Charles University

Faculty of Science

Molecular and Cell Biology, genetics, and virology

Kateřina Švehlová

Development and characterization of light-producing deoxyribozymes

Vývoj a charakterizace deoxyribozymů produkujících světlo

Doctoral thesis

Supervisor:

Edward Curtis, PhD

Institute of Organic Chemistry and Biochemistry

Czech Academy of Sciences

Prague, 2022

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

Praha, 2022 Kateřina Švehlová

I hereby declare that I wrote the thesis independently, and I cited all informational sources. This work or a substantial part of it was not presented to obtain another academic degree or equivalent.

Prague, 2022 Kateřina Švehlová

Acknowledgments

I would like to express my gratitude to my supervisor, Edward Curtis, for my excellent PhD project that I could become truly passionate about. I especially appreciate the time he dedicated to our discussions about *in vitro* selection and research generally and his moral support when my experiments did not go as planned.

I would also like to acknowledge all my lab colleagues for the friendly and stimulating work environment as well as all the fine people of my "adoptive laboratory", the Hanus team. I am grateful for all the time we spent together inside and outside of work.

However, my deepest appreciation goes to my family and partner, Jéňa, who has always supported me in pursuing my passion in research and my career as a scientist. I could not have done it without you.

Abstract

Light-producing protein enzymes such as luciferase play important roles in both applied and basic research. In this study, we used an *in vitro* selection to isolate deoxyribozymes that catalyze a chemiluminescent reaction by dephosphorylation of the commercial substrate CDP-Star. One of the most active variants, named Supernova, was further improved and characterized using a combination of random mutagenesis, *in vitro* reselection, high-throughput sequencing, comparative sequence analysis, and optimization of reaction conditions. Supernova produces light up to 6,500-fold more efficiently that the background reaction and folds into an unusual triple-helical structure. Moreover, we characterized in detail the buffer requirements including pH, the effect of various ions, substrate and Supernova concentrations, and the presence of crowding agents. Finally, we showed that Supernova can be turned into an allosteric sensor by rational design. We anticipate that this deoxyribozyme can be used as the signaling component in light-producing allosteric deoxyribozyme sensors that respond to a wide variety of stimuli and will complement existing methods that utilize radioactive, fluorescent, and colorimetric readouts.

Abstrakt

Chemiluminiscenční proteinové enzymy jako je například luciferáza mají mnoho využití jak v aplikovaném tak i základním výzkumu. V této práci jsme využili metody *in vitro* selekce, abychom identifikovali deoxyribozymy, které katalyzují chemiluminiscenční reakci. Tyto DNA enzymy světlo produkují pomocí defosforylace komerčně dostupného substrátu CDP-Star. Jednu z nejaktivnějších variant, kterou jsme pojmenovali Supernova, jsme dále vylepšili a popsali pomocí kombinace náhodné mutageneze, *in vitro* reselekce, sekvenování nové generace, komparativní sekvenové analýzy a optimalizace reakčních podmínek. V současnosti Supernova produkuje světlo 6.500-krát účinněji než reakce CDP-Star bez přidaného enzymu a zaujímá nezvyklou strukturu obsahující trojšroubovici DNA. V rámci charakterizace reakčních podmínek Supernovy jsme dále popsali její nároky na pH, koncentraci Supernovy i substrátu a přítomnost iontů či zahušťovacích činidel. V neposlední řadě jsme ukázali, že Supernova může být pomocí racionálního designu přeměněna na alosterický senzor. Předpokládáme, že námi vyvinutý deoxyribozym bude využíván jako signální část v alostericky regulovaných senzorech, které detekují rozmanité ligandy, a doplní tak ostatní metody využívající radioaktivitu, fluorescenci, či změnu barvy.

Table of Contents

1 Abbreviations

A	adenine
ABTS	2,2'-azinobis(3-ethylbenzothiazoline-6-sulfonic acid)
ADP	adenosine diphosphate
ATP	adenosine triphosphate
bp	base pair
C	cytosine
c-di-GMP	cyclic diguanosyl monophosphate
cDNA	complementary DNA
CPEB3	cytoplasmic polyadenylation element-binding protein 3
CSPD	chloro-5-substituted adamantyl-1,2-dioxetane phosphate
CTP	cytidine triphosphate
dNTP	deoxynucleotide triphosphate
ELISA	enzyme-linked immunoassay
FACS	fluorescence-activated cell sorting
Frc6P	fructose-6-phosphate
Fwd	forward
Fwd PBS	forward primer-binding site
G	guanine
GlcN6P	glucosamine-6-phosphate
glmS	glutamine-fructose-6-phosphate amidotransferase
Gln	glutamine
Glu	glutamate
GTP	guanosine diphosphate
GTP	guanosine triphosphate
HDV	hepatitis delta virus
HHR	hammerhead ribozyme
HPLC	high performance liquid chromatography
MRE	molecular recognition element
mRNA	messenger RNA
NC	negative control
NGS	next-generation sequencing
nt	nucleotide
NTP	nucleoside triphosphate
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PC	positive control
PCR	polymerase chain reaction
pre-mRNA	precursor messenger RNA

RBS	ribosome-binding site
RCA	rolling circle amplification
Rev	reverse
Rev PBS	reverse primer-binding site
RNase	ribonuclease
rRNA	ribosomal RNA
SAM	S-adenosylmethionine
SARS-CoV-2	severe acute respiratory syndrome coronavirus 2
SBS	splint-binding site
snRNA	small nuclear RNA
T	thymine
tRNA	transfer RNA
U	uracil
UTP	uridine triphosphate

2 Introduction

"The double helix is indeed a remarkable molecule. Modern man is perhaps 50,000 years old, civilization has existed for scarcely 10,000 years and the United States for only just over 200 years; but DNA and RNA have been around for at least several billion years. All that time the double helix has been there, and active, and yet we are the first creatures on Earth to become aware of its existence."

- Francis Crick

2.1 The dawn of catalytic nucleic acids

At the time of its discovery in 1869 by the Swiss doctor Friedrich Miescher[1], DNA was considered an inert molecule used by cells to store phosphorus in their nuclei. Although its involvement in fertilization was soon suspected, it was not until 1944 that the trio of scientists Oswald Avery, Colin MacLeod, and Maclyn McCarty reported that DNA is indeed the genetic agent responsible for pneumococcal transformation[2]. Received with certain hesitation by some scientists, the theory was finally confirmed by Alfred Hershey and Martha Chase and their classic experiment using radiolabeled T2 phage[3]. These discoveries in turn propelled the investigation of DNA structure, a model of which was proposed by James Watson and Francis Crick in 1953 with support from experiments performed by Maurice Wilkins and Rosalind Franklin[4].

Despite this sudden push into the spotlight of biology, the perception of DNA as a passive bearer of genetic information did not change for another 30 years. In fact, in his Nobel lecture in 1946, James Sumner insisted that "All enzymes are proteins […]"[5]. It was not until Sidney Altman and Thomas Cech postulated – somewhat heretically at the time – that RNA, a molecule structurally very close to DNA, can catalyze chemical reactions in living organisms similarly to proteins. In particular, Altman and his team showed that the catalytic moiety of *Escherichia coli* and *Bacillus subtilis* ribonuclease (RNase) P is in fact an RNA molecule[6], while Cech and his group demonstrated the enzymatic activity of a self-splicing intron in the 26S ribosomal RNA in the ciliate *Tetrahymena thermophila*[7].

RNAse P is a ribonucleoprotein complex (it contains both RNA and protein). Like the ribosomal RNA subunit, it is present in all known organisms, suggesting its evolution predates the origin of the modern cell. It is also one of the two natural ribozymes known to act in *trans,* making it a true multiple-turnover enzyme, the other being ribosomal RNA. Although RNase P is associated mainly with precursor tRNA maturation where it hydrolyzes the phosphodiester bond 5' of the mature tRNA, it was also implicated in the processing of other various substrates. Examples include

a bacterial precursor rRNA[8], mRNA of polycistronic *his* operon[9], viral genomic RNA[10], transfer messenger RNA[11], and certain riboswitch motifs[12].

Similarly, the group I intron ribozyme discovered by the Cech group has been since identified in all domains of life – prokaryotes, archaea, and eukaryotes. However, unlike RNAse P, its distribution among organisms is sporadic rather than ubiquitous. This may be due to the fact that group I intron ribozymes show a low primary sequence conservation, and their considerable size precluded a database-wide search for its secondary structure. In fact, the first group I intron ribozymes were identified in archaea only recently[13], suggesting that many such ribozymes are yet to be discovered in other organisms.

Although group I intron ribozymes have been implicated in various biological processes, the most comprehensively studied example is that of the nuclear ribosomal RNA of eukaryotes. The ribozyme itself is a part of a larger group I intron sequence. This intron is excised by the ribozyme from its flanking exons through a two-step transesterification reaction during the process of rRNA maturation (Figure 1a).

2.2 Naturally occurring ribozymes

Only three years after the discovery of the group I intron and RNase P, another type of self-splicing catalytic RNA was discovered by the Cheng lab in an intron of mitochondrial pre-mRNA in *Saccharomyces cerevisiae*[14]. Named the group II intron, the ribozyme does not require an external guanosine molecule as a co-factor. Instead, the reaction is initiated by an internal adenosine nucleotide, and the excised product forms a lariat structure (Figure 1b). Since its discovery, the group II intron has been found in each of the three domains of life. Its structural and mechanistic similarity to spliceosomes, which are ribonucleoprotein complexes involved in nuclear mRNA processing, has lead scientists to hypothesize that the two processes share a common evolutionary ancestor (Figure 1c)[15].

Since Cech's and Altman's discovery of ribozymes, many other natural ribozymes have been identified. To date, experimentally confirmed ribozymes can be divided into two main categories: those that perform reactions involving phosphoester chemistry, and the ribosomal RNA that synthesizes peptide bonds (Figure 2). The first category can be further divided into four classes based on the product of the reaction: self-cleaving ribozymes, self-splicing ribozymes, ribozymes performing RNA hydrolysis, and an RNA motif that modifies another RNA molecule.

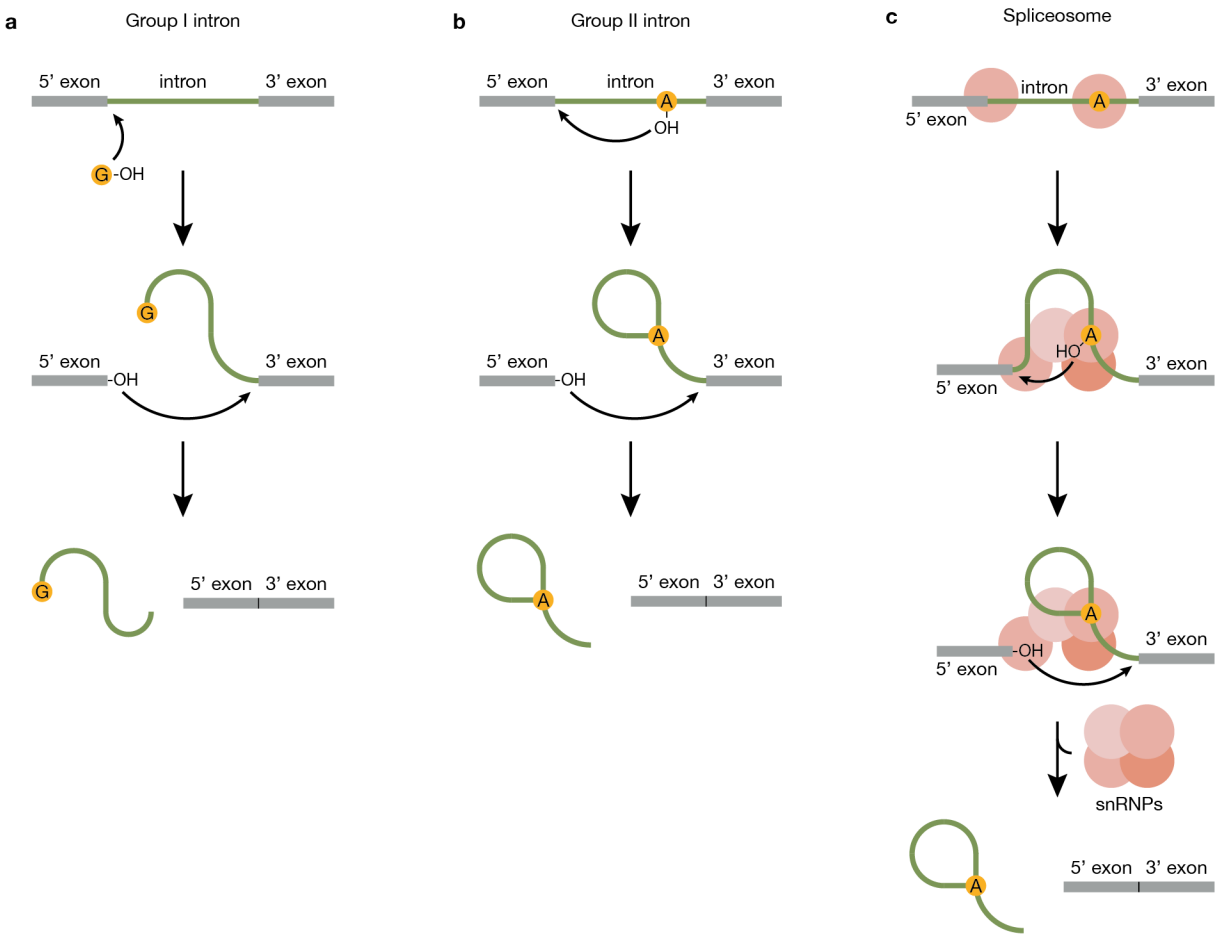


Figure 1. Overview of reaction mechanisms of self-splicing ribozymes and the spliceosome. (a) Group I intron. An exogenous guanosine cofactor docks at a G-docking site of the ribozyme and attacks the 5' splicing site. In the second step, the 3' -OH of the upstream exon cleaves the 3' splice site while ligating the two exons. (b) Group II intron. Similar mechanism as in (a), but the attacking nucleophile is the 2' -OH group of an internal adenosine. The net result is spliced exons and an excised intron in the form of a lariat. (c) Spliceosome. The same mechanism as in (b). Note that contrary to (b), a portion of the structural and catalytic RNA is provided in *trans* as small nuclear RNAs which are part of the spliceosome complex[16]. Adapted from Alberts et al. 1994[17].

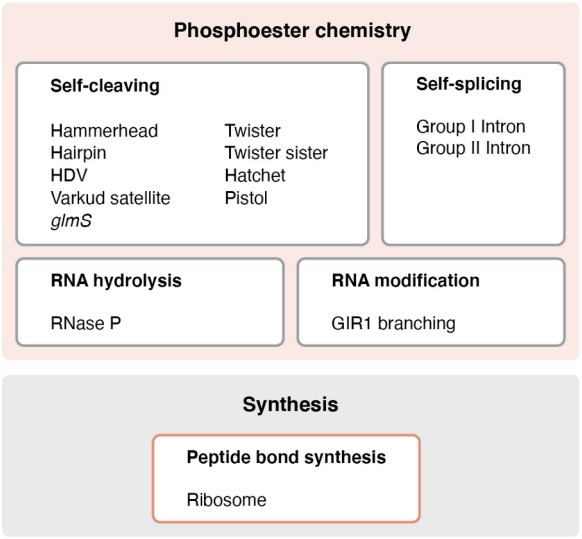


Figure 2. Functional classes of naturally occurring ribozymes.

2.2.1 Site-specific self-cleaving ribozymes

The category of self-cleaving ribozymes encompasses a variety of diverse ribozymes that cleave their own phosphodiester nucleic acid backbone in a sequence-specific manner. They catalyze a reaction where the nucleophilic 2'-oxygen attacks the adjacent 3'-phosphate. This in turn results in the cleavage of the phosphodiester bond and generates 2',3'-cyclic phosphate and 5'-hydroxyl RNA products. Although self-cleaving ribozymes accelerate the same transesterification reaction, their distinct conformational folds and sporadic occurrence among unrelated organisms suggest multiple independent origins. Only a few examples have been linked to a specific biological function. These include ribozymes involved in rolling-circle amplification in RNA pathogens[18], processing of repetitive DNA[19], regulation of gene expression [20,21], and retrotransposition[22].

2.2.1.1 Hammerhead ribozyme

Hammerhead ribozymes (HHRs) are arguably the most thoroughly studied self-cleaving RNA motifs. They are short catalytic RNAs and, like all self-cleaving ribozymes, catalyze cleavage of their own internal phosphodiester backbone. They were first discovered in RNA transcripts of the avocado sunblotch viroid[18]. The precursor RNA molecules are produced from a circular RNA template by a mechanism called rolling circle amplification (RCA). RCAs produce linear multimeric RNA molecules which then self-cleave to produce monomer RNAs[23]. Hammerhead ribozymes all share a characteristic secondary structure which consists of a highly conserved catalytic core flanked by three short helices. The resemblance of this secondary structure to the head of the hammerhead shark lent these ribozymes their name.

Other examples of HHRs have since been found in species of all kingdoms of life, including humans[24]. In lower eukaryotes, for instance, they are mostly found in repetitive regions, suggesting a function associated with retrotransposon propagation[25], while highly conserved HHRs in amniotes map to intronic sequences, indicating a possible role in mRNA biogenesis[24].

2.2.1.2 Varkud satellite and hairpin ribozyme

The Varkud satellite ribozyme was first identified in mitochondrial RNA in certain strains of *Neurospora* fungi and is responsible for RNA transcript maturation[19], while the hairpin ribozyme was discovered in satellite RNAs of the tobacco ringspot virus[26]. Similar to some hammerhead ribozymes, the hairpin ribozyme plays a crucial role in RCA and the processing of viral genomic RNA. Both Varkud satellite and hairpin ribozymes share a feature with group I introns in that they not only cleave the multimeric transcript RNA into monomer intermediates, but they can also ligate the products, unlike other self-cleaving ribozymes.

2.2.1.3 HDV and CPEB3 ribozymes

Similar in function but distinct in structural requirements to the better-known hammerhead ribozyme, the hepatitis delta virus (HDV) ribozyme was discovered as a catalytic element in the hepatitis delta virus[27]. Like the HHRs, they are responsible for cleaving the RNA transcripts generated by RCA during viral RNA replication (Figure 3). Subsequently, a similar self-cleaving motif, the CPEB3 ribozyme, was identified using *in vitro* selection and a human genomic library. It is highly conserved among all placental mammals and marsupials, but absent in non-mammalian vertebrates, which suggests that the ribozyme may have appeared as recently as ~200 million years ago. Furthermore, the HDV and CPEB3 ribozymes share a complex set of structural and sequence requirements that are unlikely to have developed independently. This led the authors to hypothesize that the HDV ribozyme originated in fact by horizontal transfer from the human transcriptome to the HDV RNA genome in recent evolutionary history[28]. Since then, HDV- and CPEB3-like ribozymes have been confirmed in diverse species of eukaryotes, including humans, bacteria, and viruses[22]. These findings seem to challenge the previous hypothesis of CPEB3's *de novo* origin in early mammals, and further studies will be needed to determine the evolutionary path of HDV-like ribozymes.

Taking into account the genetic context of HDV-like ribozymes, they have been predicted to be involved in many biological events, including retrotransposition, RNA processing, and posttranscriptional and translational control[29].

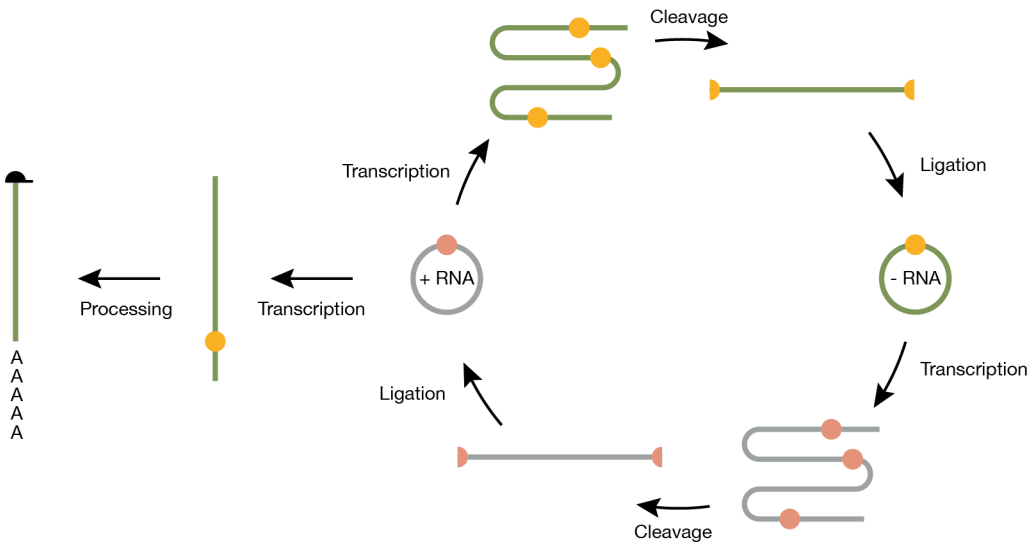


Figure 3. HDV genomic RNA life cycle. The circular genomic RNA (+RNA) is transcribed by RCA into a multimeric antigenomic RNA (green line). This is cleaved by the antigenomic HDV ribozyme (yellow circle) into monomeric transcripts, which are in turn ligated into a circular antigenomic RNA (-RNA). Similarly, the circular antigenomic RNA is transcribed by RCA into multimeric genomic RNA (grey line), processed into monomeric units by the genomic HDV ribozyme (pink circle), and ligated again into new circular genomic RNAs. Note that a similar mechanism takes place during RCA of viruses containing hammerhead and hairpin ribozymes [30]. Adapted from Riccitelli and Lupták 2013[29].

2.2.1.4 Ribozymes discovered using bioinformatics

In the 2010s, with the availability of increasingly powerful computational resources, examples of self-cleaving ribozymes were discovered using solely bioinformatic tools. These searches revealed a new class of self-cleaving RNA molecules, named twister ribozymes, in bacteria and a diverse set of eukaryotes. Since they often share their genetic context with the hammerhead ribozymes, it is possible that their function also might be interchangeable with that of the HHRs[31]. In bacteria and bacteriophages, certain genes are often associated with hammerhead and twister ribozymes. This led the Breaker lab to search for conserved structured RNA elements in the vicinity of these genes. Indeed, only one year later, three additional classes of self-cleaving ribozymes were revealed: a twister sister, hatchet, and pistol ribozyme[32].

2.2.2 Allosteric ribozymes in nature

The first identified naturally occurring allosteric ribozyme was the glmS ribozyme. It is located in the 5' untranslated region of the *glmS* gene, which codes for glutamine-fructose-6-phosphate amidotransferase (glmS) in Gram-positive bacteria. The product of this enzyme, glucosamine-6-phosphate (GlcN6P), binds to the ribozyme and triggers a self-cleavage reaction, which renders the mRNA transcript subject to

degradation by RNases. In this way, the GlcN6P molecule serves as a messenger in a negative feedback loop that controls its own production[20] (Figure 4a, b).

Currently, the only other allosteric ribozyme found in nature is the bacterial riboswitch-controlled group I intron[33]. The double RNA motif is embedded in an intron portion of a putative virulence gene of *Clostridium difficile* and comprises two functional elements: a structured RNA region that binds the cellular signaling molecule, cyclic diguanosyl monophosphate (c-di-GMP), followed by a catalytic group I intron. In the presence of c-di-GMP, the molecule binds the riboswitch, which in turn shifts the intron's 3' scission site several nucleotides upstream, creating an optimal ribosome-binding site for translation initiation. Conversely, without the bound c-di-GMP, the intron proceeds to trim the pre-mRNA at a site further downstream, producing an mRNA transcript lacking a functional ribosome-binding site[34] (Figure 4c, d). Such a transcript cannot be bound by ribosomes and is degraded by RNases.

Interestingly, there are no other known examples of allosterically regulated ribozymes in living cells. Although many such RNA devices have been constructed *in vitro*, they seem to be jarringly lacking in nature. One possible explanation is their relatively high information content compared to individual ribozymes or riboswitches which makes them vulnerable to disruptions by a single mutation[35], a common evolutionary event.

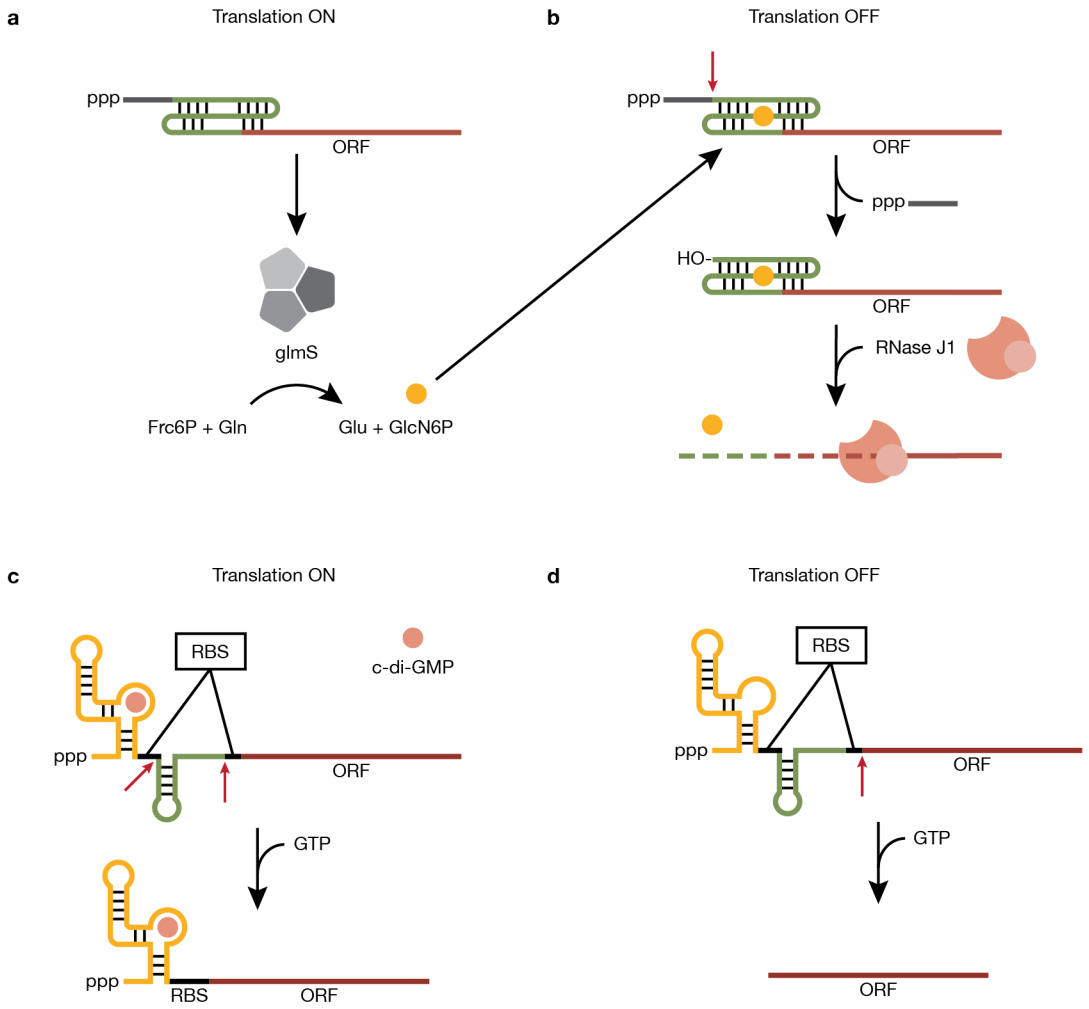


Figure 4. Gene expression control by *glmS* and c-di-GMP ribozyme. (a) The *glmS* mRNA transcript consists of a triphosphate cap (ppp), leader sequence (dark grey), *glmS* ribozyme (green), and ribosome-binding site and protein-coding sequence (red). At low concentrations of GlcN6P, the ribozyme is inactive, and the mRNA is translated into the glmS enzyme, which uses fructose-6-phosphate (Frc6P) and glutamine (Gln) to synthesize GlcN6P (yellow circle). (b) On the other hand, if the GlcN6P concentration in the cell is high, the GlcN6P binds the ribozyme in the mRNA transcript which then cleaves off the leader sequence portion of the mRNA molecule (red arrow). This leaves a hydroxyl group at the 5' end of the mRNA which in turn marks it for degradation by RNase J1. Adapted from Ferré-D'Amaré 2010[36]. (c) The c-di-GMP allosteric group I intron consists of a riboswitch (yellow), and a two-part ribosome-binding site (RBS, thick black line) disrupted by the group I intron ribozyme (green) and followed by the protein-coding gene. At high c-di-GMP concentrations, the c-di-GMP (pink circle) binds the riboswitch which causes the group I intron to excise itself (red arrows) and create an optimal RBS. (d) At low c-di-GMP concentration, the group I intron excises itself by an alternative GTP attack (red arrow) leaving an mRNA transcript lacking an RBS, which cannot be translated into protein. Adapted from Chen et al. 2011[34].

2.3 The RNA world hypothesis and the origin of life

In the 1980s, during the time when Cech and Altman were still working on their revolutionary discoveries, a widely accepted hypothesis presumed that both proteins and RNA were necessary for the self-replication of molecules at the origin of life: the RNA stored the information while proteins catalyzed the chemical replication. However, following Cech's and Altman's discoveries, Walter Gilbert hypothesized that, since RNA can also act as a catalyst, it is in fact possible that only the RNA molecule was necessary[37]. Moreover, the excision and ligation reaction that RNA was clearly capable of would permit a simple recombination and exchange of exons or whole genes between RNA molecules. Only later, RNA molecules would start using activated amino acids or short peptides as cofactors until they would eventually synthesize a full protein similar to the RNA core of modern ribosomes. Since proteins proved to be better catalysts than RNA molecules, they would later become the dominant catalysts in modern organisms. Finally, according to Gilbert, RNA was displaced by DNA as a genetic information storage molecule due to DNA's better chemical stability, and RNA became relegated to only be a messenger between the hereditary information of DNA and the executional function of proteins. He called his hypothesis the 'RNA world'. We can still see remnants of this RNA world in the catalytic RNA core of modern ribosomes: RNase P, spliceosomes, or self-splicing introns in rRNA.

Although the RNA world hypothesis has been tentatively accepted by some scientists, it has in fact introduced another egg-and-chicken conundrum: how did the RNA replicase come to be? Currently, there are several theories addressing this question. Some propose that the first replicators arose from short oligonucleotides that formed by the non-enzymatic condensation of activated mononucleotides[38]. Others argue that a templated ligation of short oligomers by early RNA replicators is more probable, and that only later the system shifted to polymerization with mononucleotides as substrates[39].

Still others have rejected the RNA world hypothesis altogether and proposed an alternative "peptide-RNA world" scenario. In this scenario, both simple peptides and RNA co-evolved from the very beginning, setting the stage for the modern genetic code[40].

Taken together, there are multiple more-or-less plausible theories about what exactly transpired at the origin of life. Moreover, the RNA world hypothesis *per se* cannot be proven. Even if scientists were eventually able to reconstruct an RNA-based life in the lab, the fact that it is plausible does not prove that it did in fact happen. Conversely, researchers cannot easily disprove the hypothesis either. The fact that a crucial ribozyme is not isolated using *in vitro* selection only shows that the current protocol might be limited by the number or length of the sampled RNA molecules or the conditions that are being used.

2.4 *In vitro* selection

In 1967, Sol Spiegelman's lab performed an experiment that would later be described as the first *in vitro* selection. Through a process they called a "Darwinian experiment", they isolated optimal RNA substrates for a bacteriophage RNA-dependent RNA polymerase over 72 rounds of selection[41].

However, it was not until 20 years later that other researchers began to follow up on Spiegelman's experiments. Several crucial findings had to be made, however, to set the stage for modern *in vitro* selection experiments as we know them today. Specifically, it was the discovery of reverse transcriptase in retroviruses by Howard Temin and David Baltimore in the 1970s[42,43]. In the 1980s, John Milligan described how to use a recombinant T7 RNA polymerase to produce large amounts of RNA using a synthetic DNA template[44], while Mullis published his work on *in vitro* DNA amplification using polymerase chain reaction (PCR)[45]. Moreover, in 1971 Gutte and Merrifield reported an efficient method of solid-phase synthesis of peptides, which was then further developed for DNA oligonucleotide synthesis by Beaucage and Caruthers[46,47]. These were all important advances that enabled for the artificial evolution experiments to proceed in controlled laboratory settings.

These findings were instrumental for another couple of researchers: Debra Robertson and Gerald Joyce. In their work from 1990, they described a method, later named *in vitro* selection, that they used to change the substrate specificity of the self-splicing I intron described by Cech and colleagues. By two rounds of selection and amplification of several mutants, they were able to isolate an RNA enzyme that cleaved a DNA substrate and attached it to its 3' end[48].

Concurrently, a similar approach was employed by two other research groups to isolate RNA molecules that non-covalently bind specific ligands, analogous to the way protein antibodies bind their targets. Craig Tuerk and Larry Gold randomized an 8-nt loop of a T4 DNA polymerase mRNA and looked for sequences that would efficiently bind its product, the T4 DNA polymerase, as part of a cellular negative-loop regulation[49]. Similarly, Andrew Ellington and Jack Szostak isolated RNA molecules, which they named aptamers (from Latin "apta", to fit), that specifically bound several molecular dyes[50]. Unlike Tuerk and Gold, whose starting pool of molecules consisted of a mostly constant sequence with only a short region of random nucleotides, Ellington and Szostak used a completely randomized pool of 100 nucleotides flanked by constant primer-binding sites without any inherent constraints on the pool's structure or primary sequence.

The method itself relies on a series of processes similar to Darwinian evolution: mutation, selection, and amplification (Figure 5). In the case of catalytic nucleic

acids, the genotype is conveniently linked to its phenotype since it is the hereditary information molecule that also performs the function.

During *in vitro* selection experiments, most diversity comes from the design of the starting library. This library usually contains between 10^{14} and 10^{16} RNA or DNA sequences and consists of a random region – each position has a 25% probability of containing either adenine, thymine, cytosine, or guanine base – and flanking primer-binding sites. The starting library is typically synthesized using an automated DNA synthesizer. Ideally, the substrate phosphoramidites should be premixed at a ratio that compensates for their different coupling efficiency during the synthesis and ensures the correct representation in the pool.

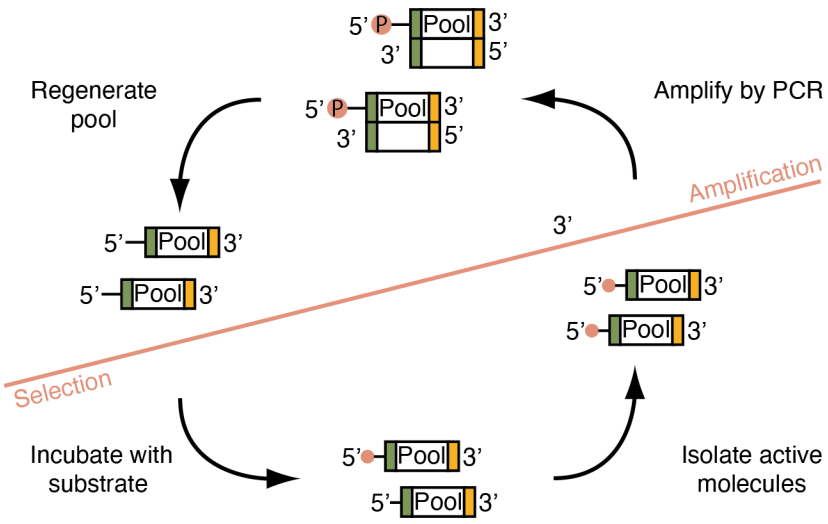


Figure 5. A general overview of *in vitro* selection. Starting pool of DNA or RNA molecules is incubated with a substrate. Only catalytically active molecules carry a chemical tag (pink circle) used for their isolation. These molecules are then transcribed into complementary DNA by reverse transcriptase (in the case of RNA) and amplified by PCR using constant regions (green and orange boxes) as primer-binding sites. The pool is then regenerated by separating the strands in the case of DNA or by transcription in the case of RNA. This pool enriched for active sequences is used for another round of selection. Note that in selections for RNA catalysts, the PCR step is sometimes omitted since the sequences are also amplified during the transcription.

Theoretically, in a library of 10^{15} molecules with a random region of 25 nucleotides, each DNA sequence is represented once on the average. In libraries with random regions longer than 25 nucleotides, a pool of 10^{15} molecules can no longer contain all the possible sequences and as such does not cover the complete sequence space. This sequence space can be partially accessed by a standard error introduction by common DNA polymerases during the amplification step or even by using error-prone DNA polymerases[51]. In either case, it is usually safe to proceed with a pool having incomplete sequence space coverage, since there are typically more than one of the active sequences in the pool. Moreover, many molecules share

the same secondary structure, which is generally more important for catalysis most of the primary sequence of the enzyme[52].

This pool is then subjected to a selection step: after an incubation period when the pool molecules are allowed to react, catalytically active molecules are isolated. Typically, the isolation procedure relies on a chemical tag attached or lost by the active molecules during incubation. This can be done either by ligation to a short oligonucleotide which allows to separate the active molecules based on size by gel electrophoresis[53], or by using a small chemical group which can be captured using an appropriately coated matrix, as is the case for biotin/(strept)avidin[54] or thiophosphate/thiopropyl sepharose[55]. Conversely, catalysts that cleave RNA are often isolated based on their smaller size and increased electrophoretic mobility[56] or by their release from an affinity column[57]. An essential drawback of this method is that it relies on chemical modification of the active catalysts, which means that the isolated molecules are inherently single-turnover. In some cases, however, it is possible to re-engineer isolated *cis*-acting (deoxy)ribozymes into *trans*-acting ones by simply breaking the molecule into a substrate and enzymatic part, as was done by Breaker and Joyce[57]. Nevertheless, while this may be possible for some nucleic acid enzymes, there is no direct selection pressure during the *in vitro* isolation itself that would favor the enrichment of catalysts acting in *trans*.

The main challenge in the selection of *trans*-acting ribozymes is the fact that the active molecules do not change their chemical structure or size, which makes them inseparable from the inactive ones. The dissolution of genotype from phenotype can only be overcome by connecting these two in a different way. Indeed, a protocol adapted by the Ellington lab for selection of RNA ligases addressed this problem by attaching both a substrate RNA oligonucleotide and a DNA sequence of the pool on a single bead at a ratio of one gene per one bead. They then enclosed the beads in water-in-oil emulsion droplets together with T7 RNA polymerase, nucleotide triphosphates, and a second unattached fluorescently labeled RNA oligonucleotide substrate. Only those beads that carried a DNA template for an active ribozyme bore the ligated fluorescent product and could be easily isolated using fluorescence-activated cell sorting (FACS)[58].

In the third step of *in vitro* selection, oligonucleotides that survived the selection step are amplified. This step differs for the isolation of ribozymes made of RNA and deoxyribozymes made of DNA. Ribozymes are first transcribed from RNA into complementary DNA (cDNA) using reverse transcriptase. Usually, the cDNA can be amplified by PCR at this point. However, this is not always necessary as it is followed by another amplification step during transcription to RNA by RNA polymerase. In contrast, PCR is the only amplification step for deoxyribozymes, after which the double-stranded product is separated to yield single-stranded molecules for the next round of selection.

Importantly, the PCR amplification can introduce an unwanted bias into the selection for sequences that are easier to amplify than others in the pool. This is especially true in later rounds when the active pool members tend to acquire more stable secondary structures. It is therefore critical to keep the number of PCR cycles low to reduce the advantage of easy PCR templates.

During the course of the selection, the pool is probed for catalytic activity. When active catalysts are observed, the selection typically continues until a plateau is reached where no increase in the portion of active pool members is detected between two consecutive rounds. At this point, the most abundant oligonucleotides are identified using either Sanger or next-generation sequencing (NGS) and tested for activity.

2.4.1 Analysis of results

In the past, the conventional approach to identify active molecules was to perform a PCR reaction of the pool from the last round of selection, clone the products into plasmid vectors, transform these into *Escherichia coli*, and sequence the isolated plasmids using Sanger sequencing[57,59]. This protocol relies on the probability of the most abundant (and presumably the most active) sequences to be preferentially sampled rather than those that are rare in the selected pool.

However, the number of individual sequences in an evolved pool is usually tens of thousands, which makes the coverage by Sanger sequencing grossly insufficient. Moreover, several studies have shown that the abundance of a functional nucleic acid molecule at the final round often does not correlate with the highest activity. On the contrary, several studies reported that successful sequences identified in earlier rounds using next generation sequencing showed better activity compared to more prevalent ones from later rounds[60–62]. Indeed, NGS has recently become a preferred method to inspect the evolved pool[63], since a standard commercial NGS usually provides between one and five million sequences, which is sufficient for further analysis.

In particular, the quality of individual reads is assessed, reads of incorrect length or containing other sequencing artifacts are removed, paired-end reads are merged if applicable, constant-region sites like primer-binding sites and adapters are usually removed, and the sequence direction is unified (Figure 6). Although programs that compile this workflow are available[64,65], many laboratories still prefer to use their own customized in-house scripts[66].

In the next step, the best candidate sequences are identified. At this point, the researcher can choose to analyze the full-length sequences or look for repeating short motifs. These can then be assessed individually or grouped into larger families based on their edit distance (number of mutations between each sequence of a family). The individual sequences in these clusters usually differ in only a few mutations

which do not significantly change their secondary structure, but the process greatly simplifies the analysis.

Another way to search for active sequences is to sort the reads based on their predicted secondary structure. Such algorithms are included in some of the software packages available for *in vitro* selection data analysis (for instance APTANI[2], COMPAS, or AptaSUITE). However, the predicted base-pair accuracy based solely on theoretical predictions average at 65 – 70% when compared to experimentally determined secondary structures of nucleic-acid molecules[67]. Moreover, these programs do not search for more complex structures like pseudoknots, non-Watson-Crick base pairs, G-quadruplexes, or triple helices.

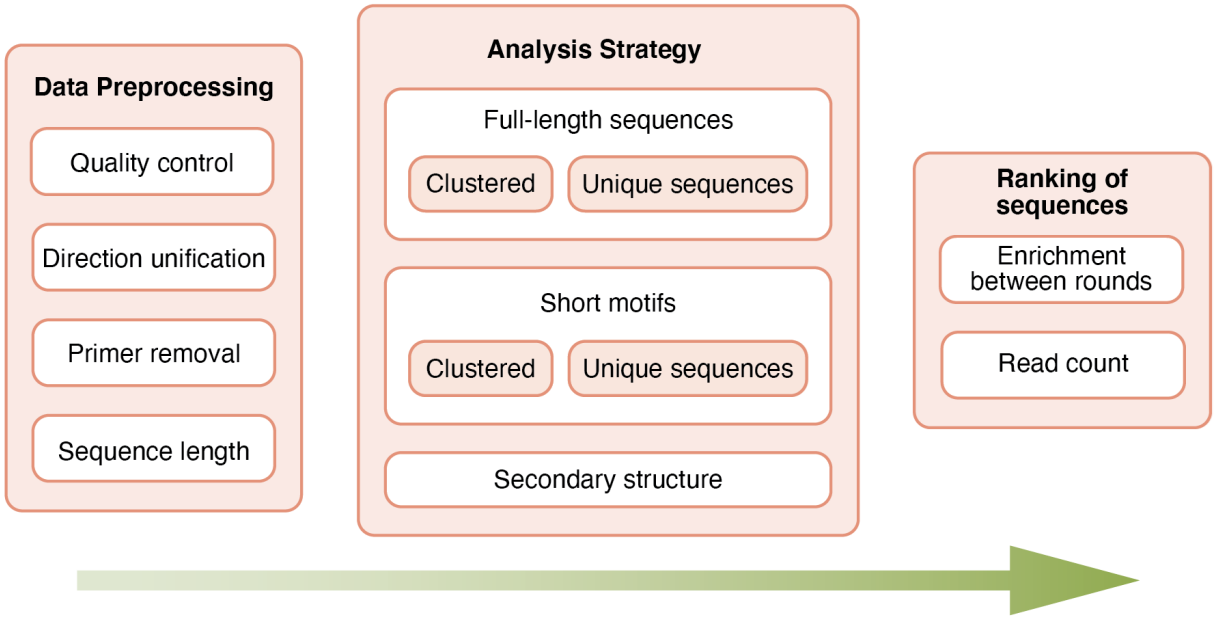


Figure 6. An overview of NGS data analysis workflow.

The final step of NGS data analysis is to evaluate the selection success of individual sequences or clusters. This can be determined based on their absolute read count at the final round. However, several studies used instead a cycle-to-cycle enrichment ratio to identify the most successful sequences[62]. Because this method also accounts for the persistence of individual sequences or clusters over the course of the selection, some argue that it can better distinguish true catalysts from molecules overrepresented due to selection bias or stochastic processes[68]. The final selection of successful sequences is then tested for activity and further characterized. However, it is not unusual for even a highly evolved pool to contain inactive sequences due to stochastic reasons or various types of selection bias.

2.4.2 Structure Determination

One characteristic of an active sequence to be determined is its secondary or tertiary structure. A known secondary structure of an active nucleic acid can help further design additional features like allosteric regulation or even improvement of catalytic rate. Many programs predicting secondary structure are based on thermodynamic parameters like Mfold[69] or ViennaRNA[70]. Others rely on probabilistic context-free grammars or conditional random fields methods[71,72]. However, these programs tend to have a rather modest accuracy of predicted versus correct base pairs, which further decreases with the increasing length of the nucleic acid[67].

Mfold, arguably the most widely used RNA-structure prediction algorithm, uses a free-energy minimization method. To test its accuracy, Doshi and coworkers analyzed a set of almost 1,500 RNA sequences, including different types of rRNA and tRNA. While the folding accuracies were reasonably good for shorter RNA molecules (~70% of correctly predicted base pairs for 5S rRNA and tRNAs), it was less reliable for the longer 16S and 23S rRNAs (both 41% of correctly predicted base pairs)[73]. Indeed, many of the energetically suboptimal structure predictions show a higher percentage of correct base pairs than those predicted to be more energetically favorable.

Organism	Sequence
	10 20
E. coli	G A A **G A G U U** U G A U C A U **G G C U C** A
H. volcanii	A U **U C C G G U** U G A U C C U **G C C G G A**
X. laevis	U A C **C U G G U** U G A U C C U **G C C A G** U
H. sapiens mitochondrial	A A U **A G G U U** U G G U C C U **A G C C U** U

Figure 7. An example of comparative sequence analysis adapted from Woese et al. 1983[74]. A section of the 16S-like ribosomal RNA sequence alignment from different organisms. All sequences form an almost identical hair-pin structure although the primary sequences differ, providing support to an evolutionarily conserved structural element. The helix at position 9-12/22-25 was confirmed by crystal structure reported by Wimberly and colleagues[75].

An alternative method that does not utilize thermodynamic information is comparative sequence analysis. This method has been used to successfully predict the secondary structure of several RNAs[76]. It is based on the principle that RNA or DNA can form structurally equivalent elements like base pairs, helices, or hairpins independent of their primary sequence. The second assumption is that the secondary and tertiary structures of nucleic acids that perform the same function should be similar, and this structure should be maintained throughout evolution, allowing for mutations that do not change this paradigm. In practice, this means that, from a structural point of view, A-T/A-U, G-C, and G-U are interchangeable, and many such

cases are observed in the phylogenetic alignments of structural RNAs[77] (Figure 7). One extraordinary example is the secondary structure of 16S-like ribosomal RNA, which was predicted using comparative sequence analysis by Woese and colleagues in 1983[74]. When its crystal structure was solved 17 years later[75], 97% of the predicted base pairs were indeed present in the crystal structure.

2.5 Artificial ribozymes

To demonstrate the plausibility of the RNA world hypothesis, many researchers set out to show the catalytic abilities of RNA with the use of the new and exciting method of *in vitro* selection. Their efforts showed that RNA can indeed catalyze a wide range of chemical reactions. The most important factor in terms of RNA-based origin of life is the ability to replicate. Pursuing this question, Bartel and colleagues described an RNA ligase that could attach a short oligonucleotide substrate to its 5' end by a 3'-5' phosphodiester bond, similar to that formed by T7 RNA polymerase[78]. They reported a rate enhancement of 7×10^6 over an uncatalyzed reaction, which is significantly slower than the naturally occurring group I intron ribozyme (rate enhancement of 10^{11} fold over the uncatalyzed reaction)[79] or that of the proteinaceous T7 RNA polymerase (i.e. 3×10^{11} fold over the uncatalyzed reaction).

To further explore the capacity of RNA to function as a true RNA polymerase, Ekland and Bartel showed that, by rational design, the original RNA ligase[78] can also work as a true, albeit poor, RNA-dependent RNA polymerase. Indeed, the ribozyme was able to add up to six nucleotides using mononucleotide triphosphates and RNA oligonucleotide as a template[80]. Through several further iterations of this enzyme by various laboratories, Attwater and colleagues were finally able to isolate a ribozyme that could replicate an RNA sequence longer than itself, specifically up to 206 nucleotides[81]. They achieved this through an in-ice *in vitro* selection and rational design based on a previously isolated ribozyme polymerase[82]. However, they also reported bias favoring templates that were G-poor and weakly structured, which is in direct contrast to a secondary structure essential for functional, self-replicating ribozymes.

To overcome this limitation, Horning and Joyce improved Attwater's ribozyme so that it was able to copy even G-rich and structured templates[83]. Moreover, they used their RNA polymerase to carry out a PCR-like reaction to amplify a short template using repeated cycles of heat denaturation and primer extension.

Another important feature of an efficient RNA replicator in a hypothetical RNA world is its ability to recognize self from non-self in order to discriminate between its replication targets and other replicative parasites. To address this problem, Cojocaru and Unrau developed an RNA polymerase ribozyme that efficiently recognized a template in *trans* using a clamping domain and a complementary primer sequence[84].

Despite these considerable advances, current artificial RNA polymerase ribozymes still fall short of both current proteinaceous RNA polymerases and hypothetical RNA replicators of the RNA world. For an artificial replicator to be able to amplify indefinitely, their fidelity needs to be improved to prevent frequent deleterious mutations in their progeny. Moreover, a double-stranded template invasion during initiation and displacement of the product strand from the template present further as of yet untackled problems[85].

Interestingly, Horning's RNA polymerase ribozyme[83] was shown to accept deoxyribonucleotides (dNTPs) as substrates in addition to the original ribonucleotides (NTPs), which were used during the *in vitro* selection[86]. Although the reaction was rather inefficient, it could indeed catalyze DNA polymerization using short RNA templates. This activity would be especially important in the hypothetical RNA world for the transition from RNA to DNA as an information storage molecule.

2.5.1 Other interesting artificial nucleic acid catalysts

An efficient RNA ligase or polymerase ribozyme is just one of the many prerequisites for the RNA world hypothesis. Therefore, researchers looked for other nucleic acid enzymes with the ability to catalyze a wide range of chemical reactions that would be indispensable during the early evolution of life.

Over the years, RNA was found to catalyze a wide variety of chemical reactions. Most artificial ribozymes manipulate a phosphate group: examples include phosphodiester cleavage of the RNA backbone[87], cyclic phosphate hydrolysis[88], RNA phosphorylation[55], tetraphosphate 5'-5' RNA ligation[89], 2'-5' and 2'-3' RNA branch formation[90], and an RNA-peptide ligation via a phosphoamide bond[91].

In addition, some ribozymes promote a carbon-carbon bond formation through Diels-Alder[92] and aldol reactions[93], and Claisen condensation, which is important for lipid synthesis[94]. Others catalyze a carbon-oxide bond formation such as the aminoacylation of an RNA molecule through transesterification, a reaction similar to that catalyzed by proteinaceous aminoacyl-tRNA synthetases[95,96]. Yet other ribozymes promote a nitrogen-carbon bond formation as in alkylation[54], peptide bond formation similar to that in a ribosome[97], and nucleotide synthesis from a phosphorylated sugar and a nucleotide base[98].

In conclusion, ribozymes are capable of a wide range of chemical reactions which both support the RNA world hypothesis and are useful as molecular tools in the laboratory. Moreover, with ~100 articles on the topic of catalytic RNAs published each year, more ribozymes catalyzing new chemical reactions will be certainly reported in the future.

2.5.2 Allosteric ribozymes and deoxyribozymes for sensing applications

In 1995, only two years after the isolation of the first *de novo* artificial ribozyme by Bartel and Szostak[78], Porta and Lizardi designed the first allosterically regulated ribozyme that is activated by the addition of a specific DNA oligonucleotide[99]. They based their construct on the sequence requirements of a self-cleaving hammerhead ribozyme published previously by Haseloff and Gerlach[100] and fused it to a 34-nucleotide effector-recognition loop and a 15-nucleotide inhibitory stretch complementary to the ribozyme's catalytic core (Figure 8). In the absence of the effector DNA oligonucleotide, the inhibitory part hybridizes to part of the ribozyme, rendering it inactive. On the other hand, if the effector oligonucleotide is added, it binds to the effector loop, which prevents the inhibitory part from interacting with the ribozyme core. This allows the ribozyme to fold into the correct conformation and cleave a substrate RNA molecule. In the absence of an effector DNA molecule, the catalytic activity decreased to 4% of a non-regulated control ribozyme, whereas after the addition of the effector DNA molecule, the catalytic activity was partially rescued to 70%.

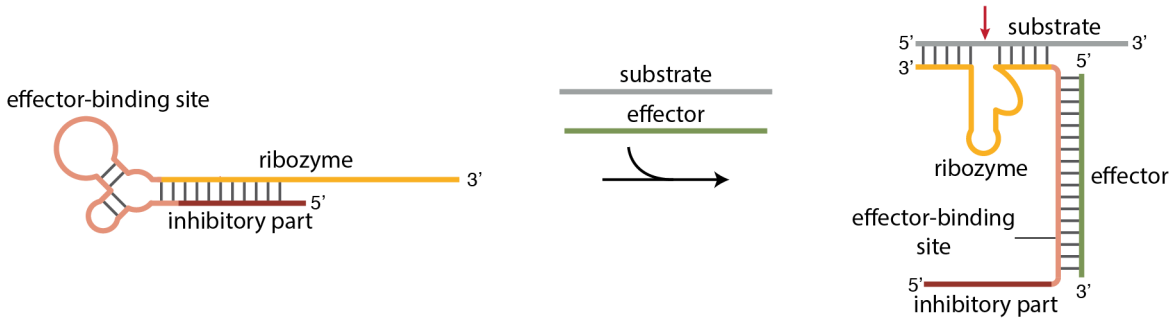


Figure 8. An overview of an allosterically regulated hammerhead ribozyme reported by Porta and Lizardi[99]. On the left, the construct is shown in the absence of an activating effector DNA oligonucleotide. This causes the hybridization of the inhibitory part to the ribozyme, which prevents its catalytic activity. After the addition of the effector DNA and RNA substrate, the effector binds the effector-binding site, displacing the inhibitory part and allowing the ribozyme to adopt the catalytically active conformation. The red arrow indicates the site of cleavage. Note: the scheme was slightly simplified compared to the original report for the purpose of clarity. Adapted from Porta and Lizardi 1995[99].

Similarly, Tang and Breaker developed another example of an allosterically regulated ribozyme by fusing a hammerhead ribozyme with an ATP/adenosine-binding aptamer[101] (Figure 9). Based on their rational design, they developed a ribozyme that is only active in the absence of ATP or adenosine, while its activity is inhibited ~180-fold upon the ligand addition. Using computer modeling and previously published structures of both the aptamer and hammerhead ribozyme, Tang and Breaker

suggested that the folded complex of the ligand-aptamer precludes the active conformation of the ribozyme, causing its drop in activity.

Moreover, the authors developed an allosteric ribozyme that is activated by the addition of the ligand, showing a five-fold increase in activity in the presence of ATP. In their design, they destabilized the stem connecting the aptamer and ribozyme units, which caused a decrease in the ribozyme activity. Only the addition of the ATP ligand stabilized the stem again, resulting in a complete rescue of its enzymatic activity. Overall, their work demonstrated that ribozymes can be both activated and inhibited by ligands using rational design.

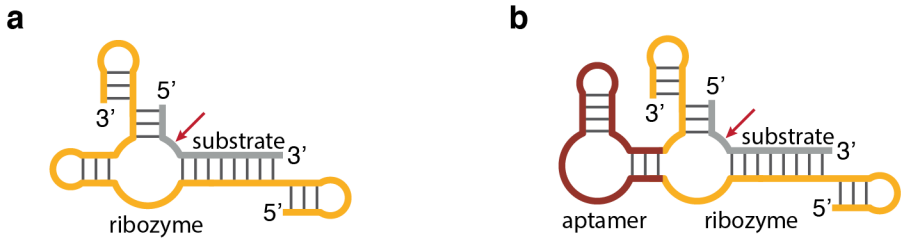


Figure 9. Design of ribozyme allosterically regulated by ATP or adenosine molecule. (a) Secondary structure of a previously described hammerhead ribozyme[102]. (b) An ATP-adenosine aptamer was fused to Stem II of the RNA enzyme to construct the allosterically regulated ribozyme. Adapted from Tang and Breaker 1997[101].

2.5.3 Deoxyribozymes as sensors

Following these first examples, most allosteric deoxyribozymes are based on a similar design: a molecular recognition element (MRE) is joined to a signal producing module. The MRE is usually an aptamer specific for the detected molecule. Currently, many deoxyribozyme-based biosensors take advantage of a peroxidase-mimicking deoxyribozyme such as PS2.M[103] (Figure 10a). This method utilizes a complex of hemin and a G-quadruplex-forming DNA molecule that catalyzes a peroxidase reaction in the presence of peroxide. The substrate can be either a chromogenic ABTS (2,2'-azinobis(3-ethylbenzothiazoline-6-sulfonic acid)), certain fluorogenic compounds[104], or a chemiluminescent luminol. In addition, the reduction of peroxide into water during the reaction also allows for a direct label-free electrochemical signal measurement[105].

Another option often used as the signaling part of a sensor is an RNA-cleaving deoxyribozyme. These DNA enzymes cleave a ribonucleotide linkage of a substrate molecule which can be either covalently attached to the catalysts or a separate molecule base-paired to the enzyme via binding arms (Figure 10b, c). Although they were first conceived as therapeutics, today they are mostly utilized in biosensing applications like environmental monitoring, medical diagnosis, or cellular and *in vivo* imaging.

While the read-out can be a simple gel electrophoresis image that detects the cleaved products, many improvements have been made to facilitate the detection: a fluorophore-quencher pair can be added on either side of the cleavage site[106], or it can be attached to the proximal ends of the deoxyribozyme and substrate[107]. In both cases, upon cleavage the substrate strands dissociate from the enzyme molecule, and fluorescence can be detected.

Due to their chemical stability and low cost of synthesis, deoxyribozymes in particular are an excellent choice for *in vitro* sensing applications.

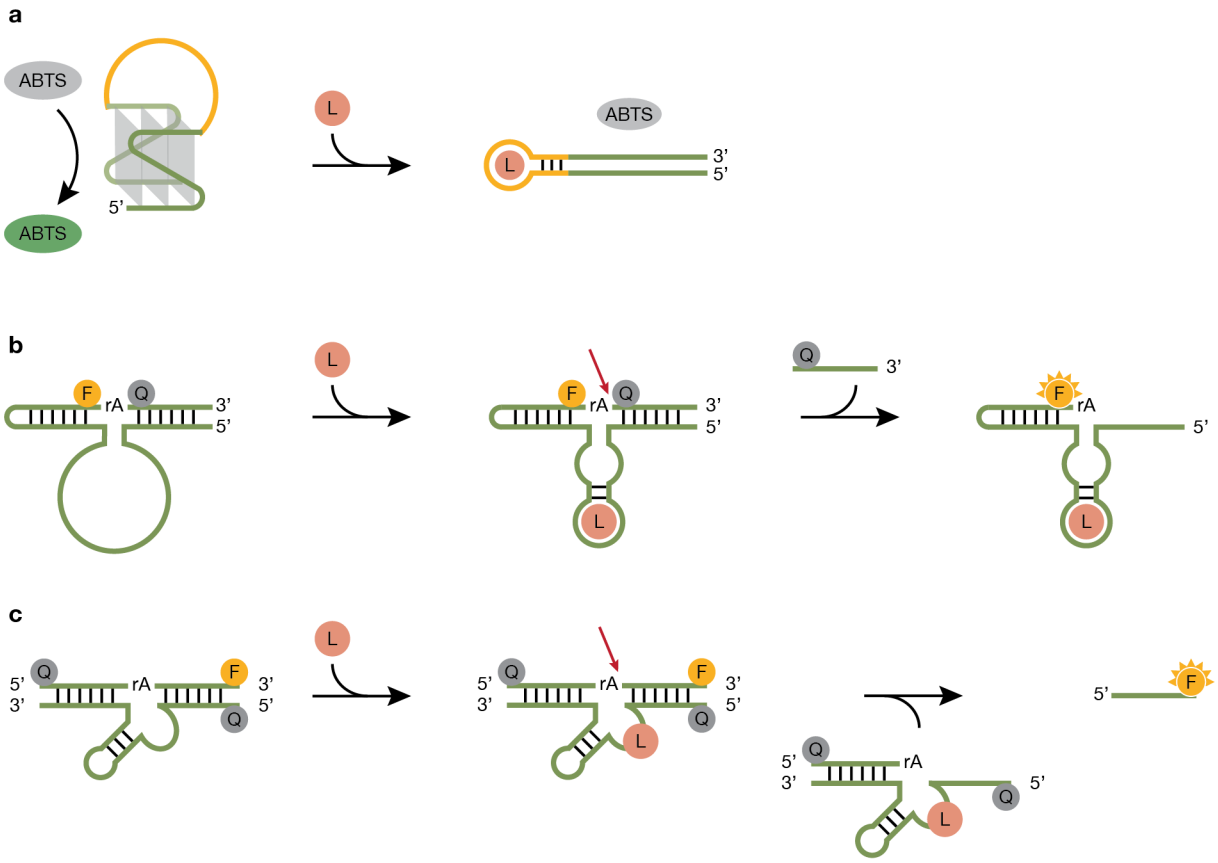


Figure 10. Examples of deoxyribozyme-based sensors. (a) Peroxidase-mimicking deoxyribozyme (green line) with a cocaine aptamer (yellow line) inserted in one of its loops. In the absence of cocaine, the aptamer sequence is unstructured and allows proper folding of the deoxyribozyme which in turn converts a colorless ABTS substrate into a green product. If cocaine (pink circle) is present, it binds the aptamer sequence which precludes a correct folding of the deoxyribozyme, and no green product is detected. Adapted from Gao et al. 2019[108]. (b) RNA-cleaving deoxyribozyme that is only properly folded and active in the presence of *Escherichia coli* extracellular mixture (pink circle). The active enzyme cleaves its own ribonucleotide linkage which uncouples a fluorophore-quencher pair (yellow and grey circle). The fluorescent signal is then detected. Adapted from Ali et al. 2011[106]. (c) A catalytic beacon constructed using a uranyl-dependent RNA-cleaving deoxyribozyme. Similar to (b), upon the ligand binding (pink circle), the enzyme cleaves the RNA linkage and releases a fluorophore-carrying oligonucleotide. As the fluorophore is removed from the vicinity of a quencher, fluorescence is detected. Adapted from Liu et al. 2007[109].

2.6 Fluorescent RNA motifs

Although fluorescent RNA aptamers, as the name suggests, are not catalytically active nucleic acids, they share a number of important features with (deoxy)ribozyme-based sensors. For this reason, I will briefly describe their function and applications here.

These RNA molecules bind a small compound which is intrinsically non-fluorescent. By doing so, the RNA stabilizes the planar dye molecule which in turn shifts its energy dissipation pathway from non-radiative (such as heat) to radiative (such as fluorescence). This simple mechanism is utilized by an assorted collection of RNA-based fluorescent tools like the malachite green aptamer[110], Spinach[111], Broccoli[112], Mango[113], Pepper[114], Corn[115], Red Broccoli[116], and Squash[117].

Similar to (deoxy)ribozymes, these RNA molecules are isolated through an *in vitro* selection method as aptamers binding a specific fluorophore. Their usefulness has been demonstrated in many cellular imaging applications. For example, Spinach RNA was successfully converted into a sensor for a variety of cellular metabolites using the same approach as for the allosteric (deoxy)ribozymes[118,119] (Figure 11). In particular, the fluorophore-binding motif is fused to an aptamer via a short stem in such a way that if the aptamer is not bound by its cognate ligand, the unfolded sequence precludes a correct folding of the fluorophore-binding motif. This unfolded state leads to no fluorescence. On the other hand, if the ligand binds the aptamer, the fluorophore-binding RNA assumes a correct conformation to bind the fluorophore, and fluorescent signal is detected. Other examples include an S-adenosylmethionine (SAM)-activated Red Broccoli, Squash, and Corn motif[115–117].

Although RNA-based sensors are powerful tools for cellular and *in vivo* imaging, DNA-based sensors are arguably more practical for *in vitro* applications due to their superior chemical stability and lower cost of synthesis.

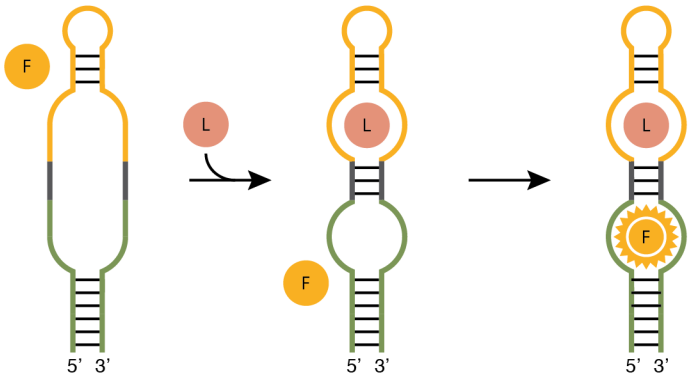


Figure 11. A mechanism of a Spinach-based sensor of small molecules as described in Paige et al. 2012[118]. Green line = fluorophore-binding sequence Spinach; dark grey line = linker sequence; yellow line = metabolite-binding aptamer (adenosine, ADP, guanine, GTP, or SAM); yellow circle = fluorophore; pink circle = target ligand.

2.7 Disadvantages of current nucleic acid sensors

Currently, a wide variety of nucleic acid-based sensors have been developed, a majority of which are made of RNA. RNA sensors are useful for *in vivo* and cellular applications like RNA-tagging[120] and protein[121] and metabolite monitoring[117], primarily because cells can be engineered to directly express such RNA sensors.

On the other hand, DNA is better suited for *in vitro* applications like diagnostics or environmental monitoring due to its better chemical stability and lower cost of synthesis. Currently, the most extensively used DNA-based signaling component are arguably the peroxidase-mimicking and RNA-cleaving deoxyribozymes (Figure 10). However, they both show a rather modest signal enhancement (Figure 12). Moreover, the RNA-cleaving deoxyribozyme is attached to a fluorophore-quencher pair which increases their cost of synthesis.

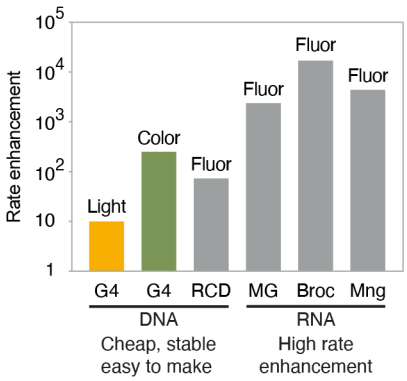


Figure 12. Comparison of current nucleic acid-based signaling systems. G4 = the DNA G-quadruplex peroxidase system that generates light in the presence of luminol (yellow bar)[122] or color in the presence of ABTS (green bar)[123]; RCD = the RNA-cleaving fluorophore-quencher deoxyribozyme[124]; MG = the malachite green RNA aptamer[125]; Broc = the Red Broccoli RNA aptamer[116]; Mng = the Mango RNA aptamer[59]. Grey bars represent fluorogenic systems.

Preferably, the signaling part of a nucleic acid sensor produces an easily detectable signal. Generally, light is an optimal output for several reasons: it can be readily recorded by widely available instruments such as plate-readers, it shows lower background than color or fluorescence and thus allows for broader dynamic range, and it poses no safety hazards as is the case for radioactivity-based assays.

3 Aims

My PhD project was to isolate and characterize deoxyribozymes that produce a chemiluminescent signal through the dephosphorylation of a commercially available substrate CDP-Star. In addition, using proof-of-principle experiments, we showed that the deoxyribozyme can be allosterically regulated by oligonucleotide ligands. Moreover, we optimized the deoxyribozyme's performance by buffer optimization and described its buffer requirements.

The concrete steps of the project were as follows:

1. Design selection buffer and DNA pool.
2. Optimize the *in vitro* selection protocol.
3. Perform initial *in vitro* selection experiment.
4. Identify active sequences; design the reselection pool.
5. Perform *in vitro* reselection experiment.
6. Identify active sequences, and determine secondary structure.
7. Minimize the deoxyribozyme; start initial buffer optimization.
8. Determine sequence requirements and kinetic properties (k_{cat} and K_m).
9. Design, test, and characterize allosteric deoxyribozymes.
10. Determine buffer requirements, and optimize reaction conditions for better performance.

4 Results

"The only failed experiment is the one you learn nothing from."

- John G. D'Angelo

CDP-Star is a commercially available chemiluminescent substrate that is mostly used for the detection of phosphatases in solution. However, it can also be used as a signaling component for the detection of nucleic acids linked to a phosphatase in Northern and Southern blots as well as for the detection of phosphatase-linked antibodies in ELISA experiments. CDP-Star is a 1,2-dioxetane molecule that contains a phosphate group as one of its substituents. When this phosphate group is removed, the rest of the molecule becomes destabilized and decomposes while emitting a flash of light at the wavelength of 466 nm[126,127] (Figure 13). Furthermore, CDP-Star shows a very low background of light produced in the absence of an enzyme, which makes it an excellent substrate for detection applications.

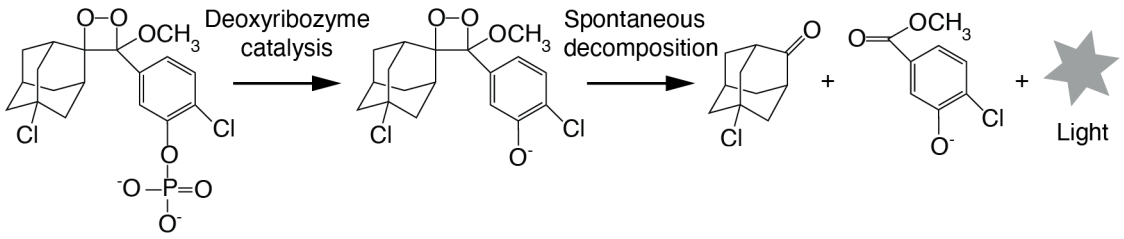


Figure 13. Reaction scheme of light production using CDP-Star. After the substrate's dephosphorylation, the intermediate undergoes a spontaneous decomposition into final products and emits a photon of light.

While the dephosphorylation reaction is normally performed by a proteinaceous phosphatase, we hypothesized that a DNA enzyme that removes the phosphoryl group from CDP-Star and attaches it to itself should also trigger the chemiluminescent reaction. Since a number of RNA and DNA molecules had been previously shown to catalyze a phosphoryl transfer[53,55,128], the possibility that a DNA molecule could use CDP-Star as its phosphate source seemed plausible.

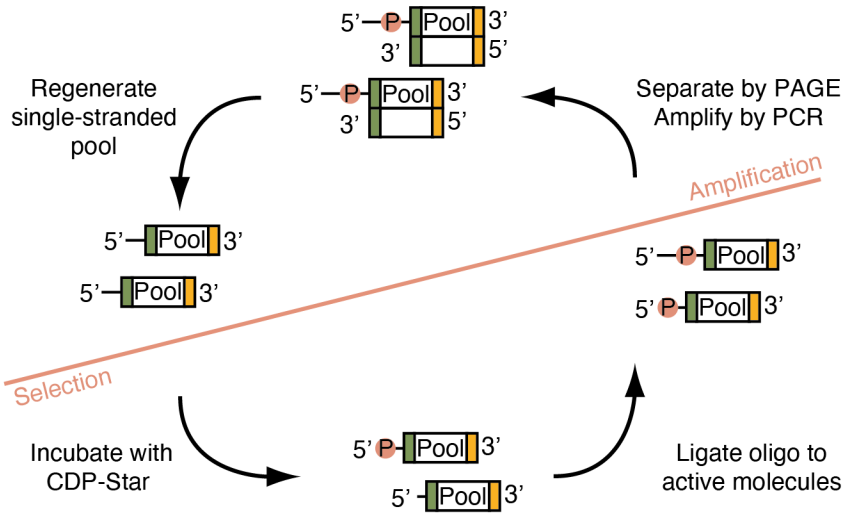


Figure 14. A schematic overview of our *in vitro* selection protocol.

To identify such a deoxyribozyme, we developed a protocol using *in vitro* selection (Figure 14). For this selection, we designed a pool of 10^{16} single-stranded DNA molecules which contained a 70-nucleotide random sequence region. This was flanked by stretches of constant sequences used for ligation at the 5’ end and as a primer-binding site at the 3’ end (Figure 15). The single-stranded pool was then incubated in selection buffer with CDP-Star for 24 hours. During this incubation period, only catalytically active molecules transferred the phosphoryl group from the CDP-Star to their 5’ hydroxyl group. These molecules were then ligated to a short oligonucleotide by splint ligation reaction and then separated using 6% PAGE (polyacrylamide gel electrophoresis). Oligonucleotides that corresponded to the length of 123 nucleotides (100 nt of the original pool plus 23 nt of the ligated oligonucleotide) were cut out and recovered from the gel. This represents the first selection step because only pool members that are monophosphorylated at their 5’ end can serve as acceptors in a ligation reaction. Next, the isolated molecules were amplified by PCR using the ligated oligonucleotide at the 5’ end and the 3’-end constant region as primer binding sites. This step served as a second selection event, since only the 5’-end phosphorylated, and thus ligated molecules, contained the correct forward-primer binding site. At the same time, the DNA molecules were amplified for the next round of selection. The forward primer contained a ribonucleotide at its 3’ end, and the reverse primer carried a monophosphate at its 5’ terminus. In this way, the region ligated to the 5’ end could be removed by base hydrolysis, while the reverse complement could be degraded by λ-exonuclease to regenerate the single-stranded pool. After one such round of *in vitro* selection, the pool was enriched for active molecules, compared to that from the previous round.

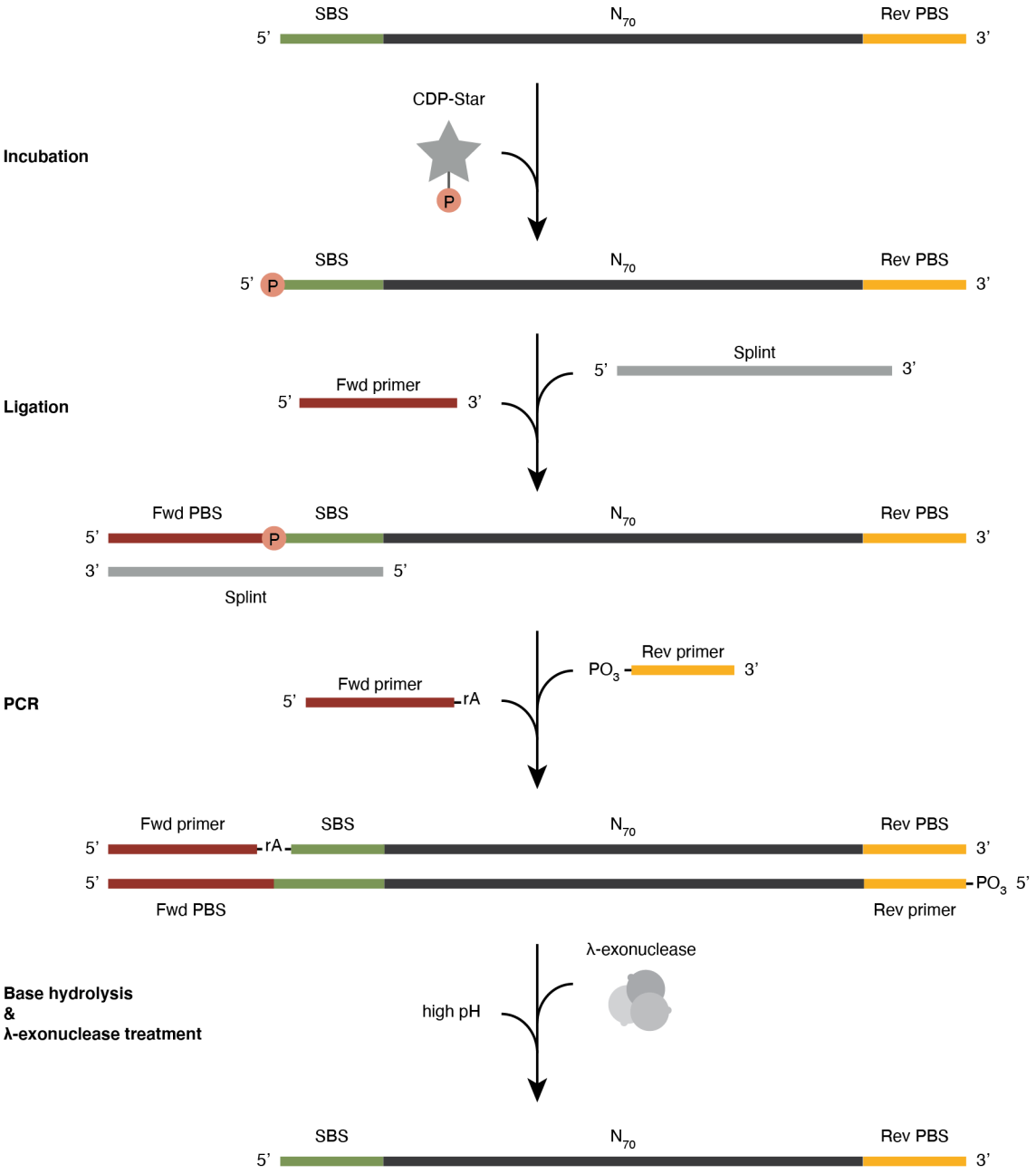


Figure 15. The life-cycle of a catalytically active DNA molecule during one round of our *in vitro* selection protocol. SBS = splint-binding site; PBS = primer-binding site; N = any nucleotide (A,C,G, or T).

4.1 Buffer selection

Optimizing the buffer conditions of an *in vitro* selection is of the utmost importance. The ions present in the buffer help fold the active nucleic acid into its correct conformation and may also serve as necessary co-factors for the catalysis itself. At the same time, it is challenging to predict which ions a future deoxyribozyme will require before the *in vitro* selection itself. For this reason, we decided to perform two parallel selections in two distinct buffers, each containing a different set of cations.

The first buffer contained a set of cations previously used in other selections of nucleic acid kinases, namely Na^{+}, K^{+}, Mg^{2+}, Ca^{2+}, Mn^{2+}, and Cu^{2+} [53,55]. Because there was a possibility that CDP-Star might simply be too stable to serve as an efficient substrate for a deoxyribozyme, we investigated a series of cations known to promote the hydrolysis of various monophosphates, in particular Ce^{4+}, Pb^{2+}, Zn^{2+}, and Co^{2+} [129–132]. Specifically, we incubated CDP-Star in the presence of these cations for 48 hours and observed the substrate's non-enzymatic hydrolysis by measuring the light production using a plate reader. Some of these cations significantly increased CDP-Star degradation (Figure 16), and we used these data to design the second selection buffer containing Ce^{4+}, Pb^{2+}, and Zn^{2+}.

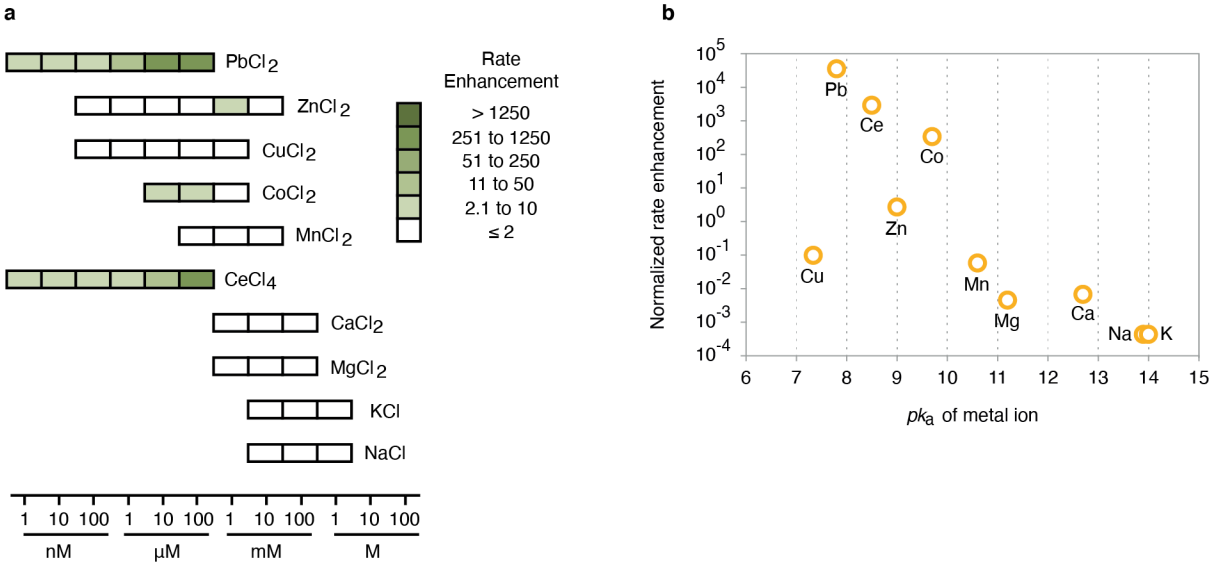
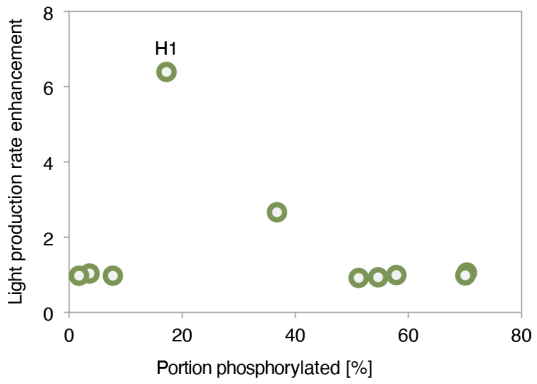


Figure 16. Effect of buffer composition on the nonenzymatic hydrolysis of CDP-Star. (a) Effects of metal ions on the nonenzymatic rate enhancement of light production by CDP-Star. The substrate was incubated in 50 mM HEPES pH 7.4 for 48 hours in the presence or absence of the indicated metal ion, and light production was measured continuously. KCl, $PbCl_2$, $ZnCl_2$, and $CeCl_4$ were the metal ions in our selection buffer. A parallel selection in a buffer containing $MnCl_2$, $CaCl_2$, $MgCl_2$, KCl and NaCl did not yield deoxyribozymes. (b) Relationship between the pKa of the metal ion and rate enhancement of light production when normalized for concentration.

4.2 *In vitro* selection experiments

The initial selection was performed in parallel in two distinct buffers as described above. After 7 rounds, the incubation time was decreased from 24 hours to ten minutes to favor the more efficient molecules in the pool. After another two rounds, only the pool incubated in buffer 2 (50 mM HEPES pH 7.4, 200 mM KCl, 1 mM $ZnCl_2$, 1 μM $Ce(SO_4)_2$, 0.1 μM $PbCl_2$) displayed any activity. This pool was cloned, and 25 clones were identified by Sanger sequencing. These clones could be assigned to ten distinct families, and from each family, one representative sequence was analyzed for catalytic activity by both ligation and light-production assay (Figure 17). Because an efficient chemiluminescent deoxyribozyme was our ultimate goal, we chose the sequence of a DNA molecule that produced light most effectively as our starting point (H1 in Figure 17) for the reselection, although it did not perform as well in the ligation assay.

a



b

	Light production [rate enhancement]	Portion phosphorylated [%]
H1	6.4	17.3
Clone 2	2.7	36.8
Clone 3	1.1	70.4
Clone 4	1.0	3.7
Clone 5	1.0	57.9
Clone 6	1.0	70.1
Clone 7	1.0	7.8
Clone 8	1.0	1.8
Clone 9	0.9	54.7
Clone 10	0.9	51.2
Starting pool	1.0	0.0

Figure 17. Activity of deoxyribozymes isolated in the initial selection. (a) Portion phosphorylated of each clone from the initial selection plotted against its light production rate enhancement. (b) Values of light production and portion phosphorylated of each clone from the initial selection. H1 = Sequence, on which the reselection pool was based.

There are two main reasons to perform a reselection (Figure 18). First, we wanted to search the surrounding sequence space of the best clone for better catalysts. Second, using the sequence data from the reselection, we could determine the deoxyribozyme's secondary structure which is, in turn, crucial for its minimization and conversion to an allosterically regulated sensor. The reselection starting pool was generated from the H1 parent sequence by mutagenesis at the rate of 21%, which means that each position has a 79% probability of being the same as in H1 and 7% of carrying each of the other three nucleotide bases. A pool of 10^{14} deoxyribozyme variants was subjected to six rounds of selection with the incubation time starting at 10 minutes (Round 1 to 3), which was then decreased to one minute

(Round 3 to 6) to favor the fastest catalysts in the pool. When robust activity was detected, the evolved pool was characterized by next-generation sequencing.

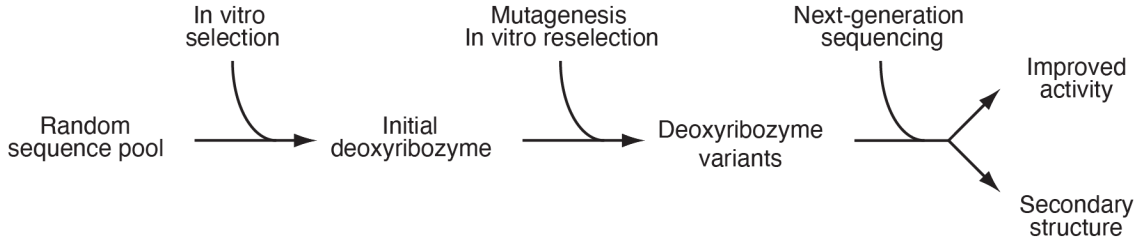


Figure 18. Workflow of artificial evolution experiments, as described in the main text.

4.3 Secondary structure determination

The sequencing identified ~135,000 distinct deoxyribozyme variants, with the most abundant variant comprising 6.3% of the reads. The 23 most abundant variants showed a robust catalytic activity (Figure 19). Furthermore, analysis of the data produced a wealth of information about the sequence requirements of the deoxyribozyme, its catalytic core, and secondary structure. For instance, a sequence alignment revealed three conserved regions (at positions 1 – 6, 33 – 42, and 61 – 82) interspersed with variable regions (Figure 20 and Figure S1). We used this information to design a minimized deoxyribozyme in which the first two variable regions were substituted by four adenosine nucleotides, and the terminal variable region was deleted. The catalytic activity of this 46-nt construct, which we named Supernova, remained comparable to that of the full-length sequence (H2 vs. H2 core in Figure 20a).

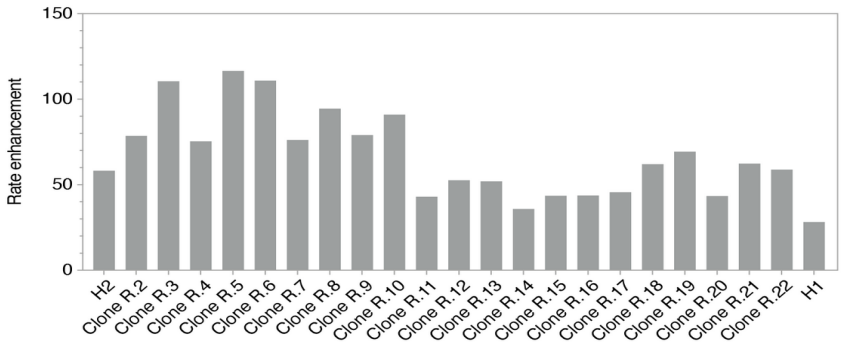
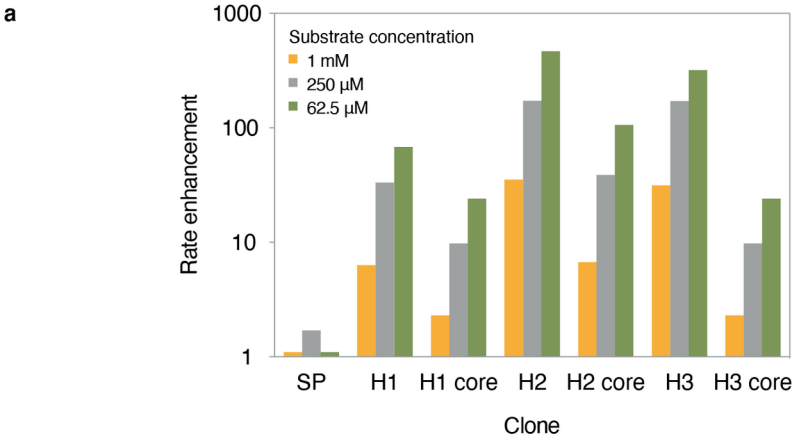


Figure 19. Rate enhancement of the 23 most abundant clones based on NGS of the reselection. Clones were tested at 1 μM deoxyribozyme and 250 μM CDP-Star concentrations in selection buffer. H1 = the most active sequence from the initial selection; H2 = the most abundant sequence from reselection.



b

```
                    Variable region 1          Variable region 2                    PBS
H1          GGAAGAGATGGCGACGACACAGGGACGATGCCGAATATCCTCAGTGCGCAGGGCCGCAGGGGGGAGTGACTTGGGATGGGGGGTCAGCTGATCCTGATGG-----
H1 core     GGAAGA----------AAAA----------GAATATCCTC-------AAAA-------GGGGAGTGACTTGGGATGGGGG----------------------
H2          GGAAGATATGGCGCGAACACAATGGACAATGCCGAATATCCCCCGCACGCAGGGCAACAAGGGGGAGTGACTTGGGATGGGGGCTGCACTAATGATCTGCCCGATG
H2 core     GGAAGA----------AAAA----------GAATATCCCC-------AAAA-------GGGGAGTGACTTGGGATGGGGG----------------------
H3          GGAAGAAATGTAGAGGAAACAGTGACTCTGCAGAATATCCTCACTGCGTAGTGGGGCAGGGGGGAGTGACTTGGGATGGGGGCTACACTAATGATCTGCCCGATG
H3 core     GGAAGA----------AAAA----------GAATATCCTC-------AAAA-------GGGGAGTGACTTGGGATGGGGG----------------------
```

Figure 20. Identification of the minimized catalytic core of Supernova. (a) Rate enhancement of light-production of full-length and minimized deoxyribozymes at three different concentrations of CDP-Star. Reactions contained 1 µM deoxyribozyme, 1.5 µM blocking oligonucleotide where necessary (i.e. the PBS is present) and selection buffer. (b) Sequence alignment of full-length and minimized deoxyribozymes. H1 = the most active sequence from the initial selection; H2 and H3 = two of the most abundant sequences from the reselection; H1 core, H2 core, and H3 core = minimized versions of these deoxyribozymes; variable region 1 and variable region 2 = variable regions identified by next-generation sequencing and comparative sequence analysis of evolved pools; PBS = primer binding site. Positions at which these deoxyribozymes differ from H1 are shown in pink. Note that H2 core is the sequence of Supernova.

In addition, comparative sequence analysis followed by mutual information assessment[133,134] revealed an unusual triple-helical structure capped by an 11-nucleotide purine-rich loop at one end and a variable region 2 at the other. Another 11-nucleotide purine-rich region, which was interrupted by variable region 1, stretched from the 5'-end of the triple helix to the 5' phosphorylated active site (Figure 21).

Indeed, we confirmed the triple-helical structure by covariation experiments, in which we designed single and double mutants of the canonical antiparallel triple C-G:G in such a way that it gradually transitioned into another canonical antiparallel triple T-A:T (Figure 22). While the single and double mutants were mostly deleterious, in most cases the activity was rescued in the T-A:T-containing mutants, confirming that the positions were not independent. To provide further evidence for the triple-helical structure, we tested another 11 triple combinations at each of the triple helix positions and plotted their activity against the number of disrupted hydrogen bonds compared to the original C-G:G triple (Figure 23). A clear correlation of mu-

tant activity with the expected number of hydrogen bonds at the triple positions further supports the triple-helical structure.

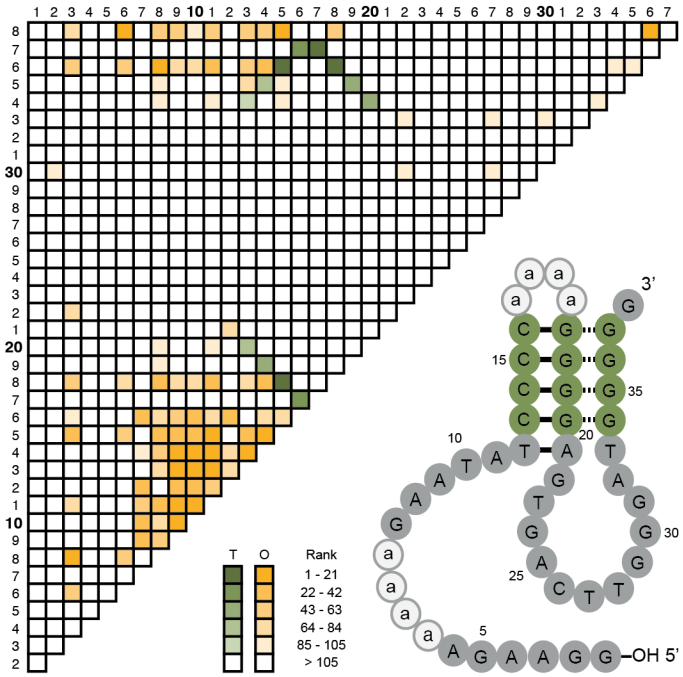


Figure 21. Correlation network and secondary structure of a minimized light-producing deoxyribozyme. Correlations are ranked by mutual information value. Those corresponding to base triples (T) are shown in green, and those corresponding to other interactions (O) are shown in orange. Nucleotides in variable regions 1 and 2 are shown in lower-case type.

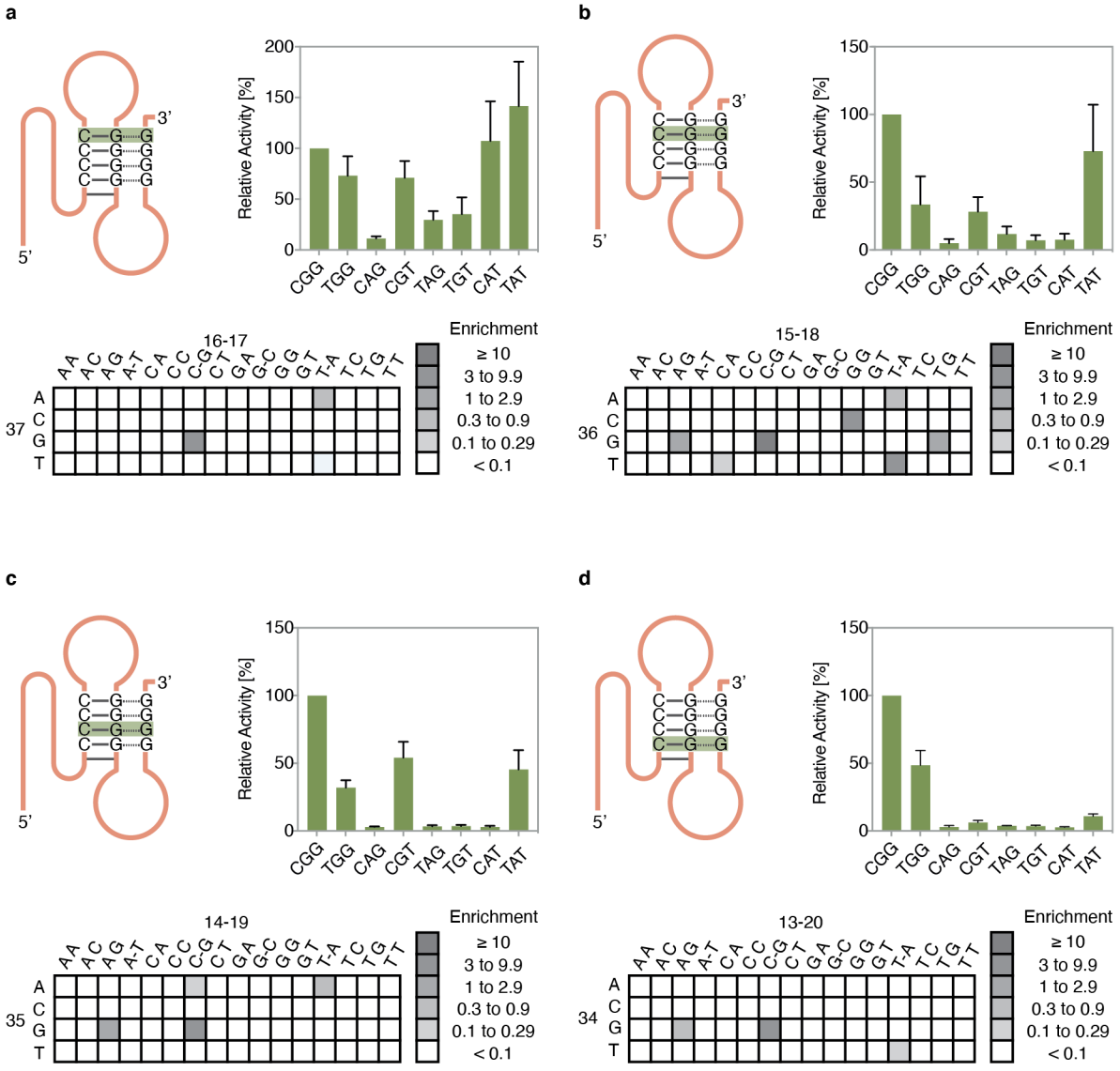


Figure 22. A purine-motif triple helix in the catalytic core of Supernova. (a) Evidence for the 16-17-37 base triple. Upper left: secondary structure model with the 16-17-37 base triple highlighted in green. Upper right: compensatory mutational effects at positions 16, 17, and 37. Activity is normalized to that of Supernova (which contains a CGG triple at this position in the triple helix). Below: nucleotide enrichments at positions 16, 17, and 37 determined by high-throughput sequencing. Enrichment values were determined by dividing the frequency of each possible sequence at positions 16, 17, and 37 in the evolved pool by its frequency in the starting pool. (b) Same, but for the 15-18-36 base triple. (c) Same, but for the 14-19-35 base triple. (d) Same, but for the 13-20-34 base triple.

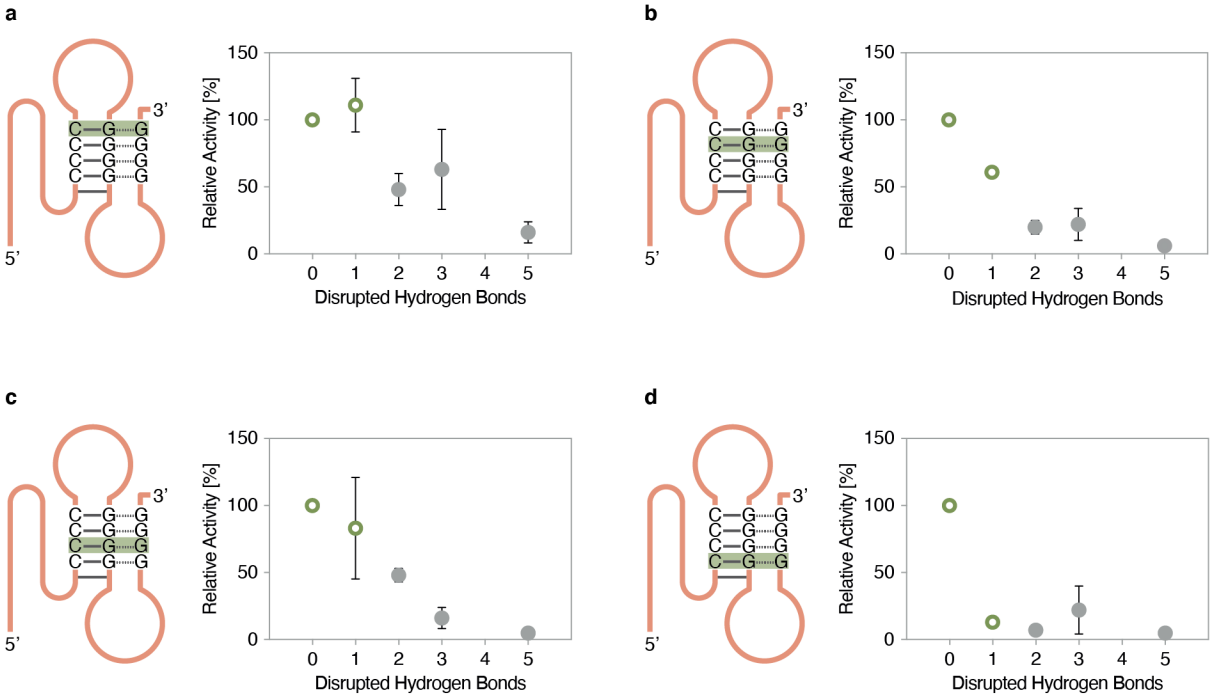


Figure 23. Relationship between the expected number of disrupted hydrogen bonds in mutated base triples and catalytic activity. This analysis assumes that each C-G:G triple in the triple helix contains three hydrogen bonds on the Watson-Crick face and two hydrogen bonds on the Hoogsteen face, and that disrupting one edge of the triple does not disrupt the other (for example, two of the original five hydrogen bonds would be disrupted in a C-G:A triple). (a) Activity of variants with partially or fully disrupted 16-17-37 base triple compared to the original C-G:G sequence. (b) Same, but for the 15-18-36 base triple. (c) Same, but for the 14-19-35 base triple. (d) Same, but for the 13-20-34 base triple. Points shown as green circles indicate canonical purine-motif base triples (C-G:G, T-A:T, or T-A:A).

4.4 Sequence requirements of Supernova

To better understand the sequence requirements of Supernova, we explored a number of its features. First, we tested if all four triples in the triple helix were required for activity by deleting one, two, or three of the triples at both ends. As Figure 24a shows, only a single triple at the 3'-end could be deleted for Supernova to partially retain its activity. This explains why this triple can be also largely mutated without significantly decreasing the catalytic activity (Figure 22a and 23a). Furthermore, inserting one additional C-G:G triple at the 3' end of the triple helix did not change its activity, suggesting that the deoxyribozyme's activity does not increase with the additional triple helix stabilization (Figure 24b).

Next, we tested if the regions that were deleted from the originally isolated deoxyribozyme (position 83 – 85 of isolated deoxyribozymes) or substituted with four adenosine nucleotides (positions 7 – 32 and 43 – 60 of isolated deoxyribozymes) could

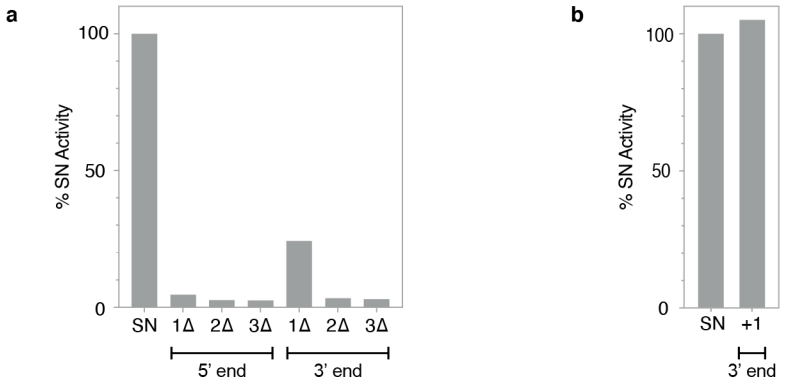


Figure 24. The effect of deleting or adding base triples on light production of Supernova. (a) Light production rate enhancements of Supernova mutants where one, two, or three triples were deleted from the triple helix from either the 5' end or 3' end. (b) Light production rate enhancement of the Supernova mutant with an additional triple at the 3' end of the triple helix. Measurements were performed at 1 μM deoxyribozyme and 250 μM CDP-Star concentrations in optimized buffer.

contain stretches of any length or sequence. We tested Supernova mutants where we inserted either 3, 10, or 30 nucleotides of a random sequence (N), or 3, 10, or 30 nucleotides of an arbitrary sequence (5 different sequences for each location and length) at each of the regions. The results indicate that extra nucleotides can be inserted at all the tested regions. While the decrease in activity is more prominent in mutants with longer inserts for all the insertion sites, the activity is never completely lost, even with inserts as long as 30 nucleotides. Moreover, variable region 1 is the most sensitive to these changes of the three tested sites (Figure 25).

In addition, to look for better catalysts, we tested minimized deoxyribozymes containing point mutations which occurred frequently in the unpaired regions of the reselected pool. As figure 26 shows, deoxyribozymes with A3C, G38C, or G38T mutations show a slight increase in activity, while most of the other mutants show a significant decrease. A possible reason is that cores containing these point mutations require the removed variable regions to catalyze the reaction efficiently.

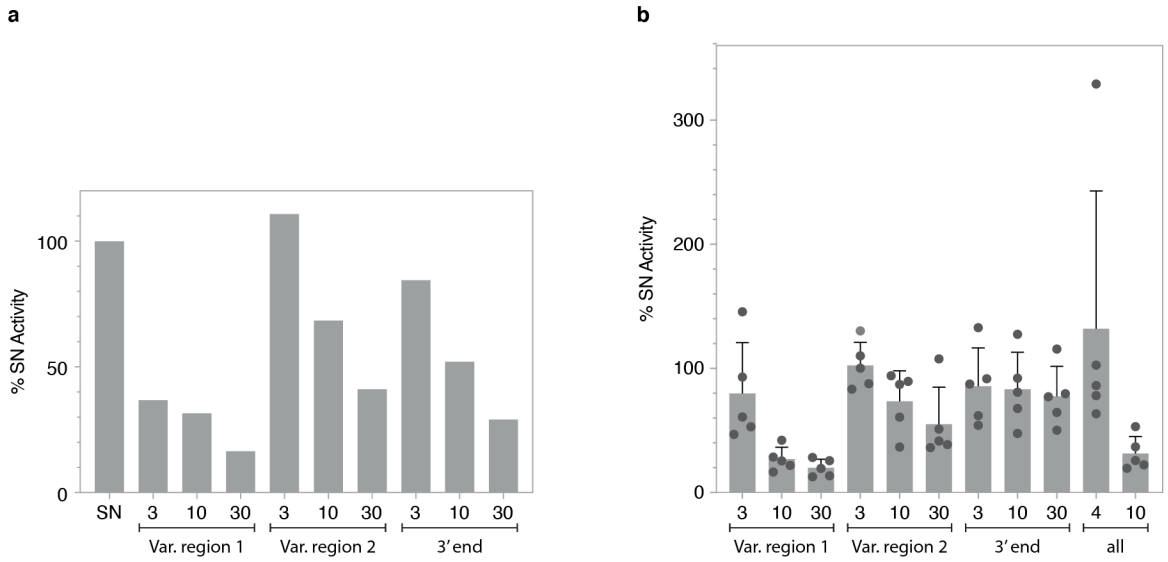


Figure 25. Light production rate enhancements of Supernova mutants with (a) inserted random sequences of varying lengths at either variable region 1, variable region 2, or at their 3' end and (b) inserted arbitrary sequences of varying lengths at either variable region 1, variable region 2, or at their 3' end. Each bar represents five different arbitrary sequence insertions. All measurements were performed at 1 μM deoxyribozyme and 250 μM CDP-Star concentrations in optimized buffer.

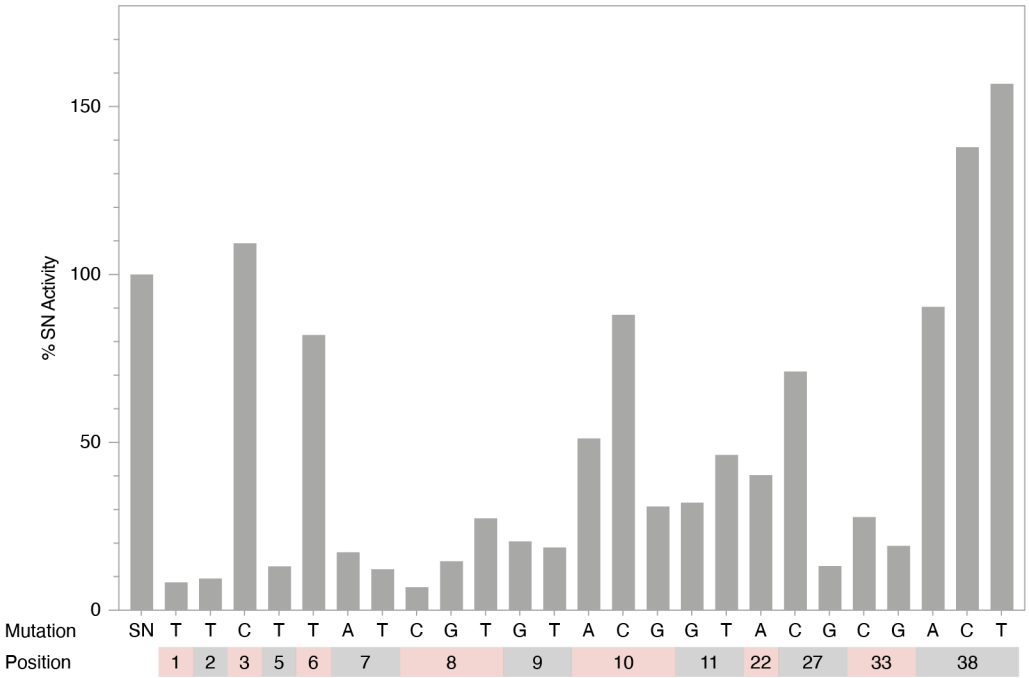


Figure 26. Relative light production rate enhancements of Supernova (SN) mutants with point mutations in unpaired regions. Measurements were performed at 1 μM deoxyribozyme and 250 μM CDP-Star concentrations in optimized buffer.

4.5 Kinetic characterization

For kinetic characterization of Supernova, we performed a CDP-Star titration at a range between 1 µM and 500 µM. At these concentrations, Supernova followed a Michaelis-Menten kinetics, with k_{cat} = 0.15 ± 0.03 min^{-1} and a K_m = 130 ± 70 µM (Figure 27a and Figure S2). A higher substrate concentration inhibited the enzymatic activity, probably due to chelation of zinc in the buffer. On the other hand, the lowest Supernova concentration detectable using our instrument set-up was at ~10 nM (Figure 27b).

Rate enhancement is the ratio of light production in the presence and absence of Supernova. To maximize it, using a low concentration of substrate is desirable. This is due to the fact that the background signal increases linearly from 1 µM (the detection limit of the plate reader) up to 3,000 µM (the highest tested concentration), while the Supernova signal follows the Michaelis-Menten curve. Thus, the highest rate enhancement would be expected below the K_m concentration where the rate increase ceases to be linear. In fact, the highest rate enhancement was detected at the CDP-Star concentration of 62.5 µM, where Supernova generated light 6,500-fold more efficiently than the non-enzymatic background reaction (Figure 27c). This means that Supernova far exceeds the rate enhancement of G-quadruplexes that produce a light or colorimetric signal in the presence of luminol or ABTS. Moreover, it is also comparable to the rate enhancement of RNA-based fluorogenic systems like malachite green [125], Red Broccoli [116], and Mango[59,113] (Figure 27d). However, the advantage of the Supernova system is that it uses DNA, which is more chemically stable, less expensive to synthesize, and easier to prepare and handle than RNA.

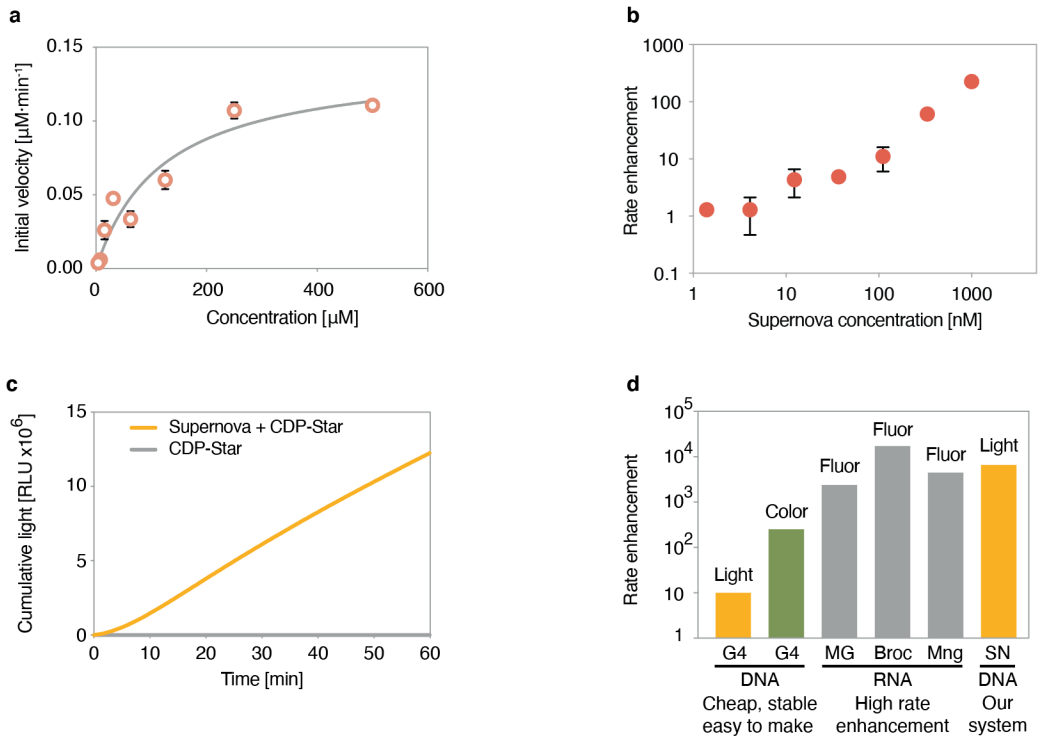


Figure 27. The rate of Supernova and its comparison to other systems. (a) Michaelis-Menten plot showing the initial velocity of Supernova at different concentrations of substrate. Velocities were measured at a Supernova concentration of 1 μM. (b) Light production at different concentrations of Supernova. Reactions contained 62.5 μM CDP-Star, and were performed in a buffer containing 50 mM HEPES pH 7.4, 20 mM KCl, and 1 mM $ZnCl_2$. Each point is the average of three measurements, with error bars indicating the standard deviation. (c) Cumulative light production by Supernova under optimal conditions. The rate enhancement peaked at ~6,500-fold after 10 minutes. Reactions contained 30 μM Supernova and 62.5 μM CDP-Star. Note that the rate at which CDP-Star decomposed after dephosphorylation is slower than the rate of the dephosphorylation reaction catalyzed by Supernova under these conditions. For this reason, the light-producing reaction took longer to reach a plateau (Figure 28). (d) Comparison of Supernova to existing nucleic acid-based methods to generate colorimetric, fluorescent, and chemiluminescent signals. G4 = the DNA G-quadruplex-peroxidase system to generate light (in the presence of luminol) or color (in the presence of ABTS); MG = the malachite green RNA aptamer; Broc = the Red Broccoli RNA aptamer; Mng = the Mango RNA aptamer; SN = the Supernova deoxyribozyme described in this work.

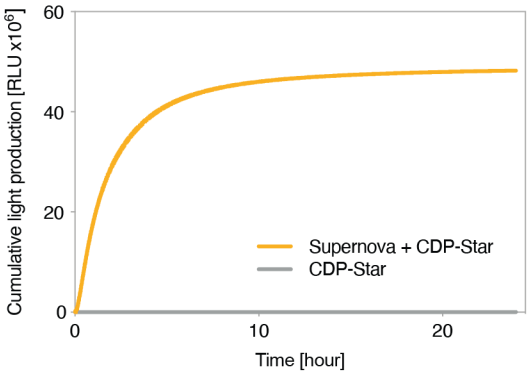


Figure 28. Light production of Supernova over 24 hours. Reactions were performed as in Figure 27c.

To further characterize Supernova, we examined its specificity for CDP-Star by testing other possible substrates, including all four ribonucleotide triphosphates and CDP-Star's closely related commercial predecessor, CSPD. The results indicate that only CDP-Star and, to a lesser extent, CSPD can be used as phosphate donors, suggesting that Supernova is relatively specific for CDP-Star (Figure 29).

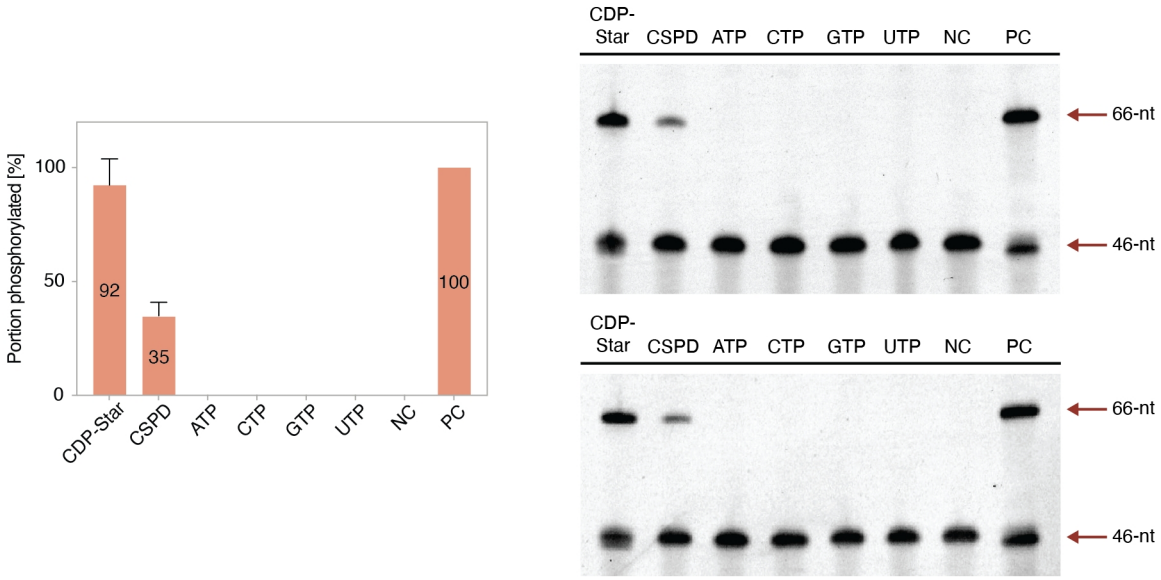


Figure 29. Substrate specificity of Supernova. Self-phosphorylation of Supernova in the presence of CDP-Star, CSPD (a substrate that differs from CDP-Star by the deletion of a single chlorine), ATP, CTP, GTP, and UTP. Reactions contained 1 μM Supernova and 100 μM substrate, and were incubated for 1 hour in a buffer containing 50 mM HEPES pH 7.4, 20 mM KCl, and 1 mM $ZnCl_2$. They were then analyzed using the ligation assay as described in Materials and Methods. NC indicates a reaction containing Supernova with no substrate, while PC indicates a 5'-phosphorylated oligonucleotide that was used as both a control for the ligation reaction and as a marker. Unreacted molecules were 46 nucleotides long, while ligated ones were 66 nucleotides long (red arrows).

4.6 Supernova as a sensor

In many applications that use DNA nanotechnology, it is desirable to link an easily detectable output to a molecular input, for example light production upon the binding of a ligand. To investigate if Supernova can be turned into such a detection system, we constructed several variants of Supernova that were able to detect the presence of short oligonucleotides in a specific manner. We utilized two sites of Supernova where additional sequences can be inserted without a significant loss of catalytic activity: variable region 2 and the 3' end (Figure 25). In the variable region 2, we inserted a part of the sequence we wished to detect, while the reverse complement of the full-length detected oligonucleotide was attached to the 3' end (Figure 30a). We hypothesized that, in the absence of the target sequence, the 3' terminus of Supernova would hybridize with the sequence inserted in variable region 2. This precludes Supernova from adopting the active conformation, and thus it results in no light production. Conversely, if the target oligonucleotide is present, it competes with variable region 2 for binding to the 3' end in a concentration-dependent manner. Following these predictions, Supernova sensor variants produced almost no light in the absence of a target sequence. Conversely, light production increased up to 38-fold in the presence of the ligand molecule at saturating concentration (Figure 30b). This corresponds to nearly the same level of light produced by unmodified Supernova at the same reaction conditions (Figure 30c).

Moreover, to examine if the sensors can reliably distinguish oligonucleotides of different sequences, we tested 5 variants of the Supernova sensor. Each Supernova sensor was designed to detect an oligonucleotide of a different sequence. Each variant only produced light in the presence of the target oligonucleotide, and no light was produced in the presence of any of the remaining four target molecules (Figure 30d). These data demonstrate that the Supernova-based oligonucleotide sensors are both versatile and specific. Moreover, their performance could possibly be further improved by an *in vitro* selection experiment that incorporated alternating rounds of positive (in the presence of the target molecule) and negative selection (in the absence of the target molecule).

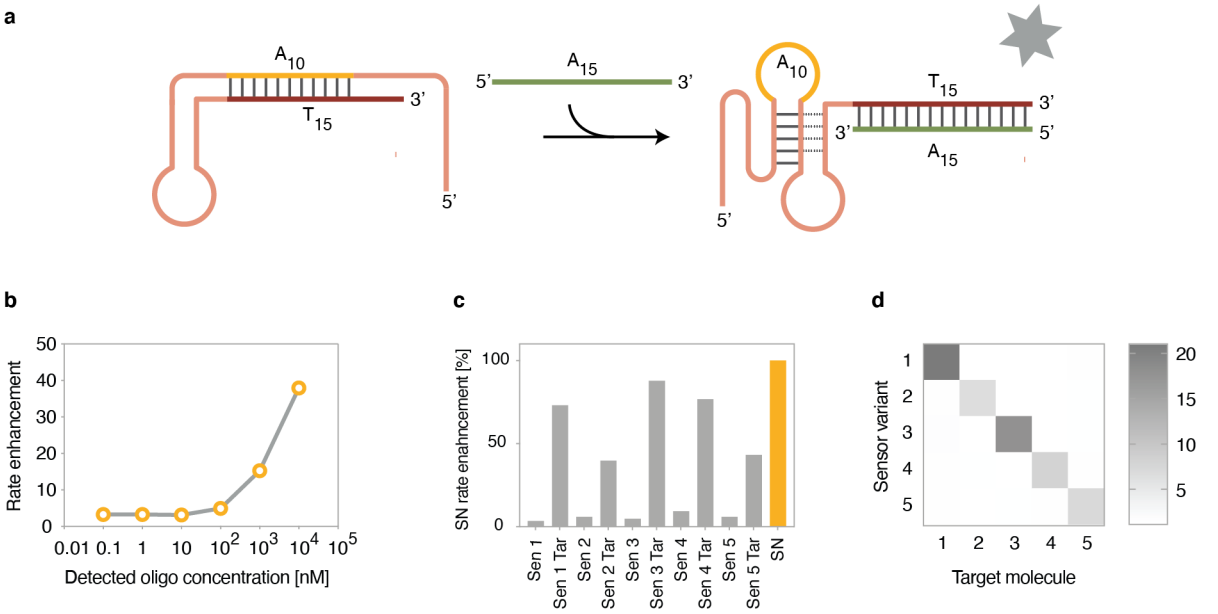


Figure 30. A programable light-producing deoxyribozyme sensor that detects oligonucleotides. (a) Schematic representation of the sensor in its OFF and ON conformations. (b) Rate enhancement of light production as a function of concentration of the target oligonucleotide. (c) Light production of unmodified Supernova (yellow bar), and five Supernova sensors (grey bars). (d) Rate enhancement of light production of five deoxyribozyme sensors in the presence of either the target oligonucleotide (for example, sensor 1 and oligonucleotide 1) or non-target oligonucleotides (for example sensor 1 and oligonucleotides 2–5). The shade of each square indicates the rate enhancement of light production for a particular combination of sensor and oligonucleotide. Reactions contained 1 μM Supernova or Supernova sensor, 10 μM target oligonucleotide (or 0 μM for unmodified Supernova), and 62.5 μM CDP-Star, and were performed in a buffer containing 50 mM HEPES pH 7.4, 20 mM KCl, and 1 mM $ZnCl_2$. See Table 1 for more information about the sequences.

4.7 Characterization and optimization of reaction conditions

As we have described in our proof-of-principle experiment above, Supernova can be turned into a sensor by rational design. For other such sensors to be developed for a wide range of applications in both basic research and diagnostics, it is essential to understand how different reaction conditions affect the ability of Supernova to produce light. For this reason, we proceeded to investigate the effects of buffer composition and the reaction conditions on the light production and self-phosphorylation of Supernova.

To briefly summarize, the Supernova deoxyribozyme exhibits the highest rate enhancement in light production at the substrate concentration of ~60 μM, which is significantly improved over that at 1 mM concentration used in the initial selection. Moreover, we observed a linear increase of rate enhancement with increasing enzyme concentration up to 30 μM Supernova, which was the highest concentration tested. While the Pb^{2+} and Ce^{4+} ions could be omitted, Zn^{2+} is essential and could

not be substituted by any other divalent ion examined. A modest increase in rate enhancement was further achieved by decreasing potassium concentration to 20 mM, combined with the addition of 15% DMSO.

Details are described in the accepted manuscript below[135]. Its supplementary information and the description of the author's contribution to this manuscript is attached in the Supplementary information section of this work (page 87 and 102, respectively).

Optimizing the chemiluminescence of a light-producing deoxyribozyme

Martin Jakubec,[a, b] Karolína Pšenáková,[a, b] Katerina Svehlova,[a, b] Edward A. Curtis*[a]

[a] M. Jakubec, K. Pšenáková, K. Svehlova, E. A. Curtis
Institute of Organic Chemistry and Biochemistry
Prague, Czech Republic
E-mail: curtis@uochb.cas.cz
[b] M. Jakubec, K. Pšenáková, K. Svehlova
Faculty of Science
Charles University in Prague
Prague, Czech Republic

Supporting information for this article is given via a link at the end of the document.

Abstract: Supernova is a chemiluminescent deoxyribozyme recently discovered in our group. It transfers the phosphate group from the 1,2-dioxetane substrate CDP-Star to its 5' hydroxyl group, which triggers a decomposition reaction and the production of light. Here we investigated the effects of reaction conditions on the ability of Supernova to generate a chemiluminescent signal (using a plate reader assay) and to phosphorylate itself (using a ligation assay). Our experiments indicate that multiple zinc ions are required for catalytic function, suggesting links between Supernova and protein enzymes that catalyze similar reactions. They also show how factors such as pH, potassium concentration, CDP-Star concentration, and DNA concentration affect the reaction. By combining information from different experiments, the rate enhancement of light production was increased by more than 1000-fold. These results should be useful for applications in which Supernova is used as a sensor.

Introduction

Chemiluminescence is light generated by chemical reactions. It occurs in phylogenetically diverse species including fireflies, squid, and worms, and can be observed in a variety of habitats, including forests, oceans, and caves.[1] In nature, chemiluminescence is typically generated by luciferase enzymes which catalyze the breakdown of substrates called luciferins.[2-4] Although a wide range of chemically distinct luciferins have been identified, most form an unstable, chemically activated 1,2-dioxetane intermediate during the reaction which produces light when it decomposes.[4] Triggered versions of 1,2-dioxetane substrate have also been developed.[5-9] These consist of a stabilized 1,2-dioxetane linked to a reactive moiety which is in turn blocked by a protecting group. Inspired by the many applications of light-producing reactions, we recently developed a deoxyribozyme that generates light using the triggered 1,2-dioxetane substrate CDP-Star (Figure 1).[10] Our deoxyribozyme, called Supernova, transfers the phosphate protecting group from CDP-Star to its 5' hydroxyl group, which triggers a decomposition reaction and the production of blue light. We also developed a programmable version of Supernova that only produce light in the presence of specific oligonucleotide sequences.[10] This shows that Supernova can be used to detect ligands in solution and raises the possibility that similar sensors could be developed for other ligands. To maximize the utility of such sensors, it is important to understand how changes in reaction conditions affect the ability of Supernova to produce light. Here we have systematically explored this question by testing the effects of a variety of reaction components and conditions (including metal ions, substrate concentration, Supernova concentration, pH, organic solvents, molecular crowding agents, urea concentration, and temperature) on both the deoxyribozyme-catalyzed and background rates of light production. Our experiments indicate that Supernova requires multiple zinc ions for activity, and suggest a link between Supernova and protein enzymes that catalyze similar reactions. They also revealed important information about reaction conditions, which made it possible to increase the rate enhancement of light production by >1000-fold.

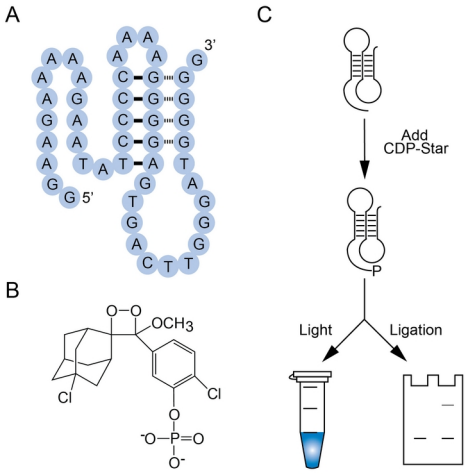


Figure 1. Production of light by the Supernova deoxyribozyme. (A) Secondary structure of Supernova. (B) Chemical structure of CDP-Star, the chemiluminescent substrate used by Supernova. (C) Workflow of the two Supernova assays used in this study. The assay shown on the left can be used to measure the rate of light production (called k in this study), while that on the right can be used to measure the rate of self-phosphorylation (called k_1 in this study).

Results and Discussion

Metal ions play important roles in nucleic acid structure and function. These include shielding of negative charge during folding and directly promoting catalysis in some ribozyme and deoxyribozyme active sites.[11-13] To better understand the roles played by metal ions in the Supernova reaction, we first investigated whether the four metal ions present in the original selection buffer (potassium, cerium, lead, and zinc) were required for activity. Removal of cerium and lead either individually or together had only a small effect on activity (Figures 2A, S2A, and S3A). This was somewhat surprising, since both of these metal ions strongly promote the nonenzymatic dephosphorylation of CDP-Star at high concentrations.[10] Removal of potassium modestly decreased the rate enhancement of light production of Supernova, while removal of zinc resulted in a complete loss of activity (Figures 2A, S2A, and S3A). A minimal buffer containing potassium and zinc was therefore used as the starting point for subsequent experiments. We next performed titrations in which

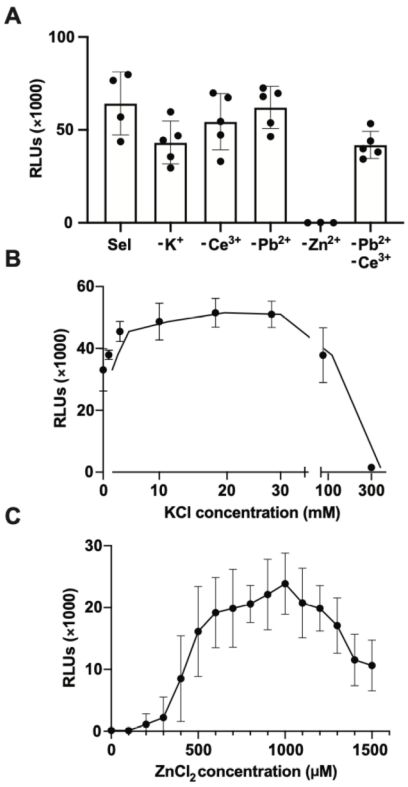


Figure 2. Importance of potassium and zinc in the Supernova reaction. (A) Activity of Supernova in a series of buffers in which a single metal ion (potassium, cerium, lead, or zinc) was omitted from the original selection buffer (labeled "Sel"), which contained 200 mM KCl, 1 μM $CeCl_3$, 0.1 μM $PbCl_2$, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. (B) Potassium titration at a constant zinc concentration. Buffers contained varying concentrations of KCl, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. (C) Zinc titration at a constant potassium concentration. Buffers contained 20 mM KCl, varying concentrations of $ZnCl_2$, and 50 mM HEPES, pH 7.4. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. Points show the average of at least three experiments, and indicate the total amount of light produced in one hour. Error bars represent one standard deviation.

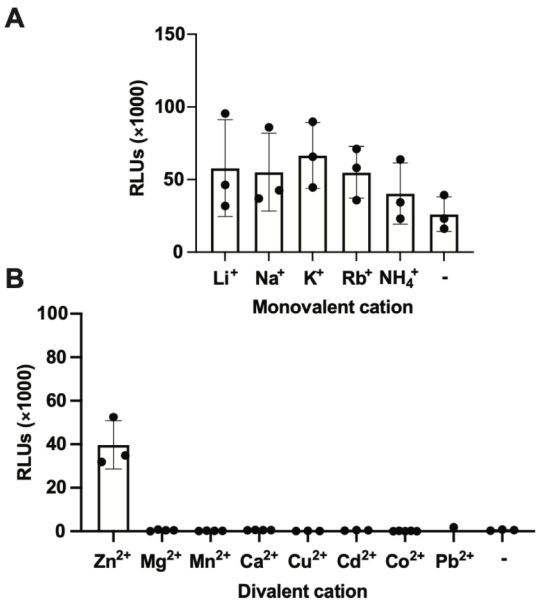


Figure 3. Metal ion requirements of Supernova. (A) Replacement of potassium with other monovalent metal ions. Buffers contained 20 mM of the indicated monovalent cation, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. (B) Replacement of zinc with other divalent metal ions. Buffers contained 20 mM KCl, 1 mM of the indicated divalent cation, and 50 mM HEPES, pH 7.4. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. Points show the average of at least three experiments, and indicate the total amount of light produced in one hour. Error bars represent one standard deviation.

the concentration of potassium was varied at a constant zinc concentration of 1 mM (Figures 2B, S2B, and S3B). The curve was bell shaped with a broad plateau from about 0.3 mM to 100 mM and an optimal potassium concentration of ~20 mM. At this concentration, the rate enhancement of light production was approximately 10-fold higher than that at the potassium concentration used in the selection (200 mM). We suggest that low concentrations of potassium stabilize the fold of Supernova by shielding negative charge, while high concentrations destabilize the purine-motif triple helix in the catalytic core of the deoxyribozyme.[14] A second titration was performed in which the concentration of zinc was varied at a constant potassium concentration of 20 mM (Figures 2C, S2C, and S3C). This curve was also bell-shaped, but with a much narrower peak at ~1000 μM zinc. The steep slope at low zinc concentrations (especially between ~100 μM and ~500 μM) suggests that Supernova requires multiple zinc ions for catalytic function (Figure S4). We note that the solubility of zinc is in the low mM range under these conditions.[15] For this reason, the reduction of Supernova activity at concentrations above 1 mM could be due to precipitation. To further probe the metal ion requirements of Supernova, we investigated the extent to which potassium could be replaced by other monovalent metal ions and zinc could be replaced by other divalent metal ions. Each of the monovalents tested stimulated catalytic activity about as well as potassium (Figures 3A, S5A, and S6A). On the other hand, zinc could not be replaced by any of the other divalents tested (Figures 3B, S5B, and S6B). This included cobalt, which can sometimes substitute for zinc in the

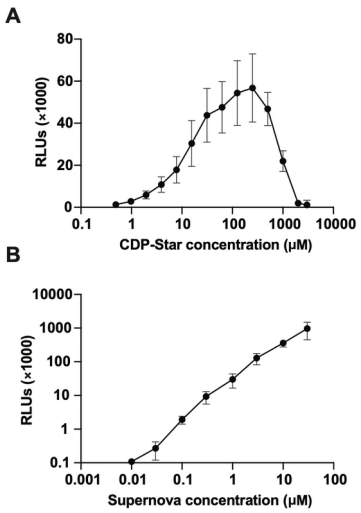


Figure 4. Effect of CDP-Star and Supernova concentration on light production. (A) CDP-Star titration. (B) Supernova titration. Experiments were performed in a buffer containing 20 mM KCl, 0.65 mM $ZnCl_2$, and 20 mM HEPES, pH 7.4. Experiments in panel A contained 1 μM Supernova and varying concentrations of CDP-Star. Experiments in panel B contained varying concentrations of Supernova and 62.5 μM CDP-Star. Points show the average of at least three experiments, and indicate the total amount of light produced in one hour. Error bars represent one standard deviation.

active sites of protein enzymes.[16] These experiments show that the amountof light produced by Supernova can be increased by reducing the concentration of potassium in the buffer from 200 mM (the concentration present in the selection) to ~20 mM. They also highlight the importance of zinc in the reaction.

Enzymes typically contain binding sites that can be saturated by substrate. In the simplest case, rates vary linearly with substrate concentration at concentrations below the K_M before reaching a plateau when the binding site becomes saturated. In contrast, such saturation was not expected to occur for the nonenzymatic reaction over the range of concentrations used in these experiments. This suggested that maximum rate enhancements of light production would be observed at concentrations below the K_M of Supernova. To determine whether this was in fact the case, we measured both the amount of light produced by Supernova and the rate enhancement of light production as a function of the concentration of chemiluminescent substrate (Figures 4A, S7A, and S8A). Three different phases of the reaction were observed. At the lowest concentrations tested, rates of light production increased linearly with substrate concentration for the enzymatic reaction but remained constant for the nonenzymatic reaction. This is probably because, at low substrate concentrations, the rate of the nonenzymatic reaction is lower than the detection limit of the plate reader (determined by the amount of light produced in the absence of CDP-Star). At higher concentrations, rates increased linearly (and by the same amount) with substrate concentration for both the enzymatic and nonenzymatic reactions, meaning that rate enhancements were constant in this phase of the reaction. Eventually, the enzymatic reaction reached a plateau with a K_M of ~100 μM and then began to decrease due to inhibition by an unknown mechanism while the nonenzymatic reaction continued to increase. As a result, rate enhancements decreased as the substrate concentration increased in this phase of the reaction. These results indicate that the rate enhancement of light production is significantly higher at 100 μM CDP-Star than at the concentration used in the original selection (1 mM).

In principle, a different way to increase light production would be to increase the concentration of Supernova in the reaction. An advantage of this approach is that, unlike the case when the substrate concentration is increased, this does not affect the background reaction. On the other hand, high concentrations can sometimes inhibit the folding of functional nucleic acid motifs. This is typically attributed to intermolecular base pairing that competes with formation of the (usually) monomeric functional structure. To determine the effect of Supernova concentration on the rate enhancement of light production, we performed titration experiments in which the concentration of Supernova was varied at a constant concentration of substrate. Light production increased linearly between 10 nM and 30 μM (Fig 4B), while the rate enhancement of light production increased linearly between 100 nM and 30 μM (Fig S7B). Robust catalytic activity at such a high DNA concentration was surprising,

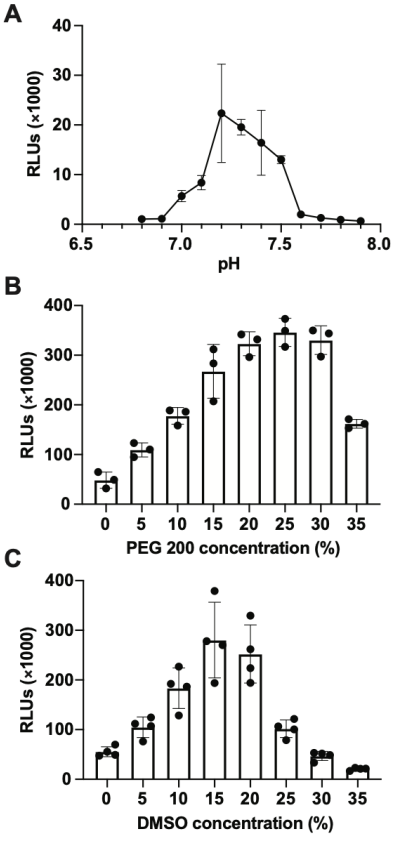


Figure 5. Effects of pH, molecular crowding agents, and organic solvents on the catalytic activity of Supernova. (A) Effect of pH on the Supernova reaction. Buffers contained 20 mM KCl, 0.65 mM $ZnCl_2$, and 20 mM HEPES at various pH values. (B) Effect of PEG 200 concentration on the Supernova reaction. Buffers contained 20 mM KCl, 0.65 mM $ZnCl_2$, 20 mM HEPES, pH 7.4, and varying concentrations of PEG 200. (C) Effect of DMSO concentration on the Supernova reaction. Buffers contained 20 mM KCl, 0.65 mM $ZnCl_2$, 20 mM HEPES, pH 7.4, and varying concentrations of DMSO. All reactions were performed in the presence of 1 μM Supernova and 62.5 μM CDP-Star. Points show the average of at least three experiments, and indicate the total amount of light produced in one hour. Error bars represent one standard deviation.

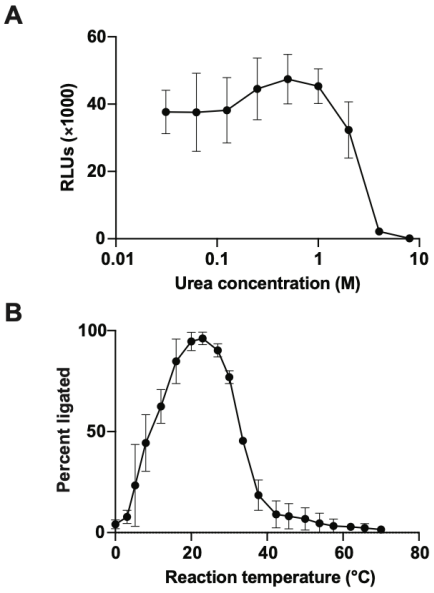


Figure 6. Effects of urea and temperature on the Supernova reaction. (A) Effect of urea on the Supernova reaction. Points indicate the total amount of light produced in a one-hour reaction. (B) Effect of temperature on the Supernova reaction. Note that, because we could not vary the temperature of the reaction using our plate reader, the reactions shown in panel B were analyzed using the ligation assay after a one-hour incubation with CDP-Star. Results are expressed relative to the amount of ligation product obtained when using an oligonucleotide containing a 5' phosphate group (normalized to a value of 100%). Buffers contained 20 mM KCl, 0.65 mM $ZnCl_2$, varying concentrations of urea (panel A), and 20 mM HEPES, pH 7.4, and were performed at either 27 °C (panel A) or varying temperatures (panel B). All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. Points show the average of at least three experiments, and error bars represent one standard deviation.

and might be related to the unusual nucleotide composition of Supernova (26 of the 32 unpaired positions in the construct used in these experiments are purines, which cannot form canonical base pairs with one another). Taken together, these experiments indicate that light production by Supernova can be maximized by using surprisingly high concentrations of DNA (30 μM).

To further investigate the buffer requirements of the reaction, we investigated how pH, organic solvents, and molecular crowding agents affect light production. pH can affect both nonenzymatic and enzyme-catalyzed light-producing reactions of 1,2-dioxetane substrates. [5,8] The rate of light production of Supernova was also pH dependent, and activity was only observed over a narrow range with a peak at approximately pH 7.2 (Figures 5A, S9A, and S10A). The narrowness of this peak is similar to one previously observed for a zinc-dependent deoxyribozyme which cleaves DNA.[17] Interpretation of the pH dependence is not straightforward. The plateau could reflect ionization of a functional group important for folding and/or catalysis with a pK_a value of ~7 (possibly a functional group in Supernova with a shifted pK_a). However, the decrease in activity at high pH values is likely due to a decrease in the solubility of zinc under these conditions.[15] Although pH can also affect the rate of the light-producing reaction after dephosphorylation of CDP-Star,[9] this does not explain the pH dependence, as similar results were obtained for reactions analyzed using the ligation assay (Figure S10A).

Organic solvents and molecular crowding agents have been reported to modestly increase the catalytic activities of a number of deoxyribozymes.[18] In addition, the quantum yield of CDP-Star is more than 10^4-fold higher in some organic solvents than in aqueous buffers.[19-21] These considerations prompted us to investigate the effects of PEG and DMSO on light production. PEG 200 increased both the amount of light produced (Figure 5B) and the rate enhancement of light production (Figure S9B), but had only small effects on reactions analyzed using the ligation assay (Figure S10B). Similarly, DMSO enhanced the rate enhancement of light production up to a concentration of ~15%, but slightly inhibited the reaction at concentrations above 30-35% (Figures 5C, S9C, and S10C). These experiments indicate that light production is modestly stimulated by organic solvents and molecular crowding agents.

For some potential applications it could be desirable to use Supernova at elevated temperatures or in the presence of denaturants. In addition, variants of nucleic acid enzymes with enhanced stabilities sometimes yield higher quality X-ray diffraction data in crystallography experiments than their less-stable counterparts.[22-24] For these reasons we investigated the stability of Supernova in the presence of the denaturant urea and over a range of temperatures. Urea had only small effects at concentrations below 1 M, but almost completely inhibited the reaction at 4 M (Figures 6A, S11, and S12). Catalytic activity was also dependent on temperature: the extent of self-phosphorylation increased from 0°C and 20°C, reached a plateau, and decreased at temperatures above 25°C (Figure 6B; note that, because we could not vary the temperature of the reaction using our plate reader, the temperature dependence of the reaction was only investigated using the ligation assay). These experiments indicate that Supernova is maximally active between ~10°C and ~35°C and at urea concentrations of less than 2 M. Reduced activity at higher urea concentrations and temperatures likely reflects denaturation under these conditions.

When taken together, our experiments revealed three parameters that can be changed relative to the conditions of the original selection to significantly improve the rate enhancement of light production: the potassium concentration in the buffer (~20 mM KCl is better than the 200 mM concentration used in the selection) (Figure S2B), the substrate concentration (~62.5 μM is better than the 1 mM used in the selection) (Figure S7A), and the concentration of Supernova (30 μM is better than the 1 μM concentration used in the selection) (Figure S7B). To determine whether combining these conditions could further improve light production, reactions were performed using each of the eight possible combinations of optimal values for these three parameters (Figure 7). Reactions also contained 15% DMSO, which was found to modestly enhance the reaction (Figure 5C). Using the best conditions identified (20 mM KCl, 62.5 μM CDP-Star, 30 μM DNA, and 15% DMSO), Supernova produced light more than 5,000-fold more efficiently than the background reaction. This is approximately 500-fold higher than the rate enhancement of the peroxidase reaction promoted by G-quadruplexes and hemin in the presence of luminol,[25-26] which is the only other method to generate light using DNA or RNA. It is also comparable to signal-to-noise ratios of less-stable RNA

aptamers such as Red Broccoli[27] and Mango,[28] which enhance the fluorescence of small molecule fluorophores.

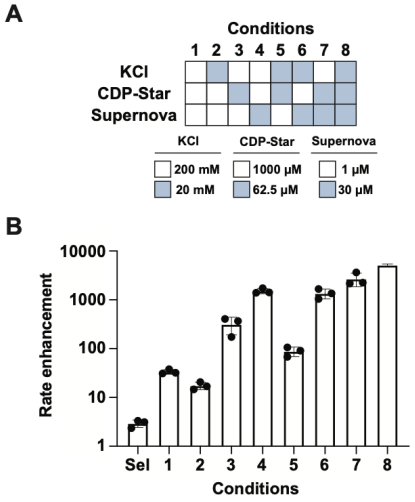


Figure 7. Optimizing the rate enhancement of light production of Supernova. (A) Optimization matrix showing the three parameters (KCl concentration, CDP-Star concentration, and Supernova concentration) varied in this experiment. (B) Rate enhancement of light production (amount of light produced in the catalyzed reaction divided by that produced in the uncatalyzed reaction) for each of the eight possible combinations of values of these parameters. "Sel" indicates the conditions of the original selection (200 mM KCl, 1000 µM CDP-Star, 1 µM Supernova, 1 µM $CeCl_3$, 0.1 µM $PbCl_2$, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4). All other reactions contained either 20 mM or 200 mM KCl, 62.5 µM or 1000 µM CDP-Star, 1 µM or 30 µM Supernova, 0.65 mM $ZnCl_2$, 20 mM HEPES, pH 7.4, and 15% DMSO. Points show the average of at least three experiments, and error bars represent one standard deviation.

Discussion

We recently developed a chemiluminescent deoxyribozyme called Supernova, and showed that it can be programmed to produce light only in the presence of a target oligonucleotide.[10] We are also working to create sensors that detect other ligands using Supernova as a starting point. The typical way to increase the amount of light produced by such a sensor would be to perform a standard selection experiment using a library of ~10^{15} sequences. Another approach could be to use a smaller, secondary structure library[29] and perform a single-step selection.[30] In this study we increased the rate enhancement of light production of Supernova by more than 1000-fold without changing its sequence. This was instead achieved by investigating the effects of a range of conditions on both the enzymatic and nonenzymatic production of light in the presence of CDP-Star. Three parameters were identified that significantly affect the rate enhancement of light production relative to the conditions used in the original selection: the potassium concentration, the substrate concentration, and the DNA concentration. Light production also depends on pH and the concentration of zinc in the buffer. Under optimal conditions, rate enhancements of light production exceeded 5,000-fold. From the perspective of rate enhancement alone, the best conditions we identified were 30 µM DNA, 62.5 µM CDP-Star, 20 mM KCl, 15% DMSO, 650 µM $ZnCl_2$, and a pH of 7.4 (column 8 of Figure 7B). The cost of a 25 µl reaction using these conditions is approximately 0.12 Euros, with the most expensive component being the HPLC-purified DNA. However, useful rate enhancements could also be achieved at lower costs. For example, when 1 µM rather than 30 µM Supernova was used (column 3 of Figure 7B), rate enhancements of ~270-fold were obtained at a cost of approximately 0.077 euros per reaction. For some applications, such conditions could be preferable.

In addition to providing information about how to maximize light production, our experiments also provide clues about the possible mechanism of the reaction. For example, CDP-Star titration experiments generate curves that can be fit using the Michaelis-Menten equation.[10] This indicates that Supernova contains a binding site for CDP-Star, and suggests that proximity between the nucleophilic 5' hydroxyl group of Supernova and CDP-Star could be one mechanism by which Supernova promotes self-phosphorylation (Figure 8). In addition, metal ion substitution and titration experiments highlight the importance of zinc in the reaction. The apparent requirement of Supernova for at least two zinc ions is intriguing when considered from the perspective of proteins that promote similar reactions.[31] For example, some phosphatase enzymes (such as alkaline phosphatase) contain two zinc ions in the active site. These increase the reactivity of the serine nucleophile that forms a covalent intermediate with the phosphate in the substrate, and also stabilize charge in the pentavalent transition state.[32-33] The requirement for zinc in the reaction catalyzed by Supernova and the sharp dependence of catalytic activity on zinc concentration raises the possibility that the active sites of Supernova and alkaline phosphatase could be similar (Figure 8A). However, our data do not exclude the possibility that zinc plays only a structural role (Figure 8B). Regardless of the role of zinc, our data suggest that Supernova contains multiple binding sites that facilitates catalytic activity when occupied by zinc but not when occupied by other metal ions. Since the nitrogen at position 7 in guanine is a preferred binding site for zinc in DNA,[34] we suggest that these functional groups in Supernova are strong candidates for the ligands in these binding sites. Indeed, such a model helps to rationalize the presence of 13 guanines among the 24 conserved unpaired positions in the secondary structure of Supernova. A high-resolution structure would likely reveal the role of zinc in the reaction as well as the location and nature of the binding sites. It would also provide a wealth of additional information about the structure and catalytic mechanism of this glowing deoxyribozyme.

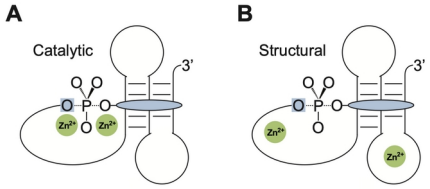


Figure 8. Possible catalytic strategies of Supernova. (A) In this model, zinc ions play catalytic roles in the reaction (perhaps using a mechanism similar to that of alkaline phosphatase). (B) In this model, zinc ions play structural rather than catalytic roles. In both models, Supernova contains a binding site for CDP-Star (represented as a blue oval) which increases its local concentration in the vicinity of the nucleophilic 5' hydroxyl group (shown in a blue square). Note that the locations of the binding sites for CDP-Star and zinc shown in the figure are hypothetical and have not yet been verified experimentally. Note also that our experiments are consistent with multiple zinc binding sites, but not necessarily exactly two as shown in the figure.

Experimental Section

Oligonucleotides: Oligonucleotides were chemically synthesized by Sigma-Aldrich, and purified using PAGE or HPLC. For the sequences of all oligonucleotides used in this study see Table S1.

Analysis of light production: Supernova was obtained from Sigma-Aldrich and purified by 6% Urea-PAGE or HPLC. Supernova was diluted in Milli-Q water to the desired final concentration, heated at 65 °C for 2 minutes, and cooled at room temperature for 5 minutes. Afterwards, 20 µl of 5× reaction buffer was added and samples were transferred to a white half-area 96-well plate (Corning). A range of buffers were used that differed with respect to metal ion composition, metal ion concentration, pH, PEG concentration, DMSO concentration, and urea concentration. CDP-Star was added to bring the final volume of the reaction to 100 µl and chemiluminescence was immediately measured for 1 hour at 27 °C (10 s orbital shaking, luminescence attenuation with 1000 ms integration) using a Tecan Spark plate reader (Tecan Group). Final concentrations were 1 µM Supernova, 1× reaction buffer, and 100 µM CDP-Star unless stated otherwise. The amount of light produced in each sample was measured along with a negative control (which contained all reaction components except Supernova) and a buffer control (which contained all reaction components except Supernova and CDP-Star). Total light production at each time point was defined as the cumulative luminescence counts (relative light units) of the sample minus the cumulative luminescence of the buffer control. Some experiments were also analyzed using data in which the amount of light produced in the buffer control was not subtracted from the amount of light produced in the Supernova-catalyzed reaction. This had a negligible effect on enzymatic rates, while nonenzymatic rates were up to 4-fold higher than those in which background was subtracted. Rate enhancements were determined by dividing the amount of light produced in the presence of Supernova by that produced in the absence of Supernova.

Analysis of phosphorylation: Standard Supernova reactions were prepared as follows: 2 µg of Supernova was diluted in Milli-Q water, heated at 65 °C for 2 minutes, and cooled at room temperature for 5 minutes. Afterwards, 5× reaction buffer and CDP-Star were added as described above to give final concentrations of 1 µM Supernova, 1× buffer, and 100 µM CDP-Star unless stated otherwise. The reaction was incubated for 1 hour at room temperature unless stated otherwise and stopped by the addition of EDTA to a final concentration of 25 mM. The oligonucleotide was then purified using SigmaSpin Sequencing Reaction Clean-Up Columns (Sigma-Aldrich) and concentrated by ethanol precipitation. Phosphorylated Supernova molecules were then ligated to a short oligonucleotide (FWD) using a splint (Splint, Table S1) and T4 DNA ligase (Jena Bioscience). Reactions were incubated at 37 °C for 30 minutes. Final conditions were 2.5 µM (1 µg) Supernova 2.5 µM FWD, 2.5 µM splint, 1× T4 DNA ligase buffer, and 0.5 Weiss Unit (0.2 µl)/µg DNA of T4 DNA ligase. Phosphorylated (i.e. ligated) and non-phosphorylated molecules were then separated by 6% Urea-PAGE. Gels were stained with 1× GelRed Nucleic Acid Gel Stain (Biotium) for 20 minutes in the dark. Ligation yields were determined using the ImageQuant TL (GE Healthcare LifeSciences) densitometry tool, and normalized to a 5' phosphorylated control (defined as 100% ligation).

Curve fitting: Supernova transfers the phosphate group from the chemiluminescent substrate CDP-Star to its 5' hydroxyl group with a rate of k_1. In the second step of the reaction, the dephosphorylated substrate spontaneously decomposes with a rate of k_2 to generate light by chemically initiated electron exchange luminescence (CIEEL). Although this is a two-step reaction, we found that most enzymatic time courses could be fit using equation 1:

$$(1)\ L = F_{max} \times (1\text{-}e^{-kt})$$

in which L is the cumulative amount of light produced, F_{max} is the maximum cumulative amount of light produced, k is the rate constant of light production (note that this is different than k_1 and k_2 as defined above), and t is time. An example of a time course of light production fit using this equation using Prism 9 software (GraphPad) is shown in Figure S1A. Although this approach could be used to follow some reactions to completion, it was not suitable for slow enzymatic reactions or nonenzymatic reactions. Furthermore, nonenzymatic reactions were too slow to detect when only early time points were analyzed. For this reason, reactions were analyzed after incubating with CDP-Star for one hour. This approach slightly underestimates the amount of light generated in the fastest reactions (with rates of light production of ~0.03 min^{-1}) relative to comparisons that use initial rates because fast reactions are beginning to plateau after one hour. However, this approach makes it possible to rapidly compare fast and slow reactions (which is important for applications in which Supernova is used as a sensor). Some reactions were also analyzed using a method in which background (the amount of light produced in the buffer control) was not subtracted, and initial rates calculated from the slope of the reaction in the first five minutes (which was linear in all cases). This yielded similar results to the approach used here. Time courses of self-phosphorylation analyzed using the ligation assay were fit using equation 2:

$$(2)\ P = F_{max} \times (1\text{-}e^{-k1t})$$

in which P is the percent of ligated deoxyribozyme, F_{max} is the maximum percent of ligated deoxyribozyme, k_1 is the rate of self-phosphorylation, and t is time. An example of a time course of self-phosphorylation fit using this equation is shown in Figure S1B. In most cases, results were similar for experiments analyzed using assays for light production and self-phosphorylation (Figure S1C). However, rates of the fastest reactions were up to 15-fold higher when the ligation assay was used, likely because k_1 (the rate of self-phosphorylation) is faster than k_2 (the rate of decomposition of CDP-Star after dephosphorylation) under these conditions.

Acknowledgements

This work was supported by a GAČR grant (19-20989S) awarded to E.A.C. and GAUK grants awarded to M.J. (290321) and K.S (102216).

Keywords: Deoxyribozyme • in vitro selection • Supernova • chemiluminescence • sensor

[1] J. W. Hastings, *J. Mol. Evol.* **1983**, *19*, 309–321.
[2] J. R. De Wet, K. V. Woodt, D. R. Helinski, M. Delucat, *Biochemistry* **1985**, *82*, 7870–7873.
[3] E. Conti, N. P. Franks, P. Brick, *Structure* **1996**, *4*, 287–298.
[4] Z. M. Kaskova, A. S. Tsarkova, I. V. Yampolsky, *Chem. Soc. Rev.* **2016**, *45*, 6048–6077.
[5] I. Bronstein, B. Edwards, J. C. Voyta, *J. Biolumin. Chemilumin.* **1989**, *4*, 99–111.
[6] P. Trayhurn, M. E. A. Thomas, J. S. Duncan, D. Black, J. H. Beattie, D. V. Rayner, *Biochem Soc. Trans*. **1995**, *23*, 494S.
[7] O. Green, T. Eilon, N. Hananya, S. Gutkin, C. R. Bauer, D. Shabat, *ACS Cent. Sci.* **2017**, *3*, 349–358.
[8] N. Hananya, D. A Shabat, *Angew. Chemie Int. Ed.* **2017**, *56*, 16454–16463.
[9] N. Hananya, J. P. Reid, O. Green, M. S. Sigman, D. Shabat, *Chem. Sci.* **2019**, *10*, 1380–1385.
[10] K. Svehlova, O. Lukšan, M. Jakubec, E. A. Curtis, *Angew. Chemie Int. Ed.* **2022**, *61*, e202109347.
[11] M. Yarus, *FASEB J.* **1993**, *7*, 31–39.
[12] R. K. O. Sigel, A. M. Pyle, *Chem. Rev.* **2007**, *107*, 97–113.
[13] W. Zhou, R. Saran, J. Liu, *Chem. Rev.* **2017**, *117*, 8272–8325.
[14] A. J. Cheng, M. W. van Dyke, *Nucleic Acids Res.* **1993**, *21*, 5630.

[15] W. J. Moon, Y. Yang, J. Liu, *ChemBioChem.* **2021**, *22*, 779-789.
[16] Maret, B. L. Vallee, *Methods Enzymol.* **1993**, *226* (C), 52–71.
[17] Y. Xiao, M. Chandra, S. K. Silverman, *Biochemistry* **2010**, *49*, 9630.
[18] S. I. Nakano, N. Sugimoto, *Biophys. Rev.* **2016**, *8*, 11–23.
[19] A. P. Schaap, M. D. Sandison, R. S. Handley, *Tetrahedron Lett.* **1987**, *28*, 1159–1162.
[20] I. Bronstein, J. C. Voyta, O. J. Murphy, R. Tizard, C. W. Ehrenfels, R. L. Cate, *Methods Enzymol.* **1993**, *217* (C), 398–414.
[21] G. B. Schuster, B. Dixon, J. Y. Koo, S. P. Schmidt, J. P. Smith, *Photochem. Photobiol.* **1979**, *30*, 17–26.
[22] F. Guo, T. R. Cech, *Nat. Struct. Biol. 2002 911* **2002**, *9*, 855–861.
[23] F. Guo, A. R. Gooding, T. R. Cech, *Mol. Cell* **2004**, *16*, 351–362.
[24] F. Guo, A. R. Gooding, T. R. Cech, *RNA* **2006**, *12*, 387–395.
[25] Y. Xiao, V. Pavlov, R. Gill, T. Bourenko, I. Willner, *ChemBioChem* **2004**, *5*, 374–379.
[26] J. Kosman, B. Juskowiak, *Anal. Chim. Acta*. **2011**, *707*, 7-17.
[27] X. Li, L. Mo, J. L. Litke, S. K. Dey, S. R. Suter, S. R. Jaffrey, *J. Am. Chem. Soc.* **2020**, *142*, 14117–14124.
[28] A. Autour, S. C. Y. Jeng, A. D. Cawte, A. Abdolahzadeh, A. Galli, S. S. S. Panchapakesan, D. Rueda, M. Ryckelynck, P. J. Unrau, *Nat. Commun.* **2018**, *9*, 1–12.
[29] R. Sgallová, E. A. Curtis, *Molecules* **2021**, *26*, 1671.
[30] T. Streckerová, J. Kurfürst, E. A. Curtis, *Nucleic Acids Res.* **2021**, *49*, 6971.
[31] K. A. McCall, C. C. Huang, C. A. Fierke, *J. Nutr.* **2000**, *130,* 1437S-1446S.
[32] K. M. Holtz, B. Stec, E. R. Kantrowitz, *J. Biol. Chem.* **1999**, *274*, 8351–8354.
[33] K. M. Holtz, E. R. Kantrowitz, *FEBS Lett*. **1999**, *462*, 7-11.
[34] N. A. Frøystein, J. T. Davis, B. R. Reid, E. Sletten, *Acta Chem. Scand.* **1993**, *47*, 649–657.

5 Discussion and future directions

"Innovation is not about 'why', it is about 'why not?'"
- Darius Tanz

Since the discovery of catalytic RNA molecules in the 1980s[6,7], many artificial ribozymes and deoxyribozymes have been isolated. Many showcased the catalytic range of both RNA and DNA molecules, while others aimed to facilitate basic research. Examples of the latter include an RNA that can integrate itself into another RNA molecule[136], DNA that fluorescently labels other RNA molecules[137], or deoxyribozymes that produce a fluorogenic signal in the presence of cognate pathogen[106]. Functional nucleic acid motifs that link an easily detectable signal to a biologically interesting ligand are especially desirable. For many applications, light is a useful output for several reasons: it allows for a higher dynamic range than color or fluorescence due to low background, it does not pose health risks like radioactivity-based assays, and it is easily detectable using common lab instruments like a plate reader.

Currently, the most efficient chemiluminescent DNA molecule is a peroxidase-mimicking G-quadruplex that oxidizes luminol in the presence of hemin and peroxide[122]. However, the rate enhancement of this system is only ~10-fold, which is insufficient for most sensing applications[122,138]. This is in part due to a relatively high background light production in the absence of a deoxyribozyme.

For this reason, we decided to isolate a new nucleic acid sensor that produces a chemiluminescent signal with a higher rate enhancement. As a substrate, we selected commercially available CDP-Star, a 1,2-dioxetane stabilized by a phosphate group. Unlike functionally similar luminol, CDP-Star produces a very low background signal which is essential for a high-rate enhancement system. Because DNA has been reported to catalyze a wide range of chemical reactions[139] and is more chemically stable and cheaper to synthetize than RNA, we decided to create a deoxyribozyme-based system. Moreover, several deoxyribozymes that catalyze phosphoryl transfer from other small molecules to their 5' end have been described previously[53,140,141]. Thus, the identification of a DNA kinase motif that uses CDP-Star as a phosphate donor seemed plausible.

Indeed, in optimal reaction conditions, the rate enhancement of Supernova deoxyribozyme exceeds the current DNA-based system with luminol as a substrate by 1,000-fold (Figure 27d). Although its rate enhancement is similar to other RNA-based signaling systems, it complements rather than competes with them. While RNA motifs are best employed as cellular and *in vivo* sensors for metabolites because they can be expressed directly in the cells, DNA is better suited for *in vitro* applications like diagnostics due to its chemical stability and cost of synthesis.

5.1 Secondary structure of Supernova

Unlike most articles describing a new functional nucleic acid motif, we used the methods of comparative sequence analysis and mutual information analysis to determine Supernova's secondary structure (Figure 21). As previously mentioned, although more laborious than the computational algorithms, it is considerably more reliable, especially when coupled with mutational analysis to support the predicted structures. In fact, the structure-prediction algorithm mfold[69] returned a structure that did not share a single base pair with the one predicted by our comparative sequence analysis (Figure 31).

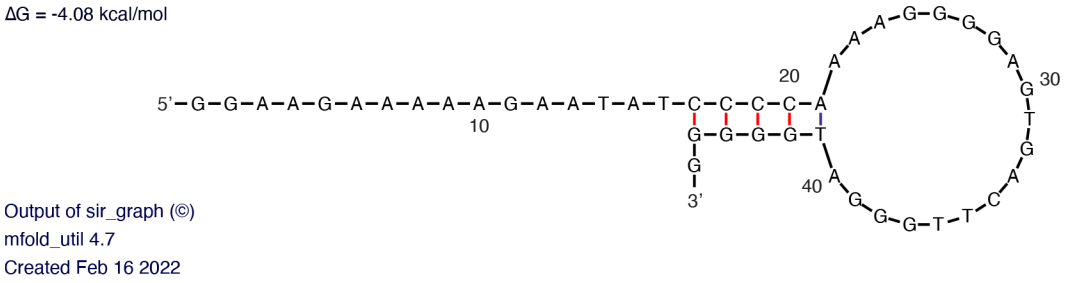


Figure 31. Secondary structure of Supernova predicted by the mfold algorithm[69]. Note that the structure shares no base pairs with our predicted secondary structure of Supernova (Figure 21).

This discrepancy can stem from the fact that minimum-energy structure prediction algorithms take into account neither the folding pathways nor the tertiary-structure interactions, both of which can alter the resulting conformation. In addition, most of them do not include non-canonical structures like triple helices or G-quadruplexes in their algorithms.

Importantly, comparative sequence analysis can only be used if the number of sequences with the same secondary structure is sufficiently high in the NGS dataset. This is the case for reselection experiments in which all catalytic molecules are based on a single parental sequence and presumably share the same structure. It can also be used when high-throughput sequencing is used to analyze sequences selected from a small starting pool[142].

In conclusion, comparative sequence analysis appears to be a superior way to predict a secondary structure of a newly isolated functional nucleic acid.

5.2 Supernova as a sensor: future possibilities

In our proof-of-principle experiment, we used a variant of Supernova to detect specific oligonucleotides. This demonstrated that Supernova can be used as a sensor. There are many previous examples of nucleic acid motifs fused to an aptamer in such a way that they only generate a signal in the presence of the cognate ligand. In particular, the Yingfu Li lab described numerous DNA aptamers for the detection of pathogens such as *Escherichia coli*[106], *Clostridium difficile*[143], *Staphylococcus aureus*[144], and the spike protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[145,146]. Other pathogen-binding aptamers include those specific for *Neisseria meningitidis*, the causative agent of meningitis[147], and *Streptococcus pyogenes*, a bacterium responsible for scarlet fever[148].

Inspired by this, ongoing research in our lab includes developing Supernova-based sensors that employ DNA aptamers as recognition elements. There are several approaches to develop such tools. The simplest strategy is to design the construct by rational design. In this case, the aptamer sequence is inserted into a loop of the signaling module via a short linker, and the construct is then tested for activity in both the presence and absence of a ligand[149]. A different approach is to insert a random sequence region into one of the loops of the deoxyribozyme. The resulting library is then subjected to alternating rounds of selection with and without the target ligand, and sensors that are only active when the ligand is present are isolated. The advantage of this approach is that sensors specific for ligands with no known DNA aptamers can be generated. Finally, a combination of the two can be utilized where a known aptamer sequence is inserted in the signaling module's loop, and a library is generated by partial mutagenesis. This library is then used to isolate efficient sensors as described above. This strategy is advantageous for researchers who want to explore the neighboring sequence space of known aptamers to achieve the highest rate enhancement for ligand detection in the sequence background of a particular signaling (deoxy)ribozyme.

In addition, if the library size is small enough (i.e. up to 10^7 sequences) due to a short random region or partial mutagenesis of a specific sequence, the aptamer motif can be identified using a single round of selection and NGS sequencing[150]. This makes the procedure of new sensor identification significantly less time-consuming.

Figure 32. An overview of a possible high-throughput system that utilizes Supernova-based sensors. In this case, a chip containing a number of sensors specific for a variety of human pathogens is incubated with a sample from the patient. After the addition of CDP-Star, a reader determines which pathogen(s) are present based on which sensor(s) produce light. Based on the speed of the Supernova sensors specific for oligonucleotides, the whole process should not exceed 15 to 30 minutes.

Theoretically, if enough sensors are generated, a high-throughput screening system could be developed. The system would comprise a collection of Supernova-based sensors on a chip that could screen a selection of ligands in a sample in one quick measurement. The ligands could range from human pathogens, cancer biomarkers[151–153], to other interesting substances like *Clostridium botulinum* neurotoxin[154], cocaine[155], or C-reactive protein, an infection marker commonly used in point-of-care diagnostics[156] (Figure 32).

5.3 Optimization of Supernova performance

During the initial *in vitro* selection experiment, we isolated a deoxyribozyme with the light production rate enhancement of 6-fold. After another reselection from a library based on this parent sequence, we were able to increase the rate enhancement to 35-fold. Although the improvement by the reselection seemed rather modest, we were able to increase the rate enhancement up to 6,500-fold above background through the optimization of reaction conditions. Namely, we decreased the substrate concentration and increased the concentration of the deoxyribozyme in the reaction. Interestingly, we were also able to omit the Pb^{2+} and Ce^{4+} ions without activity reduction. Importantly, the presence of Zn^{2+} ions was found to be essential, and the steep slope between concentrations 100 and 500 μM suggests that multiple zinc ions are required for catalysis. This is interesting especially in comparison with some proteinaceous phosphatases that require two zinc ions in their active site[157]. Although it is tempting to draw parallels, it is equally possible that the zinc ions only play a structural role. Our laboratory is currently working on acquiring a high-resolution structure, using both nucleic magnetic resonance and X-ray crystallography,

that would reveal the nature of the zinc binding sites, as well as a wealth of other information about the mechanism and structure of Supernova.

The light production by CDP-Star dephosphorylation is a two-step process (Figure 13). The first reaction is the dephosphorylation by the deoxyribozyme, resulting in an unstable intermediate anion. The second step involves a spontaneous degradation of the intermediate, accompanied by the release of a photon. Importantly, this second step is extremely slow in low pH and accelerates at pH above 9[126]. In principal, we could increase the resultant light production by optimizing either step one, step two, or both.

5.3.1 Faster variants

An obvious way to increase the first step of the reaction and thus the light production is to search for Supernova variants with higher k_{cat}. This effort is currently ongoing, and I have already isolated a variant that is significantly faster than the original Supernova enzyme using a single-step selection protocol. The advantage of this approach is that there is no need to change the current conditions of the reaction.

5.3.2 Light production versus pH

Another way to increase the signal output is to accelerate the second step of the reaction – the light-producing intermediate degradation. This is most easily done by increasing the pH of the reaction to 9 or higher. Unfortunately, Supernova has a rather narrow pH optimum between 7.1 and 7.5, with no detectable activity above the pH of 7.6 (Figure 5a, page 52). However, our preliminary experiments showed a significant light burst if we increase the buffer pH after 10 minutes of a standard Supernova reaction (at pH 7.4), allowing for faster decomposition of the light-producing intermediate. Therefore, one of our current avenues for optimizing light production is to isolate new Supernova variants that catalyze a reaction in higher pH.

5.3.3 Direct selection for light

Due to the mechanism of our deoxyribozyme isolation protocol, the variants are selected based on their ability to transfer a phosphate group from the CDP-Star substrate to their 5' terminus rather than the ability to produce light. This is a considerable drawback of the *in vitro* selection process, where the active molecules have to carry a certain tag to be separable from the inactive ones. As one adage of artificial evolution says, "you get what you select for"[158]. In this case, it means that we might be selecting for the best kinases, while missing molecules that are more efficient in substrate dephosphorylation, which in turn means that they could pro-

duce more light. In fact, in 2005, the Ellington lab used FACS to isolate *trans*-acting ligase RNAs through fluorescently tagged substrate molecules[58]. A similar approach could be used to isolate a better Supernova variant, where the method would rely directly on the light production instead of kinase activity.

Moreover, because each sequence of the pool is enclosed in a discrete compartment, active molecules are easily isolated without the need for a tag to separate them. This way, the selection pressure to attach the phosphate group to their 5' end is lifted, and there is a high probability that some of the isolated deoxyribozyme would be multiple-turnover, further increasing their light production rate enhancement.

6 Conclusion

In our work, we used *in vitro* selection to identify deoxyribozymes that catalyze phosphoryl transfer from a commercially available chemiluminescent substrate CDP-Star to their 5' terminus. We then performed a reselection from a pool based on the sequence that produced light most efficiently. By this second selection, we isolated more efficient variants of the deoxyribozyme and determined their secondary structure using comparative sequence analysis and mutual information analysis. Based on the secondary structure, we identified a minimized version of the most abundant deoxyribozyme from the reselection. This resulted in a variant that was only 46-nucleotide long, about twice the size of a PCR primer. Kinetic characterization revealed a K_M of 130 ± 70 μM and k_{cat} of 0.15 ± 0.03 min^{-1}. Moreover, we used rational design to construct a sensor that detected short oligonucleotides in a concentration-dependent manner with up to a 38-fold rate enhancement over reactions with no detected oligonucleotide.

Furthermore, we tested the efficiency of light production and Supernova phosphorylation in various reaction conditions. Our data show that Supernova requires the presence of zinc, while lead, cerium, and potassium can be omitted without a significant loss of activity. The optimization also revealed that lower CDP-Star (62.5 μM) and higher Supernova (30 μM) concentrations result in a higher rate enhancement, along with the addition of a crowding agent like DMSO or PEG-200. The pH optimum is rather narrow, between 7.1 and 7.5, and consistent with the conditions used during the *in vitro* selection. Surprisingly, our deoxyribozyme showed robust activity at a rather wide range of temperatures, between 10 and 35 °C, and up to a 2 M concentration of urea.

In conclusion, we believe that Supernova is a valuable addition to the current toolbox of nucleic acid signaling molecules and will be useful in many applications, including basic research, medical diagnostics, or environmental monitoring.

7 Materials and Methods

Oligonucleotides

Oligonucleotides were chemically synthesized by IDT or Sigma-Aldrich, and purified using PAGE or HPLC. See Table 1 for the sequences of all oligonucleotides used in this study.

Pool design

The initial library consisted of a random sequence region of 70 nucleotides flanked by primer binding sites of constant sequence. The sequence of the 5' end of the library was chosen based on one used in a previous selection that utilized a ligation step[53]. Once deoxyribozymes were identified, a second library was synthesized based on the most active variant from the initial sequence (called H1; see Figure 19b and Table 1). The 5' primer binding site and the 70 nucleotides derived from the randomized region were mutated at a rate of 21% per position by chemical synthesis, while the 3' primer binding site was changed to a different constant sequence (Table 1). This was done so that if H1 molecules in the lab contaminated the second library, they would not be amplified by PCR. To facilitate robust ligation, a splint with a constant sequence was used in the reselection (Table 1). Although this imposed a selective pressure for library variants to maintain complementarity to the splint, it did not prevent us from obtaining information about the sequence requirements of the 5' end of the deoxyribozyme: 9.3% of the H1 variants obtained in the reselection contained at least one mutation relative to H1 in this part of the sequence. We note that a strategy which makes it possible to mutagenize the 5' end of these deoxyribozymes without requiring complementarity to the splint could be useful, since the reaction site of these deoxyribozymes is the 5' hydroxyl group.

Initial Selection

The single-stranded DNA pool (Pool1) and blocking oligonucleotide (REV1) were heated in water at 65 °C for 2 minutes and cooled at room temperature for 5 minutes. They were then incubated for 24 hours (rounds 1 – 7) or 10 minutes (rounds 8 – 9) with the CDP-Star substrate (Novex) as follows: 1 μM DNA pool, 1.5 μM REV1, 1× selection buffer (50 mM HEPES pH 7.4, 200 mM KCl, 1 mM $ZnCl_2$, 1 μM $Ce(SO_4)_2$, 0.1 μM $PbCl_2$), and 1 mM CDP-Star. After incubation, DNA was precipitated in ethanol, and ligated to a short oligonucleotide (FWD1) using a splint (Splint1) and T4 DNA ligase (Jena Bioscience). The ligation reaction was incubated for 5 minutes to reduce the likelihood of isolating kinases that use ATP rather than CDP-Star as a substrate. All oligonucleotides were at a concentration of 2.5 μM. Molecules were then separated by 6% Urea-PAGE, and those of the length 120 nt were purified.

These were amplified by PCR using Taq Polymerase (Jena Bioscience) and the FWD1r and REV1 primers. The FWD1r primer contained a single RNA linkage at its 3' end. After another ethanol precipitation, double-stranded PCR products were heated at 65 °C for 2 minutes and cooled at room temperature for 5 minutes. 10× base hydrolysis buffer was then added (1× buffer: 20 mM Trizma base, 4 M KOH, 40 mM EDTA), and the sense strands were base hydrolyzed at 90 °C for 10 minutes. The two strands were then separated by 6% Urea-PAGE to regenerate the single-stranded DNA pool. After Round 9, the pool was amplified by PCR and cloned using a TOPO TA Cloning Kit (Invirogen). Plasmids were introduced into competent *Escherichia coli* cells (strain DH5α; Invitrogen) by chemical transformation. Plasmids from 48 selected colonies were then sequenced using Sanger sequencing (Eurofins Genomics).

Reselection

The single-stranded DNA pool (Pool2) was mixed with the blocking oligonucleotide (REV2) in water and heated at 65 °C for 2 minutes. It was then cooled at room temperature for 5 minutes. The mixture was then incubated for 10 minutes (rounds 1 – 3) or 1 minute (rounds 4 - 7) with CDP-Star at the following final concentrations: 1 μM DNA pool, 1.5 μM REV2, 1× selection buffer, and 1 mM CDP-Star. After incubation, DNA was ethanol precipitated and ligated to a short oligonucleotide (FWD2) using a splint (Splint2) and T4 DNA ligase (Jena Bioscience). As before, the ligation reaction was incubated for 5 minutes, and all oligonucleotides were at a concentration of 2.5 μM. Molecules were then separated by 6% Urea-PAGE, and those of the length 125 nt were purified. These were amplified by PCR using Q5 Hot Start Polymerase (NEB) and the FWD2r and REV2p primers. The FWD2r primer contained a single RNA linkage at its 3' end, and the REV2p primer contained a phosphate at its 5' end. After cleaning up the reaction using the NucleoSpin Gel and PCR Clean-up kit (Macharey-Nagel), the reverse strand was digested using λ-exonuclease (NEB). After another purification with the NucleoSpin Gel and PCR Clean-up kit, the remaining sense strands were base-hydrolyzed and purified by 6% Urea-PAGE to regenerate the single-stranded DNA pool.

Analysis of Light Production

Oligonucleotides corresponding to individual sequences from evolved libraries were ordered from Sigma-Aldrich, and purified using 6% Urea-PAGE or HPLC. Light production was assessed as follows: the oligonucleotide being tested was mixed with the blocking oligonucleotide (if necessary) in water, heated at 65 °C for 2 minutes, and cooled at room temperature for 10 minutes. Afterwards, 5× selection buffer was added, and samples were transferred to a white half-area 96-well plate (Corning).

CDP-Star was added, and chemiluminescence was immediately measured for 1 hour using a Tecan Spark plate reader (Tecan Group). Final concentrations were 1 μM of the oligonucleotide being tested, 1.5 μM of the blocking oligonucleotide, 1× selection buffer, and 1 mM CDP-Star unless stated otherwise.

Analysis of Phosphorylation

Oligonucleotides corresponding to individual sequences from evolved libraries were ordered from Sigma-Aldrich and purified by 6% Urea-PAGE. Each tested oligonucleotide was mixed with the appropriate blocking oligonucleotide (if necessary) in water, heated at 65 °C for 2 minutes and cooled at room temperature for 10 minutes. Afterwards, 5× selection buffer and CDP-Star were added. Final concentrations were 1 μM of the oligonucleotide being tested, 1.5 μM blocking oligonucleotide, 1× selection buffer, and 1 mM CDP-Star, unless stated otherwise. The reaction was incubated for a specific time at room temperature in the dark, and then stopped by the addition of EDTA to a final concentration of 25 mM. The reaction was then cleaned up using SigmaSpin Sequencing Reaction Clean-Up columns (Sigma-Aldrich), and ethanol-precipitated. The oligonucleotide being tested was then ligated to a short oligonucleotide as in the selection, with the difference that the incubation time was 30 minutes. Reacted and unreacted molecules were then separated by 6% Urea-PAGE, and the ligation yield was analyzed using the densitometry tool ImageQuant TL (GE Healthcare LifeSciences).

Next generation sequencing and data analysis

Sequencing analysis of the randomly mutagenized pool of deoxyribozyme variants and six subsequent rounds of selection was performed on an Illumina HiSeq instrument (2×150 bp, paired-end) at GATC Biotech (Konstanz, Germany). Raw reads were processed with the cutadapt tool to remove adapter and primer sequences, to perform quality trimming, and to filter low quality reads. Paired-end reads were first oriented, then merged using the program fastq-join, and finally aligned using Clustal Omega. Data quality was evaluated using the FastQC tool. In each library, we calculated the frequencies of unique sequences and generated sequence logos using the DiffLogo Bioconductor package. The secondary structure model was generated by mutual information analysis[30] using in-house scripts. Detailed information about all tools used is provided in Table 2.

Oligonucleotide detection using an engineered version of Supernova

The sensor variant being tested was mixed with the target oligonucleotide in water, heated at 65 °C for 2 minutes, and cooled at room temperature for 10 minutes. Afterwards, 5× optimized buffer (250 mM HEPES pH 7.4, 100 mM KCl, 5 mM $ZnCl_2$)

was added, and samples were transferred to a white half-area 96-well plate (Corning). CDP-Star was added, and chemiluminescence was immediately measured for 1 hour using a Tecan Spark plate reader (Tecan Group). Final concentrations were 1 μM of the sensor, 10 μM of the target oligonucleotide, 1× selection buffer, and 250 μM CDP-Star unless stated otherwise. Light production in the absence of the target oligonucleotide was also measured for each sensor.

Kinetics measurements and analysis

Kinetics were measured using ligation assay as follows: Supernova was mixed with water, incubated at 65 °C for 2 minutes, and cooled at room temperature for 10 minutes. Afterwards, 5× optimized buffer and CDP-Star were added. The reactions were then stopped by adding EDTA to a final concentration of 25 mM at time-points that corresponded to the linear phase of the Supernova reaction. Final concentrations were 1 μM Supernova, 1× optimized buffer, and 1 μM to 500 μM CDP-Star. Samples were then cleaned up, ethanol-precipitated, and ligations were performed as described in the section Analysis of Phosphorylation. The results were analyzed using Prism 9 software (Graphpad Software). For the calculation of k_{cat} and K_m we used the equations $V_0 = V_{max} \cdot [S] / (K_m + [S])$ and $k_{cat} = V_{max} / [E]$.

Calculation of rate enhancement values of other detection systems (Figure 12)

For fluorescent aptamers, rate enhancements were determined by comparing fluorescence quantum yields in the absence and presence of an aptamer, and in other cases, they indicate the signal in the presence of the motif divided by the signal in the absence of the motif using optimized conditions. Rate enhancements are from references [59,116,122–125].

Table 1. Sequences used in experiments described in this work.

Name	**5' → 3' nucleotide sequence**
Supernova	identical to H2 core
Pool1	GGAAGAGATGGCGACN$_{70}$AGCTGATCCTGATGG
REV1	CCATCAGGATCAGCT
FWD1	GAATTCTAATACGACTCACTATA
FWD1r	GAATTCTAATACGACTCACTATrA
Splint1	GTCGCCATCTCTTCCTATAGTGAGTCGTATTAG
H1	GGAAGAGATGGCGACGACACAGGGACGATGCCGAATATCCTCAGTGCGCAGGGCCGCAGGGGGGAGTGACTTGGGATGGGGGGTCAGCTGATCCTGATGG
Clone 2	GGAAGAGATGGCGACAGAGCTGACGGTCGGGATAGCGAAGGCGGTGGTCGAGTGAATCGCCTCCCTTTAGGGAGTGGGGGTGGGCAGCTGATCCTGATGG
Clone 3	GGAAGAGATGGCGACTCGGGATGATAGACTTCGTATGTCCGAGTCGCGGCAATATATTCGGGGGTGTTCGACCAGGCGGGGAGGCAGCTGATCCTGATGG
Clone 4	GGAAGAGATAGCGACGATTAGTCTGCTGCGTCCGTATGGCTTGTGAGGCGACGTCAGTTCGGGGGTGTTCGACCGGGCGGGGAGTAGCTGATCCTGATGG

Name	5’ → 3’ nucleotide sequence
Clone 5	GGAAGAGATGGCGATTAGCTAAATGTTGTAGTGTAAACTGTAGAGGGAT-GACGTCTGTTCGGGGGTGTTCGACTGGGCGGGGAGTAGCTGATCCTGATGG
Clone 6	GGAAGAGATGGCGACAAATGGCGAAGTTCGGGGACGTAATGCTATC-GACAGCAATTCGGGGGTGTGTGACCGGGCGGGGAGTAGCAGCTGATCCT-GATGG
Clone 7	GGAAGGGATGGCGACGGTATGTAAGTTTGTGTAATGAGGTAGCT-GTTTGCTCGGGGGTGTTTGACTGGGCGGGGAGTGTGTGAGCTGATCCTGATGG
Clone 8	GGAAGAGATGGCGACCGAGCTATGTTGAACCGAGCTAGCAGTCGTGTC-TATCGGGGGTGTCTGACTGGGCGGGGAGTTTTAGGAGCTGATCCTGATGG
Clone 9	GGAAGAGATGGCGACTGGGTATCATGGGCCGCAGTCTGCTTC-GATTCGGGGGTGTTCGACCGGGCGGGGAGTATCAGGGCTGTTGAGCTGATC-CTGATGG
Clone 10	GGAAGAGATGGCGACGAGTCCTTTCCGGGTAAAGGCGGACTGGTGAGCC-TATCGTTCGGGGGTGTCCGACCGGGCGGGGAGTGCGAGCTGATCCTGATGG
Pool2	GGAAGAGATGGCGACGACACAGGGACGATGCCGAATATCCTCAGT-GCGCAGGGCCGCAGGGGGGAGTGACTTGGGATGGGGGGTCCACTAAT-GATCTGCCCGATG (underlined NTs were mutagenized at the rate of 21%)
REV2	CATCGGGCAGATCATTAGTG
FWD2	ACCGCTCAGGTGTAGTATCA
FWD2r	ACCGCTCAGGTGTAGTATCrA
REV2p	pCATCGGGCAGATCATTAGTG
Splint2	GTCGCCATCTCTTCCTGATACTACACCTGAGCGGT
H1 core	GGAAGAAAAAGAATATCCTCAAAAGGGGAGTGACTTGGGATGGGGG
H2	GGAAGATATGGCGCGAACAAATGGACAATGCCGAATATCCCCCG-CACGCAGGGCAACAAGGGGGAGTGACTTGGGATGGGGGCTGCACTAAT-GATCTGCCCGATG
H2 core	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
H3	GGAAGAAATGTAGAGGAAACAGTGACTCTGCAGAATATCCTCACTGCG-TAGTGGGGCAGGGGGGAGTGACTTGGGATGGGGGCTACACTAATGATCTGCC-CGATG
H3 core	identical to H1 core
Clone R.2	GGAAGAAATGACGAGGCCACACGGACGAGGCTGAATATCCTCACTGCGGTGT-GCCCCAGGGAGGAGTGACTTGGGATGGTGGGTTCACTAATGATCTGCCCGATG
Clone R.3	GGAAGATAGTGTGTCGATAAAGGGACAATGCTGAATTTCCTCCCTGAGCCG-GCGCGTAGGGGGGAGTGACTTGGGATGGGGGGTGCACTAATGATCTGCCC-GATG
Clone R.4	GGAAGACTTGACGAGCGCACAGCGCTTGTTCAGAATATCACCCGTCATGT-GTTCGTGAAGGGGGAGTGACTTGGGATGGGGGTTCCACTAATGATCTGCCC-GATG
Clone R.5	GGAAGAAATGGTGCGACACAGGGGCGCTCCAGAATATCCTCATTGCGATATGC-CGCAGGGAGGAGTGACTTGGGATGGTGGGTCCACTAATGATCTGCCCGATG
Clone R.6	GGAAGAAATGTAGAGGAAACAGTGACTCTGCAGAATATCCTCACTGCG-TAGTGGGGCAGGGGGGAGTGACTTGGGATGGGGGCTACACTAATGATCTGCC-CGATG

Name	5’ → 3’ nucleotide sequence
Clone R.7	GGAAGAAACGCGACGACACAGTAATGATGCGGAATATCCCCCCTTCATAGGGC-GAAAGGGGGGAGTGACTTGGGATGGGGGTTACACTAATGATCTGCCCGATG
Clone R.8	GGCAGAGATGGCAACGTCAACATGAGGATGCCGCATATCCCCAGTGCA-CACTTGGGCAGTGGGGAGTGACTTGGGATGGGGGAAACACTAATGATCTGCC-CGATG
Clone R.9	GGAAGATATGGCGGTGACTCTTCGAAGCTGCCGAATATCCCCCGTGCGCG-GAGCCTCAGGGGGGAGTGACTTGGGATGGGGGGTCCACTAATGATCTGCCC-GATG
Clone R.10	GGAAGAAATGCCGACCTCCCAGGGTGGAGGCTGAATATCCACCATGGTCCG-GAACGGATGGGGGAGTGACTTGGGATGGGGGAACCACTAATGATCTGCCC-GATG
Clone R.11	GGAAGACAAGTGGACGGCAGCGGGCCTACGCCGAATATCCTCCG-TATATATGGGTTCAGGGAGGAGTGACTTGGGATGGTGGGTACACTAATGATCT-GCCCGATG
Clone R.12	GGAAGAATGGTCGAGGACACAGGGACGAAGCCGAATATCCTCT-GCCAGCGGGGACGCGGGGAGGAGTGACTTGGGATGGTGGGTACACTAAT-GATCTGCCCGATG
Clone R.13	GGAAGAAAGGACGACGACGCAAGCAGCGTCCTGAATATCCTCCGTGCTCAAG-GCAGCATGGGGGAGTGACTTGGGATGGGGGGAGCACTAATGATCTGCCC-GATG
Clone R.14	GGAAGAGATCTGGAAGAACCAGTGATGTAGATGAATATCCCCATTGCTCTTG-TACGCAGGGGGGAGTGACTTGGGATGGGGGGTTCACTAATGATCTGCCCGATG
Clone R.15	GGAAGATGACCTCAGGTCAAATGGACGAGGTCGAATATCCGCCGTGCG-CATTGCAGTACGGGGGAGTGACTTGGGATGGCGAATACACTAATGATCTGCCC-GATG
Clone R.16	GGAAGATAAGGCGTCCACAGAATAATGACGCAGAATATCCTCCGTGCTCA-GATAAGCAGGGGGGAGTGACTTGGGATGGGGGGCACACTAATGATCTGCCC-GATG
Clone R.17	GGAAGAGATTGTGGCTACATCGGCAGACGGCCGATAGTAACCAGAGCATTTCC-CCACTGGGGGGAGTGACTTGGGATGGGGGGTCCACTAATGATCTGCCCGATG
Clone R.18	GGAAGATAGGGCGTCGATACAGAGACGTTTCCGAATATCCCCCGTGC-TATGGGCCGCATGGGGGAGTGACTTGGGATGGGGGGTCCACTAATGATCTGC-CCGATG
Clone R.19	GGAAGAAGTGCTGTCGACGCAGGGACGATGCAGAATATCCTCACTGAGTAAG-GTCGCAGGGGGGAGTGACTTGGGATGGGGGGTTCACTAATGATCTGCCCGATG
Clone R.20	GGAAGAAAGCACTAGGACTCAGGGACGAGGGTGAATATCCCCTTTGGGCTGT-GTCTCAATGGGGAGTGACTTGGGATGGGGGGAACACTAATGATCTGCCCGATG
Clone R.21	GGAAGAAAGCACTAGGACTCAGGGACGAGGGTGAATATCCCCTTTGGGCTGT-GTCTCAATGGGGAGTGACTTGGGATGGGGGGAACACTAATGATCTGCCCGATG
Clone R.22	GGAAGAAAGCACTAGGACTCAGGGACGAGGGTGAATATCCCCTTTGGGCTGT-GTCTCAATGGGGAGTGACTTGGGATGGGGGGAACACTAATGATCTGCCCGATG
SN triple mutants	
Top triple 1	GGAAGAAAAAGAATATCCCTAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAAGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGTG
	GGAAGAAAAAGAATATCCCTAAAAAGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCTAAAAGGGGAGTGACTTGGGATGGGTG
	GGAAGAAAAAGAATATCCCCAAAAAGGGAGTGACTTGGGATGGGTG

Name	5’ → 3’ nucleotide sequence
	GGAAGAAAAAGAATATCCCTAAAAAGGGAGTGACTTGGGATGGGTG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCTAAAAAGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCGAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCGAAAAGGGGAGTGACTTGGGATGGGCG
	GGAAGAAAAAGAATATCCCCAAAACGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCGAAAAAGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCCAAAAAGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCAAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCAAAAAAGGGAGTGACTTGGGATGGGTG
	GGAAGAAAAAGAATATCCCGAAAAAGGGAGTGACTTGGGATGGGTG
	GGAAGAAAAAGAATATCCCAAAAAAGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCAAAAATGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCGAAAACGGGAGTGACTTGGGATGGGAG
	GGAAGAAAAAGAATATCCCGAAAATGGGAGTGACTTGGGATGGGAG
Top triple 2	GGAAGAAAAAGAATATCCTCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCTCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCTCAAAAGAGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCTCAAAAGGGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCCCAAAAGAGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCTCAAAAGAGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGAGG
	GGAAGAAAAAGAATATCCTCAAAAGAGGAGTGACTTGGGATGGAGG
	GGAAGAAAAAGAATATCCGCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCGCAAAAGGGGAGTGACTTGGGATGGCGG
	GGAAGAAAAAGAATATCCCCAAAAGCGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCGCAAAAGAGGAGTGACTTGGGATGGAGG
	GGAAGAAAAAGAATATCGCCAAAAGGAGAGTGACTTGGGATGAGGG
	GGAAGAAAAAGAATATCCCCAAAAGAGGAGTGACTTGGGATGGAGG
	GGAAGAAAAAGAATATCCACAAAAGAGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCGCAAAAGAGGAGTGACTTGGGATGGTGG
	GGAAGAAAAAGAATATCCACAAAAGAGGAGTGACTTGGGATGGAGG
Top triple 3	GGAAGAAAAAGAATATCTCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGAGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCTCCAAAAGGAGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCTCCAAAAGGGGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCCCCAAAAGGAGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCTCCAAAAGGAGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGAGGG
	GGAAGAAAAAGAATATCTCCAAAAGGAGAGTGACTTGGGATGAGGG
	GGAAGAAAAAGAATATCGCCAAAAGGGGAGTGACTTGGGATGGGGG

Name	5’ → 3’ nucleotide sequence
	GGAAGAAAAAGAATATCGCCAAAAGGGGAGTGACTTGGGATGCGGG
	GGAAGAAAAAGAATATCCCCAAAAGGCGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATGCCCAAAAGGGAAGTGACTTGGGATAGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGAGAGTGACTTGGGATGAGGG
	GGAAGAAAAAGAATATCACCAAAAGGAGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCGCCAAAAGGAGAGTGACTTGGGATGTGGG
	GGAAGAAAAAGAATATCACCAAAAGGAGAGTGACTTGGGATGAGGG
Top triple 4	GGAAGAAAAAGAATATTCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGAAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATTCCCAAAAGGGAAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATTCCCAAAAGGGGAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGAAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATTCCCAAAAGGGAAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATAGGGG
	GGAAGAAAAAGAATATTCCCAAAAGGGAAGTGACTTGGGATAGGGG
	GGAAGAAAAAGAATATGCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATGCCCAAAAGGGGAGTGACTTGGGATCGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGCAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGAAGTGACTTGGGATAGGGG
	GGAAGAAAAAGAATATACCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATACCCAAAAGGGAAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATGCCCAAAAGGGAAGTGACTTGGGATTGGGG
	GGAAGAAAAAGAATATACCCAAAAGGGAAGTGACTTGGGATAGGGG
SN single-nucleotide mutants	
	TGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GTAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGCAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAATAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGTAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAAAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAATAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGCATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGGATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGTATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAGTATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGATTATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAAAATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAACATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAAGATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATGTCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATTTCCCCAAAAGGGGAGTGACTTGGGATGGGGG

Name	**5’ → 3’ nucleotide sequence**
	GGAAGAAAAAGAATATCCCCAAAAGGGGAATGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACCTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACGTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGAGGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGACGGGGG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGT
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGC
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGA
Triple deletions and addition in SN	
	GGAAGAAAAAGAATACCCCAAAAGGGGGTGACTTGGGAGGGGG
	GGAAGAAAAAGAATACCCAAAAGGGGTGACTTGGGAGGGG
	GGAAGAAAAAGAATACCAAAAGGGTGACTTGGGAGGG
	GGAAGAAAAAGAATATCCCAAAAGGGAGTGACTTGGGATGGGG
	GGAAGAAAAAGAATATCCAAAAGGAGTGACTTGGGATGGG
	GGAAGAAAAAGAATATCAAAAGAGTGACTTGGGATGG
	GGAAGAAAAAGAATATCCCCCAAAAGGGGGAGTGACTTGGGATGGGGG
Loop insertion SN mutants	
Var. region 1	GGAAGANNNGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGANNNNNNNNNNGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGAATATCCC-CAAAAGGGGAGTGACTTGGGATGGGGG
Var. region 2	GGAAGAAAAAGAATATCCCCNNNGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCCNNNNNNNNNNGGGGAGTGACTTGGGATGGGGG
	GGAAGAAAAAGAATATCCCC-NNNNNNNNNNNNNNNNNNNNNNNNNNNNGGGGAGT-GACTTGGGATGGGGG
3’ end	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGNNN
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGT-GACTTGGGATGGGGGNNNNNNNNNN
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGT-GACTTGGGATGGGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNN
Loop insertion SN mutants	
Var. region 1	GGAAGAAGAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAACTGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGATAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAACCGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGACCAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAGGGATGACCAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAACGTAATCGGGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGACCTCGCTACCGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
	GGAAGAGGGGGATCTTGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG

Name	5’ → 3’ nucleotide sequence
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGATACC-CTTGT
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGCGATCT-GACG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGC-GAAGTTGGCAAACATTATAAGTTGGGGAG
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGCAAG-GACTTTATACAATGCTCGCGGGTCGA
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGTGGAG-CATGCATTCTTGGACAAAATCTTGT
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGGTGC-TACTTAAATTCTGTGTGTGACTTTAA
	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGGTTGGCT-GCCGTCCCAGGCGGCACGATAGAC
All	GGAAGATCCTGAATATCCCCTTGAGGGGAGTGACTTGGGATGGGGGCTAG
	GGAAGAGAACGAATATCCCCACCTGGGGAGTGACTTGGGATGGGGGCGAA
	GGAAGATGCCGAATATCCCCGTCGGGGGAGTGACTTGGGATGGGGGTACG
	GGAAGAGCGAGAATATCCCCACAGGGGGAGTGACTTGGGATGGGGGACGG
	GGAAGACCCTGAATATCCCCCCGTGGGGAGTGACTTGGGATGGGGGGCTC
	GGAAGACTTTACCTACGAATATCCCCGTAGGCACTCGGGGAGT-GACTTGGGATGGGGGACGCTCCCAT
	GGAAGATCGATGAGGCGAATATCCCCAGATCACCTAGGGGAGT-GACTTGGGATGGGGGAGCGGCACCG
	GGAAGACTCCTTTTATGAATATCCCCCTGTCTTTAAGGGGAGT-GACTTGGGATGGGGGCAGATCGCTA
	GGAAGATCGACCGCCAGAATATCCCCTCAGCGCGAGGGGGAGT-GACTTGGGATGGGGGAGAAGGAAGT
	GGAAGATAACGTAGGGGAATATCCCCCCAGAAGGTAGGGGAGT-GACTTGGGATGGGGGCCCCGGACGT
Sensor polyA	GGAAGAAAAAGAATATCCCCAAAAAAAAAAGGGGAGT-GACTTGGGATGGGGGAATTTTTTTTTTTTTT
Target polyA	AAAAAAAAAAAAAA
Sensor 1	GGAAGAAAAAGAATATCCCCCAATCGTGCGGGGAGT-GACTTGGGATGGGGGAAGCACGATTGACCTTC
Sensor 2	GGAAGAAAAAGAATATCCCCCTCTTAAGAGGGGAGT-GACTTGGGATGGGGGAATCTTAAGAGAAGGGA
Sensor 3	GGAAGAAAAAGAATATCCCCGGCACTGATGGGGAGT-GACTTGGGATGGGGGAAATCAGTGCCCAGTGG
Sensor 4	GGAAGAAAAAGAATATCCCCATGATCGGAGGGGAGT-GACTTGGGATGGGGGAATCCGATCATCCGAAG
Sensor 5	GGAAGAAAAAGAATATCCCCAGTAATAGCGGGGAGT-GACTTGGGATGGGGGAAGCTATTACTTATCTT
Target Oligo 1	GAAGGTCAATCGTGC
Target Oligo 2	TCCCTTCTCTTAAGA

Name	5’ → 3’ nucleotide sequence
Target Oligo 3	CCACTGGGCACTGAT
Target Oligo 4	CTTCGGATGATCGGA
Target Oligo 5	AAGATAAGTAATAGC

Table 2. Bioinformatic tools used for data analysis.

Tool	Version	Source
cutadapt	1.15	https://cutadapt.readthedocs.io/en/stable/
fastq-join	1.3.1	https://github.com/brwnj/fastq-join
Clustal Omega	1.2.4	https://www.ebi.ac.uk/Tools/msa/clustalo/
FastQC	0.11.5	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
DiffLogo	2.8.0	https://www.bioconductor.org/packages/release/bioc/html/DiffLogo.html
ViennaRNA	2.4.12	https://www.tbi.univie.ac.at/RNA/

8 References

[1] F. Miescher-Rüsch, *Med. Untersuchungen* **1871**, *4*, 441–460.
[2] O. T. Avery, C. M. MacLeod, M. McCarty, *J. Exp. Med.* **1944**, *79*, 137–158.
[3] A. D. Hershey, M. Chase, *J. Gen. Physiol.* **1952**, *36*, 39–56.
[4] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
[5] J. B. Sumner, *J. Wash. Acad. Sci.* **1948**, *38*, 113–117.
[6] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
[7] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.
[8] K. A. Peck-Miller, S. Altman, *J. Mol. Biol.* **1991**, *221*, 1–5.
[9] P. Alifano, F. Rivellini, C. Piscitelli, C. M. Arraiano, C. B. Bruni, M. S. Carlomagno, *Genes Dev.* **1994**, *8*, 3021–3031.
[10] C. Guerrier-Takada, A. van Belkum, C. W. A. Pleii, S. Altman, *Cell* **1988**, *53*, 267–272.
[11] Y. Komine, M. Kitabatake, T. Yokogawa, K. Nishikawa, H. Inokuchi, *Proc. Natl. Acad. Sci.* **1994**, *91*, 9223–9227.
[12] S. Altman, D. Wesolowski, C. Guerrier-Takada, Y. Li, *Proc. Natl. Acad. Sci.* **2005**, *102*, 11284–11289.
[13] E. P. Nawrocki, T. A. Jones, S. R. Eddy, *Nucleic Acids Res.* **2018**, *46*, 7970–7976.
[14] C. L. Peebles, P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell, H. L. Cheng, *Cell* **1986**, *44*, 213–223.
[15] T. R. Cech, *Cell* **1986**, *44*, 207–210.
[16] Z. Guo, K. S. Karunatilaka, D. Rueda, *Nat. Struct. Mol. Biol.* **2009**, *16*, 1154–1159.
[17] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, J. D. Watson, *Molecular Biology of the Cell (3rd Ed.)*, New York, NY: Garland Science, **1994**.
[18] C. J. Hutchins, P. D. Rathjen, A. C. Forster, R. H. Symons, *Nucleic Acids Res.* **1986**, *14*, 3627–3640.
[19] B. J. Saville, R. A. Collins, *Cell* **1990**, *61*, 685–696.
[20] W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, *Nature* **2004**, *428*, 281–286.
[21] M. Martick, L. H. Horan, H. F. Noller, W. G. Scott, *Nature* **2008**, *454*, 899–902.
[22] C. H. T. Webb, N. J. Riccitelli, D. J. Ruminski, A. Lupták, *Science* **2009**, *326*, 953.
[23] J. A. Daròs, J. F. Marcos, C. Hernández, R. Flores, *Proc. Natl. Acad. Sci.* **1994**, *91*, 12813–12817.
[24] M. de la Peña, I. García-Robles, *EMBO Rep.* **2010**, *11*, 711–716.
[25] C. Hammann, A. Luptak, J. Perreault, M. De La Peña, *RNA* **2012**, *18*, 871–885.
[26] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, G. Bruening, *Science* **1986**, *231*, 1577–1580.
[27] L. Sharmeen, M. Y. P. Kuo, G. Dinter-Gottlieb, J. Taylor, *J. Virol.* **1988**, *62*,

2674–2679.
[28] K. Salehi-Ashtiani, A. Lupták, A. Litovchick, J. W. Szostak, *Science* **2006**, *313*, 1788–1792.
[29] N. Riccitelli, A. Lupták, *Prog. Mol. Biol. Transl. Sci.* **2013**, *120*, 123–171.
[30] R. H. Symons, *Nucleic Acids Res.* **1997**, *25*, 2683–2689.
[31] A. Roth, Z. Weinberg, A. G. Y. Chen, P. B. Kim, T. D. Ames, R. R. Breaker, *Nat. Chem. Biol.* **2014**, *10*, 56–60.
[32] Z. Weinberg, P. B. Kim, T. H. Chen, S. Li, K. A. Harris, C. E. Lünse, R. R. Breaker, *Nat. Chem. Biol.* **2015**, *11*, 606–610.
[33] E. R. Lee, J. L. Baker, Z. Weinberg, N. Sudarsan, R. R. Breaker, *Science* **2010**, *329*, 845–848.
[34] A. G. Y. Chen, N. Sudarsan, R. R. Breaker, *RNA* **2011**, *17*, 1967–1972.
[35] S. S. S. Panchapakesan, R. R. Breaker, *Nat. Chem. Biol.* **2021**, *17*, 375–382.
[36] A. R. Ferré-D'Amaré, *Q. Rev. Biophys.* **2010**, *43*, 423–447.
[37] W. Gilbert, *Nature* **1986**, *319*, 618.
[38] W. Taifeng, L. E. Orgel, *J. Am. Chem. Soc.* **1992**, *114*, 317–322.
[39] K. D. James, A. D. Ellington, *Orig. life Evol. Biosph. 1999 294* **1999**, *29*, 375–390.
[40] C. W. Carter, P. R. Wills, *Mol. Biol. Evol.* **2018**, *35*, 269–286.
[41] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 217–224.
[42] H. M. Temin, S. Mizutani, *Nature* **1970**, *226*, 1211–1213.
[43] D. Baltimore, *Nature* **1970**, *226*, 1209–1211.
[44] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
[45] K. Mullis, F. Faloona, S. Scharf, R. Saiki, G. Horn, H. Erlich, *Cold Spring Harb. Symp. Quant. Biol.* **1986**, *51*, 263–273.
[46] B. Gutte, R. B. Merrifield, *J. Biol. Chem.* **1971**, *246*, 1922–1941.
[47] S. L. Beaucage, M. H. Caruthers, *Tetrahedron Lett.* **1981**, *22*, 1859–1862.
[48] D. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467–468.
[49] C. Tuerk, L. Gold, *Science* **1990**, *249*, 505–510.
[50] A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822.
[51] R. C. Cadwell, G. F. Joyce, *Genome Res.* **1992**, *2*, 28–33.
[52] K. Pobanz, A. Lupták, *Methods* **2016**, *106*, 14–20.
[53] Y. Li, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 2746–2751.
[54] C. Wilson, J. W. Szostak, *Nature* **1995**, *374*, 777–782.
[55] J. R. Lorsch, J. W. Szostak, *Nature* **1994**, *371*, 31–36.
[56] M. Chandra, A. Sachdeva, S. K. Silverman, *Nat. Chem. Biol.* **2009**, *5*, 718–720.
[57] R. R. Breaker, G. F. Joyce, *Chem. Biol.* **1994**, *1*, 223–229.
[58] M. Levy, K. E. Griswold, A. D. Ellington, *RNA* **2005**, *11*, 1555–1562.
[59] A. Autour, S. C. Y. Jeng, A. D. Cawte, A. Abdolahzadeh, A. Galli, S. S. S. Panchapakesan, D. Rueda, M. Ryckelynck, P. J. Unrau, *Nat. Commun.* **2018**, *9*, 656.
[60] D. van Simaeys, D. López-Colón, K. Sefah, R. Sutphen, E. Jimenez, W. Tan, *PLoS One* **2010**, *5*, e13770.

[61] A. Berezhnoy, C. A. Stewart, J. O. Mcnamara, W. Thiel, P. Giangrande, G. Trinchieri, E. Gilboa, *Mol. Ther.* **2012**, *20*, 1242–1250.
[62] M. Cho, Y. Xiao, J. Nie, R. Stewart, A. T. Csordas, S. S. Oh, J. A. Thomson, H. T. Soh, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 15373–15378.
[63] D. M. Dupont, N. Larsen, J. K. Jensen, P. A. Andreasen, J. Kjems, *Nucleic Acids Res.* **2015**, *43*, e139.
[64] W. H. Thiel, P. H. Giangrande, *Methods* **2016**, *97*, 3–10.
[65] J. Hoinka, T. Przytycka, *Methods* **2016**, *106*, 82–85.
[66] K. F. Tjhung, M. N. Shokhirev, D. P. Horning, G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 2906–2913.
[67] E. Rivas, *Wiley Interdiscip. Rev. RNA* **2021**, *12*, e1649.
[68] N. Komarova, D. Barkova, A. Kuznetsov, *Int. J. Mol. Sci.* **2020**, *21*, 8774.
[69] M. Zuker, *Science* **1989**, *244*, 48–52.
[70] A. R. Gruber, S. H. Bernhart, R. Lorenz, *Methods Mol. Biol.* **2015**, *1269*, 307–326.
[71] M. Andronescu, R. Aguirre-Hernández, A. Condon, H. H. Hoos, *Nucleic Acids Res.* **2003**, *31*, 3416–3422.
[72] R. D. Dowell, S. R. Eddy, *BMC Bioinformatics* **2004**, *5*, DOI 10.1186/1471-2105-5-71.
[73] K. J. Doshi, J. J. Cannone, C. W. Cobaugh, R. R. Gutell, *BMC Bioinformatics* **2004**, *5*, DOI 10.1186/1471-2105-5-105.
[74] C. W. Woese, R. R. Gutell, R. Gupta, H. F. Noller, *Microbiol. Rev.* **1983**, *47*, 621–669.
[75] B. Wimberly, D. Brodersen, W. J. Clemons, R. Morgan-Warren, A. Carter, C. Vonrhein, T. Hartsch, V. Ramakrishnan, *Nature* **2000**, *407*, 327–339.
[76] E. A. Curtis, D. P. Bartel, *Nat. Struct. Mol. Biol.* **2005**, *12*, 994–1000.
[77] R. R. Gutell, *RNA Biol.* **2014**, *11*, 254–272.
[78] D. P. Bartel, J. W. Szostak, *Science* **1993**, *261*, 1411–1418.
[79] D. Herschlag, T. R. Cech, *Biochemistry* **1990**, *29*, 10159–10171.
[80] E. H. Ekland, D. P. Bartel, *Nature* **1996**, *382*, 373–376.
[81] J. Attwater, A. Wochner, P. Holliger, *Nat. Chem.* **2013**, *5*, 1011–1018.
[82] A. Wochner, J. Attwater, A. Coulson, P. Holliger, *Science* **2011**, *332*, 209–212.
[83] D. P. Horning, G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 9786–9791.
[84] R. Cojocaru, P. J. Unrau, *Science* **2021**, *371*, 1225–1232.
[85] L. K. L. Cheng, P. J. Unrau, *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002204.
[86] B. Samanta, G. F. Joyce, *Elife* **2017**, *6*, e31153.
[87] V. K. Jayasena, L. Gold, *Proc. Natl. Acad. Sci.* **1997**, *94*, 10612–10617.
[88] T. Pan, O. C. Uhlenbeck, *Nature* **1992**, *358*, 560–563.
[89] K. B. Chapman, J. W. Szostak, *Chem. Biol.* **1995**, *2*, 325–333.
[90] T. Tuschl, P. A. Sharp, D. P. Bartel, *EMBO J.* **1998**, *17*, 2637–2650.
[91] S. Baskerville, D. P. Bartel, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 9154–9159.
[92] B. Seelig, A. Jäschke, *Chem. Biol.* **1999**, *6*, 167–176.

[93] S. Fusz, A. Eisenführ, S. G. Srivatsan, A. Heckel, M. Famulok, *Chem. Biol.* **2005**, *12*, 941–950.
[94] Y. Ryu, K. J. Kim, C. A. Roessner, A. I. Scott, *Chem. Commun.* **2006**, *0*, 1439–1441.
[95] M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus, *Science* **1995**, *267*, 643–647.
[96] A. Jenne, M. Famulok, *Chem. Biol.* **1998**, *5*, 23–34.
[97] B. Zhang, T. R. Cech, *Nature* **1997**, *390*, 96–100.
[98] P. J. Unrau, D. P. Bartel, *Nature* **1998**, *395*, 260–263.
[99] H. Porta, P. M. Lizardi, *Biotechnology. (N. Y).* **1995**, *13*, 161–164.
[100] J. Haseloff, W. L. Gerlach, *Nature* **1988**, *334*, 585–591.
[101] J. Tang, R. R. Breaker, *Chem. Biol.* **1997**, *4*, 453–459.
[102] M. Fedor, O. Uhlenbeck, *Biochemistry* **1992**, *31*, 12042–12054.
[103] P. Travascio, Y. Li, D. Sen, *Chem. Biol.* **1998**, *5*, 505–517.
[104] S. Nakayama, H. O. Sintim, *Mol. Biosyst.* **2009**, *6*, 95–97.
[105] H. Zhang, B. Jiang, Y. Xiang, Y. Chai, R. Yuan, *Analyst* **2012**, *137*, 1020–1023.
[106] M. Monsur Ali, S. D. Aguirre, H. Lazim, Y. Li, M. M. Ali, S. D. Aguirre, H. Lazim, Y. Li, *Angew. Chemie Int. Ed.* **2011**, *50*, 3751–3754.
[107] J. Liu, Y. Lu, *Angew. Chemie Int. Ed.* **2007**, *46*, 7587–7590.
[108] H. Gao, J. Zhao, Y. Huang, X. Cheng, S. Wang, Y. Han, Y. Xiao, X. Lou, *Anal. Chem.* **2019**, *91*, 14514–14521.
[109] J. Liu, A. K. Brown, X. Meng, D. M. Cropek, J. D. Istok, D. B. Watson, Y. Lu, *Proc. Natl. Acad. Sci.* **2007**, *104*, 2056–2061.
[110] D. Grate, C. Wilson, *Proc. Natl. Acad. Sci.* **1999**, *96*, 6131–6136.
[111] J. S. Paige, K. Y. Wu, S. R. Jaffrey, *Science* **2011**, *333*, 642–646.
[112] G. S. Filonov, J. D. Moon, N. Svensen, S. R. Jaffrey, *J. Am. Chem. Soc.* **2014**, *136*, 16299–16308.
[113] E. V. Dolgosheina, S. C. Y. Jeng, S. S. S. Panchapakesan, R. Cojocaru, P. S. K. Chen, P. D. Wilson, N. Hawkins, P. A. Wiggins, P. J. Unrau, *ACS Chem. Biol.* **2014**, *9*, 2412–2420.
[114] X. Chen, D. Zhang, N. Su, B. Bao, X. Xie, F. Zuo, L. Yang, H. Wang, L. Jiang, Q. Lin, M. Fang, N. Li, X. Hua, Z. Chen, C. Bao, J. Xu, W. Du, L. Zhang, Y. Zhao, L. Zhu, J. Loscalzo, Y. Yang, *Nat. Biotechnol.* **2019**, *37*, 1287–1293.
[115] H. Kim, S. R. Jaffrey, *Cell Chem. Biol.* **2019**, *26*, 1725–1731.
[116] X. Li, L. Mo, J. L. Litke, S. K. Dey, S. R. Suter, S. R. Jaffrey, *J. Am. Chem. Soc.* **2020**, *142*, 14117–14124.
[117] S. K. Dey, G. S. Filonov, A. O. Olarerin-George, B. T. Jackson, L. W. S. Finley, S. R. Jaffrey, *Nat. Chem. Biol.* **2022**, *18*, 180–190.
[118] J. S. Paige, T. Nguyen-Duc, W. Song, S. R. Jaffrey, *Science* **2012**, *335*, 1194.
[119] R. L. Strack, W. Song, S. R. Jaffrey, *Nat. Protoc.* **2013**, *9*, 146–155.
[120] A. D. Cawte, H. Iino, P. J. Unrau, D. S. Rueda, *Methods Mol. Biol.* **2022**, *2404*, 267–280.
[121] W. Song, R. L. Strack, S. R. Jaffrey, *Nat. Methods 2013 109* **2013**, *10*, 873–875.
[122] Y. Xiao, V. Pavlov, R. Gill, T. Bourenko, I. Willner, *ChemBioChem* **2004**, *5*,

374–379.
[123] X. Cheng, X. Liu, T. Bing, Z. Cao, D. Shangguan, *Biochemistry* **2009**, *48*, 7817–7823.
[124] S. H. J. Mei, Z. Liu, J. D. Brennan, Y. Li, *J. Am. Chem. Soc.* **2002**, *125*, 412–420.
[125] J. R. Babendure, S. R. Adams, R. Y. Tsien, *J. Am. Chem. Soc.* **2003**, *125*, 14716–14717.
[126] I. Bronstein, B. Edwards, J. C. Voyta, *J. Biolumin. Chemilumin.* **1989**, *4*, 99–111.
[127] P. Trayhurn, M. E. A. Thomas, J. S. Duncan, D. Black, J. H. Beattie, D. V. Rayner, in *Biochem. Soc. Trans.*, **1995**, p. 494S.
[128] A. J. Camden, S. M. Walsh, S. H. Suk, S. K. Silverman, *Biochemistry* **2016**, *55*, 2671–2676.
[129] N. W. Luedtke, A. Schepartz, *Chem. Commun.* **2005**, *0*, 5426–5428.
[130] J. J. Butzow, G. L. Eichhorn, *J. Inorg. Biochem.* **1986**, *28*, 21–31.
[131] J. Chandrasekar, S. K. Silverman, *Proc. Natl. Acad. Sci.* **2013**, *110*, 5315–5320.
[132] S. Kuusela, H. Lönnberg, *Metal Ions in Biological Systems: Volume 32: Interactions of Metal Ions with Nucleotides: Nucleic Acids, and Their Constituents*, CRC Press, **1996**.
[133] R. R. Gutell, A. Power, G. Z. Hertz, E. J. Putz, G. D. Stormo, *Nucleic Acids Res.* **1992**, *20*, 5785–5795.
[134] E. H. Ekland, D. P. Bartel, *Nucleic Acids Res.* **1995**, *23*, 3231–3238.
[135] M. Jakubec, K. Pšenáková, K. Svehlova, E. A. Curtis, *ChemBioChem* **2022**, e202200026.
[136] R. M. Kumar, G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9738–9743.
[137] D. A. Baum, S. K. Silverman, *Angew. Chemie Int. Ed.* **2007**, *46*, 3502–3504.
[138] T. Li, E. Wang, S. Dong, *Anal. Chem.* **2010**, *82*, 1515–1520.
[139] S. K. Silverman, *Trends Biochem. Sci.* **2016**, *41*, 595–609.
[140] S. A. McManus, Y. Li, *Biochemistry* **2007**, *46*, 2198–2204.
[141] W. Wang, L. P. Billen, Y. Li, *Chem. Biol.* **2002**, *9*, 507–517.
[142] L. Gu, R. Saran, W. Yan, P. J. J. Huang, S. Wang, M. Lyu, J. Liu, *ACS Omega* **2018**, *3*, 15174–15181.
[143] M. Liu, Q. Yin, J. D. Brennan, Y. Li, *Biochimie* **2018**, *145*, 151–157.
[144] M. M. Ali, R. Silva, D. White, S. Mohammadi, Y. Li, A. Capretta, J. D. Brennan, *Angew. Chemie Int. Ed.* **2022**, *61*, e202112346.
[145] Z. Zhang, J. Li, J. Gu, R. Amini, H. Stacey, J. Ang, D. White, C. Filipe, K. Mossman, M. Miller, B. Salena, D. Yamamura, P. Sen, L. Soleymani, J. Brennan, Y. Li, *Chemistry* **2022**, *28*, e202200078.
[146] Z. Zhang, R. Pandey, J. Li, J. Gu, D. White, H. D. Stacey, J. C. Ang, C. J. Steinberg, A. Capretta, C. D. M. Filipe, K. Mossman, C. Balion, M. S. Miller, B. J. Salena, D. Yamamura, L. Soleymani, J. D. Brennan, Y. Li, *Angew. Chem. Int. Ed. Engl.* **2021**, *60*, 24266–24274.
[147] K. Mirzakhani, S. L. M. Gargari, I. Rasooli, S. Rasoulinejad, *Iran. Biomed. J.*

2018, *22*, 193.
[148] C. L. A. Hamula, H. Peng, Z. Wang, G. J. Tyrrell, X. F. Li, X. C. Le, *Methods* **2016**, *97*, 51–57.
[149] T. Kato, I. Shimada, R. Kimura, M. Hyuga, *Chem. Commun.* **2016**, *52*, 4041–4044.
[150] T. Streckerová, J. Kurfürst, E. A. Curtis, *Nucleic Acids Res.* **2021**, *49*, 6971–6981.
[151] M. Izabella Abreu de Melo, C. Rodrigues Correa, P. da Silva Cunha, A. Miranda de Góes, D. Assis Gomes, A. Silva Ribeiro de Andrade, *Bioorg. Med. Chem. Lett.* **2020**, *30*, 127278.
[152] Y. Song, Z. Zhu, Y. An, W. Zhang, H. Zhang, D. Liu, C. Yu, W. Duan, C. J. Yang, *Anal. Chem.* **2013**, *85*, 4141–4149.
[153] K. Sefah, L. Meng, D. Lopez-Colon, E. Jimenez, C. Liu, W. Tan, *PLoS One* **2010**, *5*, e14269.
[154] J. B. H. Tok, N. O. Fischer, *Chem. Commun.* **2008**, *0*, 1883–1885.
[155] M. N. Stojanovic, P. de Prada, D. W. Landry, *J. Am. Chem. Soc.* **2001**, *123*, 4928–4931.
[156] C. J. Huang, H. I. Lin, S. C. Shiesh, G. Bin Lee, *Biosens. Bioelectron.* **2010**, *25*, 1761–1766.
[157] K. A. McCall, C. C. Huang, C. A. Fierke, *J. Nutr.* **2000**, *130*, 1437S-1446S.
[158] G. F. Joyce, *Annu. Rev. Biochem.* **2004**, *73*, 791–836.

9 Supplementary information

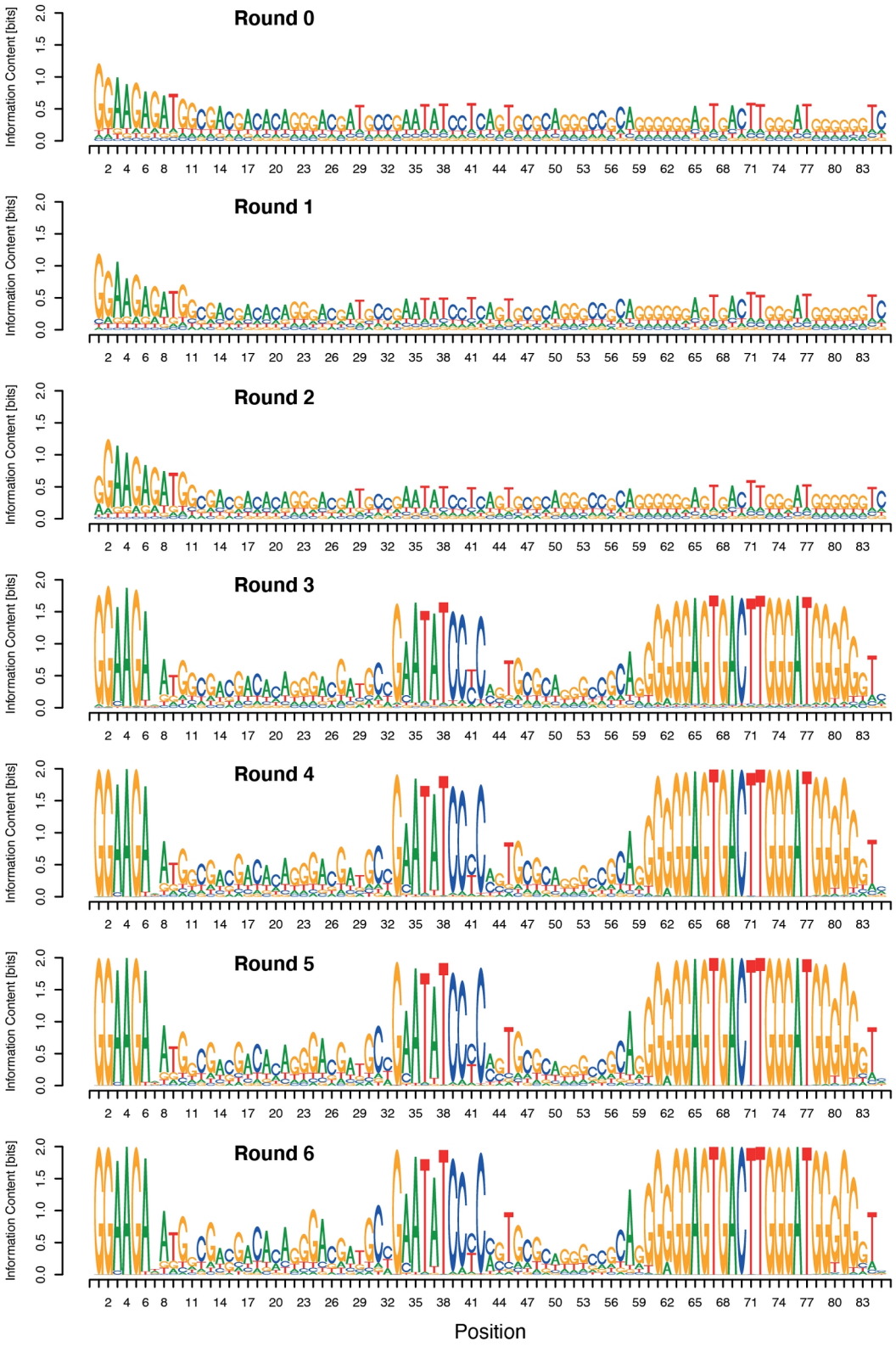
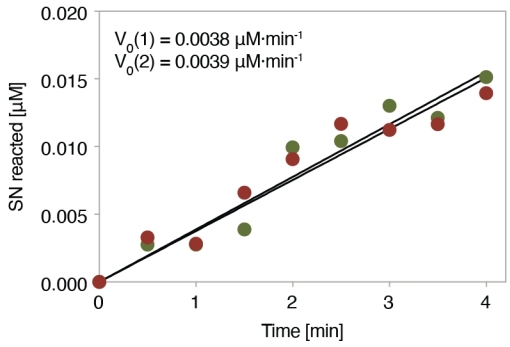
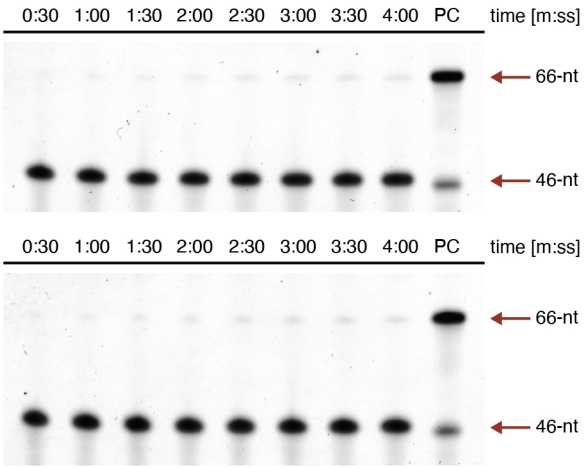


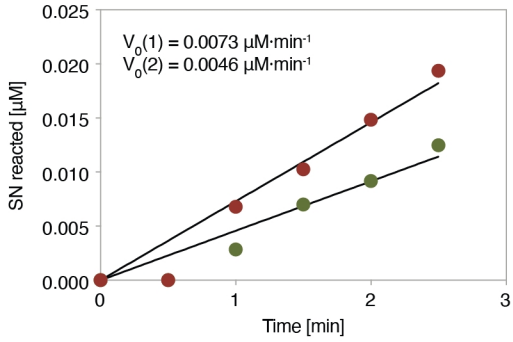
Figure S1. Identification of positions in Supernova important for catalytic activity. A library of deoxyribozyme variants was generated by randomly mutagenizing the sequence of the most active deoxyribozyme from initial selection at a rate of 21% per position. The library was characterized by high-throughput sequencing after each round of selection. The sequence logos show the extent of conservation of each position in the library after each round of selection. Positions 7-32 (variable region 1) and 43-60 (variable region 2) were replaced by AAAA spacers, and positions 83-85 as well as the 3' primer binding site (not shown) were deleted in the minimized version of the deoxyribozyme.

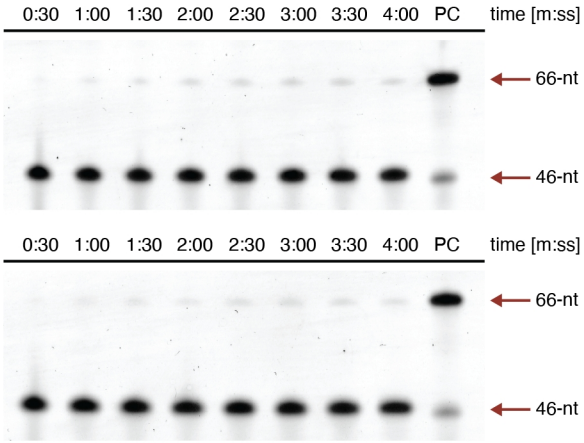
[CDP-Star] = 3.9 µM



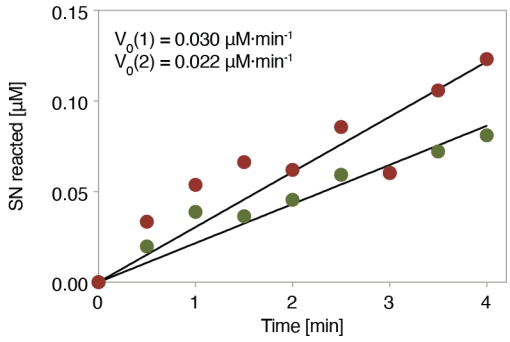


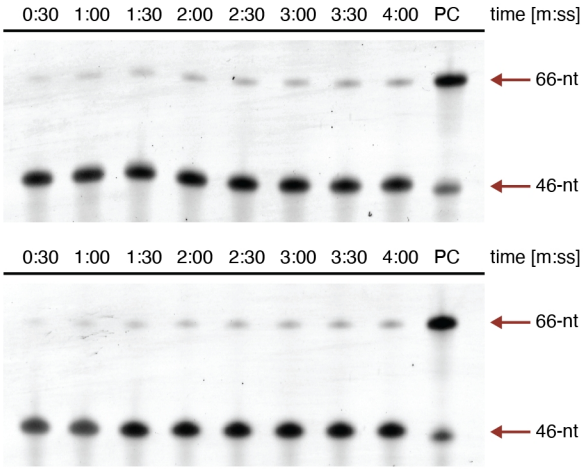
[CDP-Star] = 7.8 µM



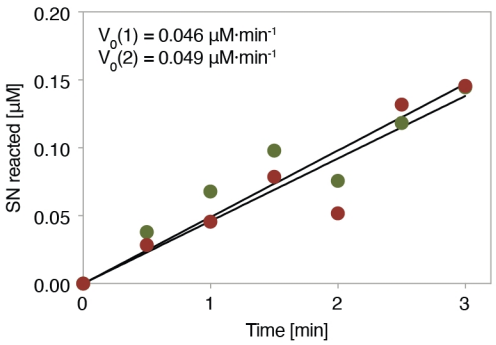


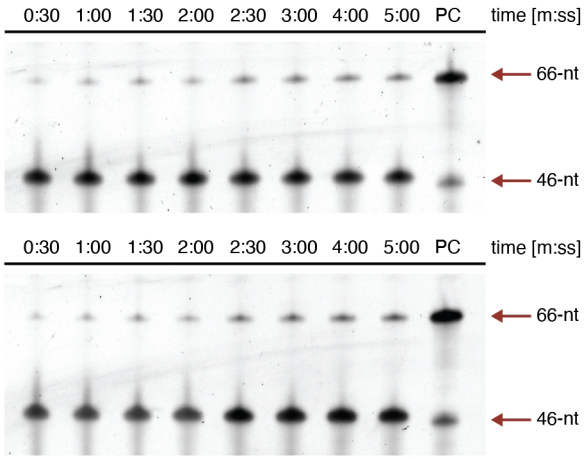
[CDP-Star] = 15.6 µM



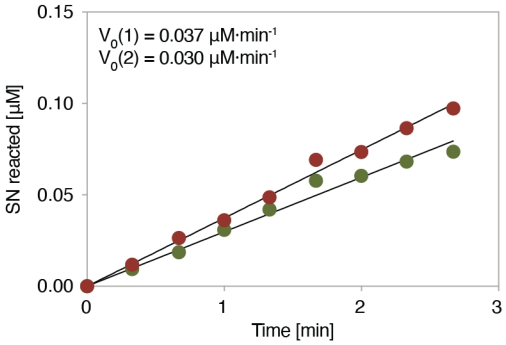


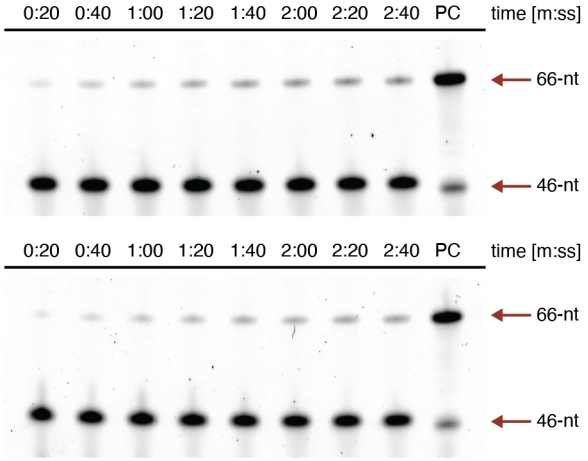
[CDP-Star] = 31.3 µM



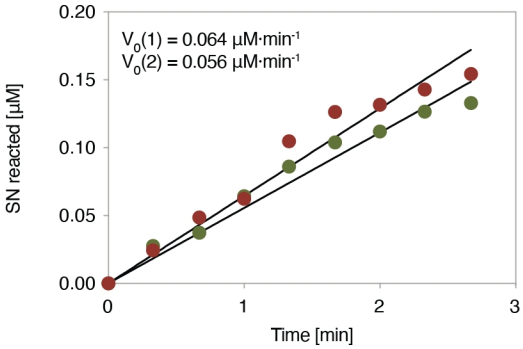


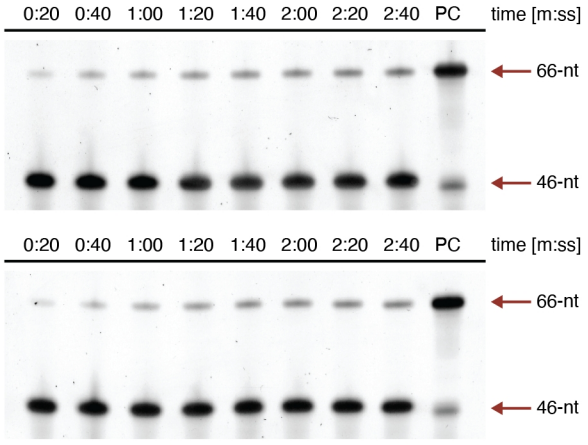
[CDP-Star] = 62.5 µM





[CDP-Star] = 125 µM





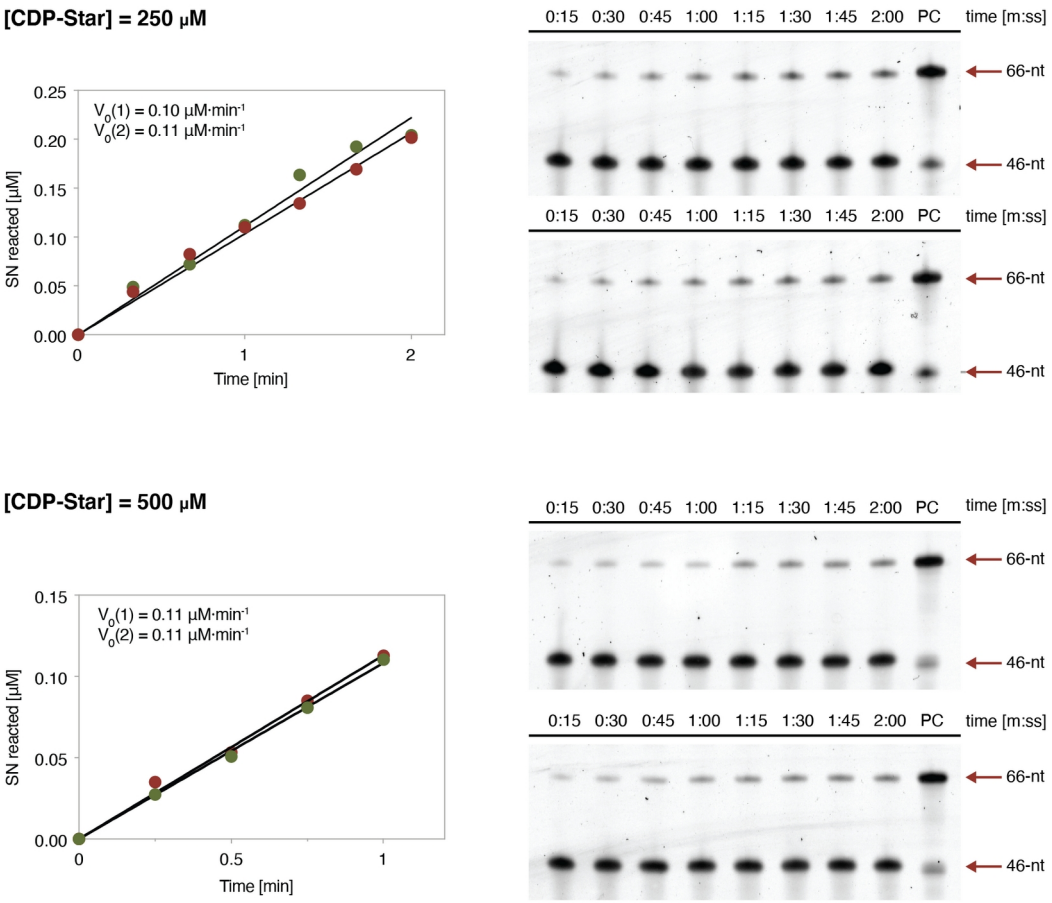


Figure S2. Determining the k_{cat} and K_M values of Supernova. Each panel shows the catalytic activity of Supernova at a different concentration of CDP-Star. Reactions contained 1 µM Supernova and were performed in a buffer containing 50 mM HEPES pH 7.4, 20 mM KCl, and 1 mM $ZnCl_2$. Reactions were stopped at different times and, after tagging the reacted (5' phosphorylated) molecules by ligation, reacted and unreacted molecules were separated by PAGE. The positive control (PC) indicates a 5'-phosphorylated oligonucleotide that was used as both a control for the ligation reaction and a marker. Unreacted molecules are 46 nucleotides long whereas ligated ones have the size of 66 nucleotides (red arrows). Initial velocities were determined as described in the methods, and plotted as a function of CDP-Star concentration to determine k_{cat} and K_M values.

Optimizing the chemiluminescence of a light-producing deoxyribozyme

Martin Jakubec,[a, b] Karolína Pšenáková,[a, b] Katerina Svehlova,[a, b] Edward A. Curtis*[a]

[a] M. Jakubec, K. Pšenáková, K. Svehlova, E. A. Curtis
Institute of Organic Chemistry and Biochemistry
Prague, Czech Republic
E-mail: curtis@uochb.cas.cz
[b] M. Jakubec, K. Pšenáková, K. Svehlova
Faculty of Science
Charles University in Prague
Prague, Czech Republic

This PDF file includes:

Figures S1 to S12
Table S1

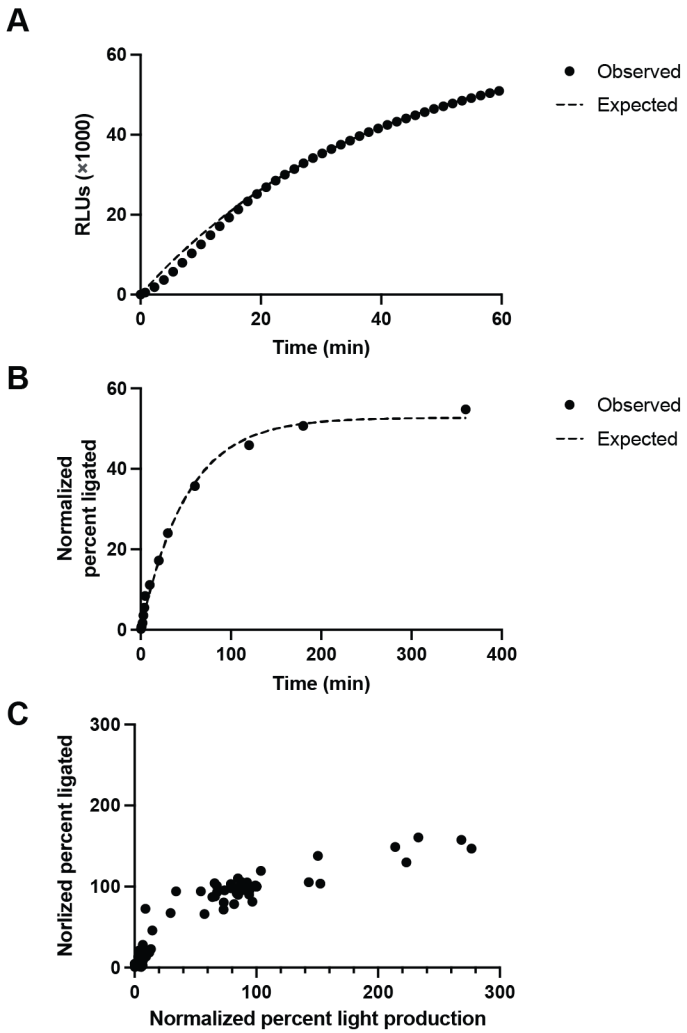


Figure S1. Comparing efficiencies of light production and self-phosphorylation.

(A) Example of a time course of light production measured using the plate reader assay. The curve was fitted using equation 1 in the Materials and Methods section of the manuscript. (B) Example of a time course of self-phosphorylation measured using the ligation assay. The curve was fit us - ing equation 1 in the Materials and Methods section of the manuscript. (C) Comparing efficiencies of light production and self-phosphorylation over a range of conditions. The *x* axis shows the amount of light produced in one hour (normalized to the activity of Supernova in an optimized buffer which contained 1 mM $ZnCl_2$, 20 mM KCl, 50 mM HEPES, pH 7.4), and the *y* axis shows the percent of ligated Supernova obtained after a one-hour incubation with CDP-Star (again, normalized to the activity of Supernova in optimized buffer).

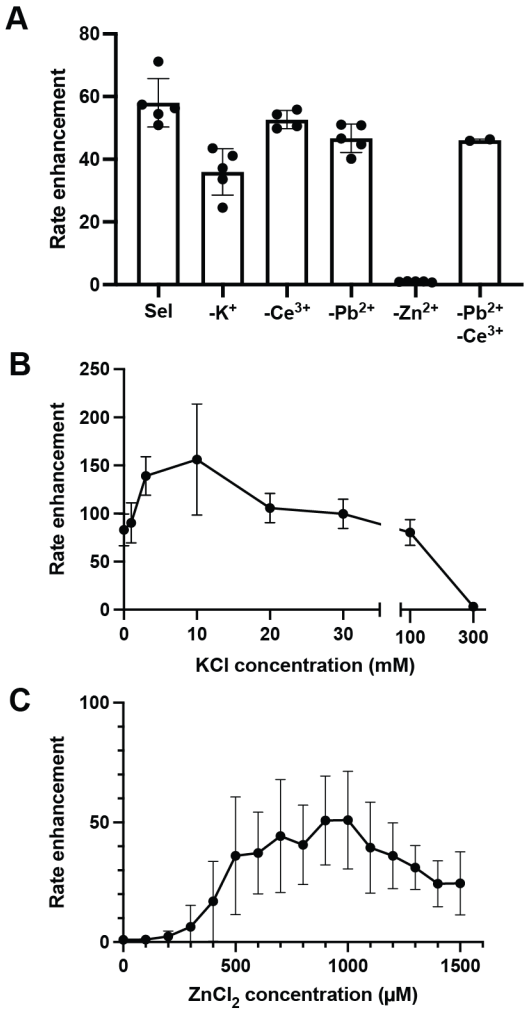


Figure S2. Importance of potassium and zinc in the Supernova reaction.

(A) Rate enhancement of Supernova in a series of buffers in which metal ions (either potassium, cerium, lead, zinc, or cerium and lead) were omitted from the original selection buffer (labeled "Sel"), which contained 200 mM KCl, 1 μM $CeCl_3$, 0.1 μM $PbCl_2$, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. (B) Potassium titration showing the rate enhancement of light production at a constant zinc concentration. Buffers contained varying concentrations of KCl, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. (C) Zinc titration showing the rate enhancement of light production at a constant potassium concentration. Buffers contained 20 mM KCl, varying concentrations of $ZnCl_2$, and 50 mM HEPES, pH 7.4. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. In each panel, rate enhancement is defined as the rate of light production in the presence of deoxyribozyme divided by the rate of light production in the absence of deoxyribozyme. Points show the average of at least three experiments, and error bars represent one standard deviation.

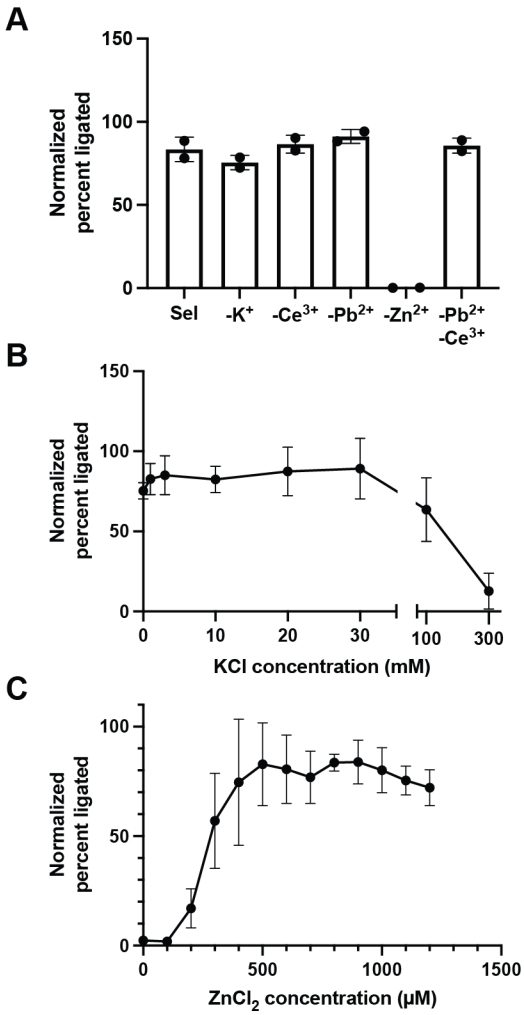


Figure S3. Importance of potassium and zinc in the Supernova reaction.

Same as Figure S2, but showing the normalized percent of phosphorylated Supernova measured using the ligation assay. Percent ligated was determined after a one-hour incubation with CDP-Star, and normalized to a control containing a 5' phosphate. Points show the average of at least three experiments, and error bars represent one standard deviation.

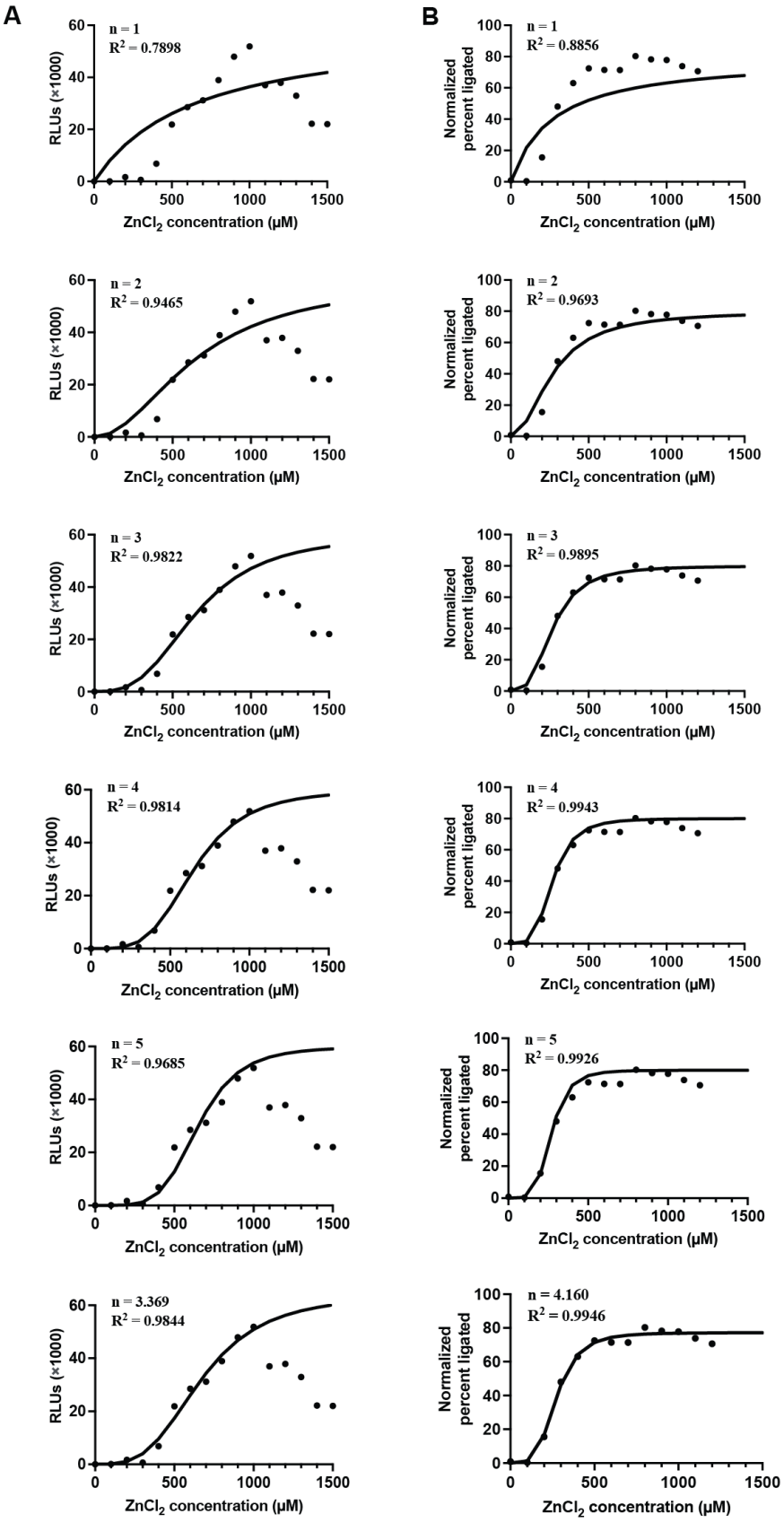

A
B
n = 1
R² = 0.7898
n = 1
R² = 0.8856
n = 2
R² = 0.9465
n = 2
R² = 0.9693
n = 3
R² = 0.9822
n = 3
R² = 0.9895
n = 4
R² = 0.9814
n = 4
R² = 0.9943
n = 5
R² = 0.9685
n = 5
R² = 0.9926
n = 3.369
R² = 0.9844
n = 4.160
R² = 0.9946
RLUs (×1000)
Normalized percent ligated
ZnCl2 concentration (μM)
0
20
40
60
80
100
500
1000
1500

Figure S4. Cooperative effect of zinc on Supernova catalytic activity.

(A) Effect of zinc on the ability of Supernova to generate light (analyzed using the plate reader assay). (B) Effect of zinc on the ability of Supernova to phosphorylate itself (analyzed using the ligation assay). Curves were fit using equation S1:

$$(S1)\ F = F_{max} \times ([Zn]^n/[K_{0.5}]^n + [Zn]^n)$$

in which F is the cumulative amount of light produced (panel A) or the normalized percent ligated (panel B) after a 60 minute incubation, F_{max} is the maximum total amount of light produced (panel A) or the maximum percent ligated (panel B), Zn is the zinc concentration, n is the Hill coefficient (1, 2, 3, 4, 5, or the best fit), and $K_{0.5}$ is the zinc concentration at which the cumulative amount of light produced (panel A) or the normalized percent ligated (panel B) is at half its maximal value. Only points from 0 μM to 1000 μM were used for curve fitting; points at higher concentrations are included to show that zinc is inhibitory at these concentrations, probably due to limited solubility.

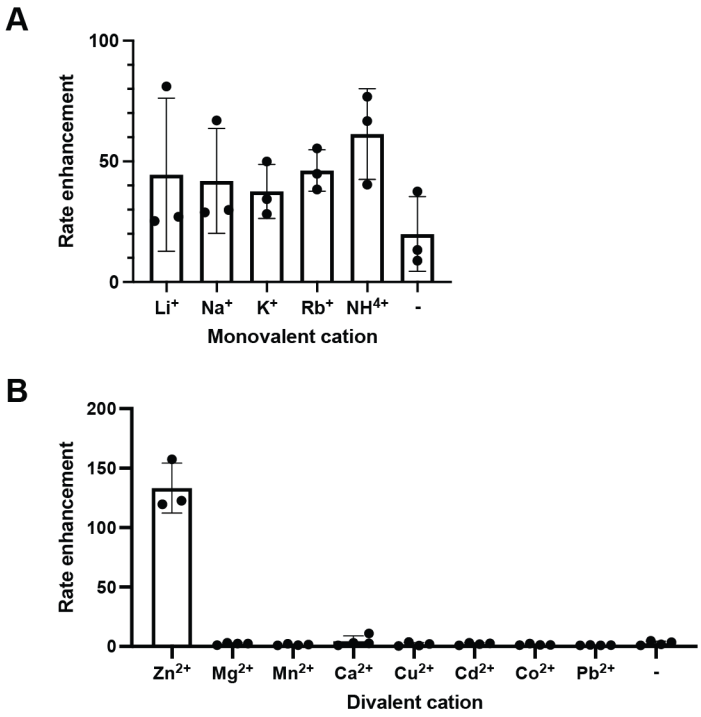


Figure S5. Metal ion requirements of Supernova.

(A) Replacement of potassium with other monovalent metal ions. Buffers contained 20 mM of the indicated monovalent cation, 1 mM $ZnCl_2$, and 20 mM HEPES, pH 7.4. (B) Replacement of zinc with other divalent metal ions. Buffers contained 20 mM KCl, 1 mM of the indicated divalent cation, and 20 mM HEPES, pH 7.4. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. In each panel, rate enhancement is defined as the rate of light production in the presence of deoxyribozyme divided by the rate of light production in the absence of deoxyribozyme. Points show the average of at least three experiments, and error bars represent one standard deviation.

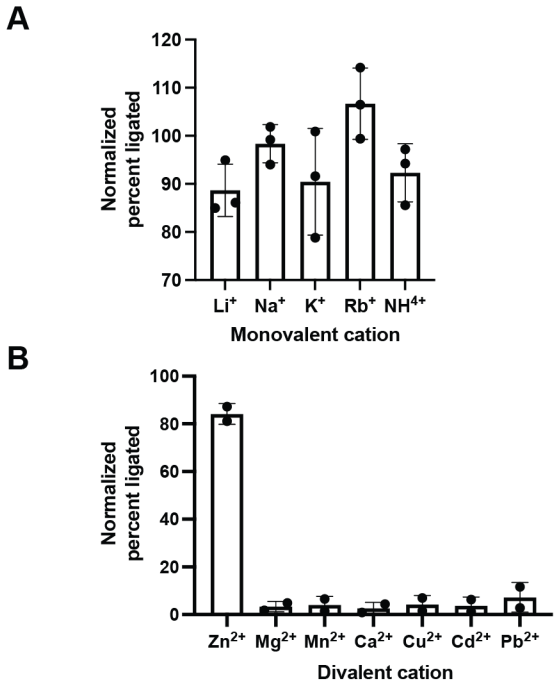


Figure S6. Metal ion requirements of Supernova.

Same as Figure S5, but showing the normalized percent of phosphorylated Supernova measured using the ligation assay. Percent ligated was determined after a one-hour incubation with CDP-Star, and normalized to a control containing a 5' phosphate. Points show the average of at least three experiments, and error bars represent one standard deviation.

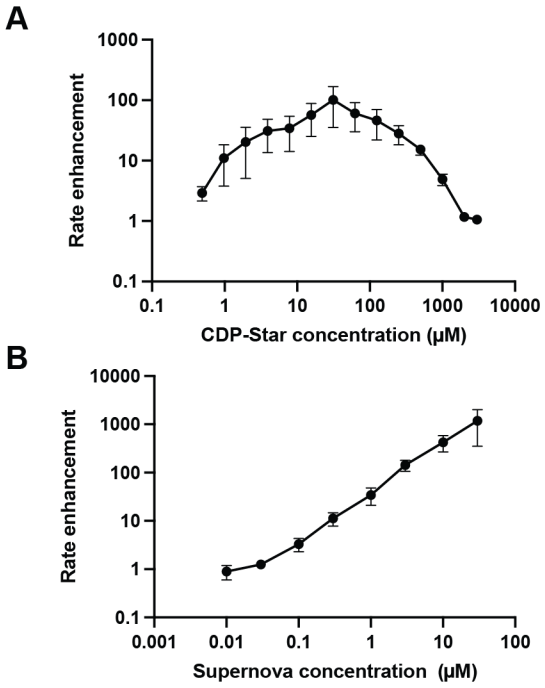


Figure S7. Effect of CDP-Star and Supernova concentration on the rate enhancement of light production.

(A) CDP-Star titration. (B) Supernova titration. Experiments were performed in a buffer containing 20 mM KCl, 0.65 mM $ZnCl_2$, and 20 mM HEPES, pH 7.4. Experiments in panel A contained 1 μM Supernova and varying concentrations of CDP-Star. Experiments in panel B contained varying concentrations of Supernova and 100 μM CDP-Star. In each panel, rate enhancement is defined as the rate of light production in the presence of deoxyribozyme divided by the rate of light production in the absence of deoxyribozyme. Points show the average of at least three experiments, and error bars represent one standard deviation.

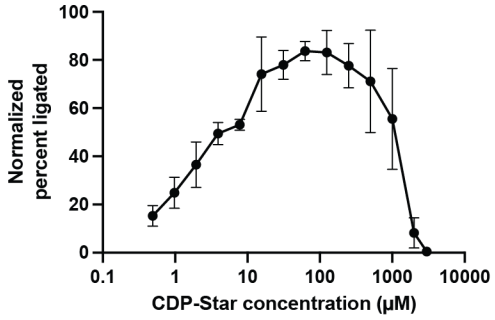


Figure S8. Effect of CDP-Star concentration on the efficiency of Supernova self-phosphorylation.

Same as Figure S7A, but showing the normalized percent of phosphorylated Supernova measured using the ligation assay. Percent ligated was determined after a one-hour incubation with CDP-Star, and normalized to a control containing a 5' phosphate. Points show the average of at least three experiments, and error bars represent one standard deviation.

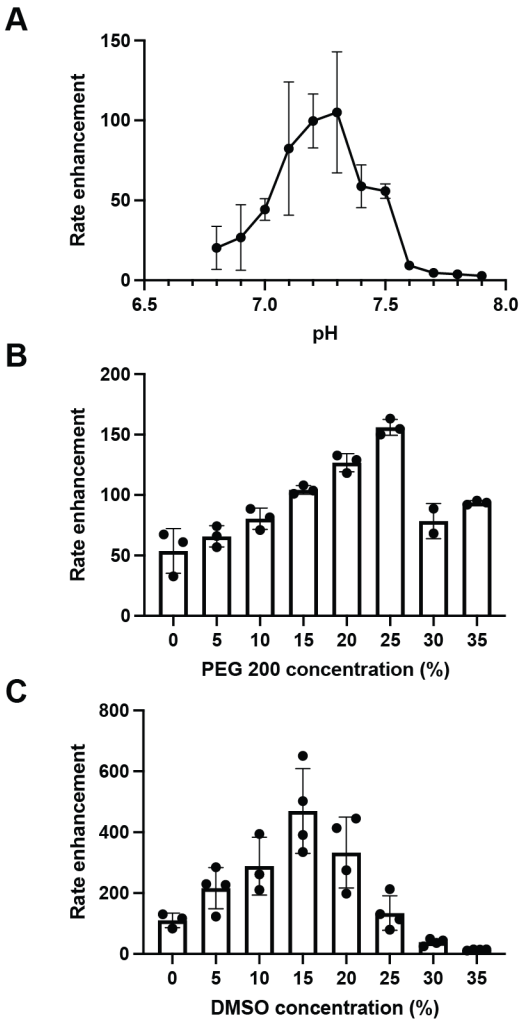


Figure S9. Effects of pH, molecular crowding agents, and organic solvents on the rate enhancement of light production.

(A) Effect of pH on the Supernova reaction. Buffers contained 200 mM KCl, 0.625 mM $ZnCl_2$, and 50 mM HEPES at various pH values. (B) Effect of PEG 200 concentration on the Supernova reaction. Buffers contained 200 mM KCl, 0.625 mM $ZnCl_2$, 50 mM HEPES, pH 7.4, and varying concentrations of PEG 200. (C) Effect of DMSO concentration on the Supernova reaction. Buffers contained 200 mM KCl, 0.625 mM $ZnCl_2$, 50 mM HEPES, pH 7.4, and varying concentrations of DMSO. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. In each panel, rate enhancement is defined as the rate of light production in the presence of deoxyribozyme divided by the rate of light production in the absence of deoxyribozyme. Points show the average of at least three experiments, and error bars represent one standard deviation.

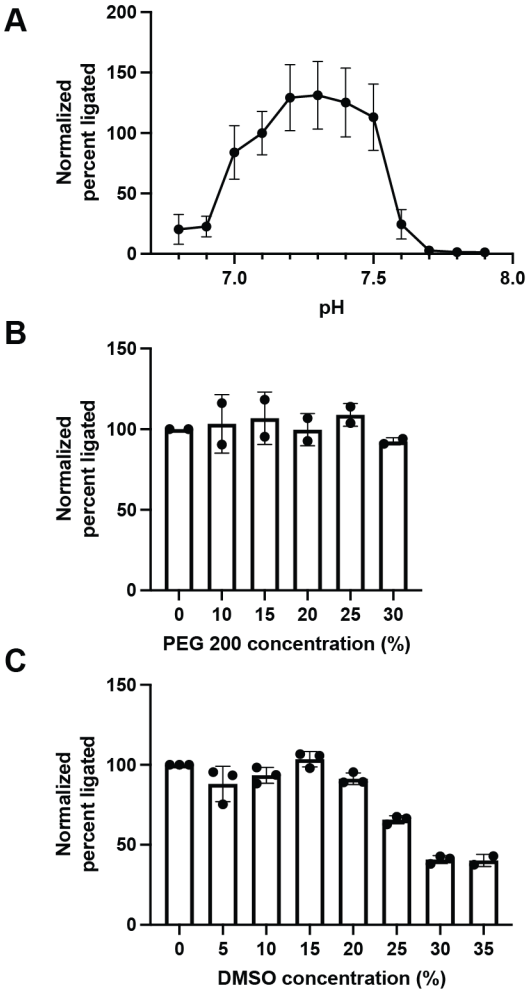


Figure S10. Effects of pH, molecular crowding agents, and organic solvents on the efficiency of Supernova self-phosphorylation.
Same as Figure S9, but showing the normalized percent of phosphorylated Supernova measured using the ligation assay. Percent ligated was determined after a one-hour incubation with CDP-Star, and normalized to a control containing a 5' phosphate. Points show the average of at least three experiments, and error bars represent one standard deviation.

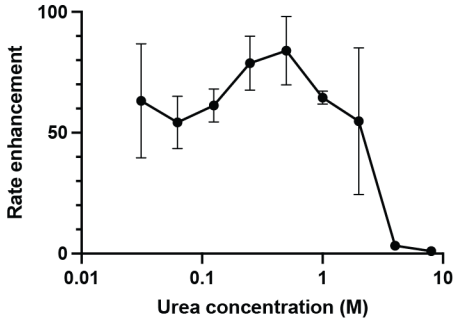


Figure S11. Effect of urea on the rate enhancement of light production.

Effect of urea on the Supernova reaction. Buffers contained 20 mM KCl, 0.65 mM $ZnCl_2$, varying concentrations of urea, and 20 mM HEPES, pH 7.4. All reactions were performed in the presence of 1 μM Supernova and 100 μM CDP-Star. Rate enhancement is defined as the rate of light production in the presence of deoxyribozyme divided by the rate of light production in the absence of deoxyribozyme. Points show the average of at least three experiments, and error bars represent one standard deviation.

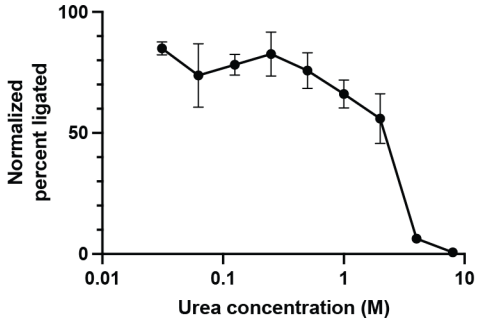


Figure S12. Effect of urea on the efficiency of Supernova self-phosphorylation.

Same as Figure S11, but showing the normalized percent of phosphorylated Supernova measured using the ligation assay. Percent ligated was determined after a one-hour incubation with CDP-Star, and normalized to a control containing a 5' phosphate. Points show the average of at least three experiments, and error bars represent one standard deviation.

Table S1. List of oligonucleotide sequences.

Name	5' → 3' nucleotide sequence
Supernova	GGAAGAAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
Splint	GTCGCCATCTCTTCCTGATACTACACCTGAGCGGT
FWD	ACCGCTCAGGTGTAGTATCA

Author's contribution to the following publications:

Svehlova, K., Lukšan, O., Jakubec, M., & Curtis, E. A. (2022). Supernova: a deoxyribozyme that catalyzes a chemiluminescent reaction. *Angewandte Chemie International Edition*, 61(3), e202109347.

The major part of the work described in this thesis was published in Angewandte Chemie in the article "Supernova: a deoxyribozyme that catalyzes a chemiluminescent reaction." Kateřina was the first author on this paper. She helped to develop the protocol and optimized all steps. She performed all experiments described in the manuscript except for bioinformatic analysis and the optimization experiments in Figure S6. She also made the figures and helped prepare the manuscript.

Jakubec, M., Pšenáková, K., **Svehlova, K.**, & Curtis, E. (2022). Optimizing the chemiluminescence of a light-producing deoxyribozyme. *ChemBioChem*, e202200026.

Kateřina was also an author on a manuscript published in ChemBioChem entitled "Optimizing the Chemiluminescence of a Light-Producing Deoxyribozyme." She optimized the protocols used for the ligation and light-production assays. She also performed preliminary experiments investigating the effects of various reaction conditions (including the pH of the buffer, the metal ion composition of the buffer, the DMSO concentration, the Supernova concentration, and the substrate concentration) on light production.