: 10.1038/s41564-022-01197-7.
- (145) Hsu, M. Y.; Inouye, M.; Inouye, S. Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: a potential transposable element encoding both reverse transcriptase and Dam methylase functions. *Proceedings of the National Academy of Sciences of the United States of America* **1990**, *87* (23), 9454–9458. DOI: 10.1073/pnas.87.23.9454.
- (146) Shimamoto, T.; Inouye, M.; Inouye, S. The formation of the 2',5'-phosphodiester linkage in the cDNA priming reaction by bacterial reverse transcriptase in a cell-free system. *Journal of Biological Chemistry* **1995**, *270* (2), 581–588. DOI: 10.1074/jbc.270.2.581.
- (147) Palka, C.; Fishman, C. B.; *et al.* Retron reverse transcriptase termination and phage defense are dependent on host RNase H1. *Nucleic acids research* **2022**, *50* (6), 3490–3504. DOI: 10.1093/nar/gkac177.
- (148) Shimamoto, T.; Shimada, M.; Inouye, M.; Inouye, S. The role of ribonuclease H in multicopy single-stranded DNA synthesis in retron-Ec73 and retron-Ec107 of *Escherichia coli*. *Journal of bacteriology* **1995**, *177* (1), 264–267. DOI: 10.1128/jb.177.1.264-267.1995.
- (149) Stein, H.; Hausen, P. Enzyme from calf thymus degrading the RNA moiety of DNA-RNA Hybrids: effect on DNA-dependent RNA polymerase. *Science* **1969**, *166* (3903), 393–395. DOI: 10.1126/science.166.3903.393.
- (150) Cerritelli, S. M.; Crouch, R. J. Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal* **2009**, *276* (6), 1494–1505. DOI: 10.1111/j.1742-4658.2009.06908.x.

References

- (151) Bobonis, J.; Mitosch, K.; *et al.* Bacterial retrons encode phage-defending tripartite toxin-antitoxin systems. *Nature* **2022**, 609 (7925), 144–150. DOI: 10.1038/s41586-022-05091-4.
- (152) Sharon, E.; Chen, S.-A. A.; *et al.* Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell* **2018**, 175 (2), 544–557.e16. DOI: 10.1016/j.cell.2018.08.057.
- (153) Lopez, S. C.; Crawford, K. D.; *et al.* Precise genome editing across kingdoms of life using retron-derived DNA. *Nature chemical biology* **2022**, 18 (2), 199–206. DOI: 10.1038/s41589-021-00927-y.
- (154) Luo, D.; Saltzman, W. M. Synthetic DNA delivery systems. *Nature biotechnology* **2000**, 18 (1), 33–37. DOI: 10.1038/71889.
- (155) Kosicki, M.; Tomberg, K.; Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat Biotechnol* **2018**, 36 (8), 765–771. DOI: 10.1038/nbt.4192.
- (156) Schubert, M. G.; Goodman, D. B.; *et al.* High-throughput functional variant screens via in vivo production of single-stranded DNA. *Proceedings of the National Academy of Sciences of the United States of America* **2021**, 118 (18). DOI: 10.1073/pnas.2018181118.
- (157) Liu, J.; Cui, L.; *et al.* Generation of DNAzyme in Bacterial Cells by a Bacterial Retron System. *ACS synthetic biology* **2024**, 13 (1), 300–309. DOI: 10.1021/acssynbio.3c00509.
- (158) Lim, D.; Maas, W. K. Reverse transcriptase-dependent synthesis of a covalently linked, branched DNA-RNA compound in *E. coli* B. *Cell* **1989**, 56 (5), 891–904. DOI: 10.1016/0092-8674(89)90693-4.
- (159) Brown, A. K.; Li, J.; Pavot, C. M.-B.; Lu, Y. A lead-dependent DNAzyme with a two-step mechanism. *Biochemistry* **2003**, 42 (23), 7152–7161. DOI: 10.1021/bi027332w.
- (160) Li, J.; Lu, Y. A Highly Sensitive and Selective Catalytic DNA Biosensor for Lead Ions. *J. Am. Chem. Soc.* **2000**, 122 (42), 10466–10467. DOI: 10.1021/ja0021316.
- (161) Moon, W. J.; Huang, P.-J. J.; Liu, J. Probing Metal-Dependent Phosphate Binding for the Catalysis of the 17E DNAzyme. *Biochemistry* **2021**, 60 (24), 1909–1918. DOI: 10.1021/acs.biochem.1c00091.
- (162) Peracchi, A. Preferential activation of the 8-17 deoxyribozyme by Ca(2+) ions. Evidence for the identity of 8-17 with the catalytic domain of the Mg5 deoxyribozyme. *Journal of Biological Chemistry* **2000**, 275 (16), 11693–11697. DOI: 10.1074/jbc.275.16.11693.
- (163) Cepeda-Plaza, M.; Peracchi, A. Insights into DNA catalysis from structural and functional studies of the 8-17 DNAzyme. *Organic & biomolecular chemistry* **2020**, 18 (9), 1697–1709. DOI: 10.1039/C9OB02453K.
- (164) Wieruszewska, J.; Pawłowicz, A.; *et al.* The 8-17 DNAzyme can operate in a single active structure regardless of metal ion cofactor. *Nat Commun* **2024**, 15 (1), 4218. DOI: 10.1038/s41467-024-48638-x.

References

- (165) Donau, C.; Späth, F.; *et al.* Active coacervate droplets as a model for membraneless organelles and protocells. *Nat Commun* **2020**, *11* (1), 5167. DOI: 10.1038/s41467-020-18815-9.
- (166) Tena-Solsona, M.; Rieß, B.; *et al.* Non-equilibrium dissipative supramolecular materials with a tunable lifetime. *Nat Commun* **2017**, *8*, 15895. DOI: 10.1038/ncomms15895.
- (167) Overbeek, J. T. G.; Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.* **1957**, *49* (S1), 7–26. DOI: 10.1002/jcp.1030490404.
- (168) van der Gucht, J.; Spruijt, E.; Lemmers, M.; Cohen Stuart, M. A. Polyelectrolyte complexes: bulk phases and colloidal systems. *Journal of colloid and interface science* **2011**, *361* (2), 407–422. DOI: 10.1016/j.jcis.2011.05.080.
- (169) Sing, C. E.; Perry, S. L. Recent progress in the science of complex coacervation. *Soft Matter* **2020**, *16* (12), 2885–2914. DOI: 10.1039/D0SM00001A.
- (170) H. G. Bungenberg; H. R. Kruyt. Coacervation (Partial miscibility in colloid systems). *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* **1929** (32), 849–856.
- (171) Abbas, M.; Lipiński, W. P.; Wang, J.; Spruijt, E. Peptide-based coacervates as biomimetic protocells. *Chemical Society Reviews* **2021**, *50* (6), 3690–3705. DOI: 10.1039/D0CS00307G.
- (172) Gabryelczyk, B.; Cai, H.; *et al.* Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. *Nat Commun* **2019**, *10* (1), 5465. DOI: 10.1038/s41467-019-13469-8.
- (173) Martin, E. W.; Holehouse, A. S.; *et al.* Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science (New York, N.Y.)* **2020**, *367* (6478), 694–699. DOI: 10.1126/science.aaw8653.
- (174) Priftis, D.; Leon, L.; *et al.* Self-assembly of α -helical polypeptides driven by complex coacervation. *Angewandte Chemie International Edition* **2015**, *54* (38), 11128–11132. DOI: 10.1002/anie.201504861.
- (175) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. *Nature Phys* **2015**, *11* (11), 899–904. DOI: 10.1038/nphys3532.
- (176) Wang, J.; Choi, J.-M.; *et al.* A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174* (3), 688-699.e16. DOI: 10.1016/j.cell.2018.06.006.
- (177) Nott, T. J.; Petsalaki, E.; *et al.* Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular cell* **2015**, *57* (5), 936–947. DOI: 10.1016/j.molcel.2015.01.013.
- (178) Leon, L.; Santiago, G. M.; Spruijt, E. Coacervation in systems chemistry. *Communications chemistry* **2024**, *7* (1), 275. DOI: 10.1038/s42004-024-01358-1.
- (179) Chen, I. A.; Salehi-Ashtiani, K.; Szostak, J. W. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.* **2005**, *127* (38), 13213–13219. DOI: 10.1021/ja051784p.

References

- (180) Monnard, P.-A.; Apel, C. L.; Kanavarioti, A.; Deamer, D. W. Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: implications for a prebiotic aqueous medium. *Astrobiology* **2002**, 2 (2), 139–152. DOI: 10.1089/15311070260192237.
- (181) Colletier, J.-P.; Chaize, B.; Winterhalter, M.; Fournier, D. Protein encapsulation in liposomes: efficiency depends on interactions between protein and phospholipid bilayer. *BMC biotechnology* **2002**, 2, 9. DOI: 10.1186/1472-6750-2-9.
- (182) Deamer, D. W.; Barchfeld, G. L. Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. *Journal of molecular evolution* **1982**, 18 (3), 203–206. DOI: 10.1007/BF01733047.
- (183) Dominak, L. M.; Keating, C. D. Polymer encapsulation within giant lipid vesicles. *Langmuir : the ACS journal of surfaces and colloids* **2007**, 23 (13), 7148–7154. DOI: 10.1021/la063687v.
- (184) Tran, M. P.; Chakraborty, T.; *et al.* Genetic encoding and expression of RNA origami cytoskeletons in synthetic cells. *Nat. Nanotechnol.* **2025**, 20 (5), 664–671. DOI: 10.1038/s41565-025-01879-3.
- (185) Chang, M.-Y.; Ariyama, H.; Huck, W. T. S.; Deng, N.-N. Division in synthetic cells. *Chemical Society Reviews* **2023**, 52 (10), 3307–3325. DOI: 10.1039/D2CS00985D.
- (186) Dora Tang, T.-Y.; Rohaida Che Hak, C.; *et al.* Fatty acid membrane assembly on coacervate microdroplets as a step towards a hybrid protocell model. *Nat. Chem.* **2014**, 6 (6), 527–533. DOI: 10.1038/nchem.1921.
- (187) Ianeselli, A.; Tetiker, D.; *et al.* Non-equilibrium conditions inside rock pores drive fission, maintenance and selection of coacervate protocells. *Nat. Chem.* **2022**, 14 (1), 32–39. DOI: 10.1038/s41557-021-00830-y.
- (188) Qiao, Y.; Li, M.; Booth, R.; Mann, S. Predatory behaviour in synthetic protocell communities. *Nat. Chem.* **2017**, 9 (2), 110–119. DOI: 10.1038/nchem.2617.
- (189) Wollny, D.; Vernot, B.; *et al.* Characterization of RNA content in individual phase-separated coacervate microdroplets. *Nat Commun* **2022**, 13 (1), 2626. DOI: 10.1038/s41467-022-30158-1.
- (190) Chollakup, R.; Smitthipong, W.; Eisenbach, C. D.; Tirrell, M. Phase Behavior and Coacervation of Aqueous Poly(acrylic acid)–Poly(allylamine) Solutions. *Macromolecules* **2010**, 43 (5), 2518–2528. DOI: 10.1021/ma902144k.
- (191) Cook, A. B.; Novosedlik, S.; van Hest, J. C. M. Complex Coacervate Materials as Artificial Cells. *Accounts of materials research* **2023**, 4 (3), 287–298. DOI: 10.1021/accountsmr.2c00239.
- (192) Choi, S.; Meyer, M. O.; Bevilacqua, P. C.; Keating, C. D. Phase-specific RNA accumulation and duplex thermodynamics in multiphase coacervate models for membraneless organelles. *Nat. Chem.* **2022**, 14 (10), 1110–1117. DOI: 10.1038/s41557-022-00980-7.
- (193) Lu, T.; Nakashima, K. K.; Spruijt, E. Temperature-Responsive Peptide-Nucleotide Coacervates. *The journal of physical chemistry. B* **2021**, 125 (12), 3080–3091. DOI: 10.1021/acs.jpcc.0c10839.

References

- (194) Wang, J.; Abbas, M.; Wang, J.; Spruijt, E. Selective amide bond formation in redox-active coacervate protocells. *Nat Commun* **2023**, *14* (1), 8492. DOI: 10.1038/s41467-023-44284-x.
- (195) Ghosh, B.; McCall, P. M.; *et al.* RNA-peptide interactions tune the ribozyme activity within coacervate microdroplet dispersions. *Nat Commun* **2025**, *16* (1), 8765. DOI: 10.1038/s41467-025-63656-z.
- (196) Mason, A. F.; Buddingh', B. C.; Williams, D. S.; van Hest, J. C. M. Hierarchical Self-Assembly of a Copolymer-Stabilized Coacervate Protocell. *J. Am. Chem. Soc.* **2017**, *139* (48), 17309–17312. DOI: 10.1021/jacs.7b10846.
- (197) Drobot, B.; Iglesias-Artola, J. M.; *et al.* Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nat Commun* **2018**, *9* (1), 3643. DOI: 10.1038/s41467-018-06072-w.
- (198) Briani, F.; Carzaniga, T.; Dehò, G. Regulation and functions of bacterial PNPase. *Wiley interdisciplinary reviews. RNA* **2016**, *7* (2), 241–258. DOI: 10.1002/wrna.1328.
- (199) Brangwynne, C. P.; Eckmann, C. R.; *et al.* Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.)* **2009**, *324* (5935), 1729–1732. DOI: 10.1126/science.1172046.
- (200) Mishra, S.; Solra, M.; Rana, S. Catalytically Active Dynamic Phase-Separated Coacervates. *ChemCatChem* **2025**. DOI: 10.1002/cctc.202500926.
- (201) Boekhoven, J.; Brizard, A. M.; *et al.* Dissipative Self-Assembly of a Molecular Gelator by Using a Chemical Fuel. *Angewandte Chemie* **2010**, *122* (28), 4935–4938. DOI: 10.1002/ange.201001511.
- (202) Boekhoven, J.; Hendriksen, W. E.; *et al.* Transient assembly of active materials fueled by a chemical reaction. *Science (New York, N.Y.)* **2015**, *349* (6252), 1075–1079. DOI: 10.1126/science.aac6103.
- (203) Leira-Iglesias, J.; Sorrenti, A.; *et al.* Supramolecular pathway selection of perylenediimides mediated by chemical fuels. *Chem. Commun.* **2016**, *52* (58), 9009–9012. DOI: 10.1039/c6cc01192f.
- (204) Sorrenti, A.; Leira-Iglesias, J.; Sato, A.; Hermans, T. M. Non-equilibrium steady states in supramolecular polymerization. *Nat Commun* **2017**, *8*, 15899. DOI: 10.1038/ncomms15899.
- (205) Te Brinke, E.; Groen, J.; *et al.* Dissipative adaptation in driven self-assembly leading to self-dividing fibrils. *Nat. Nanotechnol.* **2018**, *13* (9), 849–855. DOI: 10.1038/s41565-018-0192-1.
- (206) Machatzke, C.; Holtmannspötter, A.-L.; Mutschler, H.; Boekhoven, J. DNA affects the phenotype of fuel-dependent coacervate droplets. *Nature Communications* **2026** (17). DOI: 10.1038/s41467-026-71024-8.
- (207) Holtmannspötter, A.-L.; Machatzke, C.; Boekhoven, J.; Mutschler, H. Protocol for the recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets. *STAR Protocols* **2025**, *7* (1), 104293. DOI: 10.1016/j.xpro.2025.104293.

References

- (208) Späth, F.; Donau, C.; *et al.* Molecular Design of Chemically Fueled Peptide-Polyelectrolyte Coacervate-Based Assemblies. *Journal of the American Chemical Society* **2021**, *143* (12), 4782–4789. DOI: 10.1021/jacs.1c01148.
- (209) Rieß, B.; Grötsch, R. K.; Boekhoven, J. The Design of Dissipative Molecular Assemblies Driven by Chemical Reaction Cycles. *Chem* **2020**, *6* (3), 552–578. DOI: 10.1016/j.chempr.2019.11.008.
- (210) Donau, C.; Boekhoven, J. The chemistry of chemically fueled droplets. *Trends in Chemistry* **2023**, *5* (1), 45–60. DOI: 10.1016/j.trechm.2022.11.003.
- (211) Troll, C. J.; Kapp, J.; *et al.* A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC genomics* **2019**, *20* (1), 1023. DOI: 10.1186/s12864-019-6355-0.
- (212) Gianni, E.; Kwok, S. L. Y.; *et al.* A small polymerase ribozyme that can synthesize itself and its complementary strand. *Science (New York, N.Y.)* **2026**, *391* (6789), 1022–1028. DOI: 10.1126/science.adt2760.
- (213) Simon, A. J.; Morrow, B. R.; Ellington, A. D. Retroelement-Based Genome Editing and Evolution. *ACS synthetic biology* **2018**, *7* (11), 2600–2611. DOI: 10.1021/acssynbio.8b00273.
- (214) Liu, W.; Zuo, S.; *et al.* Retron-mediated multiplex genome editing and continuous evolution in *Escherichia coli*. *Nucleic acids research* **2023**, *51* (15), 8293–8307. DOI: 10.1093/nar/gkad607.
- (215) Grubbs, R. D. Intracellular magnesium and magnesium buffering. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* **2002**, *15* (3), 251–259. DOI: 10.1023/A:1016026831789.
- (216) Hurwitz, C.; Rosano, C. L. The Intracellular Concentration of Bound and Unbound Magnesium Ions in *Escherichia coli*. *Journal of Biological Chemistry* **1967**, *242* (16), 3719–3722. DOI: 10.1016/S0021-9258(18)95867-9.
- (217) Hayes, F.; van Melderren, L. Toxins-antitoxins: diversity, evolution and function. *Critical reviews in biochemistry and molecular biology* **2011**, *46* (5), 386–408. DOI: 10.3109/10409238.2011.600437.
- (218) Pausch, P.; Jasnauskaitė, M.; *et al.* Retron Eco2 breaks tRNAs for antiphage defense **2025**. DOI: 10.21203/rs.3.rs-7196019/v1.
- (219) Jasnauskaitė, M.; Juozapaitis, J.; *et al.* Structure and mechanism of antiphage retron Eco2. *Nature structural & molecular biology* **2026**, *33* (2), 330–340. DOI: 10.1038/s41594-026-01754-2.
- (220) McCullum, E. O.; Williams, B. A. R.; Zhang, J.; Chaput, J. C. Random mutagenesis by error-prone PCR. *Methods in molecular biology (Clifton, N.J.)* **2010**, *634*, 103–109. DOI: 10.1007/978-1-60761-652-8_7.
- (221) Muteeb, G.; Sen, R. Random mutagenesis using a mutator strain. *Methods in molecular biology (Clifton, N.J.)* **2010**, *634*, 411–419. DOI: 10.1007/978-1-60761-652-8_29.
- (222) Opalinska, J. B.; Gewirtz, A. M. Nucleic-acid therapeutics: basic principles and recent applications. *Nature reviews. Drug discovery* **2002**, *1* (7), 503–514. DOI: 10.1038/nrd837.

References

- (223) Larcher, L. M.; Pitout, I. L.; *et al.* DNAzymes: Expanding the Potential of Nucleic Acid Therapeutics. *Nucleic acid therapeutics* **2023**, 33 (3), 178–192. DOI: 10.1089/nat.2022.0066.
- (224) Zhang, J.; Xi, Z.; *et al.* Ribozyme biotechnology: emerging opportunities for medical applications in the RNA therapeutics era. *Medical review (2021)* **2026**, 6 (1), 79–82. DOI: 10.1515/mr-2025-0062.
- (225) Huang, X.; Zhao, Y.; *et al.* Intracellular selection of trans-cleaving hammerhead ribozymes. *Nucleic acids research* **2019**, 47 (5), 2514–2522. DOI: 10.1093/nar/gkz018.
- (226) Dang, C. V.; Reddy, E. P.; Shokat, K. M.; Soucek, L. Drugging the 'undruggable' cancer targets. *Nature reviews. Cancer* **2017**, 17 (8), 502–508. DOI: 10.1038/nrc.2017.36.
- (227) McCormick, F. KRAS as a Therapeutic Target. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2015**, 21 (8), 1797–1801. DOI: 10.1158/1078-0432.CCR-14-2662.
- (228) Whitfield, J. R.; Beaulieu, M.-E.; Soucek, L. Strategies to Inhibit Myc and Their Clinical Applicability. *Frontiers in cell and developmental biology* **2017**, 5, 10. DOI: 10.3389/fcell.2017.00010.
- (229) Xie, X.; Yu, T.; *et al.* Recent advances in targeting the "undruggable" proteins: from drug discovery to clinical trials. *Signal transduction and targeted therapy* **2023**, 8 (1), 335. DOI: 10.1038/s41392-023-01589-z.
- (230) Lazo, J. S.; Sharlow, E. R. Drugging Undruggable Molecular Cancer Targets. *Annual review of pharmacology and toxicology* **2016**, 56, 23–40. DOI: 10.1146/annurev-pharmtox-010715-103440.
- (231) Moore, A. R.; Rosenberg, S. C.; McCormick, F.; Malek, S. RAS-targeted therapies: is the undruggable drugged? *Nature reviews. Drug discovery* **2020**, 19 (8), 533–552. DOI: 10.1038/s41573-020-0068-6.
- (232) Singhal, A.; Li, B. T.; O'Reilly, E. M. Targeting KRAS in cancer. *Nature medicine* **2024**, 30 (4), 969–983. DOI: 10.1038/s41591-024-02903-0.
- (233) Nguyen, K.; Malik, T. N.; Chaput, J. C. Chemical evolution of an autonomous DNAzyme with allele-specific gene silencing activity. *Nat Commun* **2023**, 14 (1), 2413. DOI: 10.1038/s41467-023-38100-9.
- (234) Wang, Y.; Nguyen, K.; Spitale, R. C.; Chaput, J. C. A biologically stable DNAzyme that efficiently silences gene expression in cells. *Nat. Chem.* **2021**, 13 (4), 319–326. DOI: 10.1038/s41557-021-00645-x.
- (235) Ashton, N. W.; Bolderson, E.; *et al.* Human single-stranded DNA binding proteins are essential for maintaining genomic stability. *BMC molecular biology* **2013**, 14, 9. DOI: 10.1186/1471-2199-14-9.
- (236) Ingle, R. G.; Fang, W.-J. An Overview of the Stability and Delivery Challenges of Commercial Nucleic Acid Therapeutics. *Pharmaceutics* **2023**, 15 (4). DOI: 10.3390/pharmaceutics15041158.

References

- (237) Abbas, M.; Lipiński, W. P.; *et al.* *A Short Peptide Synthron for Liquid-Liquid Phase Separation*, 2020. DOI: 10.26434/chemrxiv.12881288.v1.
- (238) Gobbo, P.; Tian, L.; *et al.* Catalytic processing in ruthenium-based polyoxometalate coacervate protocells. *Nat Commun* **2020**, *11* (1), 41. DOI: 10.1038/s41467-019-13759-1.
- (239) Sokolova, E.; Spruijt, E.; *et al.* Enhanced transcription rates in membrane-free protocells formed by coacervation of cell lysate. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (29), 11692–11697. DOI: 10.1073/pnas.1222321110.
- (240) Purtha, W. E.; Coppins, R. L.; Smalley, M. K.; Silverman, S. K. General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. *J. Am. Chem. Soc.* **2005**, *127* (38), 13124–13125. DOI: 10.1021/ja0533702.
- (241) Scheitl, C. P. M.; Lange, S.; Höbartner, C. New Deoxyribozymes for the Native Ligation of RNA. *Molecules (Basel, Switzerland)* **2020**, *25* (16). DOI: 10.3390/molecules25163650.
- (242) Poudyal, R. R.; Keating, C. D.; Bevilacqua, P. C. Polyanion-Assisted Ribozyme Catalysis Inside Complex Coacervates. *ACS chemical biology* **2019**, *14* (6), 1243–1248. DOI: 10.1021/acschembio.9b00205.
- (243) Ameta, S.; Kumar, M.; *et al.* Multispecies autocatalytic RNA reaction networks in coacervates. *Communications chemistry* **2023**, *6* (1), 91. DOI: 10.1038/s42004-023-00887-5.
- (244) Paul, N.; Joyce, G. F. A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99* (20), 12733–12740. DOI: 10.1073/pnas.202471099.
- (245) Metz, D. H.; Brown, G. L. *Investigation of nucleic acid secondary structure by means of chemical modification with a carbodiimide reagent. I. Reaction between N-cyclohexyl-N'-β-(4 ...*, 1969.
- (246) Williams, A.; Ibrahim, I. T. Carbodiimide chemistry: recent advances. *Chem. Rev.* **1981**, *81* (6), 589–636. DOI: 10.1021/cr00046a004.
- (247) Wang, Q.; Qi, Z.; Chen, M.; Qu, D.-H. Out-of-equilibrium supramolecular self-assembling systems driven by chemical fuel. *Aggregate* **2021**, *2* (5). DOI: 10.1002/agt2.110.
- (248) Yang, Z.; Guo, W.; *et al.* Surpassing Robeson Upper Limit for CO₂/N₂ Separation with Fluorinated Carbon Molecular Sieve Membranes. *Chem* **2020**, *6* (3), 631–645. DOI: 10.1016/j.chempr.2019.12.006.
- (249) Cardona, M. A.; Prins, L. J. ATP-fuelled self-assembly to regulate chemical reactivity in the time domain. *Chemical science* **2019**, *11* (6), 1518–1522. DOI: 10.1039/C9SC05188K.
- (250) Dhiman, S.; Jain, A.; Kumar, M.; George, S. J. Adenosine-Phosphate-Fueled, Temporally Programmed Supramolecular Polymers with Multiple Transient States. *J. Am. Chem. Soc.* **2017**, *139* (46), 16568–16575. DOI: 10.1021/jacs.7b07469.
- (251) Saha, B.; Chatterjee, A.; Reja, A.; Das, D. Condensates of short peptides and ATP for the temporal regulation of cytochrome c activity. *Chem. Commun.* **2019**, *55* (94), 14194–14197. DOI: 10.1039/C9CC07358B.

References

- (252) Poprawa, S. M.; Stasi, M.; *et al.* Active droplets through enzyme-free, dynamic phosphorylation. *Nat Commun* **2024**, *15* (1), 4204. DOI: 10.1038/s41467-024-48571-z.
- (253) Aumiller, W. M.; Keating, C. D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat. Chem.* **2016**, *8* (2), 129–137. DOI: 10.1038/nchem.2414.
- (254) Biondi, E.; Maxwell, A. W. R.; Burke, D. H. A small ribozyme with dual-site kinase activity. *Nucleic acids research* **2012**, *40* (15), 7528–7540. DOI: 10.1093/nar/gks356.
- (255) Lorsch, J. R.; Szostak, J. W. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **1994**, *371* (6492), 31–36. DOI: 10.1038/371031a0.
- (256) Jovic, D.; Liang, X.; *et al.* Single-cell RNA sequencing technologies and applications: A brief overview. *Clinical and translational medicine* **2022**, *12* (3), e694. DOI: 10.1002/ctm2.694.
- (257) Saliba, A.-E.; Westermann, A. J.; Gorski, S. A.; Vogel, J. Single-cell RNA-seq: advances and future challenges. *Nucleic acids research* **2014**, *42* (14), 8845–8860. DOI: 10.1093/nar/gku555.
- (258) Meyer, M.; Kircher, M.; *et al.* A high-coverage genome sequence from an archaic Denisovan individual. *Science (New York, N.Y.)* **2012**, *338* (6104), 222–226. DOI: 10.1126/science.1224344.
- (259) Prüfer, K.; Racimo, F.; *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **2014**, *505* (7481), 43–49. DOI: 10.1038/nature12886.
- (260) Gansauge, M.-T.; Gerber, T.; *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic acids research* **2017**, *45* (10), e79. DOI: 10.1093/nar/gkx033.
- (261) Wang, P.; Fang, X.; *et al.* Principles of amino-acid–ribonucleotide interaction revealed by binding affinities between homogeneous oligopeptides and single-stranded RNA molecules. *Nano Res.* **2023**, *16* (12), 13294–13300. DOI: 10.1007/s12274-023-5971-9.
- (262) Szokoli, D.; Le Vay, K. K.; Mutschler, H. PCR Methods for the Generation of Catalytic ssDNA. *Methods in molecular biology (Clifton, N.J.)* **2022**, *2439*, 27–44. DOI: 10.1007/978-1-0716-2047-2_3.

9 Publications

9.1 Publication 1

Intracellular Expression of a Fluorogenic DNA Aptamer Using Retron Eco2

Vibhute, M.A.*; **Machatzke, C.***; Bigler, K.; Krümpel, S.; Summerer, D.; Mutschler, H.⁺

* Authors contributed equally

⁺ Corresponding author

eLife **2026**, 13:RP99554

DOI: 10.7554/eLife.99554.3

9.1.1 Personal contribution

- Design of cargo using in silico methods
- Performed all in vitro measurements for the main text and SI
- Performed all gel electrophoresis experiments for the main text and SI
- Performed the intracellular bulk fluorescent measurements in the SI with M. Vibhute and M. Dirks
- Prepared figures 3 and 4 with M. Vibhute
- Composed, edited and revised the manuscript and supporting information with M. Vibhute and H. Mutschler

Intracellular expression of a fluorogenic DNA aptamer using retron Eco2

Mahesh A Vibhute[†], Corbin Machatzke[†], Saskia Krümpel, Malte Dirks, Katrin Bigler, Daniel Summerer, Hannes Mutschler*

Department of Chemistry and Chemical Biology, TU Dortmund University, Dortmund, Germany

eLife Assessment

This study presents a method for expressing single-stranded DNA fluorescent aptamers in *E. coli* using a retron-based strategy. The evidence supporting the successful expression and folding of DNA aptamers is **solid**, with clear demonstration of fluorescence after extraction, though the aptamers do not function in living cells. The method represents an **important** technical advance that will likely become standard for DNA aptamer expression in bacterial systems.

Abstract DNA aptamers are short, single-stranded DNA molecules that bind specifically to a range of targets such as proteins, cells, and small molecules. Typically, they are utilized in the development of therapeutic agents, diagnostics, drug delivery systems, and biosensors. Although aptamers perform well in controlled extracellular environments, their intracellular use has been less explored due to challenges of expressing them in vivo. In this study, we employed the bacterial retron system Eco2 to express a DNA light-up aptamer in *Escherichia coli*. Our data confirms that structure-guided insertion of the aptamer domain into the non-coding region of the retron enables reverse transcription and biosynthesis of functional aptamer constructs in bacteria. The purified DNA aptamer synthesized under intracellular conditions shows comparable activity to a chemically synthesized control. Our findings demonstrate that retrons can be used to express short DNA aptamers within living cells, potentially broadening and optimizing their application in intracellular settings.

Introduction

Aptamers are single-stranded nucleic acids that fold into distinct structures, allowing them to bind specifically to target ligands (Dunn et al., 2017). The ligand-binding sites of aptamers are tailored to the shape and charge of the ligand, enabling precise interactions. In particular, aptamers isolated through in vitro selection protocols serve as affinity reagents, positioning them as nucleic acid analogs of antibodies (Keefe et al., 2010; Qian et al., 2022). As such, they offer distinct advantages over their protein-based counterparts such as small size, as well as rapid and cost-effective manufacturing (Qian et al., 2022). An area of particular interest is the development of fluorescent light-up aptamers (FLAPs), (Neubacher and Hennig, 2019; Zhou and Zhang, 2022; Ouellet, 2016) which specifically bind conditionally fluorescent dyes and activate their fluorescence (Chen et al., 2019). The advent of FLAPs has also significantly advanced the study of folding of aptamers, as efficient folding can be determined as a function of fluorescence output. Pioneering work by Jaffery and coworkers has led to the proliferation of numerous FLAPs. Notable examples include the RNA-FLAPs Spinach (Paige et al., 2011), Broccoli (Filonov et al., 2014), Corn (Song et al., 2017), Mango (Dolgosheina et al., 2014), Squash (Dey et al., 2022), and Pepper (Chen et al., 2019). These fluorogenic RNA aptamers serve as

*For correspondence: hannes.mutschler@tu-dortmund.de

[†]These authors contributed equally to this work

Competing interest: The authors declare that no competing interests exist.

Funding: See page 11

Sent for Review

22 May 2024

Preprint posted

24 May 2024

Reviewed preprint posted

10 July 2024

Reviewed preprint revised

27 January 2026

Version of Record published

03 March 2026

Reviewing Editor: Aaron Frank, Arrakis Therapeutics (United States), United States

© Copyright Vibhute, Machatzke et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

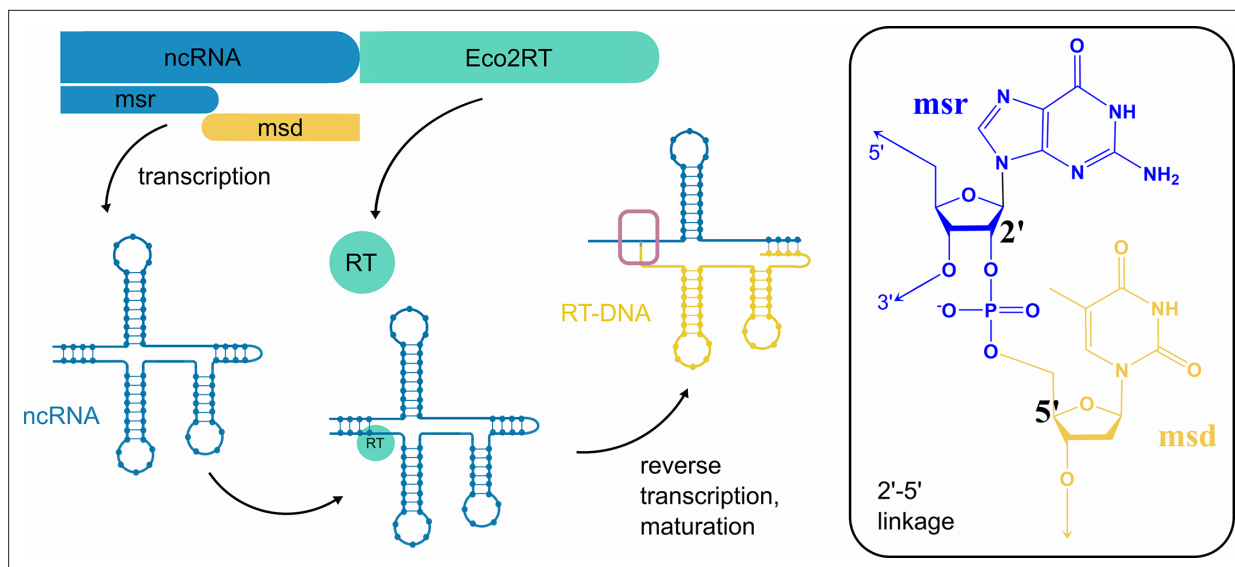


Figure 1. Eco2 retron locus and RT-DNA architecture. Schematic of the Eco2 gene locus, which encodes the non-coding RNA (ncRNA, blue) and the Eco2 reverse transcriptase gene (Eco2RT, green). The *msd* (yellow) region of the ncRNA is reverse transcribed into *msd* DNA, which is linked to the remaining *msr* via a 2'-5' linkage conserved guanosine close to the 3'-end of the *msr*, as indicated in the inset.

RNA mimics of fluorescent proteins and have transformed cellular imaging techniques to locate and understand RNA dynamics in vivo.

In contrast to RNA aptamers, DNA aptamers offer enhanced stability, greater ease of synthesis and chemical modification, as well as reduced cost, thereby facilitating more straightforward production and broader applicability in diverse fields, including therapeutics and diagnostics (Qian *et al.*, 2022; Zhu *et al.*, 2015). For example, Lettuce, a DNA FLAP that mimics the fluorescent properties of GFP, holds great promise for pathogen detection or as intracellular sensors for tumor cell recognition (VarnBuhler *et al.*, 2022; Passalacqua *et al.*, 2023; Mou *et al.*, 2023). However, since most DNA aptamer studies focus on their development and characterization in vitro, their potential for applications inside living cells remains largely unexplored. A particular challenge in this context is the difficulty of biosynthesizing DNA aptamers in cells. Indeed, the ability to transcribe single-stranded RNA from DNA-based vectors has decisively contributed to the development of diverse RNA FLAPs for in-cell applications, such as the Broccoli aptamer (Filonov *et al.*, 2014). Therefore, methods for the intracellular synthesis and functional studies of single-stranded DNA (ssDNA) aptamers are highly desirable. Only a few approaches for intracellular ssDNA synthesis have been reported, in all cases using exogenous systems such as phagemid vectors (Lin *et al.*, 2008), phages (Chen *et al.*, 2004), and eukaryotic retroviral reverse transcriptases (Li *et al.*, 2010; Chen *et al.*, 2003; Chen, 2002; Elbaz *et al.*, 2016; Alon *et al.*, 2020).

A more integrated, efficient, and potentially safer alternative to exogenous ssDNA expression systems could be the use of endogenous bacterial platforms. Retrons are prokaryotic genetic elements that are part of the bacterial phage defense machinery (Palka *et al.*, 2022; Wang *et al.*, 2022; Millman *et al.*, 2020; Bobonis *et al.*, 2020). Each retron typically consists of three components, namely a ncRNA transcript, an RT, and an effector protein. The ncRNA transcript adopts a specific secondary structure, nested between two repeat regions at its 5' and 3' ends that are complementary and hybridize (Figure 1). The RT recognizes the folded ncRNA and initiates reverse transcription from the 2'OH of a conserved guanosine, using the base-paired region as primer (Simon *et al.*, 2019). The reverse transcription proceeds until about halfway along the ncRNA. Simultaneously, the RNA template is degraded by RNase H1, except for a stretch of approximately 5–10 RNA bases, which remain hybridized to the cDNA (Palka *et al.*, 2022). This results in a hybrid RNA-DNA structure (RT-DNA), with the non-coding RNA and DNA elements *msr* and *msd*, respectively. *msr* and *msd* are covalently attached by a 2'-5' linkage at the conserved guanosine and the base paired region close to the 5' end of *msd* (3' end of *msr*). This ability of retrons to synthesize abundant ssDNA in vivo has generated considerable interest in using them as an alternative to exogenously delivered

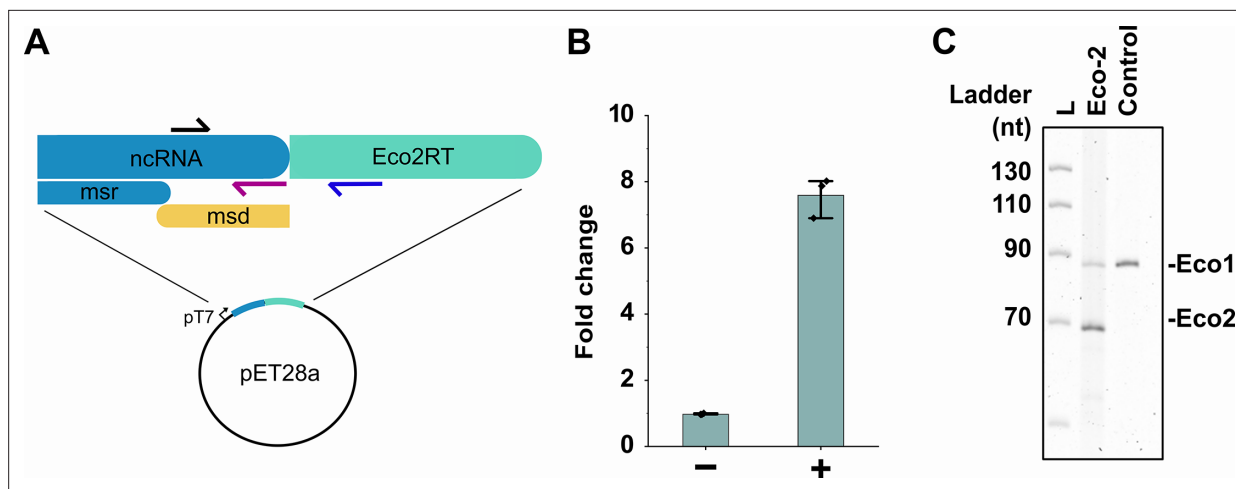


Figure 2. Recombinant expression of Eco2 RT-DNA. **(A)** Illustration of primer target sites for qPCR experiments to determine RT-DNA abundance. The black arrows indicate the forward primer that pairs either with the purple reverse primer (amplify both RT-DNA and plasmid DNA) or the blue reverse primer that only amplified the plasmid DNA. **(B)** Fold enrichment of the RT-DNA/plasmid template over the plasmid alone upon induction, as measured by qPCR; Unpaired t-test, induced versus uninduced: $p < 0.0001$; $n = 3$ biological replicates. **(C)** A TBE-Urea polyacrylamide gel, stained with SYBR Gold showing RT-DNA corresponding to retron Eco1 (90 nt) and retron Eco2 (70 nt).

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Raw data corresponding to **Figure 2B**.

Source data 2. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 2C**.

Source data 3. Original image file of PAGE gel displayed in **Figure 2C**.

Figure supplement 1. Determination of half-life of RT-DNA.

Figure supplement 1—source data 1. Raw data corresponding to **Figure 2—figure supplement 1**.

DNA oligonucleotides in genome engineering and genome editing (Simon et al., 2019; Simon et al., 2018; Ellington and Reisch, 2022; Schubert et al., 2021; Lopez et al., 2022; Lim et al., 2020; Sharon et al., 2018; Liu et al., 2023).

While retrons have also been proposed as a potential method for expressing DNA aptamers (Chen et al., 2003; Elbaz et al., 2016), this approach has not yet been experimentally validated. Recent studies have indicated that retrons can indeed be leveraged to express functional ssDNA in vivo. For example, Lopez et al., 2022 tested modifications to the ncRNA in case of retron Eco1 in order to boost RT-DNA abundance. More recently, Liu et al., 2024 Liu et al. also used retron Eco1 to express 10–23 DNazyme in vivo. However, it is poorly understood how the encoding of a functional ssDNA sequence in retrons will affect its activity. Such studies may lead to valuable design principles that would be especially important in the cellular context, as the RT-DNA is known to interact with not only the RT but, depending on the retron, also RNaseH and the retron associated effector proteins (Palka et al., 2022; Wang et al., 2022).

In this work, we sought to demonstrate expression of a DNA FLAP Lettuce as a model using retron Eco2 (Lampson et al., 1989). We first characterize expression of plasmid-borne wild type Eco2, determining abundance and intracellular stability of Eco2 RT-DNA. Subsequently, we test suitable integration sites of different length variants of Lettuce within the retron *msd* using in vitro prototyping. After identifying effective insertion positions, we finally demonstrate fluorogenic activity of the expressed RT-DNA, embedded within the Lettuce aptamer. Our work demonstrates proof of concept of expressing DNA aptamers using retrons.

Results

We first established an expression system for retron Eco2 by cloning the ncRNA and RT in a pET28a plasmid backbone, under the control of a T7 promoter. The plasmid was transformed in BL21AI cells that enable tight control of T7-based expression. The intracellular synthesis of Eco2 RT-DNA expression was induced with 1 mM IPTG and 0.2% arabinose ('Materials and methods') and confirmed by

denaturing PAGE of DNA extracts from induced cells (**Figure 2**). Using gel-based quantification of both RT-DNAs, we observed that Eco2 RT-DNA expression levels were approximately 3–4 times higher than those of genomically encoded Eco1 RT-DNA. To quantify expression levels, we further characterized Eco2 RT-DNA expression using qPCR (**Figure 2B**). Upon induction with IPTG and arabinose, we observed that the copy-number ratio between amplicon 1 (which can arise from amplification of both single-stranded RT-DNA and plasmid DNA, as shown in **Figure 2A**) and amplicon 2 (which can only be amplified from the plasmid construct) increased by almost eightfold in induced compared to uninduced cells. Assuming an average copy number of 15–20 for the pET28b expression, which utilizes a pBR322 origin (*Jahn et al., 2016*), we estimate that 100–200 single-stranded RT-DNA molecules are present in induced cells. With an average cytoplasmic volume of 7×10^{-16} L of an *Escherichia coli* cell (*Phillips et al., 2012*), this corresponds to an intracellular concentration of approximately 250–500 nM.

Because no data are available on the stability of single-stranded RT-DNA in cells, we next determined the half-life of Eco2 RT-DNA. To this end, induced cells were thoroughly washed to remove residual IPTG and arabinose, thereby preventing de novo expression of RT-DNA. The washed cells were then resuspended in fresh medium and allowed to continue growing. RT-DNA levels were quantified by qPCR at defined time points. The gradual decrease in RT-DNA abundance could thus be attributed to either active degradation (e.g., by intracellular nucleases) or dilution due to cell growth and division. Intriguingly, no evidence for active RT-DNA degradation was observed as the decrease in RT-DNA could be entirely explained by dilution (**Figure 2—figure supplement 1**), indicating exceptional intracellular stability.

Having established the inducible expression of plasmid-encoded Eco2 in *E. coli*, we next sought to explore its suitability as a scaffold to host a functional Lettuce aptamer. Expression of ssDNA using retrons is typically achieved by encoding a cargo gene in the *msd* region of the ncRNA. This approach has been successfully demonstrated for recombineering donors (*Lopez et al., 2022*), protein-binding DNA sequences (*Lee and Kim, 2023*), and DNAzymes (*Liu et al., 2024*). However, it was not clear as to what extent the folding of the cargo aptamer sequence would be affected due to the extensive secondary structure of the Eco2 *msr-msd* RNA-DNA hybrid.

We speculated that the position of the Lettuce aptamer domain in the *msd* region would significantly affect its ability to fold into the functional three-dimensional structure. To test this hypothesis, we used a structure-guided approach to insert different variants of the Lettuce aptamer as cargo into four different positions in the *msd* region of retron Eco2 (**Figure 3A, B**). To identify four suitable insertion sites in the *msd*, we used the minimal free energy structure predicted for the Eco2 wild type *msd* by RNAfold (*Lorenz et al., 2011*) with energy parameters for DNA as guide to identify single stranded and/or loop regions where an insertion of the Lettuce domain was expected to not interfere with the native fold of the remaining *msd* sequence. We omitted the *msr*-overlapping region in the input sequences as it is known to form a duplex with the RTDNA as well as the *msr* RNA, as to the best of our knowledge, there is no algorithm capable of simulating mixed RNA-DNA sequences that are connected via a 2'–5' bond. The positions v1 and v3 are in two different predicted loop regions. The insertion site v2 is located in a single stranded that connects two hairpin loops, whereas position v4 is directly upstream of the predicted base-paired region between the *msd* and *msr* sequences.

The P1 stem of the original lettuce aptamer can be truncated down to four base pairs without significantly affecting its fluorescence output upon DFHBI-1T binding (*VarnBuhler et al., 2022*). We made use of this property in order to minimize the impact of Lettuce insertion on RT-DNA maturation while maintaining the ligand binding activity of Lettuce. Initially, we probed if truncated aptamer variants with P1 stem lengths of 4, 8, and 11 base pairs (designated 4L, 8L, and 11L) along with the full-length aptamer (FL) can still fold after embedding them into the integration sites of the emulated Eco2 RT-DNA.

Using chemically synthesized RT-DNA-Lettuce fusion constructs as prototypes, we first tested the fluorogenic potential of 4L after its insertion into the above-mentioned positions v1 to v4 of the Eco2 RT-DNA (resulting in the constructs 4LEv1 – 4LEv4) using in-gel staining with DFHBI-1T (**Figure 3—figure supplement 1**). We observed only very low levels of in-gel fluorescence compared to a standard full-length Lettuce (FL) or 4L alone, indicating that 4L folding was hampered upon embedding in the retron *msd* scaffold. In contrast, we observed a pronounced fluorescence signal for 8L when it was inserted position v4 (**Figure 3C**), demonstrating that this insertion site was compatible with Lettuce

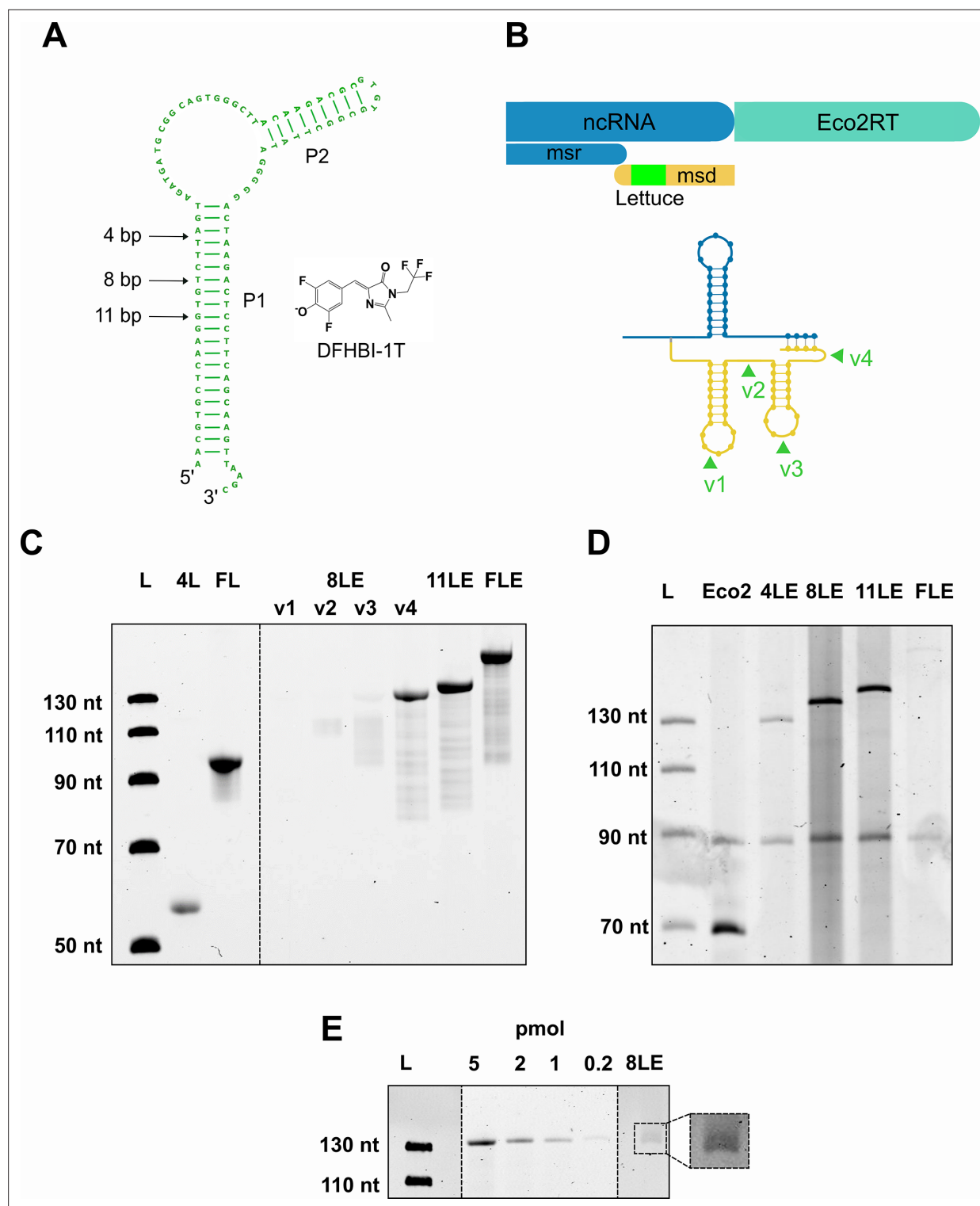


Figure 3. Expression of Lettuce-Eco2 fusion constructs in *E. coli* cells. **(A)** Schematic of Lettuce aptamer. The arrows indicate the length to which the P1 stem can be shortened without significant loss of FLAP functionality. **(B)** Schematic of insertion of Lettuce aptamer sequence in 4 distinct positions in the *msd* region of retron Eco2. The ssDNA structure was simulated using Vienna RNA fold with DNA parameters. Lettuce aptamer sequence was inserted at single-stranded and/or loop regions to minimize interference with the native fold of the *msd* sequence. **(C)** DFHBI-1T stained TBE urea-PAGE showing fluorescence of different oligonucleotides mimicking the Lettuce-Eco2 (LE) fusion constructs with varying P1 lengths: 8LE v1-v4 (134 nt), 11LE (140 nt), and FLE (166 nt). Free Lettuce with either the 4 nt P1 stem (4 L) or full-length P1 (FL) served as positive controls. **(D)** SYBR stained TBE urea-PAGE of extracted Eco2 wild type and Eco2-Lettuce RT-DNA fusions after expression in *E. coli*. In all variants, the lettuce aptamer was inserted into position v4

Figure 3 continued on next page

Figure 3 continued

of the *msd* scaffold. (E) DFHBI-1T stained TBE urea-PAGE showing fluorescence of 8LEv4 RT-DNA purified from *E. coli* cells. Lanes with different loading amounts in pmol of the synthetic RT-DNA standards are shown as comparison.

The online version of this article includes the following source data, source code, and figure supplement(s) for figure 3:

Source data 1. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 3C**.

Source data 2. Original image file of PAGE gel displayed in **Figure 3C**.

Source data 3. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 3D**.

Source data 4. Original image file of PAGE gel displayed in **Figure 3D**.

Source data 5. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 3E**.

Source data 6. Original image file of PAGE gel displayed in **Figure 3E**.

Figure supplement 1. DFBI-1T staining of 4LE oligonucleotides.

Figure supplement 1—source data 1. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 3—figure supplement 1B**.

Figure supplement 1—source data 2. Original image file of PAGE gel displayed in **Figure 3—figure supplement 1B**.

Figure supplement 2. Fluorescence titration experiments using chemically synthesized Lettuce-Eco2 v4 variants.

Figure supplement 2—source data 1. Raw data corresponding to **Figure 3—figure supplement 2**.

Figure supplement 2—source code 1. Source code used for fitting data from **Figure 3—figure supplement 2**.

Figure supplement 3. Detection of extracted RT-DNA constructs.

Figure supplement 3—source data 1. PDF file containing original PAGE gel indicating relevant bands, corresponding to **Figure 3—figure supplement 3**.

Figure supplement 3—source data 2. Original image file of PAGE gel displayed in **Figure 3—figure supplement 3**.

Figure supplement 4. Bulk in vivo fluorescence measurements of 8LEv4 with DFHBI-1T.

Figure supplement 4—source data 1. Raw data corresponding to **Figure 3—figure supplement 4**.

Figure supplement 5. Comparison of fluorogenic activity of Lettuce (DNA) and Broccoli (RNA) aptamers.

Figure supplement 5—source data 1. Raw data corresponding to **Figure 3—figure supplement 5**.

folding given a sufficient length of P1. Similarly, we observed strong in-gel fluorescence signals for 11LEv4 and FLEv4 comparable with native full-length Lettuce.

To cross-validate these results, we determined the dissociation constants between DFHBI-1T and the chemically synthesized Eco2-Lettuce v4 surrogate constructs using the reported Lettuce binding buffer (40 mM HEPES pH 7.5, 100 mM KCl, and 1 mM MgCl₂) (VarnBuhler et al., 2022). When increasing concentrations of each oligonucleotide were incubated in binding buffer containing a constant concentration of DFHBI-1T, we obtained well-defined fluorescence responses for all v4 constructs (**Figure 3—figure supplement 2**). When fitting the binding isotherms with a quadratic association curve, we obtained dissociation constants (K_d) for 8LE, 11LE, and FLE that were comparable to the unembedded full-length aptamer (FL) ~ 0.12 μ M (0.01, 2.46, 95% CI). Binding was weaker for the synthesized ssDNA mimic 4LE-v4, with a K_d of ~ 6 μ M (4.92, 7.18, 95% CI), which was well in agreement with our in-gel fluorescence experiments.

Having confirmed that the v4 surrogate of 8LE, 11LE, and FLE constructs maintained the light-up aptamer properties of the embedded Lettuce sequences, we investigated whether the different length variants could be expressed in *E. coli* after integration into the plasmid-borne *msd* region of retron Eco2. PAGE analysis of RT-DNA from induced *E. coli* BL21-AI expressing the Eco2-Lettuce fusion constructs confirmed complete synthesis of the 8LE (134 nt) and 11LE (140 nt) species, and low-level production of FLE (166 nt), despite their *msd*-encoded RT-DNAs being more than twice as long as the wild-type (67 nt). Overall, the insertion of the Lettuce aptamer domain into the *msd* region of retron Eco2 resulted in a decrease in RT-DNA level compared to wild-type Eco2: While the length-corrected intensity of wild-type Eco2 was threefold higher than that of the endogenous Eco1 band, the 4LE, 8LE, 11LE, and FLE constructs showed between $\sim 25\%$ (FLE) and ~ 65 – 90% (4LE, 8LE, and 11LE) of the length-corrected relative band intensities (**Figure 3D**).

Next, we tested whether heterologously expressed 8LEv4 RT-DNA retains its fluorogenic properties, as it remained unknown whether retron-derived aptamer-DNAs produced under intracellular conditions are functionally equivalent to chemically synthesized ssDNA oligonucleotides. Because

column-based purification could not be scaled efficiently due to the large amounts of plasmid DNA suppressing column binding of the short ssDNAs, we developed a modified TRIzol-based extraction protocol optimized for ssDNA isolation while minimizing co-purification of plasmid DNA. Although the new method yielded considerably cleaner RT-DNA preparations, overall recovery was lower, necessitating the use of larger culture volumes to obtain sufficient material for analysis. The new extraction method was evaluated in pilot experiments for the 8LEv1-4 variants, and PAGE analysis of the isolated RT-DNA from pilot-scale experiments confirmed full-length synthesis of all four position variants (**Figure 3—figure supplement 3**). This approach was then applied to larger volumes for preparative extractions. TBE-urea PAGE analysis with DFHBI-1T staining indeed revealed that the purified, intracellular ssDNA retains fluorogenic activity upon DFHBI-1T binding (**Figure 3E**).

While induced cultures expressing 8LEv4 did not show measurable fluorescence above DFHBI-1T-independent, expression-related autofluorescence (**Figure 3—figure supplement 4**), the intracellular production of a fluorogen-binding RT-DNA supports the idea that retron-encoded ssDNA can, in principle, generate functional aptamers in cells. The limited in vivo signal likely reflects both the moderate RT-DNA abundance (~ 100 – 200 molecules per cell) and the substantially lower fluorogenic efficiency of Lettuce compared with the RNA light-up aptamer Broccoli (~ 100 -fold, **Figure 3—figure supplement 5**), which is optimized for in vivo imaging of RNA (*Filonov et al., 2014*). Nevertheless, the detection of fluorogenic activity in purified material establishes a proof-of-principle that retron-derived ssDNA can be used to display DNA aptamers inside living cells.

Discussion

Bacterial retrons such as Eco2 provide a versatile platform for intracellular reverse transcription-based synthesis of single-stranded DNA. To our knowledge, our present study is the first to show that functional DNA aptamers that recognize small molecule ligands can also be expressed via retrons. Expanding the repertoire of retrons toward DNA aptamer expression promises to significantly expand the repertoire of functional non-coding DNAs and enable the creation of new sensors and regulators with a wide range of intracellular applications. At the same time, retron platforms provide a direct way to tailor DNA aptamers specifically to intracellular conditions. This is particularly important as our data show that both the low expression levels in combination with the moderate fluorogenic activity of Lettuce provide ample room for, for example, evolutionary improvement of retron-encoded DNA aptamer performance under intracellular conditions. We observed a significant decrease in Eco2 expression levels after integration of the Lettuce domain, suggesting that reverse transcription is hindered by the changes in the *msd* sequence (*Lopez et al., 2022*).

In Eco2, the RT is directly fused to a topoisomerase-primase-like (Toprim) effector domain, whose activity causes abortive infection (*Wang et al., 2025; Pausch et al., 2025*). Recent structural studies elucidated the structure of the Eco2 nucleoprotein complex and revealed that each RT-Toprim fusion protein forms a 1:1 complex with the 2'–5' linked DNA-RNA hybrid that further assembles into a trimer (*Wang et al., 2025*). Interestingly, while the *msr* portion of each complex bridges adjacent RT-Toprim monomers and is integral to the structural integrity of the complex, the electron density of the *msd* beyond the stem downstream of the branched 2'–5' linkage remained largely undefined, suggesting a strong degree of flexibility. The proximal *msd* stem is proposed to sterically block the nuclease active site of the Toprim effector domain; upon activation by the phage-encoded ssDNA nuclease DenB from T4-related phages, this inhibition is relieved, enabling Toprim to cleave tRNA and thereby inducing translational shutdown (*Wang et al., 2025; Pausch et al., 2025*). The electron density of the distal *msd* region was largely unresolved in the reported cryo-EM structure, indicating that it is not an integral structural element of Eco2 and likely tolerates mutations or insertions without preventing trimerization. In our study, we indeed observed no major growth defects upon insertion of the aptamer domain at positions v1 to v4, confirming the high degree of tolerance of the *msd* region distal to the 2'–5' branch point. This suggests that alternative functional sequences could be similarly integrated into the distal *msd* segment, provided that flanking *msr* and *msd* regions do not disrupt folding. A major advantage of retron-based intracellular synthesis of ssDNA in vivo is its simplicity. The desired sequence can be cloned into the *msd* region of plasmid-borne or genomically encoded retron Eco2. The copy number of the aptamer can also, in principle, be tuned by inducer concentration, promoter strength, and plasmid copy number. An ever-increasing number of DNA aptamers have been generated against a wide range of targets, from small molecules, macromolecules to whole

cells (Qian et al., 2022; Sharma et al., 2017). The retron system used in this work could enable the exploration of cellular applications for this growing repertoire of DNA aptamers. A further advantage of expressing DNA aptamers with retrons is the exceptional intracellular stability of the *msd*-derived ssDNA in the retron complex (Figure 2—figure supplement 1). Finally, expression of related retron systems has been demonstrated in yeast (Miyata et al., 1992) and mammalian cells (Lopez et al., 2022; Mirochnitchenko et al., 1994), thus potentially widening the range of in vivo applications of DNA aptamers.

In conclusion, retron-based aptamer expression platforms present a promising avenue for the development of novel DNA-based applications. Future research aimed at optimizing intracellular synthesis may elucidate the conditions under which larger single-stranded DNAs can be synthesized at higher concentrations without compromising reverse transcription and may also identify the key factors influencing the optimal in vivo folding of cargo sequences.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Gene (<i>Escherichia coli</i>)	Ec67 (Eco2)	https://doi.org/10.1016/j.cell.2020.09.065	Ec67 (pAA11)	Supplementary file 1a
Strain, strain background (<i>E. coli</i>)	BL21(AI)	Thermo Fisher Scientific	Cat. #: C607003	Chemically competent cells
Chemical compound, drug	Ready-Lyse Lysozyme Solution	Lucigen, Biozym	Cat. #:R1804M	
Chemical compound, drug	TRIzol Reagent	Thermo Fisher Scientific	Cat. #:15596026	
Commercial assay or kit	Luna Universal qPCR Master Mix	New England Biolabs	Cat. #:M3003L	
Commercial assay or kit	Monarch Spin Plasmid Miniprep Kit	New England Biolabs	Cat. #:T1110L	
Chemical compound, drug	DFHBI-1T	Lucerna	Cat. #:410-10mg	Fluorophore
Commercial assay or kit	ssDNA/RNA Clean & Concentrator	Zymo	Cat. #:D7011	
Chemical compound, drug	SYBR Gold Nucleic Acid Gel Stain (10,000X Concentrate in DMSO)	Invitrogen	Cat. #:S11494	

Constructs and strains

The retron Eco2 sequence (*msr-msd* and RT) was obtained from Millman et al., 2020, and purchased from IDT as gblock, which was inserted into a pET28a vector under control of T7 promoter and *lac* operator. Plasmid assembly was performed using the HiFi DNA Assembly (New England Biolabs) according to the supplier’s instructions. The resulting plasmids were transformed into chemically competent *E. coli* Top 10 cells. Plasmid sequences were verified by colony PCR and Sanger sequencing (Microsynth AG). Verified plasmids were then transformed into chemically competent BL21AI cells for further experiments. These cells harbor endogenous retron Eco1, which serves as an internal control and a cassette encoding T7 polymerase driven by arabinose-inducible *araBAD* promoter. Lettuce variant sequences were cloned into retron Eco2 plasmid using Q5 Site-Directed Mutagenesis Kit (New England Biolabs). Primer sequences and annealing temperatures were generated using the NEBaseChanger tool. The sequences of Retron Eco2 (Ec67), oligonucleotides (Eco2-Lettuce constructs), and qPCR primers are listed in Supplementary file 1a,b,c respectively.

RT-DNA extraction

Miniprep method

5 ml LB cultures supplemented with 50 µg/mL kanamycin were inoculated from single colonies of BL21AI cells transformed with the respective pet28b *msd*-Lettuce constructs and grown at 37°C with shaking at 220 rpm for 90 min. Cultures were then induced with 1 mM IPTG and 0.2% arabinose and incubated overnight for expression. The OD₆₀₀ of the overnight cultures was measured, and an amount of culture corresponding to 2 mL at OD₆₀₀=1 was used for miniprep. RT-DNA was extracted using the Monarch Plasmid Miniprep Kit (New England Biolabs) and eluted in 40 µL Milli-Q water. The

eluate was treated with RNase Cocktail Enzyme Mix (Invitrogen) at 37°C for 30 min and subsequently purified using the ssDNA/RNA Clean & Concentrator kit (Zymo).

The purified RT-DNA was analyzed on a 10% TBE-Urea polyacrylamide gel with a 1x TBE running buffer. Gels were stained with SYBR Gold (Invitrogen) and imaged on Sapphire Biomolecular Imager (Azure Biosystems). Band intensities were normalized by length of RT-DNA. RT-DNA abundance of Eco2 and Lettuce variants was quantified by determining fold change compared to the Eco1 band. Band intensity analysis was performed with ImageJ (*Schindelin et al., 2012*).

TRIzol-based extraction method

25 mL pre-cultures supplemented with 50 µg/mL kanamycin were inoculated from single colonies of BL21AI cells transformed with the respective pet28b *msd*-Lettuce constructs and grown at 37°C with shaking at 220 rpm until they reached an OD₆₀₀ of 0.5–0.8. These pre-cultures were used to inoculate 750 mL cultures, which were then grown at 37°C and 150 rpm for 2 h. The cultures were then induced with IPTG (1 mM) and arabinose (0.2%) and allowed to grow overnight (~16 h) until they reached an OD₆₀₀ 2.4–2.9 and harvested by centrifugation at 4°C and 6000 × g for 30 min. The cell pellets were resuspended in TES buffer and lysozyme (Ready-Lyse Lysozyme Solution, Biozym) was added according to the manufacturer's instructions (25 mL TES +4.5 µL 30,000 U/µL Lysozyme). Samples were incubated for 15 min at room temperature and then aliquoted into twelve 50 mL conical tubes. TRIzol (20 mL) was added to each tube, and samples were gently vortexed until clear. Chloroform (4 mL) was then added, and samples were vortexed until homogeneous, followed by incubation for 5 min at room temperature. Phase separation was achieved by centrifugation for 30 min at 4°C and 12,000 × g. The aqueous phase (10 mL) was transferred to a fresh tube, mixed with 10 mL isopropanol, vortexed, and incubated for 2 h at –20°C. Nucleic acids were pelleted by centrifugation at 30,000 × g and 4°C for 1 h. The supernatant was removed, and the pellet was washed twice with 10 mL of 70% ethanol, each time centrifuging for 5 min at 4°C and 10,000 × g. Pellets were air-dried for 15 min and resuspended in 1000 µL nuclease-free water.

In-gel staining of the RT-DNA

Following the crude extraction of chimeric RT-DNA/RNA using the above-described TRIzol-based method, 650 µL of the total RNA/retron material (1000 µL) was incubated with 250 mM NaOH and 6.25 mM EDTA at 80°C for 1 h. After incubation, the pH was neutralized by the addition of 250 mM HCl, after which the now RNA-free, single-stranded RT-DNA was purified using the ssDNA/RNA Clean & Concentrator Kit (Zymo Research) following the manufacturer's instructions. Typically, a total of three columns was used for ssDNA isolation with an elution volume of 9 µL of nuclease-free H₂O. The eluate from column 1 was used to elute ssDNA from columns 2 and 3 to enhance the total RT-DNA concentration. The entire sample (approximately 7.5 µL) was mixed with 15 µL of loading dye (10 mM EDTA in formamide) and heated to 95°C for 3 min, followed by immediate incubation on ice for 5 min. The sample was loaded into a single well of a 10% TBE-urea PAGE gel and electrophoresed for 80 min at constant power (20 W). The gel was washed twice with ddH₂O to remove residual urea and then stained for 30 min in 50 mL DFHBI-1T staining solution (40 mM HEPES, pH 7.5, 100 mM KCl, 1 mM MgCl₂, 10 µM DFHBI-1T). After a final 5 min wash in ddH₂O to remove excess dye, the gel was imaged on a Sapphire Bioimager using 488/518 nm for DFHBI-1T fluorescence and 520/565 nm for the Cy3-tagged ladder.

qPCR

qPCR experiments were performed using Techne Prime Pro 48 Real-Time qPCR system, according to comparative C_T method, (*Livak and Schmittgen, 2001; Schmittgen and Livak, 2008*) as described by *Lopez et al., 2022*. For qPCR analysis, cell cultures grown for ~16 h were used. 25 µL of bacterial culture at an OD₆₀₀=1 was mixed with 25 µL water and incubated at 95°C for 5 min to lyse cells. 0.3 µL of this boiled culture was used as templates in 30 µL qPCR reactions using Luna Universal qPCR Master Mix (New England Biolabs). qPCR quantification of RT-DNA abundance was performed using a set of three primers. Two primers were complementary to the *msd* region of Retron Eco2, and the third was an outnest primer complementary to non-retron elements of the expression plasmid. Two templates were thus amplified: one corresponding to an RT-DNA sequence that could use both the RT-DNA and the plasmid as templates, and a second that could use only the plasmid-encoded Eco2 as

a template. The difference between cycle threshold (ΔC_T) of the inside (RT-DNA) and outside (plasmid) primer sets was then calculated. This ΔC_T was then subtracted from the ΔC_T of uninduced samples. Fold enrichment of RT-DNA/plasmid over plasmid alone was then calculated as $2^{-\Delta\Delta C_T}$, for each biological replicate. Presence of RT-DNA results in a fold change >1 .

Equilibrium binding measurements

Equilibrium binding measurements were performed on a BMG Labtech Clariostar plus plate reader. Oligonucleotides corresponding to different variants of Lettuce embedded in the retron scaffold were synthesized by IDT (**Supplementary file 1b**). Sample fluorescence was measured in a Greiner 384-well black-bottom plate with excitation at 470 ± 8 nm and emission at 515 ± 20 nm. The samples were prepared by mixing 40 mM HEPES pH 7.4, 100 mM KCl, 1 mM $MgCl_2$ with different concentrations of DNA. Samples were then heated to $90^\circ C$ for 2 min, followed by incubating at room temperature for 5 min. After that 1 μM of DFHBI-1T was added to the mix. The samples were then briefly vortexed and spun down and pipetted in triplicate into a 384-well plate. Samples were then incubated for 60 min at RT in darkness, followed by fluorescence measurements on the microplate reader.

The datasets were fitted using **Equation 1**:

$$F = C_1 \cdot \frac{(A_t + B_t + k_D) - \sqrt{(A_t + B_t + k_D)^2 - 4 \cdot (A_t \cdot B_t)}}{2 \cdot B_t} + C_0 \quad (1)$$

where F is the measured fluorescence, A_t is the concentration of DNA, B_t is the concentration of DFHBI-1T, and C_0 and C_1 are the lower and upper fluorescence plateaus. To account for uncertainty in the DFHBI-1T stock concentration, B_t was also fitted under strong constraints. Curve fitting was performed using the `least_squares` function from the `scipy.optimize` package (<https://docs.scipy.org/doc/scipy/reference/optimize.html>) with shared global parameters C_0 , C_1 , and B_t for all data series.

Determination of half-life of Eco2 RT-DNA

Retron RT-DNA forms a phage surveillance complex with the associated RT and effector protein (**Rousset and Sorek, 2023; Gao et al., 2020; Carabias et al., 2024; Wang et al., 2024**). Owing to its unique 'closed' structure (**Wang et al., 2022**) (with the ends of *msr* and *msd* joined by a 2'-5' linkage and a base-paired region) and its non-coding function, the intracellular stability of RT-DNA is of particular interest. To assess the stability of Eco2 RT-DNA, we determined its half-life by qPCR.

We first induced retron Eco2 RT-DNA expression in BL21AI cells with 1 mM IPTG and 0.2% arabinose overnight. On the following day, cells were pelleted and washed twice with LB medium to remove residual IPTG and arabinose. The washed cells were then resuspended and used to inoculate a fresh culture in LB medium without inducers at an initial OD_{600} of 0.2. The culture was grown at $37^\circ C$, and aliquots were taken at defined timepoints. For each time point, a cell suspension with $OD_{600}=0.1$ was prepared and incubated at $95^\circ C$ for 5 min. 1 μL of this boiled, diluted culture was used as input for a 20 μL qPCR reaction. qPCR experiments were performed as described above.

Assuming RT-DNA degradation would occur by active degradation mechanisms (nuclease-mediated degradation) and dilution (cell growth and division), we determined the rate of degradation by the following equation:

$$C(t) = 1 + A \cdot e^{-kt} \frac{OD_{600}(t_0)}{OD_{600}(t)} \quad (2)$$

where k is the degradation rate constant and the ratio $\frac{OD_{600}(t_0)}{OD_{600}(t)}$ is the dilution factor accounting for dilution caused by cell division. $OD_{600}(t)$ was determined by fitting the OD_{600} values to the following equation describing logistic growth (**Figure 2—figure supplement 1A**):

$$OD_{600}(t) = \frac{OD_{max}}{1 + e^{-r \cdot (t-t_0)}} \quad (3)$$

After substituting $OD_{600}(t)$ with the function in **Equation 3** we fitted the experimental data for the fold change of the RT-DNA to **Equation 2** (**Figure 2—figure supplement 1B**). The best fit (red line) was obtained with a rate constant k converging towards zero, consistent with an Eco2 RT-DNA

half-life that is beyond the detection limit of our assay. To put our estimates into context, we note that typical RNA half-lives are on the order of minutes in growing *E. coli* cells (Milo and Phillips, 2016). We therefore refitted the data using fixed half-lives of 15 and 30 min. In both cases, the resulting best-fit curves deviated substantially from the experimental data, consistent with an RT-DNA half-life that is far longer than the *E. coli* doubling time under these optimal growth conditions. While we cannot exclude ongoing RT-DNA production due to promoter leakiness, we expect this effect to be minor, as RT-DNA expression requires both IPTG and arabinose, which were removed prior to inoculating the growth medium with the starter culture. Overall, our data support exceptional RT-DNA stability.

Bulk live-cell fluorescence measurements

5 mL LB cultures supplemented with 50 µg/mL kanamycin were inoculated from single colonies of BL21AI cells transformed with the respective pet28b *msd*-Lettuce constructs and grown at 37°C with shaking at 220 rpm for 90 min. Cultures were induced with 1 mM IPTG and 0.2% arabinose, followed by growth for ~16 h. A culture volume corresponding to 5 mL at OD₆₀₀=1 was pelleted and washed twice with 1 mL 1× PBS. The pellet was then resuspended in 150 µL 1× PBS and 40 µM DFHBI-1T was added to the cell suspension, followed by incubation at room temperature for 20 min. The stained cells were then used for plate reader measurements (3 × 50 µL in a 96-well plate). Fluorescence was measured on a BMG Labtech Clariostar Plus plate reader (Ex/Em 470/520 nm). To correct for pipetting and cell-density variation, cultures were diluted in 1× PBS and OD₆₀₀ was determined after fluorescence acquisition. OD₆₀₀ values were measured on a NanoDrop using a 25-fold dilution and used to normalize the fluorescence signal.

Acknowledgements

The authors thank Indrayani Phadtare and members of the Mutschler lab for fruitful discussions. HM acknowledges support by the European Research Council (ERC) under the Horizon 2020 research and innovation program (grant agreement ID: 802000, RiboLife).

Additional information

Funding

Funder	Grant reference number	Author
European Research Council	802000, RiboLife	Hannes Mutschler

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Mahesh A Vibhute, Conceptualization, Resources, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – original draft, Writing – review and editing; Corbin Machatzke, Resources, Data curation, Software, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing – review and editing; Saskia Krümpel, Resources, Investigation; Malte Dirks, Resources, Investigation, Methodology; Katrin Bigler, Resources, Methodology; Daniel Summerer, Conceptualization, Validation; Hannes Mutschler, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Visualization, Methodology, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Mahesh A Vibhute  <https://orcid.org/0000-0002-5230-5290>

Hannes Mutschler  <https://orcid.org/0000-0001-8005-1657>

Peer review material

Reviewer #1 (Public review): <https://doi.org/10.7554/eLife.99554.3.sa1>

Reviewer #2 (Public review): <https://doi.org/10.7554/eLife.99554.3.sa2>

Author response <https://doi.org/10.7554/eLife.99554.3.sa3>

Additional files

Supplementary files

Supplementary file 1. Nucleic acid sequences used in this work.

MDAR checklist

Data availability

All data generated or analyzed during this study are included in the manuscript and supporting files; source data files have been provided for figures.

References

- Alon DM**, Voigt CA, Elbaz J. 2020. Engineering a DNAzyme-based operon system for the production of DNA nanoscaffolds in living bacteria. *ACS Synthetic Biology* **9**:236–240. DOI: <https://doi.org/10.1021/acssynbio.9b00415>, PMID: 31992039
- Bobonis J**, Mateus A, Pfalz B, Garcia-Santamarina S, Galardini M, Kobayashi C, Stein F, Savitski MM, Elfenbein JR, Andrews-Polymenis H, Typas A. 2020. Bacterial retrons encode tripartite toxin/antitoxin systems. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.06.22.160168>
- Carabias A**, Camara-Wilpert S, Mestre MR, López-Méndez B, Hendriks IA, Zhao R, Pape T, Fuglsang A, Luk SHC, Nielsen ML, Pinilla-Redondo R, Montoya G. 2024. Retron-Eco1 assembles NAD⁺-hydrolyzing filaments that provide immunity against bacteriophages. *Molecular Cell* **84**:2185–2202. DOI: <https://doi.org/10.1016/j.molcel.2024.05.001>, PMID: 38788717
- Chen Y**. 2002. A novel single-stranded DNA expression vector. *Expert opinion on biological therapy* **2**:735–740. DOI: <https://doi.org/10.1517/14712598.2.7.735>, PMID: 12387672
- Chen Y**, Ji YJ, Conrad C. 2003. Expression of ssDNA in mammalian cells. *BioTechniques* **34**:167–171. DOI: <https://doi.org/10.2144/03341dd04>, PMID: 12545555
- Chen F**, Wang R, Li Z, Liu B, Wang X, Sun Y, Hao D, Zhang J. 2004. A novel replicating circular DNAzyme. *Nucleic Acids Research* **32**:2336–2341. DOI: <https://doi.org/10.1093/nar/gkh547>, PMID: 15115797
- Chen X**, Zhang D, Su N, Bao B, Xie X, Zuo F, Yang L, Wang H, Jiang L, Lin Q, Fang M, Li N, Hua X, Chen Z, Bao C, Xu J, Du W, Zhang L, Zhao Y, Zhu L, et al. 2019. Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature Biotechnology* **37**:1287–1293. DOI: <https://doi.org/10.1038/s41587-019-0249-1>, PMID: 31548726
- Dey SK**, Filonov GS, Olarerin-George AO, Jackson BT, Finley LWS, Jaffrey SR. 2022. Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature Chemical Biology* **18**:180–190. DOI: <https://doi.org/10.1038/s41589-021-00925-0>, PMID: 34937909
- Dolgoshchina EV**, Jeng SCY, Panchapakesan SSS, Cojocaru R, Chen PSK, Wilson PD, Hawkins N, Wiggins PA, Unrau PJ. 2014. RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS Chemical Biology* **9**:2412–2420. DOI: <https://doi.org/10.1021/cb500499x>, PMID: 25101481
- Dunn MR**, Jimenez RM, Chaput JC. 2017. Analysis of aptamer discovery and technology. *Nature Reviews Chemistry* **1**:1–16. DOI: <https://doi.org/10.1038/s41570-017-0076>
- Elbaz J**, Yin P, Voigt CA. 2016. Genetic encoding of DNA nanostructures and their self-assembly in living bacteria. *Nature Communications* **7**:11179. DOI: <https://doi.org/10.1038/ncomms11179>, PMID: 27091073
- Ellington AJ**, Reisch CR. 2022. Efficient and iterative retron-mediated in vivo recombineering in *Escherichia coli*. *Synthetic Biology* **7**:7. DOI: <https://doi.org/10.1093/synbio/ysac007>
- Filonov GS**, Moon JD, Svensen N, Jaffrey SR. 2014. Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* **136**:16299–16308. DOI: <https://doi.org/10.1021/ja508478x>
- Gao L**, Altae-Tran H, Böhning F, Makarova KS, Segel M, Schmid-Burgk JL, Koob J, Wolf YI, Koonin EV, Zhang F. 2020. Diverse enzymatic activities mediate antiviral immunity in prokaryotes. *Science* **369**:1077–1084. DOI: <https://doi.org/10.1126/science.aba0372>, PMID: 32855333
- Jahn M**, Vorpahl C, Hübschmann T, Harms H, Müller S. 2016. Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. *Microbial Cell Factories* **15**:211. DOI: <https://doi.org/10.1186/s12934-016-0610-8>, PMID: 27993152
- Keefe AD**, Pai S, Ellington A. 2010. Aptamers as therapeutics. *Nature Reviews. Drug Discovery* **9**:537–550. DOI: <https://doi.org/10.1038/nrd3141>, PMID: 20592747
- Lampson BC**, Sun J, Hsu MY, Vallejo-Ramirez J, Inouye S, Inouye M. 1989. Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* **243**:1033–1038. DOI: <https://doi.org/10.1126/science.2466332>, PMID: 2466332
- Lee G**, Kim J. 2023. Engineered retrons generate genome-independent protein-binding DNA for cellular control. *bioRxiv*. DOI: <https://doi.org/10.1101/2023.09.27.556556>
- Li J**, Wang N, Luo Q, Wan L. 2010. The 10-23 DNA enzyme generated by a novel expression vector mediate inhibition of taco expression in macrophage. *Oligonucleotides* **20**:61–68. DOI: <https://doi.org/10.1089/oli.2009.0217>, PMID: 20059315
- Lim H**, Jun S, Park M, Lim J, Jeong J, Lee JH, Bang D. 2020. Multiplex generation, tracking, and functional screening of substitution mutants using a CRISPR/Retron system. *ACS Synthetic Biology* **9**:1003–1009. DOI: <https://doi.org/10.1021/acssynbio.0c00002>

- Lin C, Rinker S, Wang X, Liu Y, Seeman NC, Yan H. 2008. In vivo cloning of artificial DNA nanostructures. *PNAS* **105**:17626–17631. DOI: <https://doi.org/10.1073/pnas.0805416105>
- Liu W, Zuo S, Shao Y, Bi K, Zhao J, Huang L, Xu Z, Lian J. 2023. Retron-mediated multiplex genome editing and continuous evolution in *Escherichia coli*. *Nucleic Acids Research* **51**:8293–8307. DOI: <https://doi.org/10.1093/nar/gkad607>
- Liu J, Cui L, Shi X, Yan J, Wang Y, Ni Y, He J, Wang X. 2024. Generation of DNAzyme in bacterial cells by a bacterial retron system. *ACS Synthetic Biology* **13**:300–309. DOI: <https://doi.org/10.1021/acssynbio.3c00509>
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**:402–408. DOI: <https://doi.org/10.1006/meth.2001.1262>, PMID: 11846609
- Lopez SC, Crawford KD, Lear SK, Bhattarai-Kline S, Shipman SL. 2022. Precise genome editing across kingdoms of life using retron-derived DNA. *Nature Chemical Biology* **18**:199–206. DOI: <https://doi.org/10.1038/s41589-021-00927-y>, PMID: 34949838
- Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**:26. DOI: <https://doi.org/10.1186/1748-7188-6-26>, PMID: 22115189
- Millman A, Bernheim A, Stokar-Aviail A, Fedorenko T, Voichek M, Leavitt A, Oppenheimer-Shaanan Y, Sorek R. 2020. Bacterial retrons function in anti-phage defense. *Cell* **183**:1551–1561. DOI: <https://doi.org/10.1016/j.cell.2020.09.065>
- Milo R, Phillips R. 2016. *Cell Biology by the Numbers*. New York NY: Garland Science Taylor & Francis Group.
- Mirochnitchenko O, Inouye S, Inouye M. 1994. Production of single-stranded DNA in mammalian cells by means of a bacterial retron. *The Journal of Biological Chemistry* **269**:2380–2383 PMID: 7507924.
- Miyata S, Ohshima A, Inouye S, Inouye M. 1992. In vivo production of a stable single-stranded cDNA in *Saccharomyces cerevisiae* by means of a bacterial retron. *PNAS* **89**:5735–5739. DOI: <https://doi.org/10.1073/pnas.89.13.5735>, PMID: 1378616
- Mou Y, Yin P, Chen M, Wei C, Zhang Y, Zhang J, Zhao Y, Luo X, Wang Y. 2023. Engineering of an aptamer-functionalized fluorescent DNA Sensor for Cu(II) responding in living tumor cells. *Analytical Chemistry* **95**:8348–8356. DOI: <https://doi.org/10.1021/acs.analchem.3c01008>, PMID: 37194189
- Neubacher S, Hennig S. 2019. RNA structure and cellular applications of fluorescent light-up aptamers. *Angewandte Chemie* **58**:1266–1279. DOI: <https://doi.org/10.1002/anie.201806482>, PMID: 30102012
- Ouellet J. 2016. RNA fluorescence with light-up aptamers. *Frontiers in Chemistry* **4**:29. DOI: <https://doi.org/10.3389/fchem.2016.00029>, PMID: 27446908
- Paige JS, Wu KY, Jaffrey SR. 2011. RNA mimics of green fluorescent protein. *Science* **333**:642–646. DOI: <https://doi.org/10.1126/science.1207339>, PMID: 21798953
- Palka C, Fishman CB, Bhattarai-Kline S, Myers SA, Shipman SL. 2022. Retron reverse transcriptase termination and phage defense are dependent on host RNase H1. *Nucleic Acids Research* **50**:3490–3504. DOI: <https://doi.org/10.1093/nar/gkac177>, PMID: 35293583
- Passalacqua LFM, Banco MT, Moon JD, Li X, Jaffrey SR, Ferré-D'Amaré AR. 2023. Intricate 3D architecture of a DNA mimic of GFP. *Nature* **618**:1078–1084. DOI: <https://doi.org/10.1038/s41586-023-06229-8>, PMID: 37344591
- Pausch P, Jasnauskaitė M, Juozapaitis J, Liegutė T, Grigaitis R, Skorupskaitė A, Steinchen W, Mikšys A, Truncaitė L, Kazlauskaitė K, Jiménez MT, Khochar S, Dudas G, Bange G, Malinauskaitė L, Songailienė I. 2025. Retron Eco2 breaks tRNAs for antiphage defense. *Research Square*. DOI: <https://doi.org/10.21203/rs.3.rs-7196019/v1>
- Phillips R, Kondev J, Theriot J, Garcia HG, Orme N. 2012. *Physical Biology of the Cell*. Garland Science.
- Qian S, Chang D, He S, Li Y. 2022. Aptamers from random sequence space: accomplishments, gaps and future considerations. *Analytica Chimica Acta* **1196**:339511. DOI: <https://doi.org/10.1016/j.aca.2022.339511>, PMID: 35151405
- Rousset F, Sorek R. 2023. The evolutionary success of regulated cell death in bacterial immunity. *Current Opinion in Microbiology* **74**:102312. DOI: <https://doi.org/10.1016/j.mib.2023.102312>, PMID: 37030143
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nature Methods* **9**:676–682. DOI: <https://doi.org/10.1038/nmeth.2019>, PMID: 22743772
- Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols* **3**:1101–1108. DOI: <https://doi.org/10.1038/nprot.2008.73>, PMID: 18546601
- Schubert MG, Goodman DB, Wannier TM, Kaur D, Farzadfard F, Lu TK, Shipman SL, Church GM. 2021. High-throughput functional variant screens via in vivo production of single-stranded DNA. *PNAS* **118**:e2018181118. DOI: <https://doi.org/10.1073/pnas.2018181118>, PMID: 33906944
- Sharma TK, Bruno JG, Dhiman A. 2017. ABCs of DNA aptamer and related assay development. *Biotechnology Advances* **35**:275–301. DOI: <https://doi.org/10.1016/j.biotechadv.2017.01.003>, PMID: 28108354
- Sharon E, Chen SAA, Khosla NM, Smith JD, Pritchard JK, Fraser HB. 2018. Functional genetic variants revealed by massively parallel precise genome editing. *Cell* **175**:544–557. DOI: <https://doi.org/10.1016/j.cell.2018.08.057>, PMID: 30245013
- Simon AJ, Morrow BR, Ellington AD. 2018. Retroelement-based genome editing and evolution. *ACS Synthetic Biology* **7**:2600–2611. DOI: <https://doi.org/10.1021/acssynbio.8b00273>, PMID: 30256621

- Simon AJ**, Ellington AD, Finkelstein IJ. 2019. Retrons and their applications in genome engineering. *Nucleic Acids Research* **47**:11007–11019. DOI: <https://doi.org/10.1093/nar/gkz865>, PMID: 31598685
- Song W**, Filonov GS, Kim H, Hirsch M, Li X, Moon JD, Jaffrey SR. 2017. Imaging RNA polymerase III transcription using a photostable RNA-fluorophore complex. *Nature Chemical Biology* **13**:1187–1194. DOI: <https://doi.org/10.1038/nchembio.2477>, PMID: 28945233
- VarnBuhler BS**, Moon J, Dey SK, Wu J, Jaffrey SR. 2022. Detection of SARS-CoV-2 RNA using a DNA aptamer mimic of green fluorescent protein. *ACS Chemical Biology* **17**:840–853. DOI: <https://doi.org/10.1021/acscchembio.1c00893>, PMID: 35341244
- Wang Y**, Guan Z, Wang C, Nie Y, Chen Y, Qian Z, Cui Y, Xu H, Wang Q, Zhao F, Zhang D, Tao P, Sun M, Yin P, Jin S, Wu S, Zou T. 2022. Cryo-EM structures of *Escherichia coli* Ec86 retron complexes reveal architecture and defence mechanism. *Nature Microbiology* **7**:1480–1489. DOI: <https://doi.org/10.1038/s41564-022-01197-7>
- Wang Y**, Wang C, Guan Z, Cao J, Xu J, Wang S, Cui Y, Wang Q, Chen Y, Yin Y, Zhang D, Liu H, Sun M, Jin S, Tao P, Zou T. 2024. DNA methylation activates retron Ec86 filaments for antiphage defense. *Cell Reports* **43**:114857. DOI: <https://doi.org/10.1016/j.celrep.2024.114857>
- Wang Y**, Wang C, Yin Y, Cui Y, Dai Z, Liu C, Chen Y, Guan Z, Zou T. 2025. Structural basis of the RNA-mediated Retron-Eco2 oligomerization. *Cell Discovery* **11**:73. DOI: <https://doi.org/10.1038/s41421-025-00823-y>, PMID: 40897710
- Zhou H**, Zhang S. 2022. Recent development of fluorescent light-up RNA aptamers. *Critical Reviews in Analytical Chemistry* **52**:1644–1661. DOI: <https://doi.org/10.1080/10408347.2021.1907735>, PMID: 33870782
- Zhu Q**, Liu G, Kai M. 2015. DNA aptamers in the diagnosis and treatment of human diseases. *Molecules* **20**:20979–20997. DOI: <https://doi.org/10.3390/molecules201219739>, PMID: 26610462

Publications

9.2 Publication 2

Regulating nucleic acid catalysis using active droplets

Holtmannspötter, A.L.*; **Machatzke, C.***; Begemann, C.; Salibi, E.; Donau, C.; Späth, F.; Boekhoven, J.⁺; Mutschler, H.⁺

* Authors contributed equally

⁺ Corresponding author

Angewandte Chemie International Edition **2024**, 63(47), e202412534

DOI: 10.1002/anie.202412534

9.2.1 Personal contribution

- Design of the majority of oligonucleotides used in the study
- Designed and performed all DNzyme activity assays and kinetics
- Performed measurement of partitioning coefficient via gel electrophoresis
- Preparation of Schemes and Figures together with A.-L. Holtmannspötter and J. Boekhoven
- Composed, edited and revised the manuscript and supporting information together with A.-L. Holtmannspötter, J. Boekhoven and H. Mutschler

Synthetic Biology

Regulating Nucleic Acid Catalysis Using Active Droplets

Anna-Lena Holtmannspötter⁺, Corbin Machatzke⁺, Christian Begemann, Elia Salibi, Carsten Donau, Fabian Späth, Job Boekhoven,^{*} and Hannes Mutschler^{*}

Abstract: Cells use transient membraneless organelles to regulate biological reaction networks. For example, stress granules selectively store mRNA to downregulate protein expression in response to heat or oxidative stress. Models mimicking this active behavior should be established to better understand in vivo regulation involving compartmentalization. Here we use active, complex coacervate droplets as a model for membraneless organelles to spatiotemporally control the activity of a catalytic DNA (DNAzyme). Upon partitioning into these peptide-RNA droplets, the DNAzyme unfolds and loses its ability to catalyze the cleavage of a nucleic acid strand. We can transiently pause the DNAzyme activity upon inducing droplet formation with fuel. After fuel consumption, the DNAzyme activity autonomously restarts. We envision this system could be used to up and downregulate multiple reactions in a network, helping understand the complexity of a cell's pathways. By creating a network where the DNAzyme could reciprocally regulate the droplet properties, we would have a powerful tool for engineering synthetic cells.

Introduction

Membraneless compartmentalization is a key element of cellular organization^[1] and offers a tool to spatially and temporally separate reaction pathways in its complex reaction networks.^[1b,2] Typically, these membraneless organelles comprise RNA and intrinsically disordered proteins held together by entropic effects and electrostatic interactions.^[3] Spatial separation through membraneless

organelles can occur in response to an external trigger such as stress, resulting in the selective removal of specific RNA, thereby stopping the transcription of those and the activity of their encoded downstream proteins.^[4]

Model systems of these biological condensates help us to understand the complexity of such chemical reaction networks. One of such models are the droplets formed by complex coacervates—a type of liquid phase separation typically composed of oppositely charged polyions.^[5] Several studies on coacervates in combination with reaction cycles and biocatalysts like enzymes, ribozymes, or DNAzymes exist to test how droplets regulate chemical activity.^[6] Most of these studies focused on enhancing reaction rates by coacervation, but some inhibitory systems were also described.^[7] Transient inhibition, as seen in stress granules, has not yet been achieved. Given the non-equilibrium nature of the cell, most membraneless organelles are also active and transient.^[1a,8] However, the models combining coacervates and biocatalysts typically operate close to equilibrium, lacking this active aspect of membraneless organelles.

Inspired by the active environment of the cell, chemically fueled, active droplets have been developed (Scheme 1A).^[9] In these systems, a (photo-) chemical reaction cycle regulates the ability of building blocks to phase-separate.^[10] Specifically, a molecule with high chemical potential (“fuel”) reacts with the building blocks, which activates them for phase separation (“activation”). A spontaneous back-reaction reverts the activated building block to its initial state (“deactivation”). This means that when building blocks are supplied with fuel, a population of transiently activated building blocks emerges, resulting in active droplets. Moreover, when a finite amount of fuel is supplied, the system will temporarily produce droplets until all fuel is consumed.

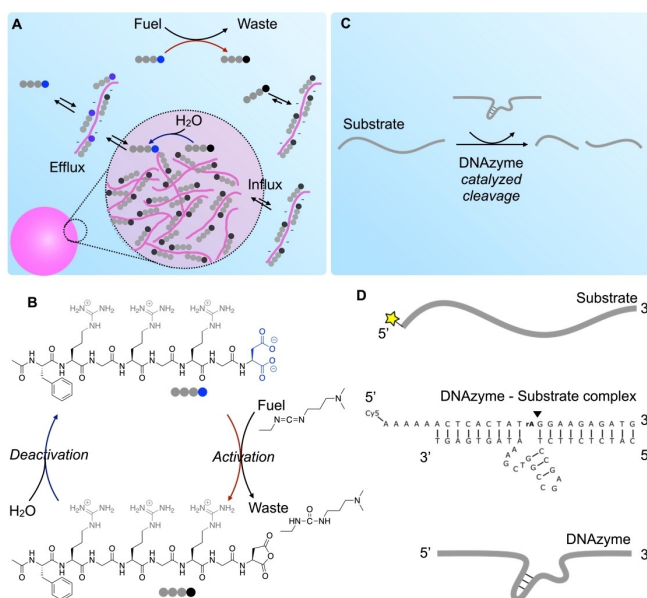
Such active droplets are a class of chemically fueled assemblies,^[11] and their fuel-dependent nature makes them an ideal model to test how transient compartmentalization affects biocatalysts in reaction networks. In this context, active assemblies have been used to host and modulate reactions and model complex biological systems. For example, chemical fuels have been used to generate vesicles, which locally increase the concentration of hydrophobic reagents, thereby increasing the rate of a Diels-Alder reaction.^[12] Alternatively, vesicle formation that contains DNA cargo is transiently controlled by UV light, mimicking the formation of membraneless organelles. Other works tackle the impact of non-equilibrium systems on catalytic activity using agarose gel-based systems.^[13] However, active coacervate droplets that control chemical reactions have yet

[*] A.-L. Holtmannspötter,⁺ C. Donau, F. Späth, J. Boekhoven
 Department of Bioscience, School of Natural Sciences, Technical
 University of Munich, Lichtenbergstrasse 4, 85748 Garching,
 Germany
 E-mail: job.boekhoven@tum.de

C. Machatzke,⁺ C. Begemann, E. Salibi, H. Mutschler
 Department of Chemistry and Chemical Biology, TU Dortmund
 University, Otto-Hahn-Strasse 4a, 44227 Dortmund, Germany
 E-mail: hannes.mutschler@tu-dortmund.de

[†] These authors contributed equally to this work.

© 2024 The Authors. Angewandte Chemie International Edition
 published by Wiley-VCH GmbH. This is an open access article under
 the terms of the Creative Commons Attribution License, which
 permits use, distribution and reproduction in any medium, provided
 the original work is properly cited.



to be studied despite their potential as a model for membraneless organelles.

In this work, we test how a dynamic environment of active droplets can regulate the activity of a biocatalyst, specifically DNAzymes. DNAzymes are catalytically active DNA oligomers and a versatile model for biocatalytic reactions due to their robustness and stability compared to catalytic RNAs or enzymes.^[14] As no naturally occurring DNAzymes exist, all known ones were isolated during *in vitro* selection experiments.^[15] We spatio-temporally control the catalytic activity of a short DNAzyme using active droplets. These active droplets emerge and decay in the presence of fuel. Upon partitioning in the active droplets, the DNAzyme unfolds and loses its ability to catalyze the cleavage of an RNA substrate. By adding fuel, we reversibly halt the DNAzyme activity by temporal partitioning into the droplets. Moreover, we show that this “on-and-off” activity can be repeated multiple times. We envision that such a system could create reaction networks that rely on the spatiotemporal separation and inactivation of catalyst and substrate.

Results and Discussion

System Design and DNAzyme Partitioning

The active droplets are based on a complex coacervate between a long RNA polyanion (polyU, 2200 monomer units) and a short cationic peptide (Scheme 1A).^[9a] The formation of droplets is regulated by a chemical reaction cycle that consumes EDC as a chemical fuel.^[16] Specifically, the C-terminus of peptide (Ac-F(RG)₃D-OH) reacts with the fuel, resulting in its corresponding anhydride (Activation, Scheme 1B). This activation increases the overall charge of the peptide from +1 to +3. The anhydride spontaneously hydrolyzes to the original peptide (Deactivation, Scheme 1B). Because of its overall charge, the peptide cannot form coacervates. However, the transient increase in charge enables the activated peptide to form complex coacervate-based droplets. Thus, when a batch of fuel is added, the chemical reaction cycle consumes it, leading to a population of transiently activated peptides that form droplets until the fuel runs out. We combined these droplets with the 17E DNAzyme—a zinc-dependent catalyst that catalyzes site-specific RNA cleavage via transesterification (Scheme 1C).^[17] We used a substrate composed of DNA, except at the cleavage site, where the presence of a 2'-OH of RNA is needed for cleavage (Scheme 1D).

For the droplets to exert control over the DNAzyme, the DNA has to partition in the droplets which we first tested (Figure 1A–D). Unless stated otherwise, we made droplets using 23 mM peptide, 4.1 mM polyU (expressed as the monomer concentration) tagged with 200 nM A₁₅-Cy5 RNA for imaging, 40 mM fuel, 1 mM ZnCl₂, 200 nM DNAzyme tagged with Cy3 for imaging, and 200 mM MES buffer (pH 5.3). We imaged the droplets after 5 minutes using a confocal microscope (Figure 1B, Figure S1). We observed poor colocalization between the droplet-forming polyU and the DNAzyme, likely due to the short length and the secondary structure of the DNAzyme.

To increase the partitioning of the DNAzyme into the droplets, we added an A₁₀ “partitioning arm” to each end of the DNAzyme (DNAzyme-A10). We hypothesized that the adenines pair with the polyU RNA inside the droplets (Figure 1A), thereby facilitating partitioning. Indeed, DNAzyme-A10 showed strong colocalization with the tagged polyU upon fuel-induced droplet formation (Figure 1C), which we quantified by measuring the fluorescence intensity of multiple regions of interest (ROIs) inside of the droplets as a measure for the concentration of DNAzyme inside the droplets. After pelleting the droplets by centrifugation, we recorded micrographs of the supernatant and determined their intensity as a measure of the DNAzyme concentrations in the bulk phase. We calculated the partition coefficient (K_p) from this data, which represents the ratio between the concentration inside the droplets and in the supernatant. Similarly, we determined the K_p of the DNAzyme-A10, the substrate, and the product of the reaction (Figure 1D). As expected, the A₁₀ partitioning arms significantly increased the partitioning of the DNAzyme, achieving more than a 300-fold enhancement ($K_p = 8600 \pm 500$). In contrast, the

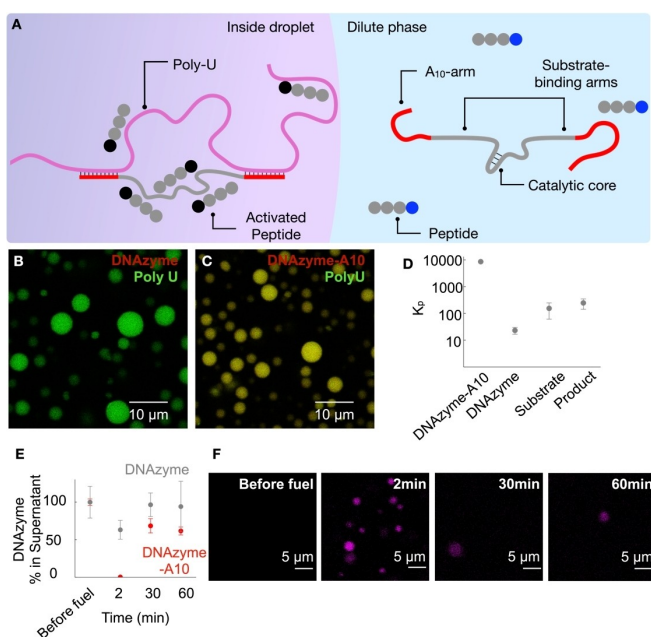


Figure 1. The molecular design of a DNAzyme controls its partitioning in transient droplets. **A.** Schematic representation of the modified DNAzyme inside and outside of the droplet. **B–C.** Confocal micrographs of a solution of complex coacervate droplets 5 minutes after fuel addition stained with a DNAzyme (Cy3, red) and a short RNA oligomer A₁₅, complementary to the poly U (A₁₅-Cy5, green). The DNAzyme with two terminal A10 arms (DNAzyme-A10) partitions preferentially inside of the droplets, whereas the DNAzyme without does not. **D.** The partition coefficients K_p of the DNAzyme, substrate, and product as calculated from the micrographs. K_p is the ratio between the concentration inside the droplets and in the supernatant. A more detailed definition can be found in the Methods section of the SI. The error bars correspond to the standard deviation of three replicates; if none are visible, they are smaller than the size of the data point. **E.** The percentage of DNAzyme and DNAzyme-A10 in the supernatant at different times after fuel addition as measured by PAGE. The error bars correspond to the standard deviation of three replicates; if none are visible, they are smaller than the size of the data point. **F.** Confocal micrographs of a solution of complex coacervate droplets 2 minutes, 30 minutes, and 60 minutes after adding fuel, stained with the Cy3-dye labeled DNAzyme-A10.

substrate ($K_p = 150 \pm 90$) and product ($K_p = 200 \pm 100$) partition inside the droplets to a lesser degree than the modified DNAzyme. No cooperative effect was observed in the partitioning of the substrate or product with DNAzyme-A₁₀. The K_p of the substrate and product remain consistent regardless of the presence of DNAzyme-A10. (Figure S8)

Next, we calculated the fraction of DNAzyme partitioned into the droplets after fuel addition (see Supporting Note 1). We determined the droplet volume of all combined droplets in a ($73.81 \mu\text{m} \times 73.81 \mu\text{m} \times 10 \mu\text{m}$) box by confocal microscopy. Five minutes after fuel addition, the droplets accounted for 0.6 % of the total volume. Combining this data with the partition coefficients, we calculated the percentage of DNAzyme outside the droplets to be 2 % for DNAzyme-A10 and 88 % for the unmodified DNAzyme. In comparison, approximately half of the substrate remained outside of the droplets (52 %). This finding demonstrates no

significant separation of the catalyst and substrate for both DNAzymes, particularly given the considerable material exchange between droplets and supernatant.

The partitioning data was further confirmed with an orthogonal method based on gel electrophoresis (Figure S2). We prepared droplets with $1 \mu\text{M}$ DNAzyme-A10 and substrate and took samples before fuel addition. After the droplet formation, we separated them from the supernatant by centrifugation. We separated the DNA species in the pellet and supernatant using polyacrylamide gel electrophoresis (PAGE) and quantified them with fluorescent gel imaging. This method confirmed that 1.5 % of the DNAzyme-A10 is present in the supernatant. This result and the confocal data support that most DNAzyme-A10 partitions into the droplets.

Given the fuel-dependent nature of droplet formation, we expect a release of the DNAzyme after the fuel depletion. Using a previously described kinetic model,^[9a] we predicted that the system consumed the fuel in roughly 30 minutes and that, during the cycle, a maximum concentration of 4.4 mM of activated peptide was reached after 2 minutes. The model predicts that after 30 minutes, most of the activated peptide had been deactivated. Therefore, we used the same gel-based method described above to determine the release of the DNAzyme into the supernatant. Surprisingly, we measured a release of only 67 % of DNAzyme-A10 30 minutes after fuel addition (Figure 1E). Moreover, this value did not increase further after 60 minutes. By analyzing the samples using confocal microscopy, we observed that the droplets had not completely dissolved 60 minutes after the addition of the fuel (Figure 1F). Thus, the apparent time delay between fuel consumption and droplet dissolution might explain why the DNAzyme-A10 is not completely released into the solution. Indeed, when we determined the volume of the droplets by confocal microscopy at 30 and 60 minutes, we found it had reduced from 0.6 % after five minutes to 0.2 % at both of those time points (Figure S4). In other words, only two-thirds of the droplet material had dissolved up to 60 minutes after fuel addition, which is in line with only two-thirds of the polyU-bound DNAzyme-A10 being released. In contrast, the unmodified DNAzyme did not partition well into the droplets after two minutes and was mostly released after 30 minutes (Figure 1E).

We suspected the millimolar amounts of Zn^{2+} ions to be responsible for the incomplete dissolution of the droplets. Measuring turbidity traces as a proxy for the presence of droplets, we found that in the absence of ZnCl_2 , the droplets dissolved after 22 minutes. In contrast, with 1 mM ZnCl_2 , we found that the turbidity decreased but did not completely disappear (Figure S3). Potentially, a change in the polyanion could optimize our system and solve the inability to dissolve in the future. We attempted to form droplets with our peptide, fuel, and polyT DNA (80mers) or genomic ssDNA as polyanions as a first test (Figure S15). Droplets formed only at very high fuel concentrations, or we observed the formation of non-dissolving aggregates, leading us to continue with the polyU-based system. In summary, we can

control the partitioning of a modified DNAzyme by the fuel-induced presence of active droplets.

Compatibility of the DNAzyme with Active Droplets

Next, we investigated the DNAzyme activity in the presence of droplets. Using gel electrophoresis, we tracked the reaction catalyzed by the modified and unmodified DNAzyme in the absence and presence of droplets (Figure 2A, Figure S9). For these activity measurements, we used 1 μM of DNAzyme or DNAzyme-A10, 4 μM of substrate-Cy5, 2 mM ZnCl_2 , and the same peptide, polyU, and buffer concentrations as described above. To ensure the sustained presence of droplets, we fueled the solution with 10 mM EDC every 5 minutes for 45 minutes.

Both DNAzymes converted 94 % of the substrate over 3 hours at the same rate, showing that the A10 modifications did not affect activity. In contrast, the presence of droplets drastically decreased the yield after three hours to only 14 % for the DNAzyme-A10. For the unmodified DNAzyme, the

decrease was only 48 %. We explain the drop in yield by the high partitioning of DNAzyme-A10 compared to the unmodified DNAzyme.

To ensure the inhibition was due to the partitioning in the droplet and not the design of the partitioning arms, we designed two similar DNAzymes with single A20 arms on either the 5' or 3' end. Our results confirmed that the design of the partitioning arms had no noticeable impact on the activity (Figure S10). This indicates that all modified DNAzymes were inhibited by the droplets. We also monitored the pH throughout the reaction and found a maximum transient increase of 0.04, which should not affect the folding or the activity of the DNAzyme (Supporting Table S5).

To understand the effect of each droplet component on the modified DNAzyme's activity, we compared the reaction yield after 60 minutes in the presence of different droplet components (Figure 2B). The fuel had no impact on the DNAzyme-A10 activity. In contrast, both the polyU and the peptide negatively affected the DNAzyme-A10 activity. Specifically, the polyU reduced the activity by 45 % and the peptide by 70 %. Combining the fuel with peptide or polyU did not further decrease the activity. When all three droplet components are together, the activity of the DNAzyme-A10 is further reduced compared to just peptide and polyU. This finding indicates that partitioning into the droplets inhibits the activity of the modified DNAzyme. Indeed, for the unmodified DNAzyme, we found no difference between the effect of polyU and peptide and all three components (Figure S5). We did unpaired t-tests to measure the significance of all results (Figure S14).

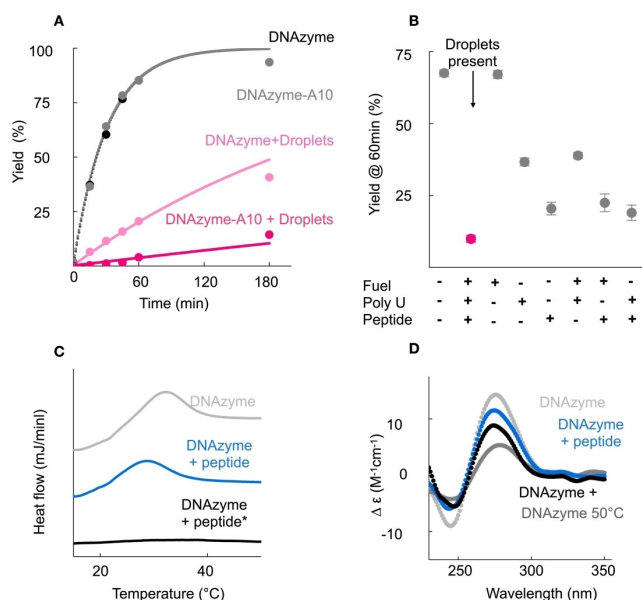


Figure 2. DNAzyme activity is inhibited inside droplets. **A.** Kinetics of DNAzyme-catalyzed reaction analyzed by PAGE (markers) over three hours. A first-order regression over the first 60 minutes was used to fit the data (line). The droplets inhibit the DNAzyme activity, which is more pronounced for DNAzyme-A10. The [DNAzyme] was 1 μM , and the [substrate] was 4 μM . The sample was fueled over the first 45 minutes with 10 mM EDC every 5 minutes. **B.** Screening of DNAzyme A10 reaction yield at 60 minutes in the presence of different droplet components analyzed by PAGE. Except for 1 μM of substrate, we used the same conditions as in A. The error bar corresponds to the standard deviation of the average of three replicates. If none are visible, they are smaller than the size of the data point. **C.** Raw heat flow against temperature as measured by DSC, using 3 mM of peptide or peptide* each with 140 μM of the unmodified DNAzyme. Peptide* is a stable mimic of the activated peptide. **D.** CD spectra of 1.6 μM unmodified DNAzyme in the presence or absence of 1 mM peptide or peptide* at 25 °C. For comparison, the spectra of the DNAzyme at 50 °C were obtained, showing thermal unfolding.

Mechanism of Inhibition

We investigated the mechanism of DNAzyme inhibition by the peptide using differential scanning calorimetry (DSC) and circular dichroism spectroscopy (CD). Specifically, we probed the stability of the folded DNAzyme in the presence and absence of both peptide ($\text{Ac-F(RG)}_3\text{D-OH}$) and a stable model of the activated peptide ($\text{Ac-F(RG)}_3\text{N-NH}_2$, peptide*). In peptide*, the C-terminal aspartic acid is substituted with an amidated asparagine, and thus, it mimics the three positive charges of the anhydride without its transient nature. Experiments with peptide* did not contain any fuel. In the DSC measurement, we used 140 μM of DNAzyme and 3 mM of peptide or peptide* to obtain a detectable signal. Without the peptide, we observed a transition at 31.5 °C, indicating the DNAzyme's unfolding (Figure 2C). With the peptide, the transition peak broadened, and its maximum shifted to 28.7 °C. Moreover, we measured a decrease in the calorimetric enthalpy from 238 to 201 kJ/mol K (Figure S6). Both observations confirm the peptide destabilizes the DNAzyme even at concentrations below the experimental conditions. In the presence of 3 mM of peptide*, mimicking the concentration of activated peptide around 5 minutes into the cycle (See Supporting Note 1), the transition disappeared completely, implying

that almost none of the DNzyme was in the catalytically active folded state (Figure 2C).

Next, we explored the impact of the peptides on DNzyme folding using CD spectroscopy. We used 1.6 μM of DNzyme for the CD spectra and 0.05 to 1 mM of peptide and peptide*. For samples containing 1 mM peptide and peptide*, we observe a decreased amplitude of the circular dichroism signal similar to a thermal unfolding of the hairpin structure of the DNzyme (Figure 2D). In a titration series, we found a dose-dependency of the unfolding to the peptide concentrations (Figure S12). As observed for the DSC measurements, peptide* impacted the unfolding characteristics of the DNzyme to a much larger extent than the peptide (Figure S12). This finding was further confirmed by repeating the activity measurements from Figure 2A with 2.5 mM peptide* and 2.5 mM peptide (Figure S13), demonstrating significantly stronger inhibition with peptide*.

These combined experiments show that the DNzyme's activity is inhibited because both peptide and peptide* unfold its catalytically active conformation. Since the concentration of peptides and DNzyme in the droplets is substantially higher than in solution, we assume the inhibitory effect is enhanced in the droplet environment, leading to complete inhibition of the DNzyme population inside the complex coacervates.

Droplets Spatiotemporally Control DNzyme Activity

Excited by the ability of our droplets to regulate the DNzyme-A10 activity, we speculated we could temporally control the activity of the DNzyme-A10 by utilizing the transient nature of the droplets (Figure 3A). We used the kinetic model to predict the time-dependent activated peptide concentration after fuel addition. Based on the CD data above, we set 0.1 mM activated peptide as a threshold above which we expected most DNzyme-A10 to be inhibited by the droplets (Figure 3A). We found that a single batch of 40 mM EDC sustained the activated peptide concentration above this threshold for 22 minutes. Confocal microscopy confirmed the fuel-induced droplet formation, the dissolution of the majority of droplets, and the release of most of DNzyme-A10 after the threshold had been reached (Figure 3B). We quantified the progress of the reaction for the DNzyme-A10 by following the product formation by PAGE after adding 40 mM fuel (Figure 3C). In the first 15 minutes, i.e., well above the threshold of activated peptide, we found 1 % of substrate conversion. The conversion had increased drastically in the following 15 minutes as an additional 2.5 % substrate was converted. That increased conversion rate was sustained for the following time points. In other words, the presence of the droplets inhibited DNzyme-A10 2.5-fold, which was recovered when most fuel was depleted.

We repeated the experiment by refueling the sample after 60 and 120 minutes. As before, DNzyme-A10 activity was drastically inhibited 15 minutes after fuel addition and mostly recovered afterward. We hypothesize that the

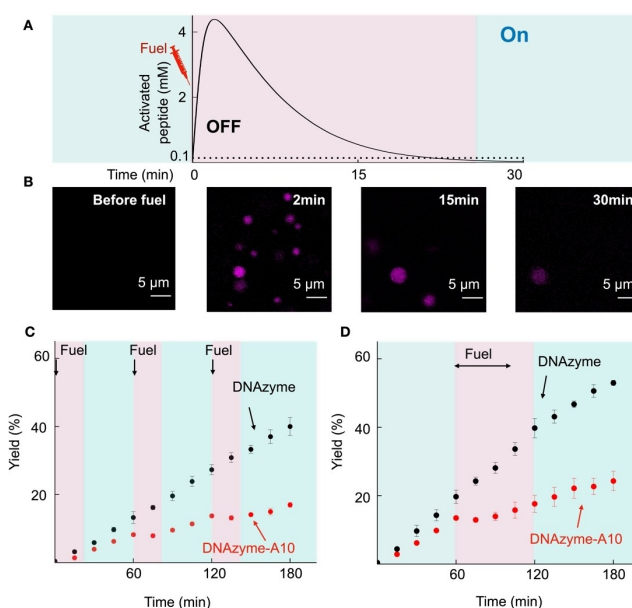


Figure 3. DNzyme activity is spatiotemporally controlled by active droplets. **A.** Activated peptide concentration after the addition of 40 mM fuel predicted by the kinetic model. 0.1 mM was set as a minimum threshold for inhibition. **B.** Confocal micrographs of active droplets containing 200 nM Cy3-tagged DNzyme-A10. The micrographs were acquired before and 5, 15, and 30 minutes after fuel addition. The color-coding correlates to the DNzyme-A10 concentration. The majority of droplets dissolve in 30 minutes. **C–D.** The yields of the DNzyme catalyzed reaction against time as measured by PAGE. Over 180 minutes, we fueled either three times every 60 minutes with 40 mM (**C**) or nine small 10 mM doses every 5 minutes starting one hour into the reaction (**D**). The period when the activated peptide is predicted to exceed the threshold is shaded purple. The error bars represent the standard deviation of the average of three replicates; if none are visible, they are smaller than the data point.

incomplete activity recovery is due to the zinc-induced persistence of the remaining droplets. To confirm this hypothesis, we measured the activity of both DNzymes after complete droplet dissolution, adding the ZnCl_2 and starting the reaction 30 minutes after fuel addition. We found no difference in the experiment with prior fuel addition compared to the ones without, confirming that incomplete DNzyme-A10 activity recovery is due to incomplete dissolution of droplets in the presence of ZnCl_2 (Figure S11).

Excitingly, when we repeated the refueling experiment with the unmodified DNzyme, i.e., the DNzyme lacking the A_{10} arms, the fuel-induced droplet formation had no effect. In other words, by adding partitioning arms, we can control whether or not the active droplets affect certain reaction pathways.

Next, we sustained the inhibition of DNzyme activity by creating a steady state of the activated peptide. We did so by adding 10 mM shots of fuel every 5 minutes (Figure 3D, Figure S7). After monitoring the DNzyme-catalyzed reaction for 60 minutes, we created a steady state of activated peptide for the next 60 minutes. The kinetic model predicted that the activated peptide concentration exceeded the

threshold value immediately and was sustained for 60 minutes. Again, the unmodified DNAzyme did not change its reaction rate after the formation of droplets. In contrast, substrate conversion by DNAzyme-A10 was strongly inhibited right after fuel addition. The DNAzyme-A10 activity slowly recovered after 30 minutes, earlier than the kinetic model predicted. We believe this to be most likely due to exhaustion of the system and accumulation of waste after extensive fuelling. Moreover, the activity of DNAzyme-A10 again did not fully recover to the initial state before the fuel was added. Taken together, we can briefly pause DNAzyme activity or sustain it for longer periods.

Conclusion

We demonstrated that active droplets regulate the activity of a DNAzyme spatiotemporally. The microenvironment of the active droplets unfolds the DNAzyme, thereby inhibiting its catalytic activity. By modifying the DNAzyme with two A10 arms, we increased the partitioning of the catalyst into the droplets, enabling us to control its activity. Given the transient nature of the droplets, the inhibition is temporary and can be modulated by the fuelling pattern.

In biology, chemical reaction networks are constantly up and downregulated. Recent findings show that sequestration inside membraneless organelles is important in this regulation.^[18a,1b,18b] We foresee that combining DNAzymes with active droplets can up or downregulate multiple reactions in a network, thereby emulating the complexity of a cell's biological pathways. These reaction networks, especially where the DNAzyme reciprocally regulates the droplet properties, are a powerful model for understanding better regulation in the cell. It will also be a powerful tool for synthetic biology and synthetic cell research.

Supporting Information

The authors have cited additional references within the Supporting Information.^[9a,16]

Acknowledgements

The BoekhovenLab is grateful for support from the TUM Innovation Network—RISE funded through the Excellence Strategy. This research was conducted within the Max Planck School Matter to Life, supported by the German Federal Ministry of Education and Research (BMBF) in collaboration with the Max Planck Society. The work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094—390783311 and the European Research Council, ERC (JB) and 802000 (HM). Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: DNAzyme · Artificial Organelles · Transient catalysis · Coacervates · Synthetic Biology

- [1] a) C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, A. A. Hyman, *Sci. J.* **2009**, 324, 1729–1732; b) S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **2017**, 18, 285–298; c) J. B. Woodruff, A. A. Hyman, E. Boke, *Trends Biochem. Sci.* **2018**, 43, 81–94; d) E. Gomes, J. Shorter, *J. Biol. Chem.* **2019**, 294, 7115–7127; e) T. Yoshizawa, R.-S. Nozawa, T. Z. Jia, T. Saio, E. Mori, *Biophys. Rev. Lett.* **2020**, 12, 519–539; f) T. Hirose, K. Ninomiya, S. Nakagawa, T. Yamazaki, *Nat. Rev. Mol. Cell Biol.* **2023**, 24, 288–304.
- [2] a) W. M. Aumiller, Jr., F. Pir Cakmak, B. W. Davis, C. D. Keating, *Langmuir* **2016**, 32, 10042–10053; b) A. J. Plys, R. E. Kingston, *Sci. J.* **2018**, 361, 329–330.
- [3] a) S. Koga, D. S. Williams, A. W. Perriman, S. Mann, *Nat. Chem.* **2011**, 3, 720–724; b) V. N. Uversky, I. M. Kuznetsova, K. K. Turoverov, B. Zaslavsky, *FEBS Lett.* **2015**, 589, 15–22; c) A. S. Holehouse, B. B. Kragelund, *Nat. Rev. Mol. Cell Biol.* **2023**.
- [4] a) J. Uniacke, W. Zerges, *J. Cell Biol.* **2008**, 182, 641–646; b) P. Anderson, N. Kedersha, *Curr. Biol.* **2009**, 19, R397–R398; c) D. S. W. Protter, R. Parker, *Trends Cell Biol.* **2016**, 26, 668–679.
- [5] a) C. E. Sing, S. L. Perry, *Soft Matter* **2020**, 16, 2885–2914; b) M. Abbas, W. P. Lipiński, J. Wang, E. Spruijt, *Chem. Soc. Rev.* **2021**, 50, 3690–3705.
- [6] a) R. R. Poudyal, R. M. Guth-Metzler, A. J. Veenis, E. A. Frankel, C. D. Keating, P. C. Bevilacqua, *Nat. Commun.* **2019**, 10, 490; b) A. Testa, M. Dindo, A. A. Rebane, B. Nasouri, R. W. Style, R. Golestanian, E. R. Dufresne, P. Laurino, *Nat. Commun.* **2021**, 12, 6293; c) W. A. Wee, H. Sugiyama, S. Park, *iScience* **2021**, 24, 103455; d) L. Zhou, Y. Fan, Z. Liu, L. Chen, E. Spruijt, Y. Wang, *CCS Chem.* **2021**, 3, 358–366; e) T. Komachi, S. Kawai, Y. Hirayama, A. Maruyama, N. Shimada, *Polym. Chem.* **2023**, 14, 4743–4747; f) K. K. Le Vay, E. Salibi, B. Ghosh, T. Y. D. Tang, H. Mutschler, *eLife* **2023**, 12, e83543; g) J. Wang, M. Abbas, J. Wang, E. Spruijt, *Nat. Commun.* **2023**, 14, 8492.
- [7] a) S. Tagami, J. Attwater, P. Holliger, *Nat. Chem.* **2017**, 9, 325–332; b) R. R. Poudyal, C. D. Keating, P. C. Bevilacqua, *ACS Chem. Biol.* **2019**, 14, 1243–1248.
- [8] A. K. Rai, J.-X. Chen, M. Selbach, L. Pelkmans, *Nature* **2018**, 559, 211–216.
- [9] a) C. Donau, F. Späth, M. Sosson, B. A. K. Kriebisch, F. Schnitter, M. Tena-Solsona, H.-S. Kang, E. Salibi, M. Sattler, H. Mutschler, J. Boekhoven, *Nat. Commun.* **2020**, 11, 5167; b) F. Späth, C. Donau, A. M. Bergmann, M. Kränzlein, C. V. Synatschke, B. Rieger, J. Boekhoven, *J. Am. Chem. Soc.* **2021**, 143, 4782–4789; c) C. Donau, F. Späth, M. Stasi, A. M. Bergmann, J. Boekhoven, *Angew. Chem. Int. Ed.* **2022**, 61, e202211905; d) A. M. Bergmann, J. Bauermann, G. Bartolucci, C. Donau, M. Stasi, A.-L. Holtmannspötter, F. Jülicher, C. A.

- Weber, J. Boekhoven, *Nat. Commun.* **2023**, *14*, 6552; e) F. Späth, A. S. Maier, M. Stasi, A. M. Bergmann, K. Halama, M. Wenisch, B. Rieger, J. Boekhoven, *Angew. Chem. Int. Ed.* **2023**, *62*, e202309318; f) S. M. Poprawa, M. Stasi, B. A. K. Kriebisch, M. Wenisch, J. Sastre, J. Boekhoven, *Nat. Commun.* **2024**, *15*, 4204.
- [10] a) W. M. Aumiller, C. D. Keating, *Nat. Chem.* **2016**, *8*, 129–137; b) G. Cheng, J. Perez-Mercader, *Chem* **2020**, *6*, 1160–1171; c) C. Donau, J. Boekhoven, *Trends Chem.* **2023**, *5*, 45–60.
- [11] a) K. Das, L. Gabrielli, L. J. Prins, *Angew. Chem. Int. Ed.* **2021**, *60*, 20120–20143; b) X. Chen, M. A. Würbser, J. Boekhoven, *Accounts of Materials Research* **2023**, *4*, 416–426.
- [12] M. A. Würbser, P. S. Schwarz, J. Heckel, A. M. Bergmann, A. Walther, J. Boekhoven, *ChemSystemsChem* **2021**, *3*, e2100015.
- [13] a) M. A. Cardona, L. J. Prins, *Chem. Sci.* **2020**, *11*, 1518–1522; b) R. Chen, S. Neri, L. J. Prins, *Nat. Nanotechnol.* **2020**, *15*, 868–874.
- [14] a) R. R. Breaker, G. F. Joyce, *Chem. Biol.* **1994**, *1*, 223–229; b) B. Cuenoud, J. W. Szostak, *Nature* **1995**, *375*, 611–614; c) E. Golub, R. Freeman, I. Willner, *Angew. Chem. Int. Ed.* **2011**, *50*, 11710–11714.
- [15] G. F. Joyce, *Annu. Rev. Biochem.* **2004**, *73*, 791–836.
- [16] M. Tena-Solsona, B. Rieß, R. K. Grötsch, F. C. Löhner, C. Wanzke, B. Käs Dorf, A. R. Bausch, P. Müller-Buschbaum, O. Lieleg, J. Boekhoven, *Nat. Commun.* **2017**, *8*, 15895.
- [17] a) J. Li, Y. Lu, *J. Am. Chem. Soc.* **2000**, *122*, 10466–10467; b) A. K. Brown, J. Li, C. M. B. Pavot, Y. Lu, *Biochem.* **2003**, *42*, 7152–7161.
- [18] a) S. Alberti, *J. Cell Sci.* **2017**, *130*, 2789–2796; b) Y. Shin, C. P. Brangwynne, *Sci. J.* **2017**, *357*, eaaf4382.

Manuscript received: July 3, 2024

Accepted manuscript online: August 9, 2024

Version of record online: October 15, 2024

Publications

9.3 Publication 3

DNA affects the phenotype of fuel-dependent coacervate droplets

Machatzke, C.*; Holtmannspötter, A.L.*; Mutschler, H.; Boekhoven, J.⁺

* Authors contributed equally

⁺ Corresponding author

Nature Communications **2026**, 17, 2953.

DOI: 10.1038/s41467-026-71024-8

9.3.1 Personal Contribution

- Designed the project with A.-L. Holtmannspötter, J. Boekhoven and H. Mutschler
- Performed all experiments related to sequencing (Figure 2 and SI)
- Prepared figure 2
- Composed, edited and revised the manuscript with A.-L. Holtmannspötter, J. Boekhoven and H. Mutschler

DNA affects the phenotype of fuel-dependent coacervate droplets

Received: 31 August 2025

Accepted: 11 March 2026

Published online: 26 March 2026

Corbin Machatzke^{1,3}, Anna-Lena Holtmannspötter^{2,3}, Hannes Mutschler¹✉ & Job Boekhoven²✉

Synthetic cells emulate fundamental biological behaviors, like growth, metabolism, and mobility, but have lacked genotype-driven selection, which is essential for Darwinian evolution. Here, we introduce libraries of short DNA sequences as genotypes into fuel-dependent peptide-RNA-based coacervate droplets, serving as synthetic cells. By sequencing, we identify sequences that partition in the droplets, revealing strong preferences for guanine-rich and adenine-rich motifs. These sequences affect the synthetic cell phenotype—adenine-rich sequences shorten droplet lifetimes through hybridization. In contrast, guanine-rich sequences kinetically trap droplets via peptide interactions, altering dissolution rates and morphology. This study demonstrates how genotype affects phenotype in synthetic cells, establishing essential design principles for achieving Darwinian evolution in minimal protocellular systems.

Synthetic cells are minimal models of living systems^{1,2}, designed from the bottom up to replicate fundamental biological behaviors such as growth^{3–7}, division^{8–13}, metabolism^{14–16}, and evolution¹⁷. A hallmark of life is the ability to sustain itself out of equilibrium through continuous energy transduction, replicate and adapt under selective pressures, ultimately enabling Darwinian evolution^{18–23}. Thus, for a synthetic cell to ultimately become synthetic life, it must sustain under non-equilibrium conditions and demonstrate Darwinian evolution, *i.e.*, replication with mutation, such that the fittest survive and spread their genotype. To address the criterion of synthetic life that requires self-sustaining properties, we recently introduced the concept of fuel-dependent synthetic cells^{24–26}. These synthetic cells are driven by a chemical reaction cycle in which a fuel-driven activation reaction activates building blocks, which then spontaneously deactivate through hydrolysis^{27–31}. Thus, the synthetic cells can grow under conditions where chemical fuel is abundant but decay and dissolve during periods of starvation (Fig. 1a). These fuel-dependent synthetic cells exhibit life-like properties as they require continuous energy input to maintain their structure and functionality²⁴. Without fuel, these synthetic cells spontaneously decay, serving as a selection pressure, especially in experiments in which fuel is periodically supplied¹².

In previous work, we used fuel-dependent synthetic cells based on lipids²⁶ and complex coacervate droplets²⁴. Complex coacervate droplets have emerged as promising protocell models due to their ability to spontaneously form membraneless compartments, concentrating biomolecules, and facilitating biochemical reactions^{32–36}. Our fuel-dependent complex coacervates for the basis of this work comprise a long RNA polyanion (poly-U, 2200 monomer units on average) and a short cationic peptide (Ac-F(RG)₃D-OH, Fig. 1b). The fuel-dependent nature of these droplets comes from this peptide's fuel-driven activation. Specifically, the peptide's C-terminus reacts with carbodiimides as fuel, forming the corresponding intramolecular anhydride and increasing the overall charge of the peptide from +1 to +3. In the deactivation reaction, the short-lived anhydride spontaneously hydrolyzes to the original peptide with a half-life of roughly 1 minute, thereby reverting the charges from +3 to +1 (Fig. 1b). The transient increase in charges enables the short-lived activated peptide to bind the RNA and assist in forming complex coacervate-based droplets. Thus, upon addition of a batch of fuel, the chemical reaction cycle initiates, leading to a population of transiently activated peptides that form droplets via phase separation. Activated peptide diffuses into these droplets, leading to growth, while deactivated peptide leaves,

¹TU Dortmund University, Dortmund, Germany. ²Department of Bioscience, School of Natural Sciences, Technical University of Munich, Garching, Germany.

³These authors contributed equally: Corbin Machatzke, Anna-Lena Holtmannspötter. ✉e-mail: Hannes.Mutschler@tu-dortmund.de;

Job.Boekhoven@tum.de

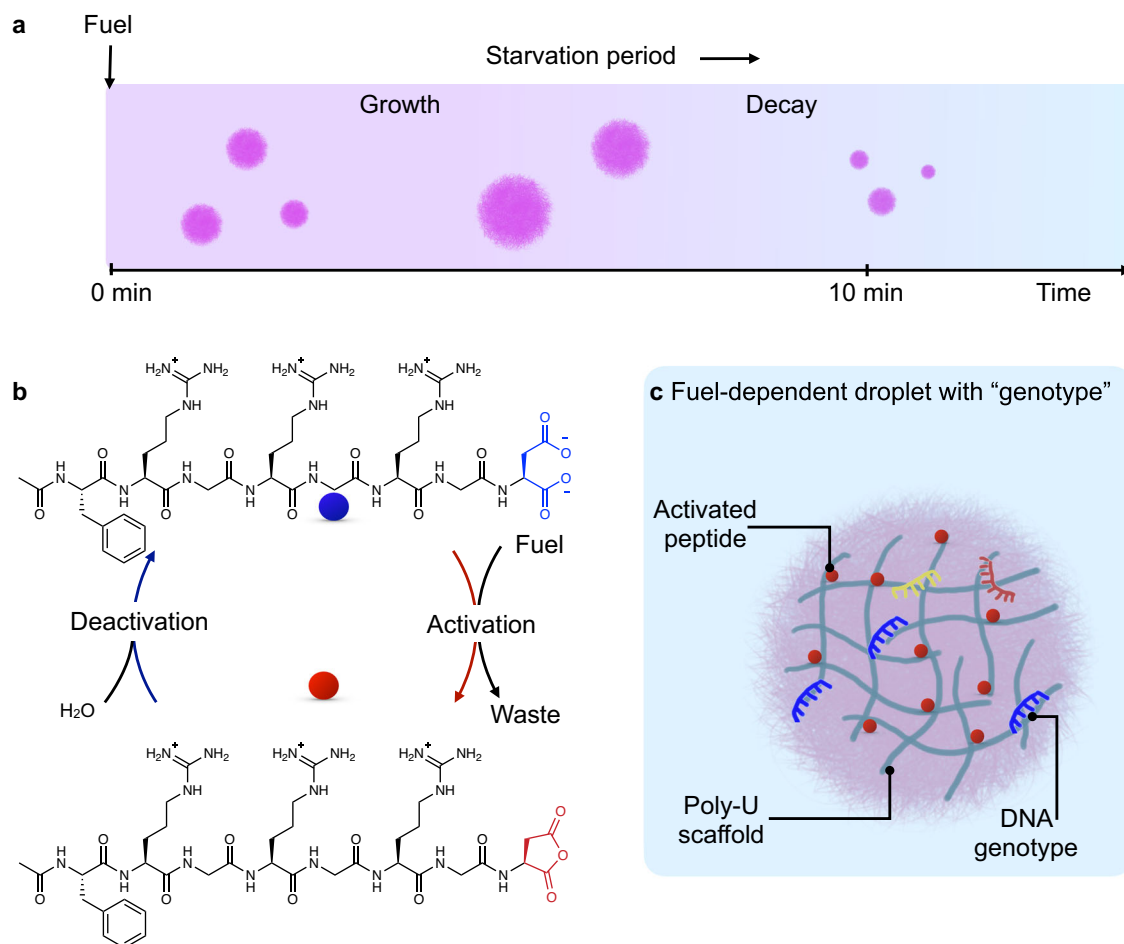


Fig. 1 | Fuel-dependent synthetic cells. **a** Upon fuel addition, droplets form, grow, fuse, and dissolve after a discrete lifetime. **b** Fuel-dependent peptide activation and deactivation cycle. Blue stands for the inactivated peptide with the aspartic acid

C-terminus. Red stands for the activated peptide with a terminal anhydride. **c** Synthetic cells are equipped with a short DNA genotype (colored strands) that can interact with both the peptide (red) and the RNA scaffold material (gray).

leading to shrinkage. This battle for survival continues until no more fuel remains and all droplets dissolve (Supplementary Movie 1). So, like life, our fuel-dependent synthetic cells require a constant influx of nutrients, and periodic fueling of our droplets can sustain them.

A critical limitation of these droplets is the absence of information-carrying elements—a genotype—which means survival is purely stochastic. In synthetic life, fuel-dependent synthetic cells would require an equivalent of a genotype: inherited molecular information that can be replicated and passed on during division. This genotype should influence the system's physical or functional traits (its phenotype), thereby enabling selection and evolution. We use these biological terms with caution—in our synthetic context, “genotype” refers to a set of molecules that are inherited, replicable, and transferable to the next generation. At the same time, “phenotype” denotes the observable physical traits of the synthetic cell. Genotype–phenotype coupling, in this context, means that these information-containing molecules directly influence the properties of the droplets. This would be a much more rudimentary analog to the biology's genotype–phenotype coupling, which operates through the central dogma of molecular biology. Despite its simplicity, it could still enable principles like natural selection or even evolution, which is what we aim for in future work.

In this work, we propose that a set of single-stranded DNA sequences may serve as a proxy for such a genotype. We reason that DNA sequences can be replicated either with or without enzymes¹⁷.

Besides, as the droplet material comprises RNA and partly peptide, we hypothesize that the DNA sequences could interact either with the RNA or peptide and thereby affect the droplet properties, i.e., their phenotype (Fig. 1c)³⁷. Thus, we endowed our synthetic cells with short, 30-base DNA strands, starting from a library of random sequences, and investigated their compatibility and interaction modes with the droplets. After we found that libraries of 30-mers biased towards one of the four nucleotides can affect droplets, we used second-generation Illumina sequencing to identify which DNA sequences preferentially partition into fuel-dependent droplets. We found patterns in the sequencing data, which we then isolated and investigated individually. Specific sequences, such as long repeats of adenosine or guanosine, can drastically alter the droplet properties, increasing their longevity, viscosity, and decreasing fusion rates; in other words, a genotype affects the droplet's phenotype. However, the genotype has not yet replicated, and in future work, we envision using self-replicating DNA sequences as a genotype that also affects phenotype^{38–40}.

Results

To establish a rudimentary genotype–phenotype coupling, we first investigated the impact of DNA libraries on the droplet's phenotype focusing on their lifetime and morphology (Fig. 2a). Throughout this study, we used the following standard reaction conditions unless stated differently: 23 mM peptide, 4.1 mM poly-U (concentration expressed per nucleotide), 30 mM EDC, and 50 μ M DNA (30-mer

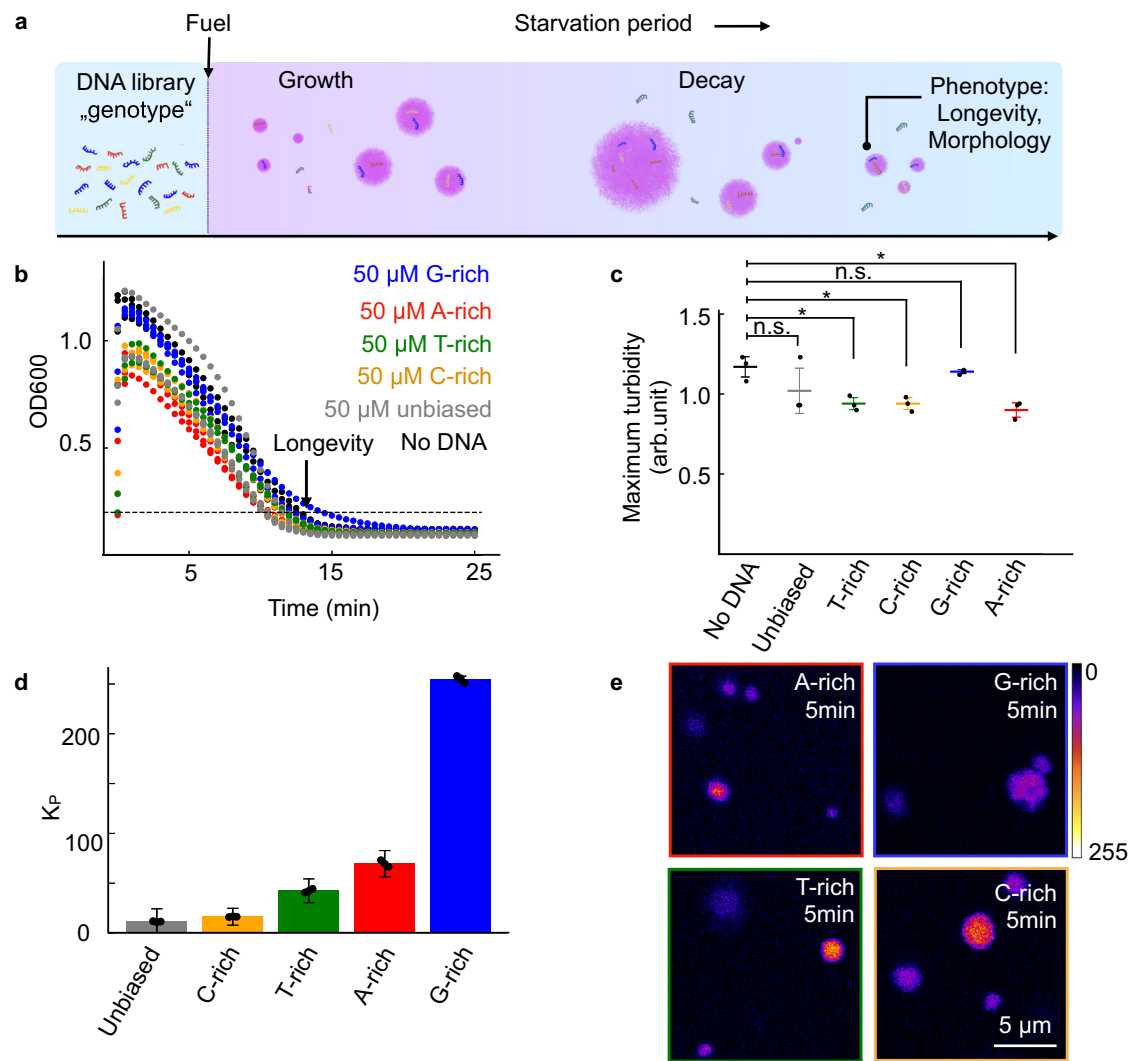


Fig. 2 | Biased DNA libraries as genotype impact the droplet phenotype.

a Schematic representation of the droplet's lifecycle. **b** Absorbance of 600 nm light (OD₆₀₀) against time as a measure of turbidity for standard conditions in the presence of 50 μ M of different DNA libraries, color indicates a bias towards one nucleobase: black meaning no DNA, gray no bias, blue a bias for guanine, red adenine, green thymine, and yellow cytosine. The same allocation applies to the other subfigures. **c** Comparison of the maximum turbidity (arbitrary unit) for different DNA libraries. The A-rich library limits coacervate formation the most, while the G-rich library has no impact. **d** Comparison of the partitioning coefficients (K_p) of biased random sequences into our droplets. The G-rich random sequences and A-rich random sequences have the highest partitioning. **e** Representative Confocal

micrographs with 500 nM Sulforhodamine. Scale bars represent 5 μ m. Error bars in Fig. 1c show the standard deviation from the average ($N=3$). N represents one biological replicate of a turbidity value. Significance was determined with a two-sided Welch's t-test with the no-DNA sample as the control group. P -values are available in Supplementary Table 1. Error bars in Fig. 1d are calculated from the standard deviation from the average ($N=3$) of three biological replicates of measured DNA concentration in the supernatant (shown datapoints) and the standard deviation from the average of the droplet volumes calculated from confocal images of water in oil droplets ($N=10$) according to the method described in the method section. Source data are provided as a Source Data file.

oligonucleotides, corresponding to a total DNA concentration of 1.5 mM when expressed per nucleotide). We used DNA libraries consisting of random 30-mer sequences, with each nucleotide position having an equal (25%) probability of containing any of the four canonical bases (A, T, C, or G). Additionally, we used biased DNA libraries, in which one specific nucleotide (A, C, T, or G) was preferentially incorporated at each position of the 30-mer sequence with a probability of 76%, while the remaining three bases each had a 8% chance of incorporation.

In a plate reader, we tracked the OD₆₀₀ of the samples over time as a measure of the presence of the droplet population (Fig. 2b). We found that most samples with DNA tended to exhibit less turbidity compared to samples without DNA, except for the unbiased and G-rich library, which showed no significant difference (Fig. 2c, Supplementary Table 1). This finding suggests fewer or smaller droplets due to the

added DNA. We defined the time at which the OD₆₀₀ fell below 0.2 as the longevity of the droplet population and found that the A-rich library showed a trend toward shorter longevity (Fig. S1, Supplementary Table 2). Taken together, among all the libraries, the A-rich library proved the most destructive for the droplets, resulting in less turbidity that was shorter-lived. We further investigated other droplet properties, such as viscosity and resistance to salt in the presence of the unbiased library, which remained unchanged (Figs. S2, S3). Lastly, we examined the impact of the biased DNA libraries on droplet morphology by confocal microscopy (Fig. 2e, Fig. S4). The droplets appeared round, rapidly fused, and remained unchanged overall, regardless of the DNA.

For DNA to act as a genotype, it must incorporate into the synthetic cells. Thus, we tested the partitioning of the biased DNA libraries into the droplets. As the partitioning strongly depends on whether a

droplet is dissolving or growing (Fig. S5), these experiments with our short-lived droplets are challenging. Therefore, we used an in-equilibrium model of our droplets to determine the partitioning coefficients. Specifically, a stable model of the activated peptide was used (Ac-F(RG)₃N-NH₂). In the model of the activated peptide, the C-terminal aspartic acid is substituted with an amidated asparagine, resulting in three positive charges, similar to the activated peptide, but without its transient nature. We replaced 3 mM of the peptide with the active peptide model to form droplets at a time point roughly midway through the cycle⁴¹. We determined the partition coefficient by measuring the DNA concentration in the supernatant using HPLC after spinning down the droplets (Supplementary Table 7) and by calculating the droplet volume using confocal microscopy. We found that completely random DNA libraries (N₃₀) partitioned weakly ($K_p = 11$) (Fig. 2d). We suspect that partitioning in this case is somewhat low due to competition with poly-U, which competes for a limited amount of anhydride. The short DNA is disadvantaged compared to the longer poly-U due to multivalency effects. The C- or T-biased DNA libraries did not partition substantially better than the unbiased DNA. In contrast, the partition coefficients increased by more than 20-fold for G-rich libraries and by more than 6-fold for A-rich libraries. We assume that partitioning for both is driven by interactions with the droplet material. While the A-rich DNA could base pair with our RNA scaffold, the G-rich DNA could interact strongly with our highly concentrated peptide. From our combined experiments, we learned that G-rich and A-rich DNA sequences have the highest partitioning coefficient. The A-rich library decreased the amount of droplet material formed and slightly decreased their longevity. In contrast, the G-rich library is the only one that does not limit droplet formation compared to the no-DNA condition.

We suspected that the minute changes in the droplet's phenotype are a result of the DNA libraries only partitioning weakly in the droplets. To put that in perspective, for the unbiased library, only 5% of the DNA sequences can be expected to be present in the droplets based on the partitioning coefficient ($K_p = 11$) and the total droplet volume. That number increases to roughly 54% percent for the G-rich library. In other words, only a fraction of our DNA library acts as a genotype, in that it partitions in the droplets and affects their properties. To understand which DNA sequences those are, and unveil patterns in these sequences, we examined the sequences of the unbiased DNA library present in the spun-down, combined droplets and those in the supernatant using Illumina sequencing (Fig. 3a). Specifically, we prepared active droplets in the presence of the unbiased DNA library at a concentration of 0.7 μ M 30-mer. After adding fuel, we waited for one minute to enable maximum droplet formation, followed by three minutes of centrifugation to separate the droplets from the supernatant. We recovered DNA from both the pellet and the supernatant and prepared the sequencing library using a custom protocol. That library was sequenced using Illumina sequencing (see Method section for detailed methodology). All subsequent analysis was performed in a comparative approach between the R1 reads of each sample, ensuring that no bias in the library preparation method leads to false results.

The sequencing results revealed no significant difference between the supernatant and the original library, further strengthening our hypothesis that only a small fraction of the library partitions in the droplets (Fig. S6). In contrast, we observed drastic differences between the sequences in the combined droplets and the supernatant. For example, we found an enrichment of the G nucleotide—of all the nucleotides in the droplets, there was a 2.5% G-enrichment compared to the supernatant (26.5% in the supernatant, 29 % in the combined droplets, Fig. 3b). This finding is in line with our finding that G-rich libraries partitioned strongly compared to unbiased 30-mer libraries. Excitingly, we could now unveil which nucleotide positions in the 30-mer these G predominantly occupied. We found a positional preference for guanines in the terminal bases: both 5' and 3' ends

contained Gs with an increased likelihood of 8% points compared to the supernatant (33% against 25%). The enrichment diminishes over the successive eight nucleotides, reaching a plateau of -1.5% (Fig. 3c). When we study the sequences with sequential Gs at their termini, we found they have a bias for more Gs in the rest of the sequence compared to all sequences from the droplet sample, e.g., a sequence starting with 3 Gs in a row, tends to have more Gs in the remaining 27 bases compared to the rest of the pool (Fig. S7). The latter finding suggests a cooperative effect—for a sequence to partition, it is more effective to have multiple Gs within a single sequence.

We calculated and compared the minimum free energy of folding for all sequences in either the combined droplets or the supernatant using a thermodynamic secondary structure prediction algorithm^{42–44}. The droplets exhibit a shift towards less folded sequences, characterized by a shifted distribution in the minimum free energy of folding. (Fig. 3d). Based on the thermodynamic secondary structure prediction, we conclude that intramolecular structures in sequences are detrimental to their partitioning into these droplets in line with previous work^{45,46}.

Finally, we searched for enriched motifs in the droplet library compared to the supernatant using STREME^{47–49}. We found 14 motifs that were statistically significantly enriched compared to the supernatant (Supplementary Table 9). Of these sequences, we present the four sequences with the highest significance in the main text as frequency logos (Fig. 3e). Two of these four are dominated by multiple Gs in consecutive stretches having a high probability of five or even nine consecutive Gs. The other two of the four logos contain stretches of consecutive As, at least five nucleotides long, with a slight trend of this motif towards the 3' or 5' ends of the sequences (Supplementary Table 9). We further analyzed the frequency of stretches of five and seven consecutive As or Gs in both the droplets and the supernatant and found a substantially higher occurrence of those in the droplets compared to the supernatant (Fig. 3f).

Based on these findings, we conclude that multiple consecutive Gs are required to interact with the droplet material. Similarly, and in line with previous work⁴⁶, consecutive A sequences strongly interact with the poly-U in the droplets. Hybridization as a recruiting mechanism and stretches of consecutive bases have previously been found in the sequencing of RNA in different coacervate systems⁵⁰.

From the data above, we conclude that sequences with stretches of A or G partition well in the droplets. Moreover, these sequences likely also affect the droplet phenotype, as observed with the slight effects of A- and G-rich biased libraries. To test which sequences specifically affect the phenotype, we ordered 30-mers containing stretches of A and measured their partitioning and impact on droplet phenotype. We started with A₃₀ as an extreme case of consecutive As (Fig. 4a). Compared to the unbiased or the A-rich library, A₃₀ resulted in a drastic decrease in droplet lifetime (Fig. 4b, Supplementary Table 3, Supplementary movie 2), which was strongly concentration-dependent above a concentration as low as 20 μ M (Fig. 4c and Fig. S8). We found that the partition coefficient was nearly five hundred-fold higher for A₃₀ compared to the A-rich random sequences, implying that over 99% of all A₃₀ oligomers are in the droplets (Fig. S9). The high partition coefficient was to be expected⁵¹, as A₃₀ can hybridize with poly-U, but it does not explain the decrease in longevity.

We hypothesize that the hybridization of the short A₃₀ with the RNA causes the domains of the poly-U to rigidify, given that these domains are chimeric double helices (Fig. 4e)⁵². As the flexibility of the RNA is essential for efficient charge screening and phase separation^{53–56}, this could explain the change in the droplet's phenotype. We examined the critical fuel concentration required for coacervate formation (Fig. 4d) and the critical salt concentration for droplet dissolution (Fig. S10). This fuel concentration corresponds to the anhydride concentration in solution required for phase separation. Without additional DNA, our coacervate-based droplets formed at

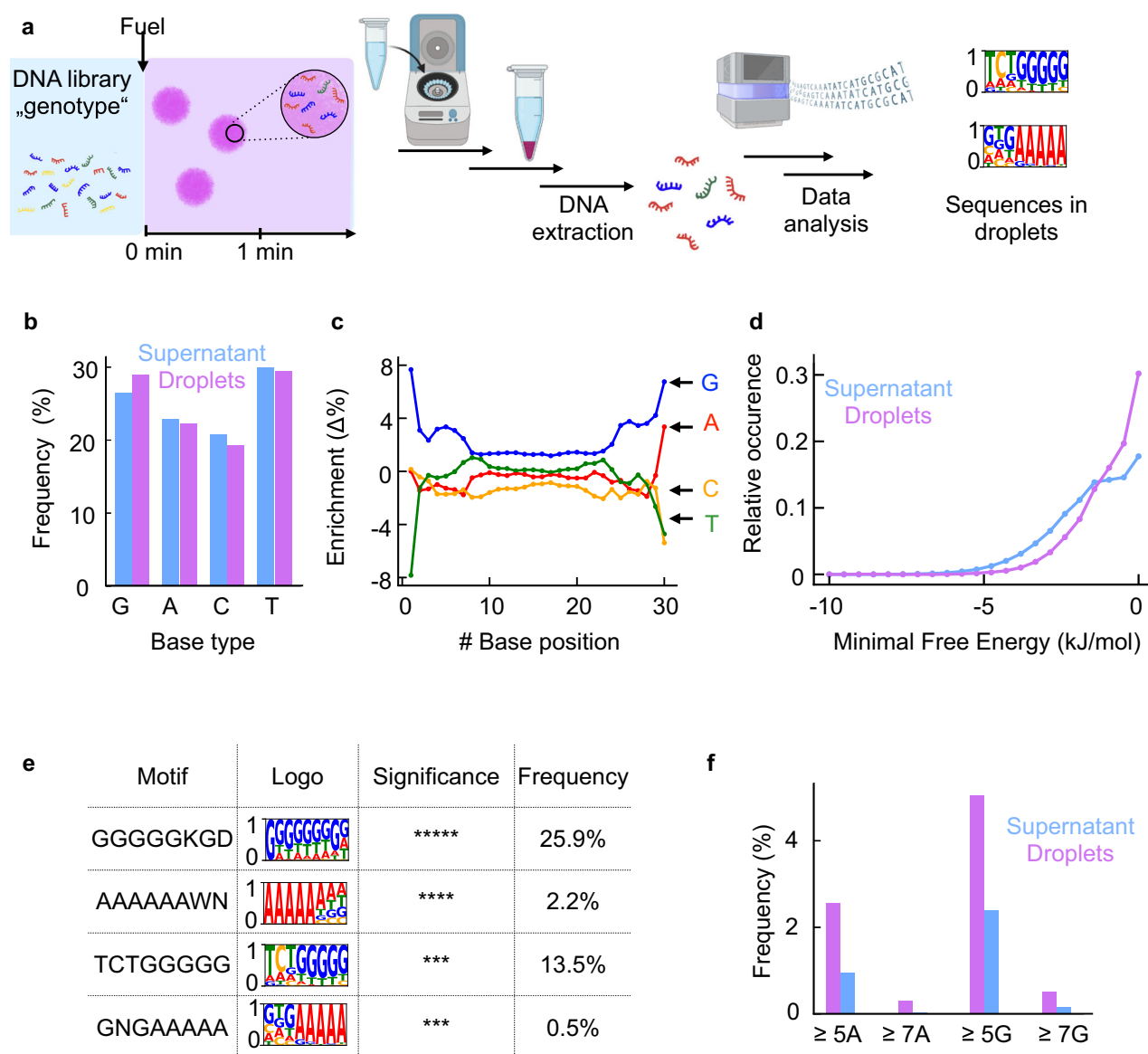


Fig. 3 | Sequencing the combined droplets genotype. **a** Schematic representation of our workflow: DNA was harvested from the combined droplets phase and the supernatant, then sequenced and analyzed. Schematic of the workflow created with <https://BioRender.com/yf9mg24>. **b** Overall base distribution in the supernatant (blue) and combined droplets phases (pink). **c** Difference in nucleotide content per position between sequences isolated from combined droplets and supernatant. Blue stands for guanine, red for adenine, green for thymine, and yellow for cytosine. The same allocation applies to subfigure **e**. **d** Minimal free energy of folding for sequences in the supernatant (blue) and combined droplets (pink) calculated with ViennaRNA package 2.0. **e** The four most overrepresented motifs in the combined

droplets sample (pink) compared to the supernatant (blue). Stretches of at least 5 Gs or As seem to be highly favored. The reported significance are *e*-values which are *p*-values for each motif appearing at random in the dataset divided by the total number of motifs. Symbols signify *e*-values where ***** are values $< 10^{-100}$, **** are values $< 10^{-50}$, *** are values $< 10^{-10}$. Exact *e*-values can be found in Supplementary Table 9. Significance was calculated with a Fisher's Exact test by MEME-Suites STREME as detailed in the software documentation. **f** The frequency of at least five or seven consecutive As or Gs in a sequence in the supernatant (blue) and in the combined droplet phase (pink). Lines in **c** and **d** are added to guide the eye. Source data are available on a public repository linked in the Data availability section.

9 mM EDC. In the presence of 50 μ M A₃₀, the threshold increased to 27 mM, indicating a lower affinity between the droplet components. To determine the critical salt concentration, we used our in-equilibrium model droplets and monitored the amount of additional NaCl required to dissolve our droplets. Without DNA or 50 μ M N₃₀, the critical salt concentration was 74 mM NaCl, which dropped to 54 mM NaCl in the presence of 50 μ M A₃₀. This lower salt threshold further indicates that the additional A₃₀ reduced the interaction strength. We confirmed that poly-U is the primary interaction partner of A₃₀ by measuring the diffusivity of both the peptide and poly-U within droplets using FRAP. The poly-U mobility was markedly reduced, while the peptide diffusivity

remained unaffected with the additional DNA as genotype (Figs. 4f, g, and S11, Supplementary Table 11), indicating a selective restriction of the RNA. These results align with previous studies on peptides with poly-A/poly-U mixed systems⁵⁷, hybridization-induced diffusivity restriction⁵⁸, and other work, which shows the polymer chain stiffness is a limiting factor on the binodal in complex coacervation^{56,59,60} and weakens the associative phase behavior of the system and its salt resistance⁶⁰. Additionally, we have confirmed the poly-U adenine hybridization via native gel electrophoresis (Fig. S12) and investigated the impact of A₃₀ on droplets containing a different, non-hybridizing polyanion. Without hybridization, the lifetime was not impacted

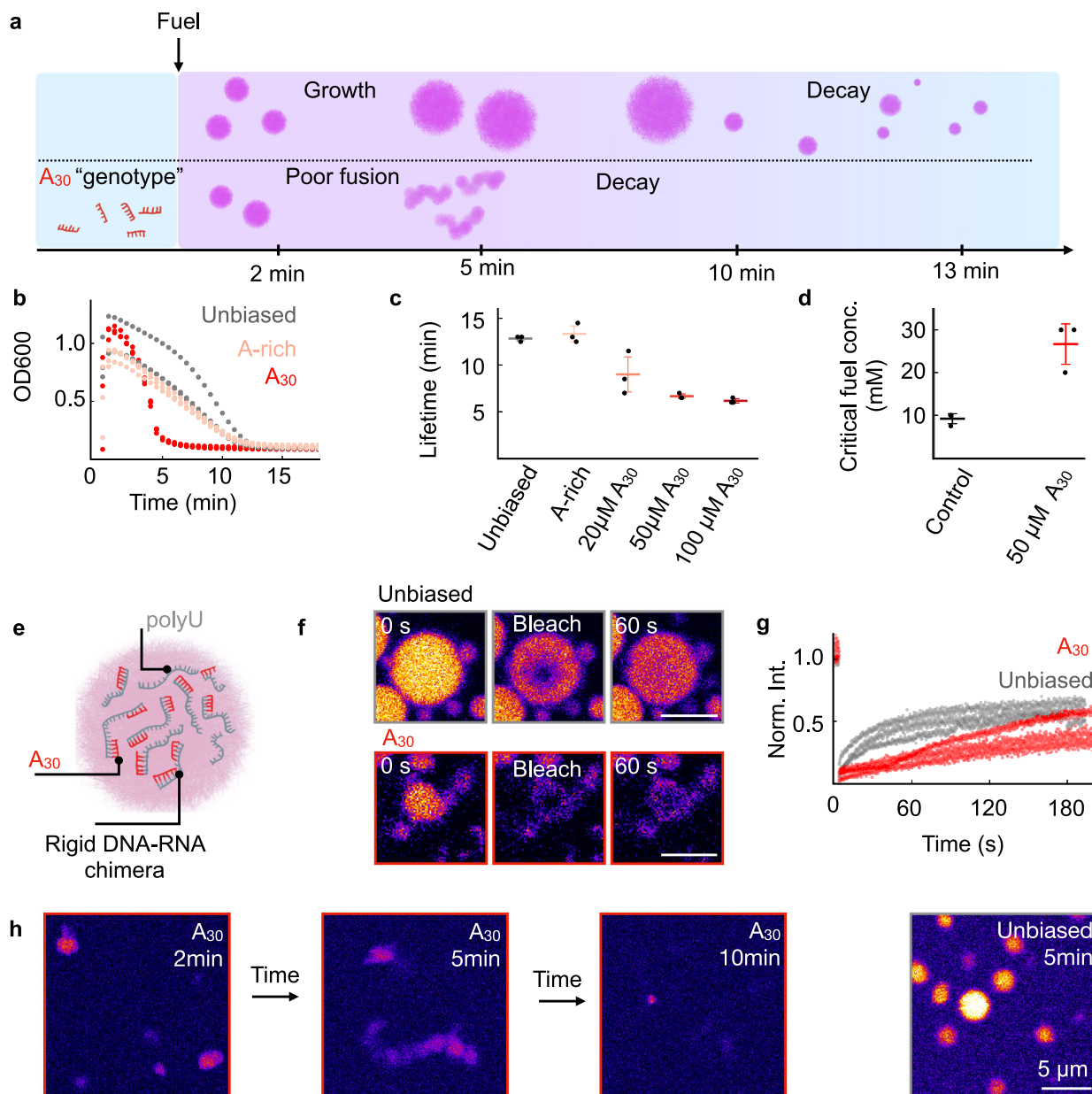


Fig. 4 | Droplet's longevity decreases with a hybridizing genotype. **a** Schematic representation of A₃₀ DNA strands (red) shortening the coacervate lifetime and changing the morphology. **b** Optical density at 600 nm (OD600) against time for droplets with an unbiased library (gray), an A-rich library (pink), or A₃₀ (red). **c** The lifetime of droplets with genotypes of unbiased DNA library (gray), A-rich library (pink), and different concentrations of A₃₀ (red, dark red). **d** The critical fuel concentration measured by tracking the OD600 upon fuel addition with (red) and without A₃₀ (black). **e** Schematic representation of hybridization of A₃₀ (red strands) with the droplet material (gray). Representative confocal micrographs and

FRAP recovery traces with normalized intensity (Norm. Int.) measured on a confocal microscope, bleaching 500 nM of Cy5-A₁₅ annealed to the poly-U in the presence of the unbiased library (gray) or A₃₀ (red). Scalebars represent 10 μm. **f, g** Confocal micrograph of select points in a fuel cycle stained with 500 nM Sulforhodamine in the presence of A₃₀ and N₃₀, scalebars represent 5 μm. **h** All error bars show the standard deviation from the average ($N = 3$). N represents one biological replicate of one analyzed turbidity trace or turbidity value upon fuel addition. Source data are provided as a Source Data file.

(Fig. S13). Adding the A₃₀ at a later point also shortened the lifetime, but did not lead to immediate dissolution (Fig. S14). Confocal microscopy over the droplet's lifetime showed that A₃₀-containing droplets fused less readily compared to their N₃₀ equivalents (Fig. 4h, Supplementary Movie. 2). Instead, they formed bead-like aggregates that are less liquid compared to our N₃₀ equivalents and resemble "pearls on a string" that settled onto the glass where they dissolved shortly after five minutes. We explain these results by the increased poly-U rigidity limiting the poly-U mobility and thereby preventing droplets from

coalescing fully and growing. Nevertheless, the mobility of the peptide facilitates the dissolution of the aggregates when fuel levels are low. Similar morphologies have been reported in poly-A/poly-U mixtures and RNA-protein networks, based on base pairing, where the peptide or protein remains mobile despite the restricted dynamics of the RNA⁶¹.

We conclude that DNA A₃₀ can hybridize with the RNA droplet material, stiffening the RNA and leading to radical changes in the droplet phenotype by rigidifying the RNA scaffold. The stark

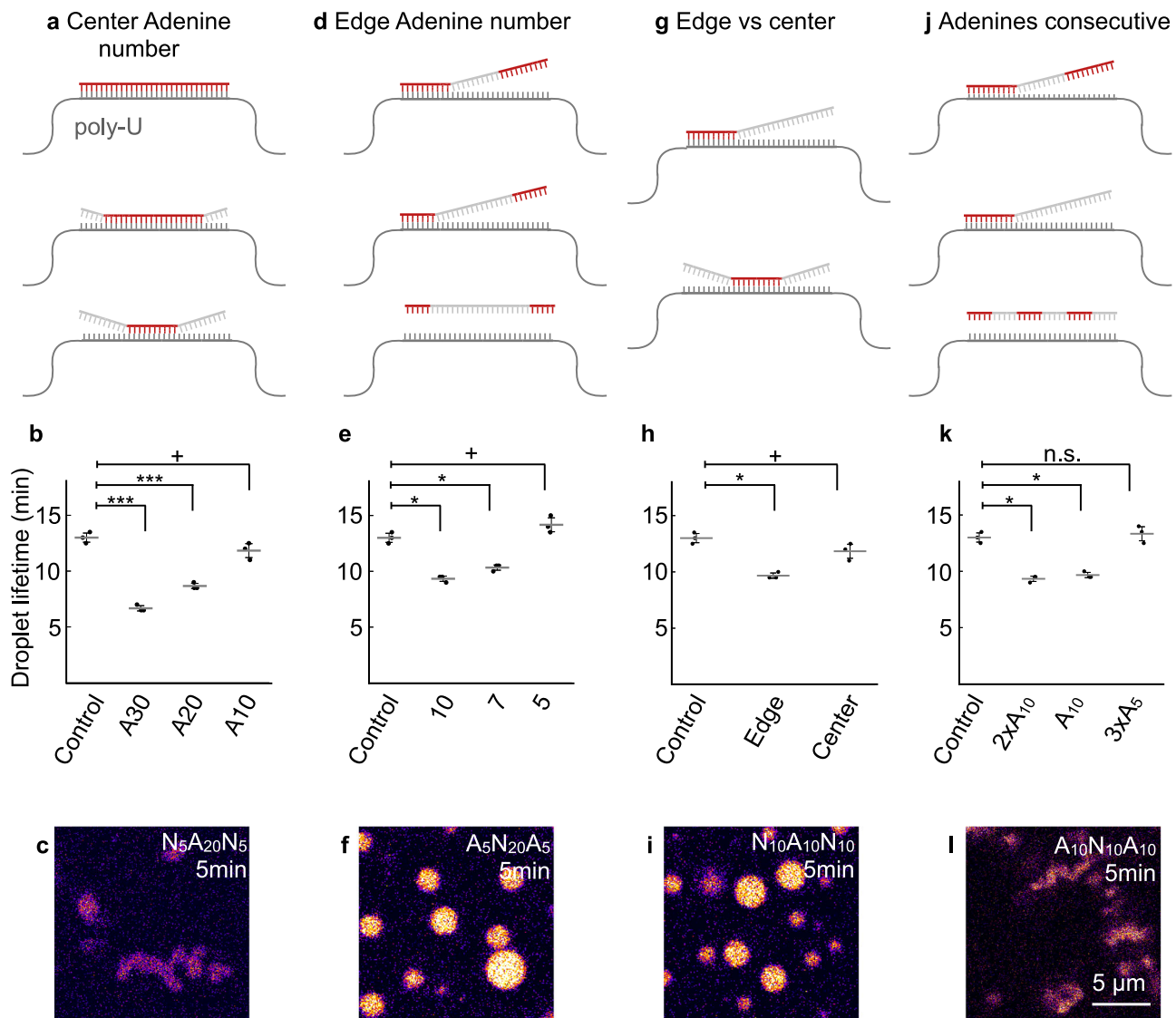


Fig. 5 | Molecular design rules for genotype-phenotype relations for adenine. Schemes of molecular design of different genotypes and their interaction with poly-U, red sections represent adenine stretches on an otherwise unbiased library strand (gray) (**a**, **d**, **g**, **j**). The lifetimes of droplets without a genotype and in the presence of specific sequences, determined by the OD 600 against time (**b**, **e**, **h**, **k**). All error bars show the standard deviation from the average ($N = 3$). N represents

one biological replicate of one analyzed turbidity trace. Significance was determined via a two-sided Welch's t -test with the no-DNA sample as the control group. P -values are available in Supplementary Table 4. Confocal micrograph in the presence of selected genotypes imaged 5 min after fuel addition on the glass using 0.5 mM NBD labeled peptide for staining (**c**, **f**, **i**, **l**). Source data are provided as a Source Data file.

difference in the droplet phenotype between A_{30} and the A-rich DNA library implies that the genotype-phenotype coupling is sensitive. Not all A-rich sequences change the phenotype. Guided by our sequencing results, we set out to test the molecular design rules that dictate how DNA sequences affect the genotype.

Guided by the sequencing data, we investigated the effects of the position and length of an adenine stretch within the sequence. First, we tested the impact of homopolymeric adenine stretches on the droplet phenotype by comparing sequences containing stretches of ten, twenty, or thirty consecutive adenines positioned centrally within otherwise random oligomers (Fig. 5a). The introduction of ten adenines resulted in a marginal reduction of the droplet's lifetime. In comparison, twenty and thirty adenines led to a more pronounced decrease (Fig. 5b). This data suggests that longer adenine stretches enhance hybridization with poly-U, thereby decreasing droplet stability and changing morphology (Figs. 5c, S15). Interestingly, a central stretch of 20 adenines had a more substantial effect than a sequence

with 76% adenine content distributed randomly, despite both having comparable total A content. This finding suggests that not only the base composition but also the local sequence arrangement matters, *i.e.*, consecutiveness is more disruptive than random distribution.

Next, we designed DNA strands with adenine stretches located at both termini (Fig. 5d). We found that a total of ten adenines (five at each end) had no significant effect on the droplet lifetime or morphology, despite the presence of five consecutive adenines being a motif observed in the sequencing (Fig. 5e, f). Increasing the total number of terminal adenines to fourteen (seven per end) reduced the lifetime to 10.3 min, while twenty terminal adenines (ten per end) further shortened it to 9.3 min and also decreased fusion rates (Figs. 5e, S16). These results reinforce the observation that, irrespective of positioning, longer adenine stretches enhance interaction with poly-U and destabilize the droplets. In contrast, shorter stretches below a critical threshold of seven are insufficient to induce any phenotypical changes.

Having established that adenine stretches promote hybridization when positioned either centrally or at the termini of the sequence, we directly compared the impact of placing ten adenines at the ends versus in the middle (Fig. 5g). Terminal placement resulted in a more pronounced reduction in droplet lifetime—9.7 minutes—compared to 12.2 min for the central configuration (Figs. 5h, i, S17).

To assess the importance of uninterrupted consecutiveness, we compared sequences containing multiple stretches of consecutive As (Fig. 5j). This comparison revealed no significant difference in lifetime or morphology when adding a second stretch of ten consecutive adenines unconnected to the first (Figs. 5k, l, S18, Supplementary Table 4). We conclude that increasing the number of adenines while interrupting the consecutive stretch does not exacerbate the adverse effect further. The second stretch of adenines is most likely unable to bind to the poly-U strand due to the central sequence mismatch, and there is no crosslinking between the strands. We observe the same lack of impact when examining a sequence with fifteen adenines distributed in three blocks of five. We cannot circumvent the threshold of seven consecutive As by distributing stretches of five consecutive As. This suggests that phenotype modulation requires not merely interaction, but sufficiently tight and continuous binding to affect the droplet material.

Taken together, to induce phenotypic changes, at least 7 consecutive adenines are required, preferably at the edge of the sequence. This requirement likely accounts for the minimal lifetime shortening of the A-rich library. With this knowledge, we calculated, using a numerical sequence generation method, that roughly one-third of sequences in the A-rich library have seven or more consecutive adenines located within one base of either end. Introducing an alternative condition of twenty consecutive adenines anywhere in the sequence, we find that only 1.4% of sequences are compliant. This analysis suggests that only about one-third of the sequences in the biased random population could hybridize with the droplet scaffold, accounting for the weak effect on droplet properties. Repeating the calculation for the entirely random N₃₀ library yielded just 0.02% of sequences fulfilling either condition, in line with the low observed partitioning and the absence of any measurable influence on droplet behavior.

To assess whether specific G-rich sequences affect the phenotype, we investigated the effects of a sequence composed of eight guanines at both termini flanking a randomized central region of fourteen bases (G₈N₁₄G₈). Additionally, we included a G₁₀A₂₀ sequence in our analysis to examine the combined effects of guanine and adenine stretches on droplet behavior. The twenty adenines are expected to enable hybridization with poly-U, while the guanine stretch may introduce additional interactions.

Excitingly, the OD600 data of the active droplets with a genotype showed evidence of a favorable phenotype i.e., with 50 μM of G₈N₁₄G₈, the droplets exhibited increased longevity. In fact, the droplets no longer dissolved (Fig. 6a, b, Supplementary Table 5). We hypothesized that G₈N₁₄G₈ irreversibly sequestered the peptide within the droplets, as guanine exhibits a high affinity for arginine and adopts higher-order structures, such as G-quadruplexes, potentially enabling network formation^{62,63}. Guanine-arginine contacts are abundant in RNA- and DNA-binding proteins due to pairwise hydrogen bonding as well as favorable cation- π stacking^{64–67}. In contrast, droplets with G₁₀A₂₀ dissolved showed a similar longevity as the unbiased library. The absence of kinetic trapping in G₁₀A₂₀ may result from the opposing effects of adenine-induced droplet destabilization counteracting guanine-mediated retention. Isothermal titration calorimetry showed a substantial difference in the dissociation constant of the peptide for both G₈N₁₄G₈ and G₁₀A₂₀ compared to the unbiased library and A₃₀, verifying the interaction between the G-rich sequences and the peptide (Fig. S19).

We tested whether the DNA's higher affinity towards the peptide would lower the critical fuel concentration and increase the critical salt

concentration (Fig. 6c, d, Supplementary Table 6). We found no difference in the critical fuel concentration required to form droplets between G₈N₁₄G₈ and the no-DNA control (8.3 vs. 9.2 mM) and an increase for G₁₀A₂₀ to 17.5 mM, consistent with the observed marginal reduction in droplet lifetime. In contrast, G₈N₁₄G₈ doubled the tolerance for additional sodium chloride, while G₁₀A₂₀ increased it even further to over 1 M. Nevertheless, the addition of salt decreased most of the turbidity with less than 100 mM additional salt, suggesting that the remaining turbidity may stem from DNA-peptide aggregates that are hard to dissolve due to the higher affinity of arginine peptides for guanine⁶⁸. The combination of an unchanged critical fuel concentration and the increased critical salt concentration required to dissolve the droplets suggests that the overall affinity of droplet components is not significantly enhanced. Despite relatively high partitioning (Fig. S9), the main droplet components, peptide and poly-U, do not change their affinity. Instead, the droplets that form under the given anhydride conditions appear to be kinetically trapped. We hypothesize this is due to restricted peptide diffusion within the dense network. If the deactivated peptide cannot diffuse freely out of these aggregates, droplet dissolution would be impeded.

Furthermore, FRAP measurements confirmed our hypothesis. The peptide diffusivity decreased in the presence of both DNA sequences (Fig. 6e, f, Supplementary Table 11), whereas the poly-U diffusivity remained essentially unchanged in the presence of G₈N₁₄G₈, but was drastically reduced in the presence of G₁₀A₂₀ (Fig. S20, Supplementary Table 11). We conclude that both sequences reduce peptide mobility. In contrast, only the sequence containing adenines (G₁₀A₂₀) additionally restricts poly-U dynamics, consistent with hybridization-induced rigidification.

Confocal microscopy was used to investigate the morphological changes of the droplets due to their genotypes. With G₈N₁₄G₈, we observed aggregates, incomplete fusion, and large network structures (Fig. 6g, h) consistent with previous reports of poly-arginine aggregation with guanine-rich RNA⁶³. After ten minutes, vacuoles formed—common during fuel starvation when droplets are large^{24,45}—but unlike the control, these droplets retained a persistent shell and resisted dissolution. We attribute the vacuolization to the peptide deactivation outpacing peptide influx due to slowed peptide diffusion in the presence of DNA, causing core collapse and kinetic trapping. G₁₀A₂₀ showed a mixed morphology between A₃₀ and G₈N₁₄G₈, i.e., the droplets did not fuse properly and formed pearl-like aggregate strings similar to those with A₃₀, while the complete dissolution is delayed relative to A₃₀ (Fig. S21).

These combined results show that specific G-rich DNA motifs can markedly modulate droplet behavior. Sequences with guanine-rich termini induced kinetic trapping and network formation, likely due to strong peptide–guanine interactions. In contrast, a hybrid sequence G₁₀A₂₀ did not show trapping but displayed features of both G- and A-rich effects, including delayed dissolution and impaired fusion.

These combined results demonstrate that DNA sequences, as a primitive genotype, influence the behavior of our fuel-dependent synthetic cells. Next, we aimed to demonstrate that these benefits impact the synthetic cells in periodic fueling-starvation experiments. We found that droplets with a G-rich DNA genotype formed kinetically trapped, semi-fused shells. Upon refueling, we observed the regrowth of the semi-fused droplet network at the beginning of the cycle. This morphology differs from that observed upon the first fuel addition and from the refueling of regular droplets without DNA (Figs. 6h, S22). We conclude that our primitive genotype can influence the droplets over multiple rounds of fueling and starvation.

Discussion

We investigated whether short DNA oligomers can function as rudimentary genotypes within fuel-dependent peptide- and RNA-based coacervate droplets. We screened random libraries, biased libraries,

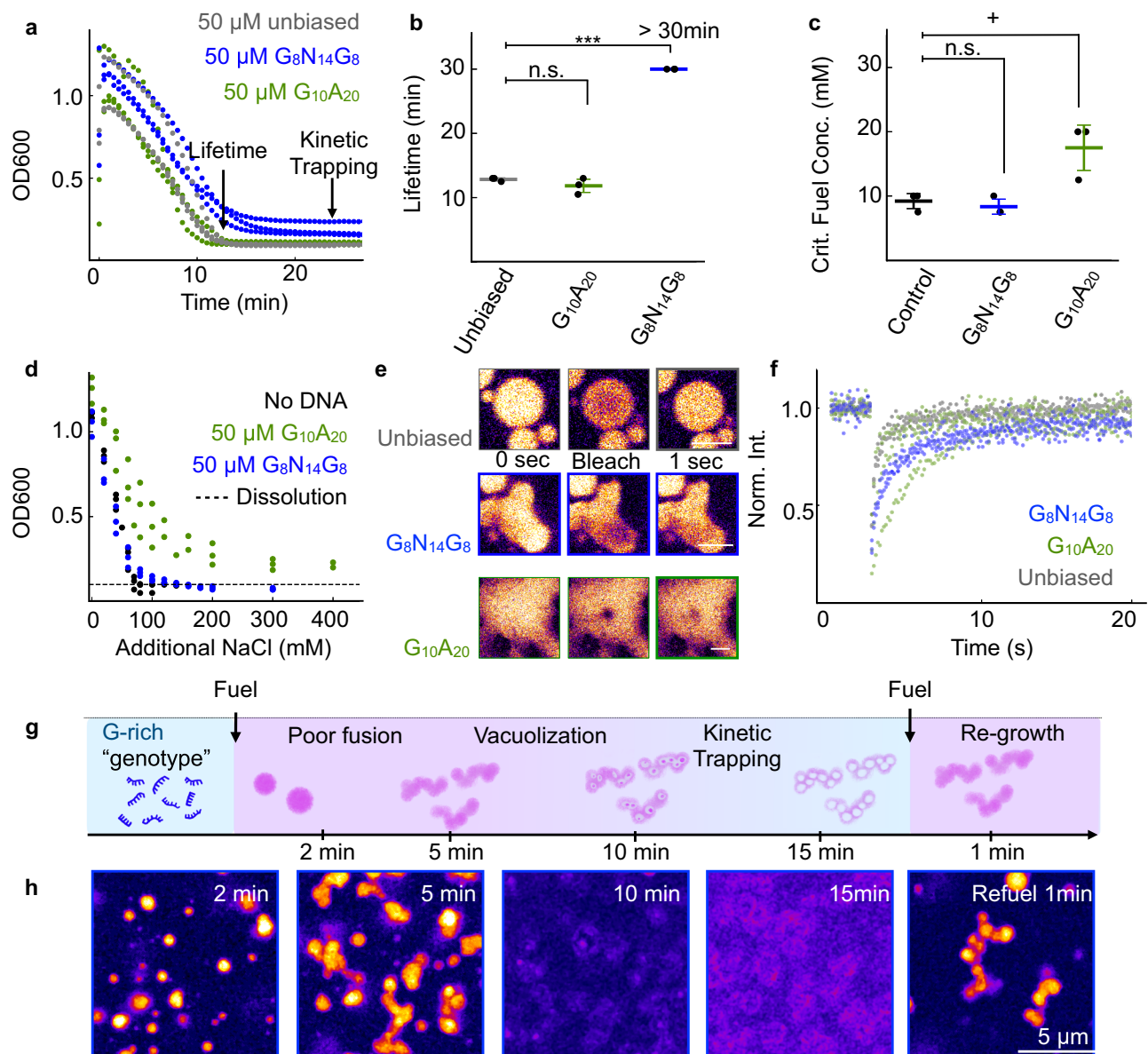


Fig. 6 | Kinetic trapping of coacervate droplets is facilitated by guanine stretches within DNA sequences. **a** Optical density at 600 nm (OD600) as a function of time for standard conditions in the presence of 50 μ M DNA sequences containing guanine stretches (8 guanines on both sides: blue, 10 guanines on one side: green). **b** The lifetimes in the presence of $G_8N_{14}G_8$ (blue) and $G_{10}A_{20}$ (green) compared to the unbiased DNA (gray), as derived from the turbidity traces measured on the UV spectrometer. **c** The critical fuel concentration in the presence of G-rich DNAs ($G_8N_{14}G_8$; blue; $G_{10}A_{20}$; green) was recorded as the onset of turbidity on the UV-VIS spectrometer during titration with fuel and compared to a no DNA control sample (black). **d** The critical salt concentration in the presence of G-rich DNAs derived from the turbidity on the UV Vis spectrometer upon addition of salt to static droplets. ($G_8N_{14}G_8$; blue; $G_{10}A_{20}$; green; no DNA: black) **e, f** Confocal micrographs and

FRAP recovery traces of the NBD- tagged precursor peptide in the presence of G-rich DNA ($G_8N_{14}G_8$; blue; $G_{10}A_{20}$; green) or an unbiased library (gray) measured on the confocal microscope. Scalebars represent 10 μ m. **g** Schematic representation of G-rich DNA kinetically trapping the assemblies and changing the morphology upon refueling. **h** Confocal micrographs imaged with 500 nM Sulforhodamine in the presence of $G_8N_{14}G_8$ at select timepoints in a fuel cycle and upon a second fuel addition. Scalebars represent 5 μ m. All error bars show the standard deviation from the average ($N = 3$). N represents one biological replicate of one analyzed turbidity trace. Significance was determined via a two-sided Welch's t -test with the unbiased DNA sample as the control group (**b**) or the no DNA samples as the control group (**c**). P -values can be found in Supplementary Tables 5 and 6. Source data are provided as a Source Data file.

and specific DNA sequences for their ability to partition in the droplets, which is crucial for acting as a genotype. We found through sequencing that A-rich and G-rich sequences, especially when these bases were consecutive, partitioned well in the droplets. We tested how these well-partitioning sequences affected key phenotypic traits, including longevity, morphology, fusion, critical fuel needed to form droplets, and the amount of salt needed to dissolve the droplets. On the one hand, A-rich sequences weakened the interactions between RNA and peptide, which was adverse for the droplet's phenotype—they

lived shorter, required more fuel to form, and were less salt resistant. Despite their negative impact on the droplet, we were able to identify these sequences in the droplets, likely because we isolated the droplets early in the cycle. On the other hand, G-rich sequences, particularly those with guanine stretches at their termini, prolonged the droplet's longevity, even to a point where they did not dissolve. These droplets resisted dissolution despite the absence of fuel, pointing to strong interactions that hindered peptide diffusion. Interestingly, sequences combining G and A motifs exhibited mixed phenotypes, balancing

trapping and hybridization. Taken together, our results demonstrate the first steps toward coupling genotype and phenotype in fuel-driven synthetic cells. We believe that hybridization, rigidification, and diffusion inhibition, as interaction modes, are also transferable to other systems of active droplets.

We recognize that the DNA sequences in this study do not replicate and, therefore, cannot be considered true genotypes in a biological sense. Nevertheless, our results show that if such sequences could replicate, they could exert a strong influence on droplet behavior. In the context of periodic fueling–starvation experiments, droplets with “favorable” sequences—those that prolong lifetime or increase resilience to low fuel or high salt—would be more likely to survive a starvation period and resume growth when fuel becomes available again, as demonstrated in the refueling experiments. If these droplets were also capable of division, as we have recently demonstrated in related systems, these favorable sequences could be inherited by daughter droplets, allowing the persistence and spread of advantageous traits¹². This scenario points toward a future in which integrating replicating genetic elements into fuel-dependent synthetic cells could enable true Darwinian evolution, with selection acting on genotype-defined phenotypes. To achieve that scenario, however, we would need our DNA sequences to replicate, which we envision is possible with rolling circle amplification or lesion-induced DNA amplification^{38,69}.

In summary, we demonstrate that short, non-replicating DNA oligomers can partition into fuel-driven, peptide–RNA coacervate droplets and modulate key phenotypic traits such as longevity, morphology, fusion behavior, and tolerance to fuel or salt stress. Systematic screening and sequencing revealed that consecutive adenine or guanine stretches—especially at sequence termini—drive strong partitioning, with A-rich sequences weakening peptide–RNA interactions and shortening lifetimes, while G-rich sequences induce kinetic trapping and markedly extend persistence. Our findings establish design rules for coupling sequence identity to droplet phenotype. We envision that, once combined with self-replicating DNA, these principles could enable droplets to survive starvation, endure environmental stress, and pass advantageous traits to progeny upon division—laying the groundwork for synthetic cells capable of Darwinian evolution.

Methods

Materials

We purchased 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), polyuridylic acid, potassium salt (poly-U), 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer, sodium chloride, polystyrene sulfonate (17 kDa), triethylammonium acetate buffer, different amino acids, and Sulforhodamine from Sigma-Aldrich and used them without any further purification. General solvents were purchased from Sigma Aldrich in analytical or synthesis grade and used without further purification. Cy5-A₁₅ was purchased from biomers.net GmbH. Peptides were either ordered from Caslo or Biocat as the lyophilized TFA salt with a purity of >98% or synthesized on a peptide synthesizer according to the method described below. Self-synthesized peptides were characterized before use with electrospray ionization mass spectrometry (ESI-MS) in positive mode, and all peptides were characterized and their concentration verified by analytical HPLC as described in Supplementary Table 8. For library preparation, we used PEG-8000 from Carl Roth, RNase Cocktail from Invitrogen, and T4 DNA ligase, T4 pkn, NEBNext Ultra II Q5 Master Mix, NEBNext Multiplex Oligos for Illumina (Index Primer Set 3), and Monarch Spin PCR & DNA cleanup kit from New England Biolabs, NucleoMag NGS Clean-up and Size Selection beads from Macherey-Nagel, and PhiX Control v3 from Illumina. All DNA oligomers used in this study were ordered from IDT or Sigma Aldrich with standard desalting. Sequences are found in Supplementary Table 10 in the Supplementary Information.

Peptide synthesis and purification. Ac-F(RG)₃D-OH (Ac-Phe-Arg-Gly-Arg-Gly-Arg-Gly-Asp-OH) was synthesized on a 0.5 mmol scale with a Liberty Blue® peptide synthesizer from CEM by standard Fmoc-SPPS. Pre-loaded Wang resin (Fmoc-Asp(OtBu) loaded, 100–200 mesh, 0.67 mmol/g) was used. A double Fmoc-deprotection and double coupling cycle with microwave heating was applied. Therefore, Fmoc-removal was performed twice per amino acid (2 × 5 mL piperidine 20 v/v% solution in DMF, 90 °C, 90 s) before each amino acid was coupled twice (2 × 5 mL of 0.2 M amino acid in DMF, 2 × 2 mL 0.5 M DIC in DMF, 2 × 1 mL 1 M Oxyma pure in DMF, 90 °C, 120 s each cycle). Arginine couplings were performed for 240 s at 90 °C with the same reagent concentrations. Phenylalanine was used pre-acetylated. Sidechain deprotection and final cleavage from the resin were achieved with a solution of 2.5% water, 2.5% triisopropylsilan, and 95% trifluoroacetic acid for two hours under continuous agitation at room temperature. Solvents were evaporated under a constant nitrogen flow, and the residue was dissolved in a mixture of water and acetonitrile. Purification was done by preparative reversed-phase HPLC on an Agilent 1260 Infinity II setup (Agilent InfinityLab ZORBAX SB-C18 column (250 mm×21.2 mm, 5 μm particle size). Separation was performed on a linear gradient from 2 % to 98% acetonitrile in Mili Q water, both with 0.1% TFA at a flow rate of 20 ml/min. The fluorescently labeled peptide with NBD was prepared by reacting the peptide on the resin with two equivalents of NBD-Cl and 1.2 equivalents of DIPEA in 4 ml DMF overnight under constant agitation. Cleavage and purification was performed as described above, but with 3 hours of shaking at room temperature during the cleavage step⁷⁰.

Sample preparation. We prepared stock solutions of the peptide and peptide model by dissolving the peptides in MQ water, after which we adjusted the pH to pH 5.3. Stock solutions of EDC were prepared by dissolving the EDC powder in MQ water. We prepared the stock solutions of 0.5 M EDC freshly. Reaction cycles were started by adding the highly concentrated EDC to the acid solution. We carried out all experiments at 25 °C. Standard active droplet conditions were 23 mM peptide, 4.1 mM poly-U (monomer conc.), 30 mM EDC, 200 mM MES buffer (pH 5.3), and 50 μM DNA (strand conc.). Control samples without DNA were prepared with 5.6 mM poly-U (monomer conc.) to keep the overall charge constant. Standard static droplet conditions were 20 mM peptide, 3 mM peptide model, 4.1 mM poly-U, and 200 mM MES buffer (pH 5.3), and 50 μM DNA (strand conc.). Control samples without DNA were prepared with 5.6 mM poly-U (monomer conc.).

DNA partitioning. We determined the partitioning of different DNA sequences by measuring the DNA concentration in the supernatant after spinning down the droplets using analytical HPLC (HPLC Thermo Vanquish, Thermo DNAPac RP column particle size 4 μm 2.1 ×100 mm) with a gradient of acetonitrile in 100 mM triethylammonium acetate buffer and a flow rate of 0.3 ml/min at 60 °C column temperature. The gradient ran from 6.25% acetonitrile to 25 % acetonitrile in 8 minutes, with a pre-equilibration of 5 minutes.

A 20 μL static droplet sample was prepared and spun down for 10 minutes. 18 μL supernatant was removed from the sample and transferred into a screw cap HPLC vial with a plastic inlet and incubated for at least 30 minutes in the HPLC. Samples of the solutions were injected without further dilution (injection volume: 1 μL) and tracked with a UV/Vis detector at 260 nm, and for the Cy3-A₃₀ additionally at 550 nm. For analysis the Chromeleon HPLC analysis software (Version 7) was used. Calibrations were made for each tested DNA library at three different concentrations. Each spin-down experiment for the libraries was performed in triplicate, apart from the additional A₃₀ and G₈N₁₄G₈. DNA partitioning coefficients were then calculated using the droplet volume determined with confocal microscopy, as described below. The final error was calculated from the standard deviation of

the measured DNA supernatant concentrations and the standard deviation of the droplet volume determination using error propagation.

Sequencing Library preparation. Detailed information can be found in the accompanying protocol paper.

Sequencing Data analysis. A detailed method for this can be found in the corresponding protocol paper. All scripts used to generate the final data and figures from the raw data can be found at <https://github.com/hmutschler/Droplet-Sequencing-Analysis> (<https://doi.org/10.5281/zenodo.18597396>).

We opted for paired-end sequencing, which means each molecule is read twice during sequencing, once from each end. Due to the short length of the library (including adapters >100 nt), this means that every position in the sequence is read twice and greatly reduces error frequency. After adapter trimming and length filtering, the R2 reads were discarded, as these reads are not present in the droplets (R1 represents all oligos from the N₃₀ library, while R2 is the reverse complement), and R2 reads were not further analyzed.

PolyG tails due to two color sequencing chemistry were accounted for in data processing, where we first trimmed adapters, usually deleting everything downstream of the 2nd adaptor in each read, as well as when filtering for reads of exactly 30 nt length.

For the motifs, we decided to use an e-value of 0.05 as the cut-off for significance. The e-value corresponds to the p-value of these motifs appearing enriched at random, divided by the total number of motifs found by the STREME software.

ITC measurements. ITC titrations were performed on a MicroCal PEAQ-ITC instrument from Malvern Panalytical. All experiments were performed at 25 °C and with a control titration of the respective peptide concentration in 200 mM MES buffer (pH 5.3) into the same buffer without any DNA. All samples (cell and titration solutions) were prepared with the same concentration of MES buffer (200 mM at pH 5.3) to avoid additional dilution effects. The data was analyzed using a non-linear least squares algorithm provided by the PEAQ-ITC analysis software. The component concentrations were chosen in a regime where no coacervation could be observed.

UV/Vis Spectroscopy/OD measurements. The UV/Vis measurements were carried out using a Multiskan GO (Thermo Fisher) microplate reader, which measures top-down absorption. Samples (either 20 μ L or 50 μ L) were directly prepared into a 384 or 96-well half-area plate (flat bottom) at 25 °C. Each experiment was performed for 30 minutes with turbidity measurements at 600 nm every 30 s and in triplicate. Lifetimes were determined at a threshold of 0.12; we define the droplet longevity at a threshold of 0.2.

Critical Salt concentration determination. We determined the critical salt concentration by titrating a 20 μ L sample of static droplets with standard static droplet conditions with 0.5, 1, or 5 M sodium chloride solution in 10 or 20 mM steps. Volumetric dilution was neglected due to being <25% at the maximum, and the same for all samples. After each addition, the sample was mixed by pipetting up and down and left to incubate for 30 seconds until air bubbles had disappeared. Afterwards, the turbidity was measured at 25 °C. The titrations were performed in triplicate and each replicate was fit linearly. For the linear fit we use all absorption values > 0.12 to simplify fitting. We determined the critical fuel concentration as the average of the three concentrations where the extrapolated fits cross the threshold of 0.1. When dissolution was not linear and a linear fit was not possible due to aggregate formation (Fig. 6), we determine the critical salt concentration as the average of the first concentration with a measured absorption value < 0.1.

Critical Fuel concentration determination. We determined the critical fuel concentration by titrating a 20 μ L sample with standard active droplet conditions with a 500 mM EDC solution in 2.5 (<12.5 mM EDC) and 5 mM steps (> 12.5 mM EDC). We determined the critical fuel concentration as the average of the first concentration where the turbidity crosses above the threshold of 0.1.

Confocal Fluorescence Microscopy. We used a Leica SP8 confocal microscope with a 63x oil immersion objective to image the droplets at different times. We prepared samples as described above, but with 500 nM Sulforhodamine B or the fluorescently tagged peptide as dye. 30 μ L of the sample was deposited on a PVA-coated IBIDI μ -slide Angiogenesis Glass bottom slide. Samples were excited with a 552 nm laser and imaged at 565–635 nm with a HyD detector. Samples with the fluorescent peptide were excited with a 488 nm laser and imaged at 499–569 nm. The pinhole was set to 1 Airy unit. Images were taken either 20 μ m above the bottom or on the glass. Micrograph images were analyzed with Fiji (Image J) and contrast and brightness adjusted for better visibility, for Fig. 6 a median filter was used for better visibility.

Droplet volume determination. The total droplet volume was determined from Z-projections of static droplet samples prepared in water in oil microreactors. These microreactors were screened as a z-stack after the droplets had mostly merged. The Z-projections of these images were analyzed with Fiji (ImageJ). From the area of the individual coacervates their volume was determined and added for the total droplet volume. The total sample volume was determined by measuring the maximum diameter of the water in oil droplet. The ratio of these values is given in percent. A total of 10 microreactors were screened and the average and standard deviation taken from that. The samples were prepared as static droplets using the standard conditions for static droplets (20 mM peptide, 3 mM peptide model, 4.1 mM poly-U, 200 mM MES (pH 5.3)).

Statistics and Reproducibility. Measurements were taken from at least three distinct samples. For representative confocal images, experiments were performed at least three times. For all experiments where significance is indicated, we have performed a two-sided Welch's t-test for independent samples using SciPy in a Python (Version 3.13) script. A p-value below 0.05 is considered significant (*), a p-value of 0.1 is a slight trend (+). The degrees of freedom are calculated with the Welch-Satterthwaite equation. Tables with the exact p-values, t values, and calculated degrees of freedom can be found in the SI. Information on the significance of the cutoff during motif discovery can be found in the Sequencing Data analysis.

Data collection and analysis. To collect our data on the confocal, we used the Leica LAS X software Version 4.6, on the UV spectrometer, we used the SkanIt Software Version 6.0, on the HPLC, we used Chromeleon Version 7.3.2, and on the ITC, we used MicroCal PEAQ-ITC Control software(2022). The numerical generator can be found on GitHub at <https://doi.org/10.5281/zenodo.18597396>. For data analysis, we have used Fiji (ImageJ), Python (Version 3.13), Chromeleon Version 7.3.2, and Excel. The FRAP analysis script can be found on GitHub at <https://doi.org/10.5281/zenodo.18597396>.

Fluorescence recovery after photobleaching (FRAP). The diffusivity of droplet components inside the coacervates was calculated from fluorescence recovery after photobleaching experiments with the above-mentioned confocal in FRAP mode. We prepared static droplets with the standard conditions for static droplets mentioned above and imaged droplets close to the glass, waiting at least 5 min for the droplets to reach an adequate size. For the peptide diffusivity, we excited the fluorescent version of the precursor peptide, NBD-G(RG)₃D-OH, at

488 nm and detected it from 499 nm–569 nm with a PMT detector for at least 70 s. For the poly-U diffusivity, we annealed a short fluorescently tagged RNA oligomer, Cy5-A₁₅, to the poly-U and bleached at 638 nm, detecting from 650 nm–750 nm with a PMT detector for at least 180 s. Images were acquired at a resolution of 256 × 256 pixels, 700x scan speed (bidirectional scan), and between 8 and 10x zoom. The time intervals for the peptide bleach were 0.1342 s and 0.1902 s for the poly-U bleach. The bleach radius was chosen at 2.5 μm for the peptide and 1.0 μm for the poly-U and corrected for the calculation based on the post-bleach frame. Recovery data was double normalized using a Fiji macro script with the following equation:

$$F(t) = \frac{(I_t - B_t)(N_0 - B_0)}{(N_t - B_t)(I_0 - B_0)} \quad (1)$$

$F(t)$ represents the normalized fluorescence intensity at time point t . I_t is the average intensity of the bleached ROI at a specific time point, N_t is the average intensity of an unbleached ROI in a neighboring coacervate, and B_t represents the average intensity of a ROI without any coacervates. N_0 , I_0 , and B_0 are the averaged pre-bleach intensities of the unbleached ROI, bleached ROI, and the background until the bleach frame. We fitted the data with the following exponential equation:

$$F(t) = a(1 - e^{-bt}) + c \quad (2)$$

With a representing the mobile fraction or the fluorescence at full recovery, b being related to the half recovery time, and c being the minimum intensity upon bleaching.

$$t_{1/2} = \frac{\ln(2)}{b} \quad (3)$$

The diffusion coefficient is calculated via the following calculation according to Soumpasis⁷¹:

$$D = \frac{0.224 r^2}{t_{1/2}} \quad (4)$$

We have used the radius r from the postbleach frame; 0.224 is a numerically determined coefficient. We have determined the diffusion coefficient for each condition in triplicate.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw sequencing data generated in this study have been deposited in the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI): <https://www.ncbi.nlm.nih.gov/sra/PRJNA1424544>. The other experimental data generated in this study are provided in the Supplementary Information/Source Data file. Source images are available upon request to Job Boekhoven due to their size; permanent access can be given with a one-month response time. Source data are provided with this paper.

Code availability

All scripts used to generate the final data and figures from the raw sequencing data, as well as the numerical generator and the FRAP analysis script, can be found on GitHub at <https://doi.org/10.5281/zenodo.18597396>.

References

- Schwille, P. et al. MaxSynBio: avenues towards creating cells from the bottom up. *Angew. Chem. Int. Ed.* **57**, 13382–13392 (2018).
- Giaveri, S. et al. Building a synthetic cell together. *Nat. Commun.* **16**, 7488 (2025).
- Zhu, T. F. & Szostak, J. W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.* **131**, 5705–5713 (2009).
- Bhattacharya, A., Brea, R. J., Niederholtmeyer, H. & Devaraj, N. K. A minimal biochemical route towards de novo formation of synthetic phospholipid membranes. *Nat. Commun.* **10**, 300 (2019).
- Nakashima, K. K., van Haren, M. H. I., André, A. A. M., Robu, I. & Spruijt, E. Active coacervate droplets are protocells that grow and resist Ostwald ripening. *Nat. Commun.* **12**, 3819 (2021).
- Cho, C. J. et al. Protocells by spontaneous reaction of cysteine with short-chain thioesters. *Nat. Chem.* **17**, 148–155 (2025).
- Ji, P. et al. Photochemical synthesis of natural lipids in artificial and living cells. *Nat. Commun.* **16**, 5068 (2025).
- Dreher, Y., Jahnke, K., Bobkova, E., Spatz, J. P. & Göpfrich, K. Division and regrowth of phase-separated giant unilamellar vesicles. *Angew. Chem.* **133**, 10756–10764 (2021).
- Dreher, Y., Jahnke, K., Schröter, M. & Göpfrich, K. Light-triggered cargo loading and division of DNA-containing giant unilamellar lipid vesicles. *Nano Lett.* **21**, 5952–5957 (2021).
- Kohyama, S., Merino-Salomón, A. & Schwille, P. In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nat. Commun.* **13**, 6098 (2022).
- Reverte-López, M. et al. Self-organized spatial targeting of contractile actomyosin rings for synthetic cell division. *Nat. Commun.* **15**, 10415 (2024).
- Wenisch, M. et al. Toward synthetic life- Emergence, growth, creation of offspring, decay, and rescue of fuel-dependent synthetic cells. *Chem* **11**, 102578 (2024).
- Zambrano, P. et al. Chemically driven division of protocells by membrane budding. *J. Am. Chem. Soc.* **146**, 33359–33367 (2024).
- Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci. USA* **101**, 17669–17674 (2004).
- Nijemeisland, M., Abdelmohsen, L. K. E. A., Huck, W. T. S., Wilson, D. A. & van Hest, J. C. M. A Compartmentalized out-of-equilibrium enzymatic reaction network for sustained autonomous movement. *ACS Cent. Sci.* **2**, 843–849 (2016).
- Bose, R. et al. Protometabolically generated NADH mediates material properties of aqueous dispersions to coacervate microdroplets. *Biomacromolecules* **27**, 112–120 (2025).
- Abil, Z. et al. Darwinian evolution of self-replicating DNA in a synthetic protocell. *Nat. Commun.* **15**, 9091 (2024).
- Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387–390 (2001).
- Pross, A. Toward a general theory of evolution: Extending Darwinian theory to inanimate matter. *J. Syst. Chem.* **2**, 1 (2011).
- Adamala, K. & Szostak, J. W. Competition between model protocells driven by an encapsulated catalyst. *Nat. Chem.* **5**, 495–501 (2013).
- Liu, K. et al. Light-driven eco-evolutionary dynamics in a synthetic replicator system. *Nat. Chem.* **16**, 79–88 (2024).
- Kriebisch, C. M. E. et al. A roadmap toward the synthesis of life. *Chem* **11**, 102399 (2025).
- Soria-Carrera, H., Kauling, L. & Boekhoven, J. Primitive genotype-phenotype coupling in fuel-dependent synthetic cells with an autocatalyst. *Chem* **12**, 102816 (2026).
- Donau, C. et al. Active coacervate droplets as a model for membraneless organelles and protocells. *Nat. Commun.* **11**, 5167 (2020).
- Späth, F. et al. Molecular design of chemically fueled peptide–polyelectrolyte coacervate-based assemblies. *J. Am. Chem. Soc.* **143**, 4782–4789 (2021).
- Zozulia, O. et al. Acyl phosphates as chemically fueled building blocks for self-sustaining protocells. *Angew. Chem. Int. Ed.* **63**, e202406094 (2024).

27. Boekhoven, J., Hendriksen, W. E., Koper, G. J. M., Eelkema, R. & van Esch, J. H. Transient assembly of active materials fueled by a chemical reaction. *Science* **349**, 1075–1079 (2015).
28. Maiti, S., Fortunati, I., Ferrante, C., Scrimin, P. & Prins, L. J. Dissipative self-assembly of vesicular nanoreactors. *Nat. Chem.* **8**, 725–731 (2016).
29. Deng, J. & Walther, A. Programmable ATP-fueled DNA coacervates by transient liquid-liquid phase separation. *Chem* **6**, 3329–3343 (2020).
30. Sun, J. et al. Carbodiimide-driven dimerization and self-assembly of artificial, ribose-based amphiphiles. *Chem. A Eur. J.* **28**, e202104116 (2022).
31. Singh, A., Parvin, P., Saha, B. & Das, D. Non-equilibrium self-assembly for living matter-like properties. *Nat. Rev. Chem.* **8**, 723–740 (2024).
32. Drobot, B. et al. Compartmentalised RNA catalysis in membrane-free coacervate protocells. *Nat. Commun.* **9**, 3643 (2018).
33. Deshpande, S. et al. Spatiotemporal control of coacervate formation within liposomes. *Nat. Commun.* **10**, 1800 (2019).
34. Liu, J. et al. Hydrogel-immobilized coacervate droplets as modular microreactor assemblies. *Angew. Chem. Int. Ed.* **59**, 6853–6859 (2020).
35. Slootbeek, A. D., van Haren, M. H. I., Smokers, I. B. A. & Spruijt, E. Growth, replication and division enable evolution of coacervate protocells. *Chem. Commun.* **58**, 11183–11200 (2022).
36. Wang, J., Abbas, M., Wang, J. & Spruijt, E. Selective amide bond formation in redox-active coacervate protocells. *Nat. Commun.* **14**, 8492 (2023).
37. Nakashima, K. K. et al. Differential stability and dynamics of DNA-based and RNA-based coacervates affect non-enzymatic RNA chemistry. *Nat. Commun.* **16**, 9296 (2025).
38. Kausar, A., Mitran, C. J., Li, Y. & Gibbs-Davis, J. M. Rapid, isothermal DNA self-replication induced by a destabilizing lesion. *Angew. Chem. Int. Ed.* **52**, 10577–10581 (2013).
39. Hänle, E. & Richert, C. Enzyme-free replication with two or four bases. *Angew. Chem. Int. Ed.* **57**, 8911–8915 (2018).
40. Park, H., Parshotam, S., Hales, S. C., Mittermaier, A. K. & Gibbs, J. M. Minimizing product inhibition in DNA self-replication: insights for prebiotic replication from the role of the enzyme in lesion-induced DNA amplification. *Chem. A Eur. J.* **29**, e202300080 (2023).
41. Chen, X., Soria-Carrera, H., Zozulia, O. & Boekhoven, J. Suppressing catalyst poisoning in the carbodiimide-fueled reaction cycle. *Chem. Sci.* **14**, 12653–12660 (2023).
42. Hofacker, I. L. et al. Fast folding and comparison of RNA secondary structures. *Monatshefte für. Chem. Chem. Monthly* **125**, 167–188 (1994).
43. Mathews, D. H. et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **101**, 7287–7292 (2004).
44. Lorenz, R. et al. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
45. Bergmann, A. M. et al. Liquid spherical shells are a non-equilibrium steady state of active droplets. *Nat. Commun.* **14**, 6552 (2023).
46. Holtmannspötter, A.-L. et al. Regulating nucleic acid catalysis using active droplets. *Angew. Chem. Int. Ed.* **63**, e202412534 (2024).
47. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
48. Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).
49. Tareen, A. & Kinney, J. B. Logomaker: beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2020).
50. Wollny, D. et al. Characterization of RNA content in individual phase-separated coacervate microdroplets. *Nat. Commun.* **13**, 2626 (2022).
51. Aumiller, W. M. Jr., Pir Cakmak, F., Davis, B. W. & Keating, C. D. RNA-based coacervates as a model for membraneless organelles: formation, properties, and interfacial liposome assembly. *Langmuir* **32**, 10042–10053 (2016).
52. Arnott, S., Chandrasekaran, R., Millane, R. P. & Park, H.-S. DNA-RNA hybrid secondary structures. *J. Mol. Biol.* **188**, 631–640 (1986).
53. Hayashi, K. et al. Influence of RNA strand rigidity on polyion complex formation with block cationomers. *Macromol. Rapid Commun.* **37**, 486–493 (2016).
54. Kwok, A., McCarthy, D., Hart, S. L. & Tagalakakis, A. D. Systematic comparisons of formulations of linear oligolysine peptides with siRNA and plasmid DNA. *Chem. Biol. Drug Des.* **87**, 747–763 (2016).
55. Cakmak, F. P., Choi, S., Meyer, M. O., Bevilacqua, P. C. & Keating, C. D. Prebiotically-relevant low polyion multivalency can improve functionality of membraneless compartments. *Nat. Commun.* **11**, 5949 (2020).
56. Zumbro, E. & Alexander-Katz, A. Polymer stiffness regulates multivalent binding and liquid-liquid phase separation. *Biophys. J.* **119**, 1849–1864 (2020).
57. Boeynaems, S. et al. Spontaneous driving forces give rise to protein–RNA condensates with coexisting phases and complex material properties. *Proc. Natl. Acad. Sci. USA* **116**, 7889–7898 (2019).
58. Majumder, S., Coupe, S., Fakhri, N. & Jain, A. Sequence-encoded intermolecular base pairing modulates fluidity in DNA and RNA condensates. *Nat. Commun.* **16**, 4258 (2025).
59. Lytle, T. K. & Sing, C. E. Tuning chain interaction entropy in complex coacervation using polymer stiffness, architecture, and salt valency. *Mol. Syst. Des. Eng.* **3**, 183–196 (2018).
60. Shakya, A. & King, J. T. DNA local-flexibility-dependent assembly of phase-separated liquid droplets. *Biophys. J.* **115**, 1840–1847 (2018).
61. Ma, W., Zhen, G., Xie, W. & Mayr, C. In vivo reconstitution finds multivalent RNA–RNA interactions as drivers of mesh-like condensates. *eLife* **10**, e64252 (2021).
62. Wang, P. et al. Principles of amino-acid–ribonucleotide interaction revealed by binding affinities between homogeneous oligopeptides and single-stranded RNA molecules. *Nano Res.* **16**, 13294–13300 (2023).
63. Vidal Ceballos, A., Geissmann, A., Favaro, D. C., Deshpande, P. & Elbaum-Garfinkle, S. RNA guanine content and G-quadruplex structure tune the phase behavior and material properties of biomolecular condensates. *Sci. Rep.* **15**, 9295 (2025).
64. Luscombe, N. M., Laskowski, R. A. & Thornton, J. M. Amino acid–base interactions: a three-dimensional analysis of protein–DNA interactions at an atomic level. *Nucleic Acids Res.* **29**, 2860–2874 (2001).
65. Egli, M. & Zhang, S. Ned Seeman and the prediction of amino acid-basepair motifs mediating protein–nucleic acid recognition. *Bio-phys. J.* **121**, 4777–4787 (2022).
66. Sarmah, M. P. & Sarma, M. Mechanistic insights into the electron attachment process to guanosine in the presence of arginine. *Phys. Chem. Chem. Phys.* **26**, 27955–27963 (2024).
67. Adlhart, M., Hoffmann, D., Polyansky, A. A. & Žagrović, B. Coding relationship links RNA G-quadruplexes and protein RGG motifs in RNA-binding protein autoregulation. *Proc. Natl. Acad. Sci. USA* **122**, e2413721122 (2025).
68. Lu, T., Nakashima, K. K. & Spruijt, E. Temperature-responsive peptide–nucleotide coacervates. *J. Phys. Chem. B* **125**, 3080–3091 (2021).
69. Sakatani, Y., Yomo, T. & Ichihashi, N. Self-replication of circular DNA by a self-encoded DNA polymerase through rolling-circle replication and recombination. *Sci. Rep.* **8**, 13089 (2018).

70. Späth, F. et al. The role of chemically innocent polyanions in active, chemically fueled complex coacervate droplets. *Angew. Chem. Int. Ed.* **62**, e202309318 (2023).
71. Soumpasis, D. M. Theoretical analysis of fluorescence photo-bleaching recovery experiments. *Biophys. J.* **41**, 95–97 (1983).

Acknowledgements

We thank Monika Wenisch for providing the image analysis Macro used to analyze the FRAP data. The Mutschler lab is grateful for support from the European Research Council through the ERC Synergy Grant 101166888 (H.M.). The BoekhovenLab is grateful for support from the TUM Innovation Network – RISE, funded through the Excellence Strategy (J.B.). This research was conducted within the Max Planck School Matter to Life, supported by the German Federal Ministry of Education and Research (BMBF) in collaboration with the Max Planck Society (J.B.). Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094 – 390783311 (J.B.). The BoekhovenLab is grateful for support from the European Research Council through ERC Starting Grant 852187 and ERC Consolidator Grant 101124380 (J.B.).

Author contributions

C.M. and A.-L.H. both did experimental work, data analysis, conceptual planning, and writing, contributed equally, and have the right to list their names first in their CV or any bibliographic list. J.B. and H.M. supervised the research. The manuscript was written through the contributions of all authors. All authors have approved the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-026-71024-8>.

Correspondence and requests for materials should be addressed to Hannes Mutschler or Job Boekhoven.

Peer review information *Nature Communications* thanks Alexandre Cheng and the other anonymous reviewers for their contribution to the peer review of this work. [A peer review file is available].

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

Publications

9.4 Publication 4

Recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets.

Holtmannspötter, A.L.*; **Machatzke, C.***; Boekhoven, J.; Mutschler, H.⁺

* Authors contributed equally

⁺ Corresponding author

STAR protocols **2026**, 7(1), 104293

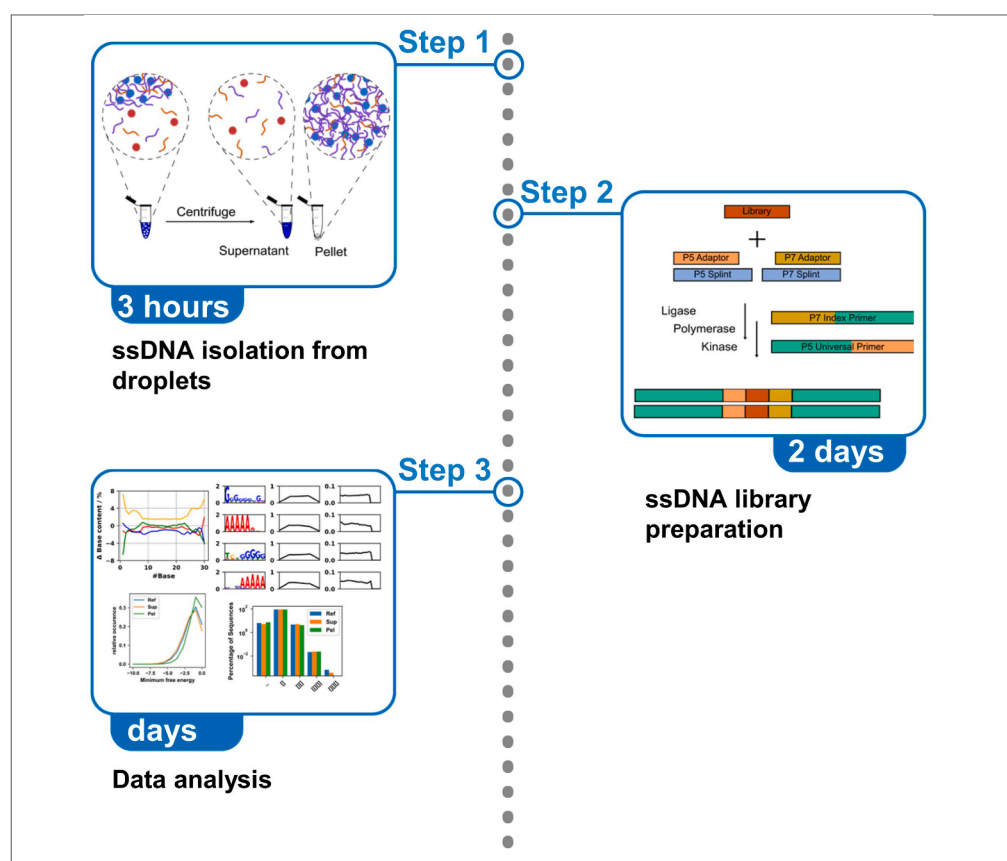
DOI: 10.1016/j.xpro.2025.104293

9.4.1 Personal Contribution

- Designed the sequencing method
- Performed all experiments in this paper
- Performed data analysis and wrote relevant code
- Prepared the figures
- Composed, edited and revised the manuscript with H. Mutschler

Protocol

Protocol for the recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets



Complex coacervate droplets are synthetic cell models that sequester nucleic acids. Here, we present a protocol for the recovery and deep sequencing of single-stranded DNA (ssDNA) from metabolically active coacervate droplets. We describe time-resolved harvesting, ssDNA recovery, and sequence-independent library preparation for NGS, along with an analysis pipeline to assess enrichment dynamics and sequence distributions. The protocol is adaptable to diverse droplet systems.

Publisher's note: Undertaking any experimental protocol requires adherence to local institutional guidelines for laboratory safety and ethics.

Anna-Lena
Holtmannspötter,
Corbin Machatzke,
Job Boekhoven,
Hannes Mutschler

corbin.machatzke@
tu-dortmund.de (C.M.)
job.boekhoven@tum.de
(J.B.)
hannes.mutschler@
tu-dortmund.de (H.M.)

Highlights

Protocol to prepare
and harvest ssDNA
from fueled complex
coacervates

Sequence-
independent,
orientation-specific
ssDNA library
preparation for NGS

Comprehensive data
analysis pipeline to
standardize results
across research
groups

Holtmannspötter et al., STAR
Protocols 7, 104293
March 20, 2026 © 2025 The
Author(s). Published by
Elsevier Inc.
<https://doi.org/10.1016/j.xpro.2025.104293>



Protocol

Protocol for the recovery and deep sequencing of short ssDNA pools from transient, fuel-dependent coacervate droplets

Anna-Lena Holtmannspötter,^{2,3} Corbin Machatzke,^{1,3,4,*} Job Boekhoven,^{2,*} and Hannes Mutschler^{1,5,*}

¹TU Dortmund University, Otto-Hahn-Strasse 4a, 44227 Dortmund, Germany

²Department of Bioscience, School of Natural Sciences, Technical University of Munich, Lichtenbergstrasse 4, 85748 Garching, Germany

³These authors contributed equally

⁴Technical contact

⁵Lead contact

*Correspondence: corbin.machatzke@tu-dortmund.de (C.M.), job.boekhoven@tum.de (J.B.), hannes.mutschler@tu-dortmund.de (H.M.)
<https://doi.org/10.1016/j.xpro.2025.104293>

SUMMARY

Complex coacervate droplets are synthetic cell models that sequester nucleic acids. Here, we present a protocol for the recovery and deep sequencing of single-stranded DNA (ssDNA) from metabolically active coacervate droplets. We describe time-resolved harvesting, ssDNA recovery, and sequence-independent library preparation for next generation sequencing (NGS), along with an analysis pipeline to assess enrichment dynamics and sequence distributions. The protocol is adaptable to diverse droplet systems.

For complete details on the use and execution of this protocol, please refer to Machatzke et al.¹

BEFORE YOU BEGIN

Rational for method development

We recently developed fuel-dependent complex coacervate droplets sustained by a chemical reaction cycle.^{2,3} These non-equilibrium compartments require continuous fuel to form and persist, and will decay without it.⁴ Under defined conditions, they can undergo growth-division cycles.⁵ While they provide a powerful platform to study life-like, out-of-equilibrium behavior, these droplets lack a heritable genotype, which would be required for them to undergo Darwinian evolution. We envision that self-replicating DNA sequences could serve as a genotype if sequence-dependent interactions modulate droplet properties.⁶ To test this hypothesis, we established a deep-sequencing workflow that links DNA sequence to droplet-level phenotypes, enabling high-throughput mapping of sequences that bias droplet formation, stability and dynamics.

Specifically, we developed a protocol to retrieve, identify, and analyze ssDNA molecules that partition into the polymer-rich phase of a transient, fuel-driven complex coacervate system and which ssDNA molecules remain in the polymer-poor phase. Because sequence determinants of partitioning are difficult to predict, this protocol enables experimental determination of enrichment and depletion rules and dynamics.^{7–9} The workflow is compatible with other synthetic and protocellular model systems, provided that ssDNA recovery and library preparation are adjusted accordingly. The protocol can also be adapted to RNA-based systems, if the extraction is modified and strand-aware library steps follow cDNA by reverse transcription. After ssDNA recovery, we present a strand-specific, sequence-agnostic library preparation method for deep



sequencing. Finally, we provide a modular analysis pipeline tailored to short (e.g., prebiotically plausible) sequences,^{10–13} quantifying nucleotide composition, predicted intramolecular structure, and motif enrichment.

Innovation

We have developed a method for the time-resolved recovery and deep sequencing of ssDNA from transient, fuel-driven complex coacervate droplets.^{2,3} Because these droplets have limited lifetimes, we have implemented a rapid quenching and harvesting protocol to capture polymer-rich and polymer-poor sequence pools with minimal bias. To generate the deep-sequencing library, we adapted the SRSly (single reaction single-stranded library) protocol¹⁴ for use with low amounts of unknown ssDNA input, thereby increasing yields while preserving strand orientation and minimizing sequence bias.

On the computational side, we offer a modular and customizable analysis pipeline tailored to short oligonucleotides, typical of synthetic coacervate droplets (~10–100 nt). The modules report quality metrics, nucleotide composition, intramolecular folding, and motif discovery. The pipeline also promotes standardization across laboratories working on synthetic/protocell systems and can be extended to alternative nucleic acid backbones such as RNA with minimal changes to recovery and library preparation.

Taken together, the protocol provides a reproducible end-to-end workflow for identifying and analyzing sequence determinants of partitioning into static and metabolically active droplet systems, as well as related compartments.

Buffer preparation

⌚ Timing: 30 min to 1 h

1. Prepare 500 mM MES pH 5.3 stock solution.
 - a. Dissolve 53.3 g MES monohydrate in 400 mL nuclease-free ddH₂O.
 - b. Adjust the pH to 5.3, then top up to 500 mL and mix thoroughly before use.
2. Prepare a 41 mM polyU (charge concentration/monomer concentration) stock solution.
 - a. Dissolve 14.15 mg polyuridylic acid potassium salt in 1 mL nuclease-free ddH₂O.
3. Prepare 155 mM RG₃ peptide stock solution.
 - a. Dissolve 500 mg Ac-FRGRGRGD in 1 mL nuclease-free ddH₂O.
 - b. Adjust pH with 1 M NaOH to pH 5.3 (~475 μ L).
 - c. Add nuclease-free ddH₂O until the total volume is equal to 2.475 mL.

⚠ CRITICAL: Do not prepare a stock solution with a peptide concentration that is too high! If the concentration is too high, the peptide will precipitate over time.

⚠ CRITICAL: Immediately adjust the pH of the peptide solution.

⚠ CRITICAL: Always verify the quality of custom peptides. Identity and purity can be confirmed by mass spectrometry and HPLC. Demonstrate correct formation of complex coacervates with no significant precipitation under the conditions described in the detailed protocol.

⚠ CRITICAL: Always store RNA oligonucleotides at –20°C or –80°C and thaw them on ice. If not stored properly, RNA can hydrolyze, leading to insufficient droplet formation.

4. Prepare a 5 M NaCl stock solution.
 - a. Dissolve 29.22 g NaCl in 70 mL nuclease-free ddH₂O.

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SRSly P7 splint:/5AmMC12/GT GAC TGG AGT TCA GAC GTG TGC TCT TCC GAT CTN NNN NNN/3AmMO/	Integrated DNA Technologies	N/A
SRSly P5 adapter Cy5 tagged (for troubleshooting):/5Cy5/ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT	Integrated DNA Technologies	N/A
SRSly P7 adapter FAM tagged (for troubleshooting):/5Phos/AG ATC GGA AGA GCA CAC GTC TGA ACT CCA GTC A/36-FAM/	Integrated DNA Technologies	N/A
Tagged RNA:/5Cy5/rArArArArCrCrArGrUrC	Integrated DNA Technologies	N/A
Software and algorithms		
Illumina Sequencing Analysis Viewer 2.4.7	Illumina	https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html
FastQC 0.12.1		https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
MultiQC 1.15		https://github.com/MultiQC/MultiQC
Cutadapt 4.3		https://cutadapt.readthedocs.io/en/stable/
Seqkit 2.4.0		https://bioinf.shenwei.me/seqkit/
Vienna RNA package, RNAfold 2.4.17		https://www.tbi.univie.ac.at/RNA/
RNAshapes 3.4		https://bibiserv.cebitec.uni-bielefeld.de/rnashapes
Python 3.8.8		https://www.python.org/
Meme Suite, streme 5.5.7		https://meme-suite.org/meme/
Other		
4150 TapeStation	Agilent	https://www.agilent.com/en/product/automated-electrophoresis/tapestation-systems/tapestation-instruments/4150-tapestation-system-297322
Illumina iSeq 100	Illumina	https://www.illumina.com/systems/sequencing-platforms/iseq.html
ProFlex PCR Systems or equivalent thermocycler	Thermo Fisher Scientific	Cat# 4483636
Micro Star 17R or equivalent centrifuge	VWR	Cat# 521-1647
DynaMag-2 or equivalent magnetic rack	Thermo Fisher Scientific	Cat# 12321D
SafeSeal reaction tubes or similar low binding reaction tubes	Sarstedt	Cat# 72.704.700

STEP-BY-STEP METHOD DETAILS

Droplet preparation and DNA recovery

⌚ Timing: 3 h

First, we describe how to prepare fuel-dependent complex coacervates in the presence of an ssDNA library. Then, we explain how to separate the phases and recover the ssDNA for downstream analysis.

1. Prepare complex coacervates and separate phases.
 - a. Thaw ssDNA library (and tagged RNA, if used) on ice.
 - b. Prepare complex coacervates mix (900 μ L).

Reagent	Stock	Final	V/ μ L
MES pH 5.3	500 mM	200 mM	400
RG3 peptide	155 mM	23 mM	150
polyU	41 mM	4.1 mM	100
N30 library	3 μ g/ μ L	9 μ g	3
Tagged RNA (optional)	100 μ M	1 μ M	10
ddH ₂ O			237 or 247

Note: Adding tagged RNA makes the pellet easier to detect after centrifugation. While this is not strictly necessary, it is recommended for initial trials. The RNA sequence is not critical, as the RNase cocktail will remove all RNAs.

- c. Add 100 μ L 400 mM EDC solution to the complex coacervates master mix. Mix thoroughly by pipetting up and down.

Note: Within seconds of mixing the sample with fuel, the solution should become visibly turbid, indicating coacervate formation. If this does not occur and the sample remains clear, refer to [problem 1](#). For representative static images and time-lapse videos of the expected turbidity and droplet morphology, see Donau et al., 2020.²

△ CRITICAL: From the addition of EDC until the supernatant is removed, it is crucial to adhere to all incubation times precisely and work quickly. The fuel-dependent system has a very short lifespan.

- d. Incubate for 1 minute at $\sim 25^{\circ}\text{C}$.
- e. Spin for 3 minutes at 16,000 g in a centrifuge set to 25°C .

△ CRITICAL: The temperature should not exceed 30°C during centrifugation. The lifetime of droplets is greatly affected by temperature. Lower temperatures may greatly extend lifetimes, whereas higher temperatures can drastically decrease them.

- f. After centrifugation, immediately remove all supernatant very thoroughly and store it for downstream analysis.

Note: Use a P1000 to remove most of the bulk of the supernatant without disturbing the pellet. Then, using a P10, remove as much liquid as possible from the pellet. Discard these small volumes.

- g. Resuspend the pellet in 10 μ L 5 M NaCl by pipetting up and down.
 - h. After resuspension, the solution should be clear. Add 40 μ L ddH₂O.
 - i. Use the Monarch Spin PCR & DNA Cleanup Kit according to the manufacturer's instructions for the "small oligo" protocol to clean up 50 μ L of pellet and 50 μ L of supernatant:
 - i. Add 100 μ L Buffer BZ to the solution.
 - ii. Add 300 μ L Isopropanol. Mix well by pipetting up and down. Do not vortex.
 - iii. Transfer 450 μ L solution onto a spin column and spin for 1 minute at 16,000 g. Discard flow-through.
 - iv. Add 500 μ L Buffer WZ to the spin column and spin for 1 minute at 16,000 g. Discard flow-through.
 - v. Repeat step iv.
 - vi. Transfer the spin column to a fresh 1.5 mL sample tube. Spin empty for 1 minute at 16,000 g to remove residual buffer.
 - vii. Transfer the spin column to a fresh sample tube. Add 10 μ L ddH₂O to the center of the matrix to elute DNA. Spin for 1 minute at 16,000 g, then discard the spin column.
2. Digest polyU RNA from both samples.
 - a. Add 1 μ L RNase cocktail to both samples.

Note: Always keep enzyme mixes in -20°C cooling blocks.

- b. Incubate the samples in a thermocycler for 2 hours at 37°C, followed by 10 minutes at 65°C with the lid set to 105°C.

Note: 2 μ L of the recovered, RNase-digested ssDNA can be used for analysis via urea-PAGE gel (20% 19:1 acrylamide: bisacrylamide). If low yields are observed in this step, refer to [problem 2](#).

Library preparation and sequencing

⌚ Timing: 2 days

This step describes how to prepare ssDNA for sequencing using sequence-agnostic, orientation-specific splinted ligations. We then provide a brief overview of how to perform sequencing using an iSeq 100. However, any Illumina-type sequencer that is compatible with the adapters can be used. If you are proceeding with in-house sequencing, remember to thaw the sequencing reagents according to the manufacturer's instructions. This process usually takes 9–18 hours.

3. Ligation of isolated ssDNA to the adapter oligonucleotides via splinted ligation.
 - a. Thaw adapter and splint oligos on ice.
 - b. Prepare 10 μ L of P5 adapter and splint mix per ssDNA sample.

Reagent	Stock	Final	V/ μ L
T4 DNA Ligase reaction buffer	10 $\times$	1 $\times$	1
SRSly P5 adapter	100 μ M	10 μ M	1
SRSly P5 splint	100 μ M	20 μ M	2
ddH ₂ O			6

- c. Prepare 10 μ L of P7 adapter and splint mix per ssDNA sample.

Reagent	Stock	Final	V/ μ L
T4 DNA Ligase reaction buffer	10 $\times$	1 $\times$	1
SRSly P7 adapter	100 μ M	10 μ M	1
SRSly P7 splint	100 μ M	20 μ M	2
ddH ₂ O			6

- d. Anneal oligonucleotides by heating the reaction mix to 95°C for 10 seconds and then cooling them down to 10°C at a ramp rate of 0.1°C/s in a thermocycler.
- e. Prepare 78 μ L Ligation mix per ssDNA sample. Mix well by pipetting up and down or vortexing.

Reagent	Stock	Final	V/ μ L
T4 DNA Ligase reaction buffer	10 $\times$	1 $\times$	8
P5 adapter and splint mix			9
P7 adapter and splint mix			9
ssDNA sample			9
PEG-8000	50%	20%	32
ddH ₂ O			11

- f. Add 1 μ L (30 U) T4 DNA Ligase and 1 μ L T4 PNK (10 U) to the reaction mix. Carefully mix by pipetting up and down.
- g. Incubate ~16 hours at 25°C in a thermocycler with the lid set to "off".

- h. Incubate 10 minutes at 65°C with the lid set to 105°C to inactivate proteins.
- i. Use the Monarch Spin PCR & DNA Cleanup Kit with [the standard cleanup protocol](#) to clean up 80 µL of each sample. Change the ratio of sample: binding buffer to 3:1:
 - i. Add 240 µL Buffer BZ to the solution. Mix well by pipetting up and down. Do not vortex.
 - ii. Transfer 320 µL solution onto a spin column and spin for 1 minute at 16,000 g. Discard flow-through.
 - iii. Add 200 µL Buffer WZ to the spin column and spin for 1 minute at 16,000 g. Discard flow-through.
 - iv. Repeat step iv.
 - v. Transfer the spin column to a fresh 1.5 mL sample tube. Spin empty for 1 minute at 16,000 g to remove residual buffer.
 - vi. Transfer the spin column to a fresh 1.5 mL sample tube. Add 17 µL ddH₂O to the center of the matrix to elute DNA. Spin for 1 minute at 16,000 g, then discard the spin column.

Note: 2 µL of the final cleaned-up product can be used for analysis via Urea-PAGE gel (20% 19:1 acrylamide: bisacrylamide) or a TapeStation or Bioanalyzer. The adapters in this work are 33 nt and 34 nt long, meaning the expected ligated product is library length + 67 nt, in the case of N₃₀, 97 nt. If low yields are observed, refer to [problem 3](#).

Pause point: At this point, it is possible to store the samples at –20°C and continue the protocol at a later point.

4. Addition of index- and sequencing-primers via PCR.
 - a. Prepare the following PCR mix:

Reagent	Stock	Final	V/µL
adapter ligated DNA			15
NEBNext Ultra II Q5 Master Mix	2×	1×	25
Index primer (NEB primer set)			5
Universal primer (NEB primer set)			5

- b. Run the following PCR cycle:

Steps	Temperature	Time	Cycles
Initial Denaturation	98°C	30 sec	1
Denaturation	98°C	10 sec	15 cycles
Annealing/Extension	65°C	75 sec	
Final extension	65°C	300 sec	1
Hold	4°C	forever	

△ CRITICAL: Do not use the same index primer for multiple samples! Make a note of which index primer is used for each sample and the correct index read of that primer. You will need to provide this information to the sequencer.

- c. Clean up using NucleoMag NGS Clean-up and Size Selection:
 - i. Vortex NucleoMag solution until all beads are resuspended.
 - ii. Add 45 µL bead solution to the finished PCR reaction, and incubate for 5 minutes.
 - iii. Put on a magnetic rack, incubate for 5 minutes, and remove the supernatant.
 - iv. Add 200 µL 80% Ethanol while on the magnetic rack, incubate 30 seconds, and remove supernatant.
 - v. Repeat step iv. Ensure that as much liquid as possible is removed.
 - vi. Air dry beads for a maximum of 5 minutes.

- vii. Remove from the magnetic rack. Add 33 μL of 0.1 TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), mix well by pipetting up and down to resuspend beads. Incubate 2 minutes.
- viii. Put on a magnetic rack, incubate 5 minutes.
- ix. Transfer 30 μL to a fresh 0.2 mL PCR tube.
- d. Confirm the success of the library preparation, e.g., using a TapeStation with a D1000 cassette if available.

Pause point: At this point, it is possible to store the samples at -20°C and continue the protocol at a later point.

5. Based on the TapeStation concentration measurements, dilute and mix your library for loading concentrations.

Note: For calculations, use the peak corresponding to the correct library length (for the same adapters and library length, this peak is at 154 bp), the peak corresponding to an adapter dimer (for the same adapters, this peak is at 124 bp) and the peak corresponding to a double insert (for the same adapters and library length, this peak is at 184 bp). Add these three peaks together to calculate the total sample library concentration.

- a. In low-binding tubes, dilute each sample library to 1 nM with 10 mM Tris, pH 8.5.
- b. Mix 10 μL of each 1 nM sample library in a low-binding tube. This yields a 1 nM mixed sample library.
- c. Prepare the final loading library by preparing this last dilution step and adding in the PhiX Control Library:

Reagent	Stock	Final	V/ μL
mixed sample library	1 nM	50 pM	5
PhiX Control V3	1 nM	50 pM	5
Tris pH 8.5	10 mM	10 mM	90

6. Prepare the iSeq 100 and sequencing cassette according to the [manufacturer's instructions](#):
 - a. Load the sequencing cassette with the flow cell and library into the iSeq 100:
 - i. Thaw the flow cell for 30 minutes at $\sim 25^{\circ}\text{C}$.
 - ii. Unpack the thawed sequencing cassette.
 - iii. Invert the sequencing cassette 5 times.
 - iv. Tap the sequencing cassette on the table 5 times.
 - v. Pierce foil protecting the library loading well with a pipette tip.
 - vi. With a fresh pipette tip, load 20 μL of the final loading library into the library loading well.
 - vii. Insert the flow cell into the sequencing cassette.
 - viii. Insert the sequencing cassette into the iSeq 100.
 - b. On the sequencer, set up a new run in Local Run Manager:
 - i. Select "Generate FastQ" as the method.
 - ii. Name your experiment.
 - iii. Choose "Custom" as the library preparation protocol.
 - iv. Choose 1 Index primer read (P7) with a read length of 6.
 - v. Choose dual-read mode with a read length of 151 each.
 - vi. In the sample table, add all the samples you prepared together with the corresponding Index read.
 - c. In the sequencing software, click on sequencing. Follow the instructions on the sequencer, choose the new experiment created in the step above, and start the sequencer.

Pause point: At this point, it is possible to store the samples at -20°C and continue the protocol at a later point.

Data analysis

⌚ Timing: Days to weeks

This final step provides a workflow for analyzing the sequencing data. First, we describe the general downstream processing, followed by multiple optional analysis modules for your specific data.

7. Quality control of sequencing itself using Illumina Sequencing Analysis Viewer.
 - a. Download or transfer the entire folder containing sequencing results onto a machine with Illumina Sequencing Analysis Viewer installed.
 - b. Open Illumina Sequencing Analysis Viewer. The software is available under the following link: https://support.illumina.com/sequencing/sequencing_software/sequencing_analysis_viewer_sav.html).
 - c. At the top, click "Browse" and navigate to the folder containing the sequencing results. Open that folder.
 - d. On the "Analysis" screen, many relevant metrics can be found, like quality statistics for reads per cycle, sequence, etc.
 - e. Changing to "Imaging" and clicking the "Scatter Plot..." button allows plotting of different characteristics in an x/y scatter plot with each tile on the flow cell as a separate data point.
 - i. Plot "% Occupied" on the x-axis.
 - ii. Plot "% Pass Filter" on y-Axis.

Note: If the flow cell was loaded correctly, you should see an approximately oval cloud of data points. If, instead, you see a diagonal line with $x \approx y$, the flow cell was underloaded. A vertical band at % Occupied 95%–100% indicates overloading.

Note: If the number of reads per sample is unexpectedly low, refer to [problems 4](#) and [5](#).

8. Quality control of FASTQ files using FastQC.¹⁵

⚠ **CRITICAL:** Quality control with FastQC should always be performed with raw FASTQ files, even after adapter trimming or filtering.

- a. If applicable, unzip the FASTQ files with gzip.

```
>#!bin/bash
>gunzip *.fastq.gz
```

- b. Install FastQC and MultiQC in a new environment using conda.

```
>#!bin/bash
>conda create -n fastqc-env fastqc multiqc -c bioconda -c conda-forge
```

- c. Activate the new environment and run FastQC.

```
>#!bin/bash
>conda activate fastqc-env
>fastqc *.fastq
```

- d. FastQC generates.html outputs for each file it analyzes. You can either look at each of these individually or run MultiQC to explore them together.¹⁶

```
>#!/bin/bash
>multiqc .
```

Note: MultiQC uses any FASTQ output in the specified folder to generate its own output, which is simply a combined report of all the individual FastQC reports. We advise using MultiQC only to group analysis examples of the same processing stage together. For example, do not use it to group samples before and after adapter trimming, as these files may differ significantly and their combined report may not be useful.

9. Remove adapters using cutadapt.¹⁷
 - a. Install cutadapt in a new environment using conda.

```
>#!/bin/bash
>conda create -n cutadapt-env cutadapt -c bioconda -c conda-forge
```

- b. Activate the new environment and run cutadapt for trimming of paired-end reads.

```
>#!/bin/bash
>source "$(conda info --base)/etc/profile.d/conda.sh"
>
># Define an array of variable names
>variables=("Ref" "Sup" "Pel")
>
># Loop through each variable
>for var in "${variables[@]}; do
>
>    # Define input and output file names based on the variable
>    input_file_1="./raw_data/${var}_R1.fastq"
>    input_file_2="./raw_data/${var}_R2.fastq"
>    trimmed_output_file_1="${var}_trim_1.fastq"
>    trimmed_output_file_2="${var}_trim_2.fastq"
>
>    # Run cutadapt to trim adapters
>    conda activate cutadapt-env
>    cutadapt -a "AGATCGGAAGAGCACACGTCTGAACTCCAGTCA" -A >"AGATCGGAAGAGCGTCGTAGGAAAGAGTGT"
--discard-untrimmed -o >"$trimmed_output_file_1" -p "$trimmed_output_file_2" >"$input_file_1"
"$input_file_2"
>
>    conda deactivate
>
>    # Run FastQC on the trimmed output file
>    conda activate fastqc-env
```

```
> fastqc "$trimmed_output_file_1"
> fastqc "$trimmed_output_file_2"
> conda deactivate
>done
>
>echo "Processing complete for all variables."
```

Note: This script assumes that the original FASTQ files are stored in the “../raw_data” folder with the naming convention “{var}_RX.fastq”, where X is either 1 or 2 (1 being the first read and 2 being the second read). Ensure that your data follows this format or adapt the script as required.

Note: This script uses the P5 and P7 adapter sequences presented in this work for adapter trimming. Ensure that you change the above sequences to match those used when running the script.

Note: The error tolerance for finding matching adapters is set to 0.2. This allows for a maximum of three mismatches when searching for adapters.

- c. Confirm if adapter trimming was successful by looking at the FastQC output.
10. Filter sequences for correct length using seqkit.¹⁸
 - a. Install seqkit using conda.

```
>#!/bin/bash
>conda create -n seqkit-env seqkit -c bioconda -c conda-forge
```

- b. Activate the new environment and run the following script to filter your sequencing reads by length.

```
>#!/bin/bash
>source "$(conda info --base)/etc/profile.d/conda.sh"
>
># Define an array of variable names
>variables=("Ref" "Sup" "Pel")
># Loop through each variable
>for var in "${variables[@]}; do
>
>    # Define input and output file names based on the variable
>    input_file_1="../cut_adap/${var}_trim_1.fastq"
>    output_file_1="${var}_fl_1.fastq"
>    input_file_2="../cut_adap/${var}_trim_2.fastq"
>    output_file_2="${var}_fl_2.fastq"
>
>
```

```
> # Run seqkit to filter for length
> conda activate seqkit-env
> seqkit seq -m 30 -M 30 ${input_file_1} > ${output_file_1}
> seqkit seq -m 30 -M 30 ${input_file_2} > ${output_file_2}
> conda deactivate
>
> # Run FastQC on the filtered output file
> conda activate fastqc-env
> fastqc "$output_file_1"
> fastqc "$output_file_2"
> conda deactivate
>done
>
>echo "Processing complete for all variables."
```

c. Confirm if adapter trimming was successful by looking at the FastQC output.

Note: All analysis modules are optional following this step and depend heavily on the samples you aim to analyze. All the scripts presented in this protocol are compatible with the data generated from sequencing the N₃₀ library that was isolated above.

d. Generate FASTA files from your FASTQ files for faster data processing.

```
>#!bin/bash
>seqkit fq2fa Ref_fl_1.fastq -o Ref_fl_1.fasta
```

Note: It is not possible to use a wildcard character here, so you must process all files manually instead.

△ CRITICAL: As our sequencing data originated from ssDNA only, we will now only process the R1 reads, which correspond to the original sequence in the droplets.

11. Calculate and plot nucleotide content in Python. ^{19–22}

a. Install and import the relevant Python packages into a Python environment.

```
>#!bin/bash
>conda install pandas -c conda-forge
>conda install matplotlib -c conda-forge
```

```
>#!/usr/bin/env python3
>import pandas as pd
>import numpy as np
>import matplotlib.pyplot as plt
```

b. Use the following Python script to load the FASTA files into a pandas library in Python.

```
>#!/usr/bin/env python3
>
>def read_fasta_to_dataframe(file_path):
>    sequences = []
>
>    with open(file_path, 'r') as file:
>        for i, line in enumerate(file):
>            if i % 2 == 1: # Every second line contains the sequence
>                sequences.append(line.strip())
>
>    # Create a DataFrame from the list of sequences
>    df = pd.DataFrame(sequences, columns=['Sequence'])
>
>    return df
>
>variants = ["Ref", "Sup", "Pel"]
>
>#Create an empty dictionary
>df_raw = {}
>
>#Store each dataframe generated by read_fasta_to_dataframe in df_raw under its variants name
>for variant in variants:
>    file_path = f"../len_fil/{variant}_fil_1.fasta"
>    df_raw[variant] = read_fasta_to_dataframe(file_path)
```

c. Calculate the nucleotide content per position for each sample.

```
>#!/usr/bin/env python3
>
>def calculate_nucleotide_content(df):
>    # Get the maximum sequence length
>    max_length = df['Sequence'].str.len().max()
>
>    # Initialize a dictionary to store nucleotide counts for each position
>    position_counts = {nucleotide: [0] * max_length for nucleotide in 'AGCT'}
>
```



```
> # Count the nucleotides at each position
> for sequence in df['Sequence']:
>     for position, nucleotide in enumerate(sequence):
>         if nucleotide in position_counts:
>             position_counts[nucleotide][position] += 1
>
> # Convert counts to percentages
> position_percentages = {nucleotide: [] for nucleotide in 'AGCT'}
> for position in range(max_length):
>     total_count = sum(position_counts[nuc][position] for nuc in 'AGCT')
>     if total_count > 0:
>         for nucleotide in 'AGCT':
>             position_percentages[nucleotide].append((position_counts[nucleotide]
[position] / total_count) * 100)
>     else:
>         for nucleotide in 'AGCT':
>             position_percentages[nucleotide].append(0)
>
> # Create a DataFrame from the position percentages
> percentage_df = pd.DataFrame(position_percentages)
>
> return percentage_df
>
>#Create an empty dictionary
>df_perc = {}
>
>#Store each dataframe generated by calculate_nucleotide_content in df_perc under its vari-
ants name
>for variant in variants:
>    df_perc[variant] = >calculate_nucleotide_content(df_raw[variant])
```

d. Plotting the nucleotide content per position for each sample in an individual plot.

```
>#!/usr/bin/env python3
>
>#Some generic settings to match the aesthetics of this paper
>plt.rcParams['font.family'] = 'Arial'
>plt.rcParams['font.size'] = 9
>plt.rcParams['font.weight'] = 'bold'
```

```
>
>for variant, perc in df_perc.items():
>    #Plots the lines for G, A, C and T
>    plt.figure(figsize=(2.5, 2))
>    plt.plot(perc.index + 1, perc["G"], label="G", color="orange")
>    plt.plot(perc.index + 1, perc["A"], label="A", color="red")
>    plt.plot(perc.index + 1, perc["C"], label="C", color="blue")
>    plt.plot(perc.index + 1, perc["T"], label="T", color="green")
>
>    #Labels and design
>    plt.grid()
>    plt.title(f"{variant} Content", weight="bold", fontsize=10)
>    plt.xlabel("#Base", weight="bold", fontsize=10)
>    plt.ylabel("Base content / %", weight="bold", fontsize=10)
>    plt.yticks(np.arange(15, 40, 5), fontsize=9, weight='bold')
>
>    #Save each figure individually
>    plt.savefig(f"{variant}_content.pdf", bbox_inches="tight", pad_inches=0.1)
>    plt.show()
```

e. Creating the corresponding legend as a separate file.

```
>#!/usr/bin/env python3
>
>#Creating the legend as a separate file
>plt.figure(figsize=(4, 1))
>handles = [
>    plt.Line2D([0], [0], color='orange', lw=4),
>    plt.Line2D([0], [0], color='red', lw=4),
>    plt.Line2D([0], [0], color='blue', lw=4),
>    plt.Line2D([0], [0], color='green', lw=4),
>]
>labels = ['G', 'A', 'C', 'T']
>plt.legend(handles, labels, loc='center', fontsize=9, ncol=4)
>
># Turn off axes and save legend figure
>plt.axis('off')
>plt.savefig("legend_content.pdf", bbox_inches="tight")
```

- f. Calculating the difference between each sample and the "Reference" sample to see the change in distribution before and after samples were taken.

```
>#!/usr/bin/env python3
>
>#Creating an empty dictionary
>df_diff = {}
>
>#Subtracting the Reference from each sample
>for variant in variants:
>    df_diff[variant] = df_perc[variant] - df_perc["Ref"]
```

- g. Plotting the new difference once again in individual plots for each sample.

```
>#!/usr/bin/env python3
>
>#Plotting the df_diff for each sample.
>for variant, diff in df_diff.items():
>    plt.figure(figsize=(2.5, 2))
>    plt.plot(diff.index + 1, diff["G"], label="G", color="orange")
>    plt.plot(diff.index + 1, diff["A"], label="A", color="red")
>    plt.plot(diff.index + 1, diff["C"], label="C", color="blue")
>    plt.plot(diff.index + 1, diff["T"], label="T", color="green")
>    plt.grid()
>    plt.title(f"{variant} Difference", weight="bold", fontsize=10)
>    plt.xlabel("#Base", weight="bold", fontsize=10)
>    plt.ylabel(r"$\Delta$ Base content / %", weight="bold", fontsize=10)
>    plt.yticks(np.arange(-8, 9, 4), fontsize=9, weight='bold')
>
>    plt.savefig(f"{variant}_difference.pdf", bbox_inches="tight", pad_inches=0.1)
```

12. Calculating folding and minimum-free energy values.
 - a. Install Vienna RNA package per suppliers' instructions^{23–25}: <https://www.tbi.univie.ac.at/RNA/documentation.html#install>.
 - b. Use the following script to calculate the minimum free energy structure using RNAfold, then extract the minimum free energy from the result file.

Note: The flag '-P DNA' flag is parsed, so DNA folding parameters are used instead of RNA parameters. The parameters used by RNAfold are from Matthews et al.²⁶

Note: The script produces four output files: the regular RNAfold output, named {var}_rna-fold.txt; a file containing the dot-bracket structure and minimum free energy values, named

{var}_deltaG.txt; and files containing either just the dot-bracket structure or just the energy values, named {var}_dot_bracket.txt and {var}_energy.txt respectively.

```
>#!/bin/bash
>
>file_list="file_list.txt"
>rnafold_output_dir="rnafold_output"
>mkdir -p $rnafold_output_dir
>
>while IFS= read -r input_fasta; do
>   filename=$(basename "$input_fasta")
>   rnafold_output="$rnafold_output_dir/${filename%.fasta}_rnafold.txt"
>   deltaG_output="$rnafold_output_dir/${filename%.fasta}_deltaG.txt"
>   RNAfold -i "$input_fasta" --noPS --noconv -P DNA > "$rnafold_output"
>   grep " ( " "$rnafold_output" > "$deltaG_output"
>   echo "Processed RNAfold and extracted deltaG values for $output_fasta -> $rnafold_output"
>
>   # Python processing block
>   python3 <<EOF
># Define file paths directly from bash variables
>deltaG_file = "${deltaG_output}"
>dot_bracket_file = ".$rnafold_output_dir/${filename%.fasta}_dot_bracket.txt"
>energy_file = ".$rnafold_output_dir/${filename%.fasta}_energy.txt"
>
># Read input file and process line by line
>with open(deltaG_file, 'r') as infile, \
>   open(dot_bracket_file, 'w') as dot_out, \
>   open(energy_file, 'w') as energy_out:
>   for line in infile:
>       # Extract first 60 characters (dot-bracket notation)
>       dot_bracket = line[:30].strip()
>       dot_out.write(dot_bracket + '\n')
>
>       # Extract energy value starting from character 61 onward,
>       # stripping spaces and brackets (retaining only numeric parts)
>       energy_value_raw = line[31:] # Start reading from character 61 onward
>       energy_value = ''.join(c for c in energy_value_raw if c.isdigit() or c == '.' or
>       c == '-')

```

```
> energy_out.write(energy_value + '\n')
>print(f"Generated {dot_bracket_file} and {energy_file}")
>EOF
done < "$file_list"
```

c. In a Python environment, import the relevant Python packages.

```
>#!/usr/bin/env python3
>import pandas as pd
>import matplotlib.pyplot as plt
>import numpy as np
```

d. Load the energy values for each sample from {var}_fil_1_energy.txt.

```
>#!/usr/bin/env python3
>
>file_list = ["Ref", "Sup", "Pel"]
>df = {}
>
>for file in file_list:
>    df[file] = >pd.read_csv(f"rnafold_output/{file}_fil_1_energy.txt", header=None)
>    df[file].rename(columns={0:"values"}, inplace=True)
```

e. Sort the values into bins, count them, and calculate the frequency of each bin with the following script.

```
>#!/usr/bin/env python3
>
>num_bins = 10
>min_value = -10
>max_value = 0
>bin_edges = np.linspace(min_value, max_value, num_bins + 1)
>x_values = np.linspace(min_value, max_value, num_bins + 2)
>
>df_counts = {}
>df_freq = {}
>
>for file in file_list:
>    df[file] ["bin"] = np.digitize(df[file] ["values"], bin_edges)
>    df_counts[file] = np.bincount(df[file] ["bin"])
>    df_freq[file] = df_counts[file] / len(df[file])
```

- f. Plot the MFE frequencies as a distribution plot.

```
>#!/usr/bin/env python3
>
>plt.rcParams['font.family'] = 'Arial'
>plt.rcParams['font.size'] = 9
>plt.rcParams['font.weight'] = 'bold'
>
>plt.figure(figsize=(2.5, 2.5))
>
>for file in file_list:
>    plt.plot(x_values, df_freq[file], label = file)
>
>plt.legend(loc="upper left")
>plt.xlabel(r"Minimum free energy", weight="bold", fontsize=10)
>plt.ylabel("relative occurrence", weight="bold", fontsize=10)
>
>plt.savefig(fname="deltaG.pdf", bbox_inches='tight', pad_inches=0.1)
>plt.show()
```

- g. Calculate the average MFE values per sample.

```
>#!/usr/bin/env python3
>
>df_avg = {}
>
>for file in file_list:
>    df_avg[file] = df[file]["values"].mean()
>    print(f"{file}: {round(float(df_avg[file]), 3)}")
```

13. Discovering enriched motifs in samples compared to a control sample.
- Install the meme-suite per suppliers' instructions^{27,28}: <https://meme-suite.org/meme/doc/install.html>.
 - Use streme from the meme suite to find enriched motifs in a sample compared to a reference.

⚠ **CRITICAL:** It is essential to run this with the `-rna` flag! Using a DNA library instead will automatically force the algorithm to generate and search the reverse complement of each sequence, which was not present in the original experiment. This will result in false positives that are the reverse complement of real motifs.

Note: Adjusting the maximum length of motifs to be discovered can drastically increase or decrease the number found. For example, we found a total of ten statistically significant

motifs with a maximum motif width (*maxw*) of 20; however, with a *maxw* of 10, we found over twenty. This is due to the way the algorithm works: it deletes parts of sequences that carry a motif so that many variations of the same motif are not found at the same sites.

Note: This algorithm is very computationally taxing as it was designed for far fewer, longer reads, and it is incapable of multi-threading. For a sample of 370,000 reads and a reference of 1,500,000 reads, we found that it took approximately one day on an i9 14900k processor and required approximately 25 GB of RAM.

```
>#!/usr/bin/env python3

>streme -p ../len_fil/pel_fil_1.fasta -o streme_output -n ../len_fil/sup_fil_1.fasta -maxw
10 --rna
```

- c. From the output files, you can get the Motif information from the streme.txt file. Copy these (starting with the line "MOTIF 1..." and paste them into a new file called motifs.txt.
- d. Import all relevant libraries in Python. Install all that you are missing analogues to the above explanation (Step 11A).²⁹

```
>#!/usr/bin/env python3

>

>import numpy as np

>import logomaker

>import pandas as pd

>import matplotlib.pyplot as plt

>import re

>from collections import defaultdict
```

- e. Use the following script to open both motifs.txt as well as sites.tsv for plotting of logos and positional information of logos.

```
>#!/usr/bin/env python3

>

># Load the text file

>with open("motifs.txt", "r") as f:

>    raw_data = f.read()

>

># Regex to split motifs

>motif_blocks = re.findall(r' (MOTIF\s+\d+.*?) (?=MOTIF\s+\d+|\Z)', raw_data, re.DOTALL)

>

>motifs = []

>

># Map index to base

>base_order = ['A', 'C', 'G', 'T']
```

```
>
>for block in motif_blocks:
> # Extract lines with matrix values only
> lines = block.strip().splitlines()
> matrix_lines = [line for line in lines if re.match(r'\s*d*\.\d+\s+', line)]
>
> motif = defaultdict(list)
>
> for line in matrix_lines:
>     probs = list(map(float, line.strip().split()))
>     for i, base in enumerate(base_order):
>         motif[base].append(probs[i])
>
> motifs.append(dict(motif))
>
>df = pd.read_csv("sites.tsv", sep="\t")
```

- f. Calculate information content to plot logos with information content on the y-axis instead of nucleotide content (column 1 of final figure).

```
>#!/usr/bin/env python3
>
>def calculate_information_content(motif):
> ic = np.log2(4) + (motif * np.log2(motif)).sum(axis=1)
> ic = ic.fillna(0) # Handle NaNs resulting from log2(0)
> return ic
```

- g. Calculate positional hits for each motif for later plotting (column 3 of final figure).

```
>#!/usr/bin/env python3
>
>position_counts = defaultdict(lambda: defaultdict(int))
>total_counts = defaultdict(int)
>
>for _, row in df.iterrows():
> # Skip rows with missing start/end
> if pd.isna(row['site_Start']) or pd.isna(row['site_End']):
> continue
>
```



```
> motif = row['motif_ALT_ID']
> try:
> start = int(row['site_Start'])
> end = int(row['site_End'])
> except ValueError:
> continue # skip if conversion fails
>
> # Only count positions within the 1-30 range
> for pos in range(start, end + 1):
> if 1 <= pos <= 30:
>     position_counts[motif][pos] += 1
>
> total_counts[motif] += 1
```

h. Build a dataframe storing all information for column 3 (positional hits of motifs).

```
>#!/usr/bin/env python3
>
># Create list of all unique motifs
>motifs_ids = list(position_counts.keys())
>
># Build table for column 1-3
>data = []
>
>for motif in motifs_ids:
> row = {'motif_ALT_ID': motif}
> total = total_counts[motif]
> row['total'] = total
>
> # Fill in positions 1-30
> for pos in range(1, 31):
>     row[f'pos{pos}'] = position_counts[motif].get(pos, 0)
>
> data.append(row)
>
># Convert to DataFrame
>result_df = pd.DataFrame(data)
```

- i. Define a function for numerical sorting of indexes (needed to correctly sort dataframes).

```
>#!/usr/bin/env python3
>
>def numeric_sort_index(df):
> # Extract the number from strings like "STREME-11"
> nums = >df.index.to_series().str.extract(r'(\d+)$').astype(int)[0]
> return df.iloc[nums.argsort()]
```

- j. Build a dataframe for column 4 (starting position of motifs).

```
>#!/usr/bin/env python3
>
># Build df_start used for column 4
>df_start = (
> df.dropna(subset=['site_Start'])
> .assign(site_Start=lambda x: x['site_Start'].astype(int))
> .query('1 <= site_Start <= 30')
> .groupby(['motif_ALT_ID', 'site_Start'])
> .size()
> .unstack(fill_value=0)
>)
>
># Ensure all positions exist
>for pos in range(1, 31):
> if pos not in df_start.columns:
>     df_start[pos] = 0
>df_start = df_start[sorted(df_start.columns)]
>
># Sort index numerically
>df_start = numeric_sort_index(df_start)
```

- k. Build a dataframe for column 5 (end sites of motifs).

```
>#!/usr/bin/env python3
>
># Build df_end used for column 5
>df_end = (
> df.dropna(subset=['site_End'])
> .assign(site_End=lambda x: x['site_End'].astype(int))
```

```
> .query('1 <= site_End <= 30')
> .groupby(['motif_ALT_ID', 'site_End'])
> .size()
> .unstack(fill_value=0)
>)
>
>for pos in range(1, 31):
> if pos not in df_end.columns:
>   df_end[pos] = 0
>df_end = df_end[sorted(df_end.columns)]
>df_end = numeric_sort_index(df_end)
```

l. Calculate fractions of hits from the three dataframes generated above.

```
>#!/usr/bin/env python3
>
># Fractions of hits at each position used for column 3
>result_fraction_df = result_df.set_index('motif_ALT_ID').copy()
>for pos in range(1, 31):
> result_fraction_df[f'pos{pos}'] = (
> result_fraction_df[f'pos{pos}'] / >result_fraction_df['total'].replace(0, np.nan)
> ).fillna(0)
>
># Sort same as df_start/df_end
>result_fraction_df = numeric_sort_index(result_fraction_df)
>
>#Fractions of start and end used for column 4 and 5
>
>df_start_frac = df_start.div(df_start.sum(axis=1), >axis=0).fillna(0)
>df_end_frac = df_end.div(df_end.sum(axis=1), axis=0).fillna(0)
```

m. Generate a final dataframe that will store all data used for plotting.

```
>#!/usr/bin/env python3
>plot_data = []
>
># Loop through motifs in the already-sorted result_df
>for mid in result_df['motif_ALT_ID']:
> plot_data.append({
```

```
> 'id': mid,
> 'logo_matrix':
>pd.DataFrame(motifs[result_df['motif_ALT_ID'].tolist().index(mid)],
> 'pos_fraction': result_fraction_df.loc[mid, [f'pos{p}' >for p in range(1,
31)]] .values,
> 'start_fraction': df_start_frac.loc[mid].values,
> 'end_fraction': df_end_frac.loc[mid].values
> })
```

- n. Plot a master figure with five columns per motif, with the first column being the logo with information content as the y-axis, the second column logo with nucleotide content as the y-axis, the third column showing positional hits (frequency that logo appears at each site), the fourth column showing starting sites of logos, and the fifth column showing ending positions of logos.

```
>#!/usr/bin/env python3
>
># Number of motifs
>num_motifs = len(plot_data)
>
># Create subplots
>fig, axes = plt.subplots(num_motifs, 5, figsize=(5 * 35 / 25.4, 18 / 25.4 * num_motifs))
>plt.subplots_adjust(wspace=0.5, hspace=0.7)
>
># Define the titles for each column
>col_titles = [
> "Logo inf",
> "Logo nuc",
> "Motif pos",
> "Motif start",
> "Motif end"
>]
>
># Position titles above the top row's axes
>for col_idx, title in enumerate(col_titles):
> ax = axes[0][col_idx]
> ax.set_title(title, fontsize=12, fontweight='bold', pad=20)
>
># Loop over motifs
>for i, motif_entry in enumerate(plot_data):
```

```
> row_axes = axes[i] # 5 axes in a row
> motif_name = motif_entry['id']
>
> # Extract precomputed data
> df_logo = motif_entry['logo_matrix']
> pos_fraction = motif_entry['pos_fraction']
> start_fraction = motif_entry['start_fraction']
> end_fraction = motif_entry['end_fraction']
>
> # Calculate information content
> ic = calculate_information_content(df_logo)
> df_ic = df_logo.mul(ic, axis=0)
>
> color_scheme = {
>     'A': 'red',
>     'C': 'orange',
>     'G': 'blue',
>     'T': 'green'
> }
>
> # Column 1: Information content logo
> logomaker.Logo(df_ic, ax=row_axes[0], color_scheme=color_scheme)
> row_axes[0].set_xticks([])
> row_axes[0].set_ylim(0, 2)
>
> # Column 2: Nucleotide content logo
> logomaker.Logo(df_logo, ax=row_axes[1], color_scheme=color_scheme)
> row_axes[1].set_xticks([])
> row_axes[1].set_ylim(0, 1)
>
> # Column 3: Position hits
> row_axes[2].plot(range(1, 31), pos_fraction, color='black')
> row_axes[2].set_ylim(0, 1)
> row_axes[2].set_xlim(1, 30)
> row_axes[2].set_xticks([])
>
> # Column 4: Start positions
> row_axes[3].plot(range(1, 31), start_fraction, color='black')
> row_axes[3].set_ylim(0, 0.1)
```

```
> row_axes[3].set_xlim(1, 30)
> row_axes[3].set_xticks([])
>
> # Column 5: End positions
> row_axes[4].plot(range(1, 31), end_fraction, color='black')
> row_axes[4].set_ylim(0, 0.1)
> row_axes[4].set_xlim(1, 30)
> row_axes[4].set_xticks([])
>
> # Y-axis styling
> for ax in row_axes:
>     ax.tick_params(axis='y', labelsiz=8)
>     for tick in ax.get_yticklabels():
>         tick.set_fontsize(8)
>         tick.set_fontweight('bold')
>
> # Save & show
> plt.savefig("combined_motifs_all_in_one.pdf", bbox_inches='tight')
> plt.show()
```

14. Calculate the abstracted structure of sequences and analyze them for enrichment of specific motifs.

a. Install RNASHAPES in a new environment using conda:³⁰

```
> #!/bin/bash
> conda create -n RNASHAPES -env RNASHAPES -c bioconda
```

b. Run the following script to generate the abstracted structure for all samples.

```
> #!/bin/bash
> source "$(conda info --base)/etc/profile.d/conda.sh"
>
> variables=("Ref" "Sup" "Pel")
>
> conda activate RNASHAPES-env
> for var in "${variables[@]"; do
>     RNASHAPES -c 0 < ../len_fil/${var}_fil_1.fasta > ${var}_abstract.txt
>     echo "${var} finished computing"
>
```

```
>done
>conda deactivate
>
>echo "Finished the script"
```

- c. Import all relevant libraries in Python. Install all that you are missing analogues to the above instructions (Step 11A).

```
>#!/usr/bin/env python3
>
>import numpy as np
>import pandas as pd
>import matplotlib.pyplot as plt
```

- d. Use the following script to load the data into a Python dataframe.

```
>#!/usr/bin/env python3
>
>def RNAsaper_to_dataframe(file_path):
>    energies = []
>    dotbrackets = []
>    abstracted_dotbrackets = []
>
>    with open(file_path, 'r') as file:
>        lines = file.readlines()
>        total_lines = len(lines)
>
>        # Loop through the file and look for header lines (those starting with '>')
>        for i in range(total_lines):
>            if lines[i].startswith('>'):
>                # Attempt to get the third line after the header (i+2)
>                if i + 2 < total_lines:
>                    line = lines[i + 2].strip()
>                    components = line.split()
>
>                    # Make sure the line has exactly 3 components
>                    if len(components) == 3:
>                        try:
>                            energy = float(components[0])
```

```
>         dotbracket = components[1]
>         abstracted_dotbracket = components[2]
>
>         energies.append(energy)
>         dotbrackets.append(dotbracket)
>         abstracted_dotbrackets.append(abstracted_dotbracket)
>     except ValueError:
>         # Skip lines where energy is not a float
>         continue
>
>     df = pd.DataFrame({
>         'Energy': energies,
>         'DotBracket': dotbrackets,
>         'Abstracted': abstracted_dotbrackets
>     })
>
>     return df
```

```
>#!/usr/bin/env python3
>
>variables = ("Ref", "Sup", "Pel")
>
>df = {}
>
>for var in variables:
>     file_path = f"{var}_abstract.txt"
>     df[var] = RNAshaper_to_dataframe(file_path)
```

e. Count the occurrences of each abstracted structure.

```
>#!/usr/bin/env python3
>
>def count_abstracted_structures(df):
>     # Count occurrences of each unique structure in the 'Abstracted' column
>     structure_counts = df['Abstracted'].value_counts()
>
>     # Create a new DataFrame with the structure and its count
>     result_df = pd.DataFrame({
>         'Structure': structure_counts.index,
```



```
>         'Occurrences': structure_counts.values
>     })
>
>     return result_df
>
> df_abs = {}
>
> for var in variables:
>     df_abs[var] = count_abstracted_structures(df[var])
```

f. Plot the occurrences of different structures in a grouped bar graph.

```
>#!/usr/bin/env python3
>
># Dynamically collect all unique "Structure" values across all df_abs DataFrames
>all_categories = sorted(
>     set().union(*[df_abs[var]["Structure"] for var in variables]),
>     key=lambda x: (len(x), x) # First sort by length, then alphabetically
>)
>
># Reindex all DataFrames in df_abs to include all categories, filling missing ones with 0
>for var in variables:
>     df_abs[var] =>df_abs[var].set_index("Structure").reindex(all_categories, fill_value=0).
>     reset_index()
>
>bar_width = 0.8 / len(variables) # Adjust bar width based on the number of variables
>x = np.arange(len(all_categories))
>
>plt.figure(figsize=(2.5, 2))
>
># Loop through variables and plot each one
>for i, var in enumerate(variables):
>     plt.bar(
>         x + (i - len(variables) / 2) * bar_width, # Shift bars >for each variable
>         df_abs[var]["Occurrences"] / len(df[var]) * 100, # Normalize occurrences by total
>         sequences
>         width=bar_width,
>         label=var.capitalize() # Use variable names as labels
>     )
```

```
>
>plt.xlabel("Abstracted Structures", size=10, fontweight="bold")
>plt.ylabel("Percentage of Sequences", size=10, fontweight="bold")
>plt.xticks(x, all_categories, rotation=45)
>plt.yscale("log")
>
>plt.legend()
>
>plt.savefig("Abstracted_Structures.pdf", bbox_inches="tight", pad_inches=0.1)
>plt.show()
```

g. Calculate the average stem length per sample and write it in a new dataframe.

```
>#!/usr/bin/env python3
>
>def count_stem_length(df):
> # Calculate the count of '(' for each entry in the 'DotBracket' column
> df['StemLength'] = df['DotBracket'].apply(lambda x: x.count('('))
> return df
>
>for var in variables:
> count_stem_length(df[var])
>
>def count_stem_length_occurences(df):
> # Count occurrences of each unique structure in the 'Abstracted' column
> structure_counts = df['StemLength'].value_counts()
>
> # Create a new DataFrame with the structure and its count
> result_df = pd.DataFrame({
>     'Length': structure_counts.index,
>     'Occurrences': structure_counts.values
> })
>
> return result_df
>
>df_stem = {}
>
>for var in variables:
> df_stem[var] = >count_stem_length(df[var]).sort_values("Length")
```

h. Plot the density of stem length per sample in a density plot.

```
>#!/usr/bin/env python3
>
># Calculate max_length dynamically from all df_stem DataFrames
>max_length = max([df_stem[var]["Length"].max() for var in variables])
>
>plt.figure(figsize=(2.5, 2.5))
>
># Loop through each variable to plot its density
>for var, color in zip(variables, ["purple", "blue", "green", "orange", "red", "cyan",
"brown"]): # Add more colors as needed
> # Normalize occurrences by total sequences and plot the line
> plt.plot(
>     df_stem[var]["Length"],
>     df_stem[var]["Occurrences"] / len(df[var]) * 100,
>     alpha=0.7,
>     label=var.capitalize(),
>     color=color
> )
> # Fill under the curve for better visualization
> plt.fill_between(
>     df_stem[var]["Length"],
>     df_stem[var]["Occurrences"] / len(df[var]) * 100,
>     alpha=0.3,
>     color=color
> )
>
># Labels and other configurations
>plt.xlabel("Stem Length", size=10, fontweight="bold")
>plt.ylabel("Percentage of Sequences", size=10, fontweight="bold")
>plt.xticks(np.arange(0, max_length+1, 2)) # Dynamically adjust x-ticks based on max_length
>plt.legend()
>
># Save and show plot
>plt.savefig("StemLength_Density.pdf", bbox_inches="tight", pad_inches=0.1)
>plt.show()
```

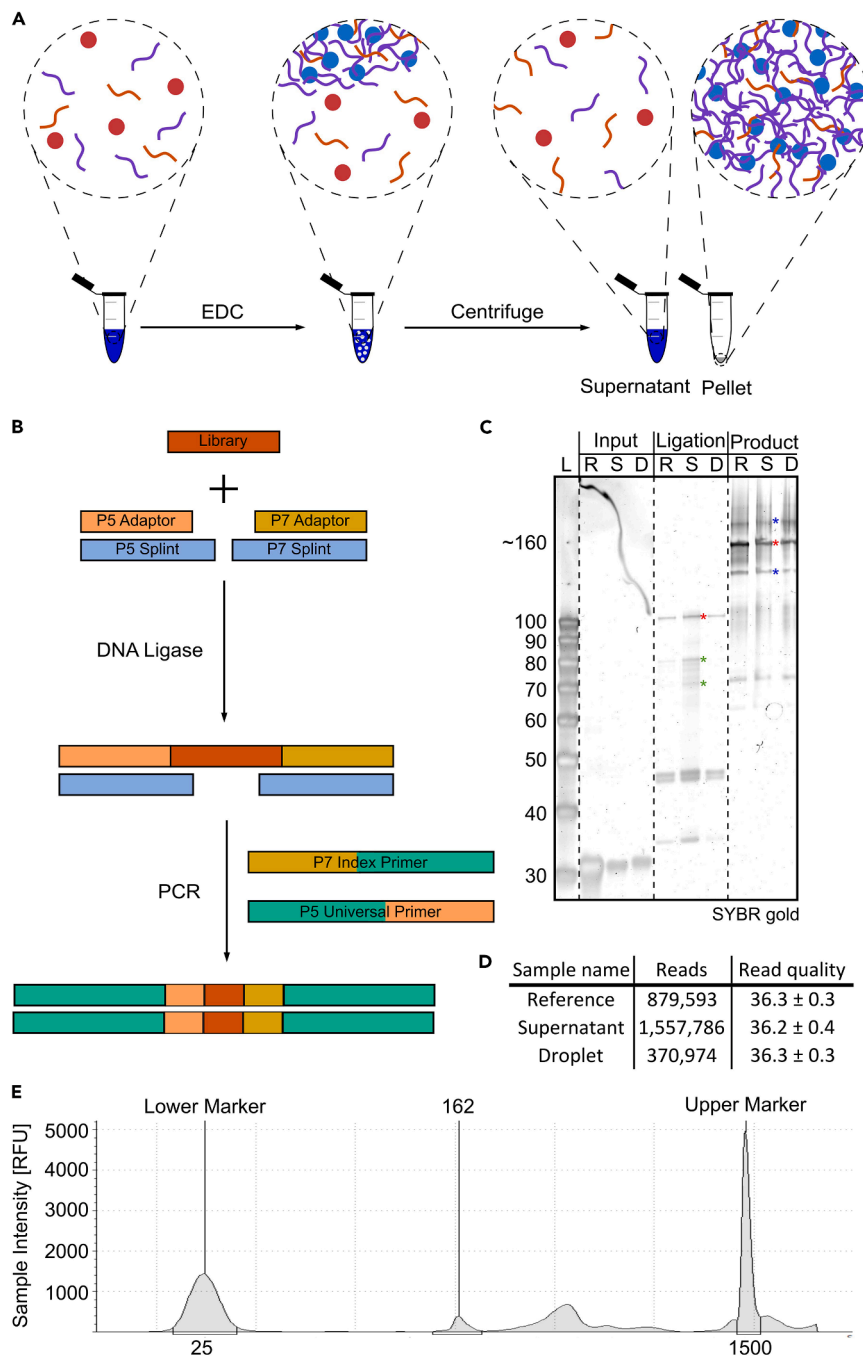


Figure 1. ssDNA isolation and library preparation

(A) Prior to the addition of fuel, the peptide, polyU RNA and ssDNA molecules are mixed in solution, and do not form supramolecular assemblies. When fuel (EDC) is added, phase separation occurs, forming coacervate droplets that sequester the ssDNA molecules from the polymer-poor phase. The phases are separated via centrifugation to yield the polymer-poor and polymer-rich phases, from which the ssDNA populations are isolated. For a complete characterization of this droplet system, please refer to Donau et al., 2020.²

(B) Library preparation relies on the orientation-specific ligation of two adaptor molecules to the recovered ssDNA oligonucleotides. After the ligation, the sequencing primers are added in a PCR using primers with overhangs.

(C) Urea-PAGE is performed following library preparation. R = Reference, S = Supernatant, D = Droplets. 'Input' corresponds to the samples obtained from the droplets, while the 'R' sample is the naïve input library. 'Ligation' corresponds to the product of the first step in B and shows a clear band (red star) corresponding to the expected

Figure 1. Continued

full-length ligation product, as well as a weaker smear (green stars) corresponding to the two partially adapter-ligated products. Bands at 45 nt correspond to the adapters and splint, and bands at ~30 correspond to unreacted input ssDNA. 'Product' shows the full-length sequencing library, with the correct PCR band being marked with a red star. Several additional bands are visible, including two prominent bands (blue stars), which are located approximately 30 nt above and below the main PCR band. These correspond to the no-insert (adapter dimer) and double-insert ligation products. The smear and bands below correspond to indexing primers.

(E) Electropherogram of a TapeStation run of a prepared library. Only the peak corresponding to the correct library size (~160 bp) is used to calculate the correct library dilution for loading onto the sequencer.

EXPECTED OUTCOMES

The results of successful isolation of ssDNA from droplets (Figure 1A) and library preparation (Figure 1B) can be verified on a urea-PAGE gel. The ssDNA isolated from complex coacervates should produce a distinct band (Figure 1C, lane 4). Similarly, the ligation reaction (Figure 1C, lanes 5–7) and the final PCR (Figure 1C, lanes 8–10) should produce a prominent product band (marked with a red star). Note that bands corresponding to the adapter dimers (no insert) and double-insert products (concatemers) are expected (Figure 1C, lanes 8–10). These consume only a small proportion of the reads in the sequencing data and can therefore be disregarded, given that their yield is significantly lower compared to the target library (Figure 1C, lanes 8–10). These reads will be removed during the filtering of sequencing results downstream. Using the library shown here with a loading concentration of 12 pM, we obtained 2.8 million reads across three samples (Figure 1D). This comparably low yield is likely due to the high PhiX spike-in concentration (50 pM), but it still provides high-quality data with sufficient depth for robust statistics.

Data analysis should produce results similar to those in Figure 2. These include base composition (Figure 2A), difference in minimum free energy of folding (Figure 2B), abstract secondary structure classes (Figure 2C), and average intramolecular hybridization length (Figure 2D). We would like to emphasize motif discovery using STREME,^{27,28} which identifies enriched sequence motifs. Using an e-value cutoff of ≤ 0.05 , we identified a total of 14 statistically significant motifs in our data (Figure 3).

We encourage researchers to analyze their dataset further to identify system-specific results. In our associated research manuscript, the results of the base-composition analysis prompted targeted analyses of reads starting or ending in guanosines (see Figure S6 in the associated research manuscript), while the motif discovery analysis informed searches for runs of G or A in droplet samples (see Figure 2F in the associated research manuscript). Follow-up experiments in the droplet system verified specific interactions between ssDNA sequences and RNA or peptide droplet components.

LIMITATIONS

This protocol is optimized for transient complex coacervates. As implementations of protocell-like compartments vary widely between, the recovery steps are system-specific and may differ in static droplet systems or vesicle-based models, for example. While these models can reuse the library preparation and analysis modules, alternative ssDNA extraction methods that are compatible with the system matrix must be developed. If RNA rather than ssDNA is used as the nucleic acid component, both recovery and library preparation require re-optimization.

Our analysis pipeline is modular but not exhaustive. It provides a baseline of base composition, predicted folding energies, and motif discovery for short oligonucleotides. Researchers with different systems (e.g., longer sequences or double-stranded sequences, or RNA) may need to extend or develop additional modules to meet their specific analysis needs.

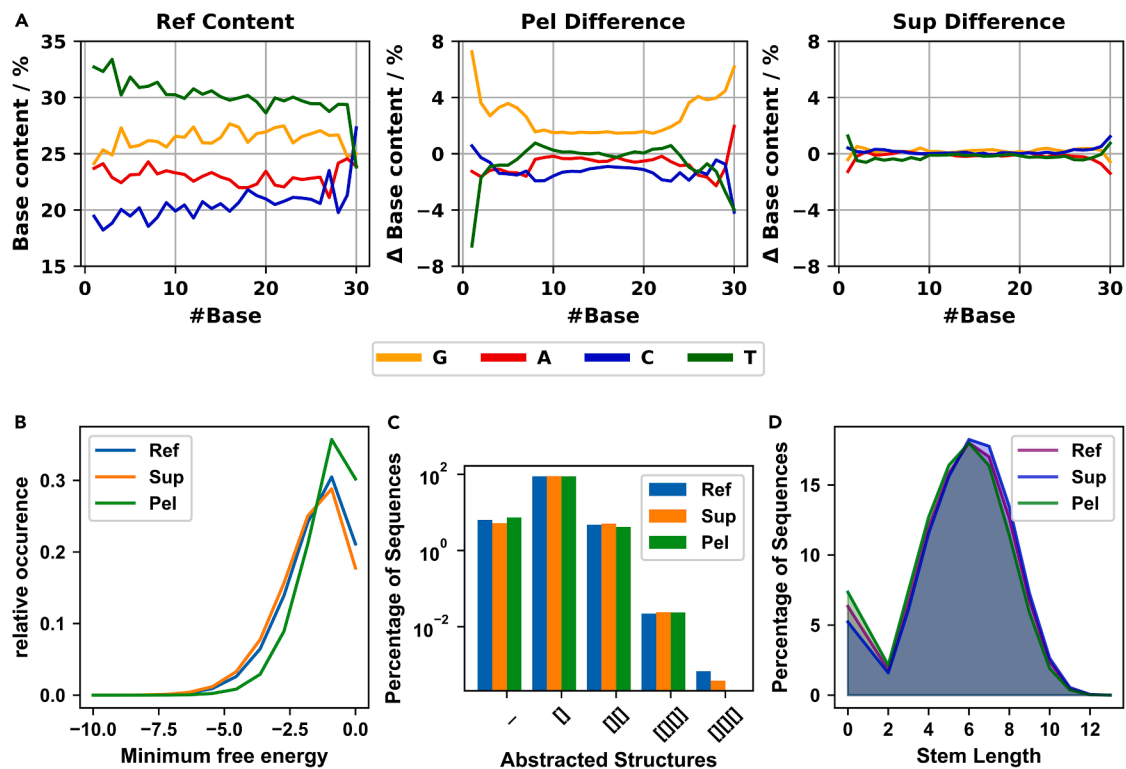


Figure 2. Results generated during data analysis

(A) The base composition of the Reference sample enables the identification of biases in the naive library; most of these result from solid-phase synthesis (left panel). Difference in base composition between the samples from the polymer-rich phase (Pel = Droplet, middle panel) and the polymer-poor phase (Sup = Supernatant, right panel).

(B) Distribution of minimum free energies of folding for sequencing reads of each sample.

(C) Occurrence of different abstracted structures in each sample. / represents a single stem, [] represents two stems, [[]] is a stem containing two secondary side stems, and [[][]] represents three separate stems.

(D) Distribution of intramolecular stem lengths in each sample.

TROUBLESHOOTING

Problem 1

The droplet solution does not turn turbid upon the addition of EDC (Step 1C).

Potential solution

- Make sure to use a fresh solution of dry EDC, as the compound is not stable in the presence of water. If the powder forms clumps, it may be too wet, and a new, pristine stock should be opened. For this reason, EDC should always be stored in completely dry conditions.
- Confirm the identity and pH of the peptide stock. Impurities can result in poor or no droplet formation or cause the peptide to precipitate, particularly at low pH values. Always store at pH 5.2 at 4°C.
- High concentrations of salts (e.g., in buffers or peptide solution due to multiple pH adjustments) can also inhibit droplet formation.
- Improperly stored polyU RNA will hydrolyze over time, which leads to inhibition of droplet formation. Make sure that the polyU RNA is always stored at -20°C.

Problem 2

Low recovery of ssDNA from the polymer-rich phase (Step 2C).

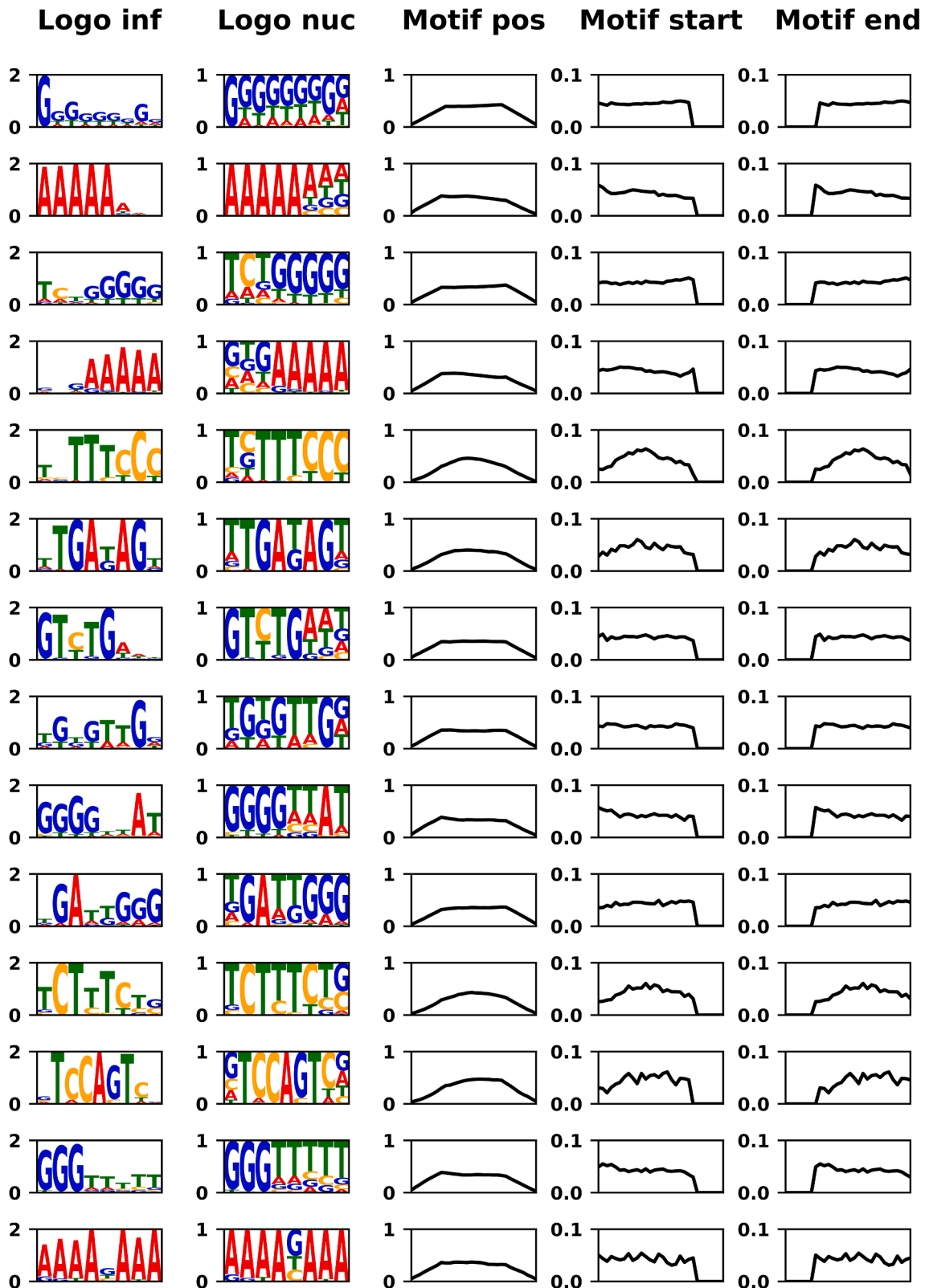


Figure 3. List of all statistically significant motifs identified with STREME

"Logo inf" shows sequence logos scaled by information content (positions with higher information content appear larger). "Logo nuc" shows sequence logos based only on nucleotide frequencies. The columns "Motif start" and "Motif end" report the counts of motif start and end positions, respectively.

Potential solution

- Recovery of ssDNA is a time-sensitive step due to the dependency of droplets on fuel. Work quickly and strictly adhere to the indicated incubation times.
- The droplet pellet formed after centrifugation can be very small and transparent. To confirm the presence of a pellet, consider adding a dye-labelled RNA oligo (e.g., a Cy5-tagged RNA) at a concentration of $\sim 1 \mu\text{M}$ to the reaction mixture. This will enable identifying a pellet by fluorescence even in the absence of specific excitation of the dye molecule. The RNA concentration of the labelled RNA (for short oligonucleotides) is less than 1% of the total RNA concentration in the system and therefore should not significantly impact partitioning results. In addition, it will be digested by RNase in later steps, rendering this a non-issue for the subsequent application.
- If working with different droplet systems, ensure that your downstream clean-up protocol is compatible with the droplet material. Complex coacervates commonly contain poly-anions and poly-cations that can bind to the column material of spin-down columns from PCR clean-up kits. This may result in oversaturation of the column material due to its limited capacity and cause the loss of ssDNA material during loading and washing.

Problem 3

Low ligation yields in the one-pot reaction with T4 DNA ligase and T4 PNK (Step 3I).

Potential solution

- Depending on the system, carryover of droplet material may inhibit T4 DNA ligase. We found that poly-styrene sulfonate and dextran sulphate can both inhibit ligase activity, but they are difficult to separate from the ssDNA due to their poly-anionic character. It is paramount that the ssDNA is recovered cleanly from your droplet sample.
- If your library contains fixed ends (even just one or two nucleotides), the P5 and/or P7 splint oligos can be modified to match these regions. This will significantly increase yields, as more compatible splints will match your library design.

Problem 4

The files for each individual sample are almost empty, with a large number of reads ending up in the 'Undetermined' file (Step 7E).

Potential solution

- Confirm that the correct index sequence was entered in the correct orientation in the sample table of the sequencer.
- If demultiplexing on the sequencer fails, it is possible to demultiplex manually using cutadapt. Follow the instructions at: <https://cutadapt.readthedocs.io/en/stable/guide.html#demultiplexing>.
- Reducing the PhiX spike-in will reduce the number of 'Undetermined' reads, as all reads from the PhiX control will be stored here. However, going too low on the spike-in might result in poor cluster formation and low-quality sequencing data.

Problem 5

The sequencing resulted in a low number of reads with a high occupancy but low pass filter on the sequencing flow cell (Step 7E).

Potential solution

- This is most likely due to overloading of the flow cell. Verify the loaded concentration, ideally with an orthogonal method (e.g., using a TapeStation or a Qubit).

- Reduce the concentration loaded. We obtained good results with concentrations as low as 8 pM on the iSeq 100, despite the recommended loading concentration being between 50 pM and 200 pM.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Hannes Mutschler (hannes.mutschler@tu-dortmund.de).

Technical contact

Technical questions on executing this protocol should be directed to and will be answered by the technical contact, Corbin Machatzke (corbin.machatzke@tu-dortmund.de).

Materials availability

This study did not generate new unique materials.

Data and code availability

Original data generated using this method are available at <https://doi.org/10.17877/TUDODATA-2025-MFE1EMF9>. The published article includes all scripts and codes needed to generate the results presented in this work. They can be found in the main text of this work or at <https://doi.org/10.5281/zenodo.17735401>.

ACKNOWLEDGMENTS

The BoekhovenLab is grateful for support from the TUM Innovation Network – RISE, funded through the Excellence Strategy. This research was conducted within the Max Planck School Matter to Life, supported by the German Federal Ministry of Education and Research (BMBF) in collaboration with the Max Planck Society, and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. The BoekhovenLab is grateful for support from the European Research Council through ERC Starting Grant 852187 and ERC Consolidator Grant 101124380. H.M. is grateful for support from the European Research Council through the ERC Synergy Grant 101166888.

AUTHOR CONTRIBUTIONS

C.M. worked on method design and development, as well as sample preparation, sequencing, and bioinformatic data analysis. A.-L.H. worked on sample preparation. J.B. and H.M. worked on conceptualizing the project. All authors worked together on writing and revising the manuscript. C.M. and A.-L.H. have contributed equally and have the right to list their name first in their CV or any bibliographical list.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Machatzke, C., Holtmannspötter, A.-L., Mutschler, H., and Boekhoven, J. (2025). DNA affects the phenotype of fuel-dependent coacervate droplets. Preprint at Chemrxiv. <https://doi.org/10.26434/chemrxiv-2025-mkd5r>.
- Donau, C., Späth, F., Sossion, M., Kriebisch, B.A.K., Schnitter, F., Tena-Solsona, M., Kang, H.-S., Salibi, E., Sattler, M., Mutschler, H., and Boekhoven, J. (2020). Active coacervate droplets as a model for membraneless organelles and protocells. *Nat. Commun.* 11, 5167. <https://doi.org/10.1038/s41467-020-18815-9>.
- Späth, F., Donau, C., Bergmann, A.M., Kränzlein, M., Synatschke, C.V., Rieger, B., and Boekhoven, J. (2021). Molecular Design of Chemically Fueled Peptide-Polyelectrolyte Coacervate-Based Assemblies. *J. Am. Chem. Soc.* 143, 4782–4789. <https://doi.org/10.1021/jacs.1c01148>.
- Bergmann, A.M., Donau, C., Späth, F., Jahnke, K., Göpfrich, K., and Boekhoven, J. (2022). Evolution and Single-Droplet Analysis of Fuel-Driven Compartments by Droplet-Based Microfluidics. *Angew. Chem. Int. Ed.* 61, e202203928. <https://doi.org/10.1002/anie.202203928>.
- Wenisch, M., Li, Y., Braun, M.G., Eylert, L., Späth, F., Poprawa, S.M., Rieger, B., Synatschke, C.V., Niederholtmeyer, H., and Boekhoven, J. (2025). Toward synthetic life—Emergence, growth, creation of offspring, decay, and rescue of fuel-dependent synthetic cells. *Chem* 11, 102578. <https://doi.org/10.1016/j.chempr.2025.102578>.
- Holtmannspötter, A.-L., Machatzke, C., Begemann, C., Salibi, E., Donau, C., Späth, F., Boekhoven, J., and Mutschler, H. (2024). Regulating Nucleic Acid Catalysis Using Active Droplets. *Angew. Chem. Int. Ed.* 63, e202412534. <https://doi.org/10.1002/anie.202412534>.
- Mirlohi, K., and Blocher McTigue, W.C. (2024). Coacervation for biomedical applications: innovations involving nucleic acids. *Soft Matter* 21, 8–26. <https://doi.org/10.1039/D4SM01253D>.
- Frankel, E.A., Bevilacqua, P.C., and Keating, C.D. (2016). Polyamine/Nucleotide Coacervates Provide Strong Compartmentalization of Mg²⁺, Nucleotides, and RNA. *Langmuir* 32, 2041–2049. <https://doi.org/10.1021/acs.langmuir.5b04462>.
- Wollny, D., Vernot, B., Wang, J., Hondele, M., Safrastyan, A., Aron, F., Micheel, J., He, Z., Hyman, A., Weis, K., et al. (2022). Characterization of RNA content in individual phase-separated coacervate microdroplets. *Nat. Commun.* 13, 2626. <https://doi.org/10.1038/s41467-022-30158-1>.
- Robertson, M.P., and Joyce, G.F. (2012). The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* 4, a003608. <https://doi.org/10.1101/cshperspect.a003608>.
- Orgel, L.E. (1998). Polymerization on the rocks: theoretical introduction. *Orig. Life Evol. Biosph.* 28, 227–234. <https://doi.org/10.1023/A:1006595411403>.
- Horning, D.P., and Joyce, G.F. (2016). Amplification of RNA by an RNA polymerase

- ribozyme. *Proc. Natl. Acad. Sci. USA* 113, 9786–9791. <https://doi.org/10.1073/pnas.1610103113>.
13. Wochner, A., Attwater, J., Coulson, A., and Holliger, P. (2011). Ribozyme-catalyzed transcription of an active ribozyme. *Science* 332, 209–212. <https://doi.org/10.1126/science.1200752>.
14. Troll, C.J., Kapp, J., Rao, V., Harkins, K.M., Cole, C., Naughton, C., Morgan, J.M., Shapiro, B., and Green, R.E. (2019). A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. *BMC Genom.* 20, 1023. <https://doi.org/10.1186/s12864-019-6355-0>.
15. (2015). FastQC. <https://www.bioinformatics.babraham.ac.uk/index.html>.
16. Ewels, P., Magnusson, M., Lundin, S., and Källér, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>.
17. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10. <https://doi.org/10.14806/ej.17.1.200>.
18. Shen, W., Sipos, B., and Zhao, L. (2024). SeqKit2: A Swiss army knife for sequence and alignment processing. *iMeta* 3, e191. <https://doi.org/10.1002/imt2.191>.
19. Python Software Foundation (2025). *Python*.
20. McKinney, W. (2010). Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference (SciPy)*, pp. 56–61.
21. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. *Nature* 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>.
22. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
23. Lorenz, R., Bernhart, S.H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F., and Hofacker, I.L. (2011). ViennaRNA Package 2.0. *Algorithm Mol. Biol.* 6, 26. <https://doi.org/10.1186/1748-7188-6-26>.
24. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.* 125, 167–188.
25. Hofacker, I.L., Fekete, M., and Stadler, P.F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.* 319, 1059–1066. [https://doi.org/10.1016/S0022-2836\(02\)00308-X](https://doi.org/10.1016/S0022-2836(02)00308-X).
26. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., and Turner, D.H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* 101, 7287–7292. <https://doi.org/10.1073/pnas.0401799101>.
27. Bailey, T.L. (2021). STREME: accurate and versatile sequence motif discovery. *Bioinformatics* 37, 2834–2840. <https://doi.org/10.1093/bioinformatics/btab203>.
28. Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. *Nucleic Acids Res.* 43, W39–W49. <https://doi.org/10.1093/nar/gkv416>.
29. Tareen, A., and Kinney, J.B. (2019). Logomaker: Beautiful sequence logos in python. *Bioinformatics* 36, 2272–2274.
30. Janssen, S., and Giegerich, R. (2015). The RNA shapes studio. *Bioinformatics* 31, 423–425. <https://doi.org/10.1093/bioinformatics/btu649>.

Eidesstattliche Versicherung (Affidavit)

Machatzke, Corbin

Name, Vorname
(Surname, first name)

177624

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

Dortmund, 29.05.26

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

Functional capabilities of short deoxy ribonucleic acids beyond storing information

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.

Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

***Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.**

Dortmund, 29.05.26

Ort, Datum
(Place, date)

Unterschrift
(Signature)