Charles University

Faculty of Science

Biology

Molecular and Cell Biology, Genetics and Virology

Mgr. Martin Volek

Development and use of deoxyribozymes that generate color and fluorescence

Vývoj a využití deoxyribozymů, které produkují barvu a fluorescenci

Doctoral thesis

Supervisor:

Edward Arthur Curtis, PhD

Prague, 2024

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

Praha, 2024

Martin Volek

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

Prague, 2024

Martin Volek

Acknowledgements:

Here I would like to thank my supervisor Edward Arthur Curtis, PhD, for leading me through my PhD studies, for his great ideas, advice and support.

I would like to thank all members of the Laboratory of functional potential of nucleic acids, for the friendly atmosphere, kind humour and useful help in solving countless problems. I am very grateful that I could work in such a team.

Last and the most important thanks is addressed to my family, my parents and especially to my wife Zuzka and daughters Anna and Eva, for their immense faith, understanding, kindness, and their support which they have given me during my studies. Without my kids I would have finished my studies earlier, but definitely less happy.

Abstract

The recent SARS-CoV-2 pandemic showed how humanity can be threatened by organisms as simple as viruses. This crisis emphasized the need for timely and reliable diagnostics, as well as the fast discovery of effective therapeutics against new diseases. We propose that both challenges can be addressed by deoxyribozymes (catalytically active DNA molecules).

The most common types of signals are light, fluorescence, color and radioactivity. Some of these signals can be generated by deoxyribozymes, but others cannot. A functional DNA molecule that generates a robust fluorogenic or chromogenic signal would have significant advantages for many applications compared to intrinsically fluorescent proteins, which are expensive and labour intensive to synthesize, and fluorescent RNA aptamers, which are more prone to degradation than DNA molecules. To broaden the spectrum of possible outcomes generated by DNA, we have used *in vitro* selection to develop deoxyribozymes, which generate strong fluorogenic and chromogenic signals.

Deoxyribozymes that generate a fluorogenic signal achieve this by dephosphorylating the coumarin substrate 4-methylumbelliferyl phosphate (4-MUP). They transfer the phosphate group from 4-MUP to their own 5' hydroxyl group. This reaction converts the non-fluorogenic 4-MUP into the fluorogenic 4-methylumbelliferone (4-MU). We named the most active deoxyribozyme identified in our studies Aurora. This enhances fluorescence by more than 700-fold and generates a detectable signal in minutes. Deoxyribozymes that produce a yellow color transfer the phosphate group from the colorless substrate 4-nitrophenyl phosphate (4-NPP) to their own 5' hydroxyl group. After dephosphorylation, 4-NPP is converted into yellow 4-nitrophenol (4-NP). We named this deoxyribozyme Apollon. It enhances the chromogenic signal by more than 100-fold.

To improve the catalytic activities of Aurora and Apollon, to identify their minimal catalytic cores, and to solve their secondary structures we used *in vitro* selection in combination with high-throughput sequencing and comparative sequence analysis. Furthermore, to obtain the best possible signal-to-noise ratio, we extensively characterized reaction conditions, including the effects of various mono- and divalent ions, different buffering agents, molecular crowding agents, and pH levels, as well as different substrate and deoxyribozyme concentrations.

Finally, we developed several ways to convert Aurora and Apollon into allosterically regulated sensors. Aurora and Apollon were designed to function as oligonucleotide sensors for specific sequences and also to act as sensors of various types of nucleases in homogeneous assays that did not require washes or purification steps. Notably, an engineered version of Aurora can detect picomolar concentrations of ribonuclease, and was used for rapid identification of novel small-molecule inhibitors of the SARS-CoV-2 ribonuclease Nsp15 in a high-throughput screen. Our results show that deoxyribozymes can generate robust fluorogenic and chromogenic signals and that such deoxyribozymes can be used in applied research to solve real-world challenges.

Abstrakt

Nedávná pandemie SARS-CoV-2 ukázala, že lidstvo mohou ohrozit i tak jednoduché organismy, jako jsou viry. Tato krize zdůraznila potřebu včasné a spolehlivé diagnostiky a také rychlého objevu účinných léků proti novým nemocem. Navrhujeme, že obě tyto výzvy lze řešit pomocí deoxyribozymů (katalyticky aktivních molekul DNA).

Nejběžnějšími typy signálů jsou světlo, fluorescence, barva a radioaktivita. Některé z těchto signálů mohou být generovány deoxyribozymy, jiné však nikoli. Funkční molekula DNA, která generuje robustní fluorogenní nebo chromogenní signál, by měla pro mnoho aplikací značné výhody ve srovnání s fluorescenčními proteiny, jejichž syntéza je nákladná a pracná, a fluorescenčními RNA aptamery, které jsou náchylnější k degradaci než molekuly DNA. Abychom rozšířili spektrum možných signálů generovaných pomocí DNA, použili jsme *in vitro* selekci k vývoji deoxyribozymů, které generují silné fluorogenní a chromogenní signály.

Deoxyribozymy, které generují fluorogenní signál, toho dosahují defosforylací kumarinového substrátu 4-methylumbelliferylfosfátu (4-MUP). Přenášejí fosfátovou skupinu z 4-MUP na svou vlastní 5' hydroxylovou skupinu. Touto reakcí se nefluorogenní 4-MUP mění na fluorogenní 4-methylumbelliferon (4-MU). Nejaktivnější deoxyribozym identifikovaný v našich studiích jsme pojmenovali Aurora. Ten zvyšuje fluorescenci více než 700krát a vytváří detekovatelný signál během několika minut. Deoxyribozymy, které produkují žlutou barvu, přenášejí fosfátovou skupinu z bezbarvého substrátu 4-nitrofenylfosfátu (4-NPP) na vlastní 5' hydroxylovou skupinu. Po defosforylaci se 4-NPP přemění na žlutý 4-nitrofenol (4-NP). Tento deoxyribozym jsme pojmenovali Apollon. Apollon zvyšuje chromogenní signál více než 100krát.

Pro zlepšení katalytické aktivity Aurory a Apollona, pro identifikaci jejich minimalizovaných verzí a k řešení jejich sekundárních struktur jsme použili *in vitr*o selekci v kombinaci s vysokokapacitním sekvenováním a komparativní sekvenční analýzou. Dále jsme pro dosažení co nejlepšího poměru signál k šumu podrobně charakterizovali reakční podmínky, včetně vlivu různých jedno- a dvojmocných iontů, různých pufrů, molekulárních zahušťovadel a různých hodnot pH, jakož i různých koncentrací substrátu a deoxyribozymu.

Dále jsme vyvinuli několik způsobů, jak přeměnit Auroru a Apollona na alostericky regulované senzory. Aurora a Apollon byly navrženy tak, aby fungovaly jako

oligonukleotidové senzory pro specifické sekvence a také jako senzory pro různé typy nukleáz v homogenních testech, které nevyžadují promývání nebo následné purifikační kroky. Upravená verze Aurory může detekovat pikomolární koncentrace ribonukleáz a byla použita k rychlé identifikaci nových nízkomolekulárních inhibitorů ribonukleázy Nsp15 ze SARS-CoV-2 ve vysoce kapacitním testování. Naše výsledky ukazují, že deoxyribozymy mohou generovat robustní fluorogenní a chromogenní signály a že takové deoxyribozymy lze použít v aplikovaném výzkumu k řešení aktuálních problémů.

Klíčová slova: deoxyribozym, *in vitro* selekce, senzor, aptazym, fluorescence, 4-methylumbelliferylfosfát, kolorimetrický, chromogenní, 4-nitrofenylfosfát, homogenní test, RNáza A, Nsp15, SARS-CoV-2

List of Abbreviations

4-MUP	4-methylumbelliferyl phosphate
4-MU	4-methylumbelliferon
4-NPP	4-nitrophenyl phosphate
4-NP	4-nitrophenol
A	adenine
AP	alkaline phosphatase
bp	base pair
C	cytosine
CSPD	disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2′-(5′-chloro)tricyclo [3.3.1.1^{3,7}] decan}-4-yl)phenyl phosphate
CDP-Star	disodium 2-chloro-5-(4-methoxyspiro [1,2-dioxetane-3,2′-(5-chlorotricyclo [3.3.1.1$^{3.7}$] decan]) -4-yl]-1-phenyl phosphate
DAB	dabigatran
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSC	differential scanning calorimetry
dsDNA	double-stranded DNA
dNTP	deoxynucleotide triphosphate
FACS	fluorescence-assisted cell sorting
FLAP	fluorescent light-up aptamers
G	guanine
GMP	guanosine monophosphate
GTP	guanosine triphosphate
GFP	green fluorescent protein
HPLC	high performance liquid chromatography
HTS	high-throughput screening
IVS	intervening sequence
ITC	isothermal titration calorimetry
NC	negative control
NGS	next-generation sequencing
NMR	nuclear magnetic resonance

nt	nucleotide
PAGE	polyacrylamide gel electrophoresis
PC	positive control
PCR	polymerase chain RNA
PTP	protein tyrosine phosphatase
RNA	ribonucleic acid
RNase	ribonuclease
SARS-CoV-2	severe acute respirator syndrome coronavirus 2
ssDNA	single-stranded DNA
T	thymine
U	uracil

Table of Contents

1. Introduction - A short story about a long molecule

We can envision DNA as one of the oldest relics of prehistoric life on our planet. This is molecule that is much older than dinosaurs or even eukaryotic cells, which is beyond my imagination. However, it is still here and is still working. This is fascinating! If DNA can do the things it does with its relative chemical simplicity compared to proteins, this suggests that perhaps we do not understand its functional limits, and we might even be able to find new applications for DNA molecules that at first glance seem impossible. To say this in another way, we should never underestimate DNA.

1.1 Once upon a time a Swiss doctor lived in a castle in Thübingen.

Once upon a time a Swiss doctor lived in a castle in Thübingen. This was not exactly once upon a time, but instead the 19th century and the year 1869. This doctor was named Friedrich Miescher and he was analysing a white substance isolated from leucocytes[1]. Because this substance was isolated using a protocol that contained a protease digestion step, it could not be protein. Since Miescher isolated this substance from the cells' nuclei, he named it nuclein, and this name is preserved in today's designation deoxyribonucleic acid.

Friedrich Miescher was a rigorous scientist and he was careful with his extraction protocol and controls during DNA isolation. He checked his starting extraction material (pus from fresh surgical bandages) for any signs of decomposition with a microscope. After nuclei isolation, he stained the samples with iodine solutions to confirm cytoplasm removal. Through elementary composition analysis, he was able to determine that nuclein contained high amounts of phosphorus and that this phosphorus was present in organic bonds rather than in an inorganic state. Based on these observations as well as the behaviour of nuclein during the isolation procedure (it precipitated at acidic pH, could be redissolved by adding base, and did not coagulate when boiled), Miescher was convinced that nuclein was different from all known types of protein and other organic molecules[2,3].

Miescher was also able to isolate nuclein from various tissues such as liver, testes, and kidney, from yeast cells, and from the sperm of salmon, carp, frogs, chicken and bulls. Isolation of large quantities of nuclein from salmon sperm provided an important clue about nuclein function, and Miescher wrote: "If one [...] wants to assume that a single substance is the specific cause of fertilization, than one should undoubtedly first and

foremost consider nuclein.[4]" Although this claim captured the situation with astonishing accuracy, Miescher never concluded that nuclein is the carrier of hereditary information. Moreover, Miescher wrongly determined an atomic weight of 500-600 for nuclein and proposed several incorrect atomic formulas for nuclein. Miescher's view of nuclein was more as a chemical substance that occurs in the nucleus and provides the correct chemical environment for fertilization than as a molecule that encodes hereditary information, in part because he thought it unlikely that the same substance could result in the diversity of different animal species[2,3].

However, Miescher aim was not to discover the molecular basis of hereditary information. His question was, what is the chemical basis of life? And in this aim Friedrich Miescher succeeded.

Figure I_1 The discovery of DNA. (A) Historic photograph of Tübingen castle. The picture was taken around the time when Friedrich Miescher stayed in Tübingen. This photograph was taken by Paul Sinner and is in the Stadtarchiv Tübingen. **(B)** The laboratory where Miescher discovered DNA. This photograph was taken by Paul Sinner and is in the Stadtarchiv Tübingen. **(C)** Photograph of young Johann Friedrich Miescher. **(D)** Glass vial containing nuclein from salmon sperm isolated during Miescher's stay at the University of Basel. This photograph was taken by Alfons Renz and is in the University of Tübingen. Adapted from Dahm 2005 and 2007[2,3].

1.2 Dead mouse, good mouse

Although shortly after its discovery DNA was considered wrongly as a phosphate storage molecule, its possible role in fertilization and in the storage of genetic information was correctly suggested soon afterwards. However, it took decades to finally prove that DNA rather than protein is the carrier of genetic information.

The first successful work in this area was Griffith's experiment in 1928[5]. In this experiment Griffith was working with two strains of *Diplococcus peneumoniae*. The III-S strain is virulent and kills infected mice, while the II-R strain is non-virulent and does not

kill infected mice. When Griffith inactivated the III-S strain by heating and mixed this with the II-R strain, the mixture killed infected mice, but these two components did not kill mice by themselves. Based on this observation, Frederick Griffith postulated a "transforming principle" by which the non-virulent II-R strain was converted into the lethal III-S strain.

The mechanism of the "transforming principle" was explained 16 years later by Oswald Avery, Colin MacLeod, and Maclyn McCarty (Figure I_2c-e) in the famous Avery-MacLeod-McCarty experiment (Figure I_2a)[6]. This was done using the same bacterial strains that were used by Griffith, but instead of just mixing the inactivated type III-S strain with the active type II-R strain, they isolated DNA from the inactivated type III-S strain (Figure I_2b)[7]. They then mixed this DNA from the type III-S strain with the non-lethal type II-R strain bacteria and they have observed transformation into the lethal strain, just like they would be using the crude inactivated bacteria. They observed same results even after they have treated the DNA sample with the proteases or ribonucleases, however when the DNA was treated with deoxyribonuclease, the transforming effect was lost. Despite the fact that proper controls were done, many scientists were still unwilling to accept that DNA, not protein, is the carrier of genetic information.

This scepticism continued until 1952 when Alfred Hershey and Martha Chase (Figure I_2g) helped to confirm that DNA is the genetic material (Figure I_2h)[8]. They used T2 bacteriophages for their experiments, and isotopically labeled bacteriophage DNA with radioactive ^{32}P and bacteriophage protein with radioactive ^{35}S. They then mixed both types of labeled T2 bacteriophages (one labelled with ^{32}P, another with ^{35}S) with unlabeled bacteria. Phage coats should remain on the outside of the bacteria, while the genetic material should enter the bacterial cells. After infection, bacterial cells where washed, centrifuged and lysed. Hershey and Chase were able to detect the signal only from ^{32}P labelled T2 bacteriophages, indicating that only the labeled DNA had entered the bacterial cells. This in turn provided additional evidence that DNA is the genetic material.
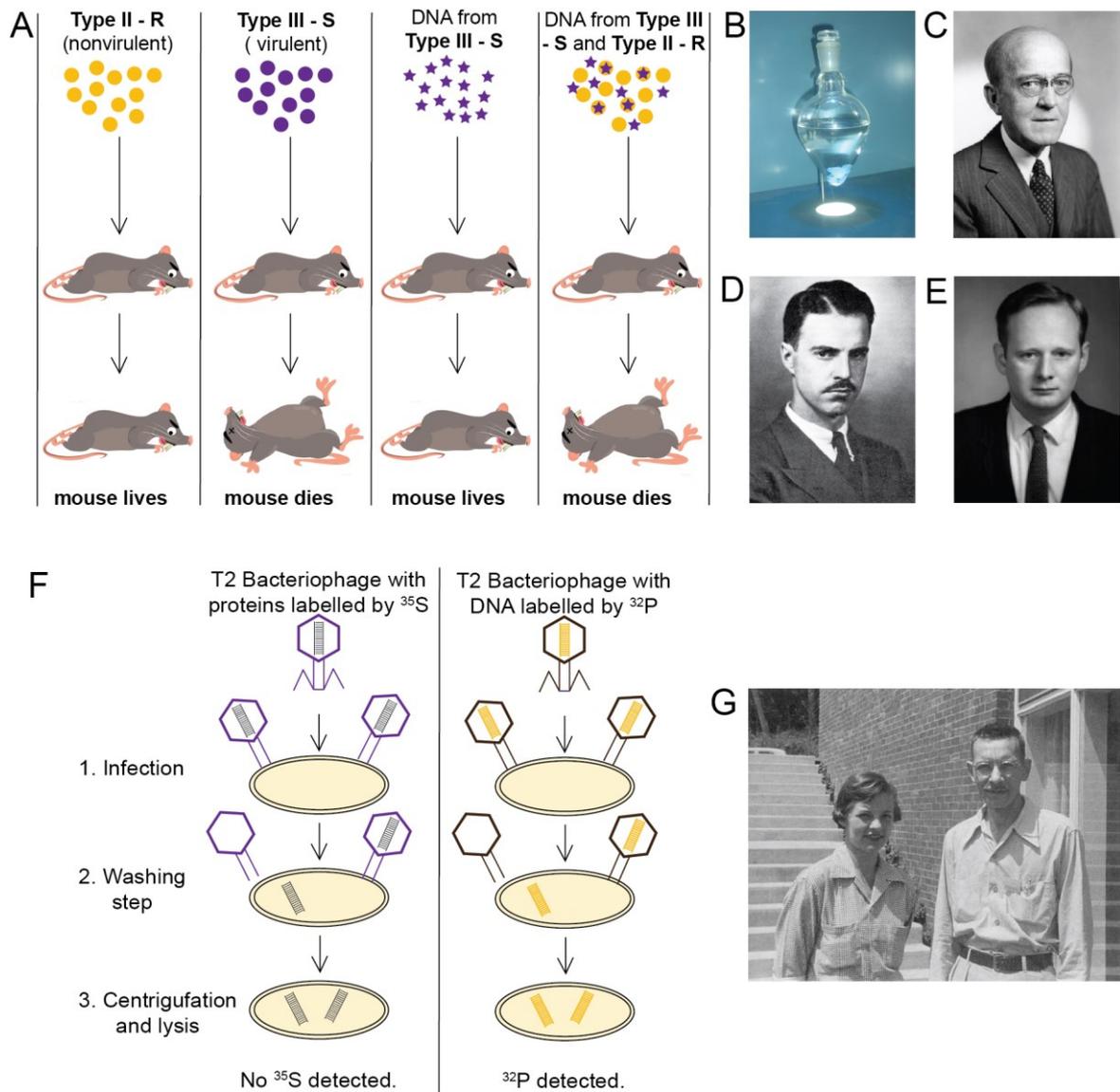
Figure I_2. DNA is the carrier of genetic information. (A) Schematic representation of the Avery-MacLoad-McCarty experiment with Type II-R and Type III-S *Streptococcus pneumoniae* bacteria. **(B)** Photograph of precipitated DNA from solutions of cell components. **(C)** Photograph of Oswald Avery. **(D)** Photograph of Colin MacLeod from around 1940. **(E)** Photograph of Maclyn McCarty. **(F)** Schematic representation of the Hershey-Chase experiment with T2 bacteriophages and selectively labeled proteins or DNA. **(G)** Photograph of Marta Chase (left) and Alfred Hershey (right).

1.3 DNA from different organisms always contains the same amount of adenine and thymine as well as the same amount of guanine and cytosine. Great job gentlemen! We are done.

From the perspective of hindsight, things appear obvious and simple, and we wonder why some discoveries were not made earlier. One such example concerns Erwin Chargaff (Figure I_3b), the excellent biochemist who discovered Chargaff's rules. These rules state

that the amount of guanine in the DNA of an organism should be equal to the amount of cytosine and the amount of adenine should be equal to the amount of thymine (Figure I_3c). This rule is valid for DNA from any organism. If we think about these rules nowadays, we immediately envision the DNA double helix and wonder why Chargaff did not propose such a model.

To put Chargaff's discoveries into context, we have to go back to Friedrich Miescher. His discovery of nuclein caused some interest in the scientific community and there were even attempts to link nuclein with the transfer of hereditary information. However, interest in this idea gradually diminished due to the so-called tetranucleotide hypothesis developed by Phoebus Levene and Hermann Steudel (Figure I_3a)[9]. This hypothesis assumed that DNA was composed of repeating tetranucleotide building blocks of the same sequence, each of which contained guanine, cytosine, adenine, and thymine . This implied that the four bases would always be present in equimolar proportions and the DNA from any organism would always contain the same amount of each of the bases. Moreover, accordingly to this hypothesis, DNA is a polymer composed of just one type of monomeric unit, and therefore not well suited for the encoding and storage of genetic information. From these reasons, many scientists started to focus on proteins, which are composed of more than 20 different amino acids and appeared to be more suitable for encoding complex information.

Erwin Chargaff's initial results were in agreement with this hypothesis. He and his co-workers determined that DNA contained as many purine (A, G) as pyrimidine bases (C, T) and that the molar ratios of the bases A/T and G/C were very close to one[10]. However, they also discovered that the ratios of A/C, A/G, T/C and T/G could deviate significantly from each other. Moreover, the relative proportion of bases in DNA was same in individuals of a single species, but differed in different species[11]. These findings finally proved that the tetranucleotide hypothesis was not correct, and they also suggested that DNA might be able to encode complex information. The work of Erwin Chargaff therefore provided valuable and correct information about DNA's chemical composition, and helped other scientists to solve the structure of DNA.
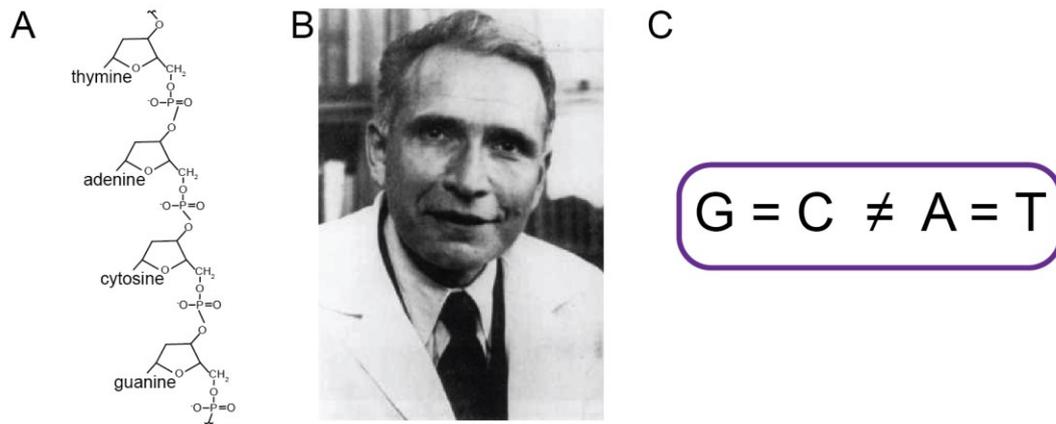
Figure I_3. Chargaff's rules. (A) Repeating tetranucleotide unit as described in the tetranucleotide hypothesis postulated by Phoebus Levene. **(B)** Photograph of Erwin Chargaff. **(C)** Schematic representation of simplified Chargaff's rules.

1.4 Can you please draw something for me, darling? Tomorrow James and I are sending our paper to Nature.

This was probably not what Francis Crick said to his wife Odile Crick before she drew the model of double helical structure of DNA. This drawing was published in Nature in April 1953 and became a symbol of molecular biology (Figure I_4b)[12].

The discovery of the double helical model of DNA came at the right time and individual pieces of previously known information came together perfectly. When James Watson and Francis Crick (Figure I_4a) started working on the structure of DNA, there was strong evidence that DNA stores genetic information. The work of Erwin Chargaff showed that the DNA of organisms contains equal numbers of A/T and G/C bases, but the breakthrough came from X-ray crystallographers who were studying organic macromolecules. Thanks to the excellent X-ray photographs of DNA obtained by Rosalind Franklin (Figure I_4c,d)[13], Watson and Crick were able to develop a model of the DNA double helix that rationalized Chargaff's rules. Watson and Crick recognized the biological implications of their structure. They also correctly predicted how DNA can function and suggested a semi-conservative mechanism of DNA replication[14].

However, many ethical controversies are also associated with this discovery. Rosalind Franklin did not give permission to Watson and Crick to use her X-ray photographs of DNA, but it is not clear if she would have solved the structure of DNA by herself. Moreover, she was not awarded the Nobel Prize in 1962 along with Watson, Crick and Wilkins. Rosalind Franklin died in 1958, but the stipulation against posthumous

awards was not instated until 1974. Something more positive is that her name is now mentioned more often together with Watson and Crick and the importance of her contribution in solving the structure of DNA is appreciated.
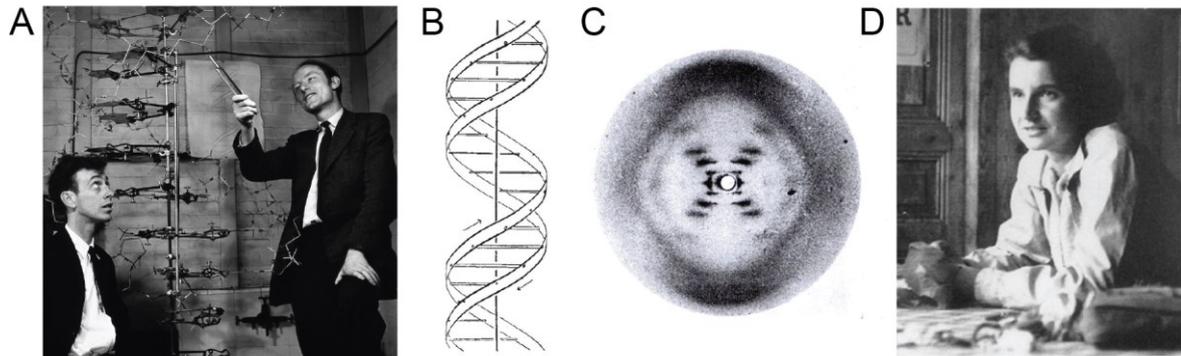
Figure I_4. Molecular structure of nucleic acids. (A) Historic photograph of James Watson (left) and Francis Crick (right) with their DNA model at the Cavendish Laboratories in 1953. This photograph was taken by A. Barrington Brown. **(B)** Schematic drawing of a DNA double helix drawn by Odine Crick. The two ribbons represent the sugar-phosphate backbones of the two DNA chains, and the horizontal rods represent base pairs that hold the chains together. The vertical line marks the fibre axis.**(C)** Franklin's X-ray diagram of the B form of sodium thymonucleate (DNA) fibres. **(D)** Historic photograph of Rosalind Elsie Franklin in Paris in 1949. This photograph was taken by Vittorio Luzzati. Source: Wikipedia Commons. Adapted from Watson and Crick 1953 and Franklin and Gosling 1953[12,13].

1.5 Have you had a drink Matthew? How many lines do you see?

After Watson and Crick solved the double helical structure of DNA, they wrote in their second Nature paper of 1953 "The double helix of antiparallel strands is, in effect a pair of templates, each of which is complementary to the other. We imagine that prior to duplication the hydrogen bonds are broken down and the two chains unwind and separate. Each chain then acts as a template from the formation onto itself of a new companion chain so that eventually we shall have two pairs of chains, where we only had one before.[14]" This hypothesis suggested a semi-conservative mechanism of DNA replication. However, it needed to be confirmed, and this was done by Matthew Meselson and Franklin Stahl in 1958[15] in an experiment which is sometimes called the most beautiful experiment in biology.

If *E. coli* cells are grown only in ^{15}N medium, their DNA become heavier than DNA of *E. coli* cells which are grown in normal ^{14}N medium. This simple observation, in combination with a method to separate heavy and light DNA strands by centrifugation in a CsCl density gradient, revealed the answer to this important biological question.

E. coli cells were first grown in ^{15}N medium. They were then collected and put into ^{14}N medium. Cells divisions were monitored by microscopic cell counting. After one cell division, the DNA appeared in only one band in the CsCl gradient, which indicated an intermediate density and ruled out a conservative replication mechanism in which the entire DNA double helix served as a template for synthesis of completely new DNA double helix. Such a model would instead predict that two bands should appear in the CsCl gradient after one cell division (one daughter DNA molecule would contain only "old" ^{15}N DNA and the other would only contain "new" ^{14}N DNA).

Another model of DNA replication is called the dispersive hypothesis. This was postulated by Max Delbrück[16]. In this model, breaks in the DNA occur every 10 nucleotides during DNA replication. At each break, a new DNA strand is synthetized and attached to the end of the old strand. Such a mechanism of replication would result in a patchwork double helix in which new DNA is mixed with old DNA every 10 nucleotides. If breaks are completely random, DNA would appear only in one band in the CsCl gradient after one round of DNA replication because each daughter DNA molecule will contain 50% old ^{15}N DNA and 50% new ^{14}N DNA. This is consistent with the results of Meselson and Stahl. After two cycles of replication, only one band should still be observed, because each daughter DNA molecule will contain 25% old ^{15}N DNA and 75% new ^{14}N DNA. However, Meselson and Stahl instead observed two distinct bands of approximately the same intensity after two cycles of replication (Figure I_5b). This is instead consistent with the semi-conservative model of replication postulated by Watson and Crick[14] (two daughter DNA molecules will contain 50% old ^{15}N DNA and 50% new ^{14}N DNA while the other two daughter DNA molecules will contain only ^{14}N DNA (Figure I_5a).
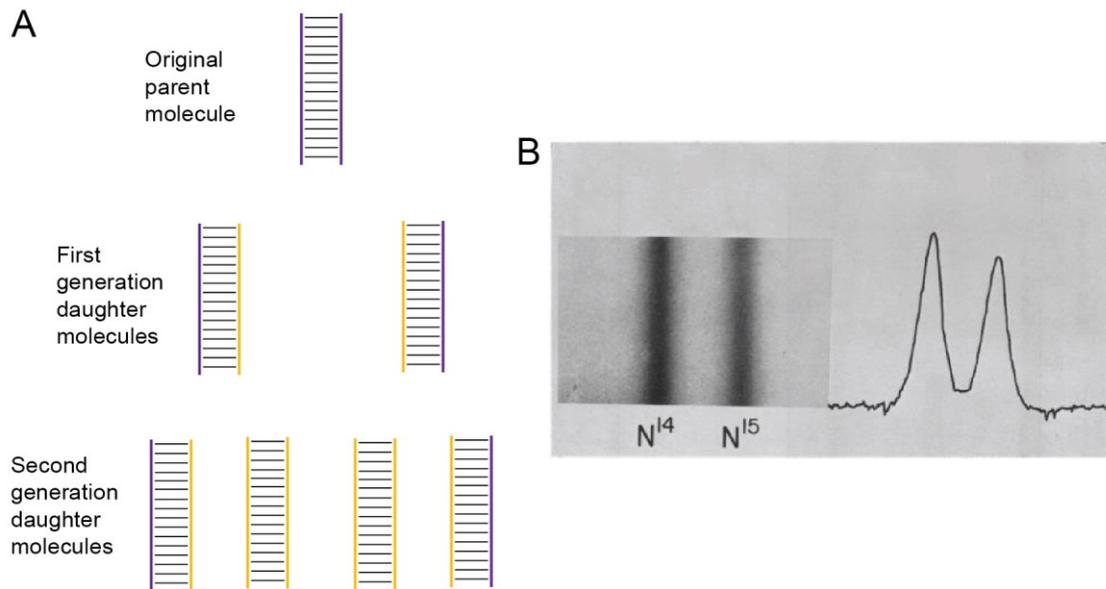
Figure I_5. The Meselson-Stahl experiment. (A) Illustration of the semi-conservative mechanism of DNA replication proposed by Watson and Crick. **(B)** Left: The resolution of ^{14}N DNA from ^{15}N DNA by density-gradient centrifugation. The photograph was taken after 24 hours of centrifugation at 44,770 RPM. Right: A microdensitometer tracing showing the DNA distribution after two generations of *E. coli* cells replication. Adapted from Meselson and Stahl 1958[15].

1.6 Ribozyme? That is a good joke, Dr. Cech.

The work of Thomas Cech can be taken as an excellent example of serendipity in science. When Cech started his own research group he was studying a different topic. He did not believe initial results showing catalytic activity of RNA, and was so determined to find the source of contamination in these reactions that he even mentioned this in his Nobel lecture[17]. But after accepting this discovery, he and his research team were able to identify and describe the mechanism of Group I self-splicing introns. Moreover, he was able to persuade the scientific community that his results were correct, even though the existing dogma was that all enzymes are proteins.

When Thomas Cech (Figure I_6a) was setting up his own research group in Boulder in 1978, he wanted to be better understand gene expression. He therefore turned his attention to the rDNA (gene for the large ribosomal RNAs) of the protozoan *Tetrahymena*. The genes encoding rRNA in *Tetrahymena* are located on small DNA molecules in the nucleoli, and they are transcribed at exceptionally high levels[18]. The plan was to isolate this gene along with its associated structural proteins as well as proteins that regulated its transcription.

The presence of an intervening sequence (IVS), or intron, which interrupted the rRNA-coding sequences of the rDNA in some strains of *Tetrahymena*[19] was not interesting to Cech at that time. However, when they were trying to prepare the rRNA in a crude cell-free system and separated the products of this *in vitro* transcription by gel electrophoresis, Cech and co-workers observed a discrete low molecular weight product (Figure I_6b, marked 0.4kb). Further analysis revealed that this small RNA product corresponded to the intervening sequence[20,21]. This was an exciting result, because each *Tetrahymena* cell contained thousands of rDNA genes and each produced unspliced pre-rRNA. Cech therefore reasoned that the nuclei might contain a high concentration of the splicing enzyme and he wanted to isolate it.

However, isolation of this splicing enzyme never happened, because after purification of unspliced pre-rRNA and incubation of this pre-rRNA in a solution containing salts and nucleotides, the same small RNA product corresponding to the IVS was observed on the gel. As Cech mentioned in his Nobel lecture, he said to his co-worker: "Well, Art, this looks very encouraging, except you must have made some mistake making up the control sample." However several careful repetitions of the experiment confirmed that the release of the IVS occurred independently of the addition of any enzyme. Further analysis of the minimum components necessary for the release of the IVS revealed that GTP was required at micromolar concentrations. This GTP molecule was transferred to the 5' end of the spliced IVS, and formed a normal 3'-5' phosphodiester bond. Addition of GTP was stoichiometric and the triphosphate was unnecessary, as GMP and even guanosine could be used in place of GTP. These findings led to the discovery of the self-splicing mechanism of Group 1 introns. Addition of guanosine to the 5' splice site occurs through a transesterification reaction. This frees the 5' exon and creates a new 3' hydroxyl group at its 3' end. A second transesterification reaction then occurs and the 3' hydroxyl group at the 3' end of the 5' exon attacks a phosphodiester bond at the 3' splice site. Thus two exons are joined together in a reaction catalyzed entirely by RNA (Figure I_6c)[22].

Even after all of this work, Cech was not fully convinced that splicing of the IVS occurs without the help of protein. He speculated that splicing might be catalysed by a protein tightly bound (or even covalently bonded) to the pre-rRNA isolated from *Tetrahymena* nuclei. The best way to rule out this possibility was to synthesize the RNA in as artificial a manner as possible. For this reason, a bacterial plasmid encoding the IVS with the flanking rRNA sequences was prepared. The plasmid was grown in *E. coli* and the

RNA was transcribed by *E. coli* RNA polymerase under conditions that were inhibitory for splicing. RNA was then purified by gel electrophoresis and upon addition of GTP, $MgCl_2$ and salt, the IVS was released, providing proof that the self-splicing RNA was not dependent on any protein from *Tetrahymena* or any protein in general[23].

As Thomas Cech wrote, the term ribozyme was coined as follows. "We held a relatively subdued celebration in the lab. Between sips of champagne we compiled a list of possible general names for RNA molecules able to lower the activation energy for specific biochemical reactions. It was then that we coined the term ribozyme, for a ribonucleic acid with enzyme-like properties.[17]"
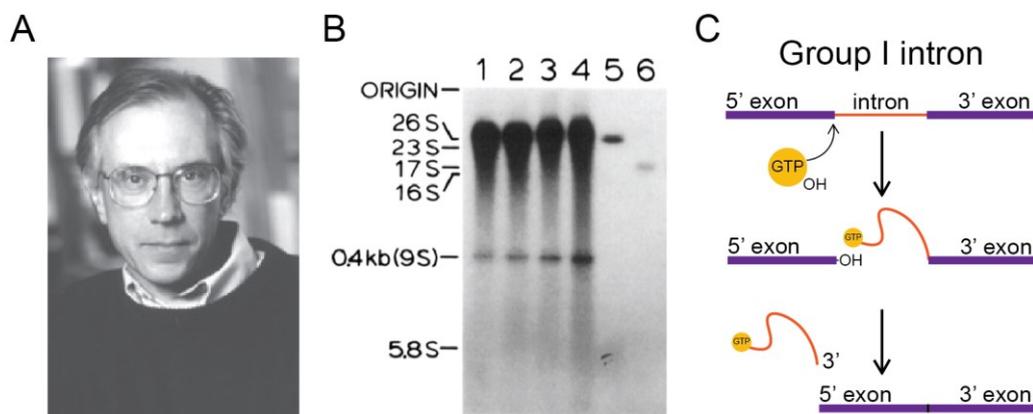
Figure I_6. The discovery of catalytic RNA. (A) Historic photograph of Thomas Cech. **(B)** *In vitro* transcription and splicing of pre-rRNA in isolated *Tetrahymena termophila* nuclei. (Lanes 1-4) RNA produced by incubation of nuclei at 30 °C for times from 5 to 60 minutes. The 0.4 kb product is the spliced IVS RNA. (Lanes 5 and 6) Molecular standards. **(C)** Self-splicing of the Group I intron from the pre-rRNA catalysed by by two consecutive transesterification reactions. Adapted from Zaug and Cech 1980[20].

1.7 A nucleic acid chemist, a mass-spectrometer manufacturer, and a rocket engineer meet in bar...

This sounds like the beginning of a bad joke, but it is in fact a description of three teams that worked together to provide new information about the possible origin of life on our planet.

In 2014, the European satellite Philae landed on the surface of comet 67P/Churyumov-Gerasimenko. The Philae satellite was equipped with a mass spectrometer and was able to send to Earth data about the chemical composition of our solar system 4.6 billion years ago[24]. Knowledge of this chemical composition provided valuable information about whether it is possible to synthesize the basic building blocks of life from

these chemical substances and gave scientists new clues to evaluate the possibility of the RNA world hypothesis.

The RNA world hypothesis was first described by Walter Gilbert[25] and it was inspired by Cech and Altman's discoveries of catalytic RNA function. Gilbert hypothesized that a primordial ribozyme encoded genotype and phenotype in a single molecule. Moreover, excision and reverse insertion reactions catalyzed by self-splicing introns enable genetic recombination, which is a much more powerful technique for information exchange than random mutagenesis and makes it possible for a useful exon to move from one replicating structure to another.

According to Gilbert, the first stage of evolution consisted of RNA molecules that could assemble themselves from a nucleotide mixture. In the next stage, RNA molecules started to use RNA cofactors such as nicotinamide adenine dinucleotide and flavin mononucleotide, and they began to develop new enzymatic activities. RNA molecules then started to use activated amino acids and short peptides as cofactors, and in this way the first proteins were formed. Gilbert suggested that modern protein enzymes do not catalyse different enzymatic reactions than primordial RNA molecules, but because they are better catalysts, they now dominate modern biology. At some point DNA appeared on the scene. It replaced RNA as the main storage molecule for genetic information because it is more stable. At this point RNA no longer played the central role it once had. It was displaced from its catalytic role by more effective protein enzymes and from its role as a storage molecule by more stable DNA. The role of RNA then became a messenger between DNA and proteins. However, it is possible that we still observe some remnants of the RNA world in the core of ribosomes, in self-splicing introns, in the spliceosome, and in various RNA modifications and nucleotide cofactors.

In regard to the RNA world hypothesis it is also worth mentioning the work of Professor Phil Holliger from Cambridge University. He and his team isolated by *in vitro* selection a polymerase ribozyme, which was efficient enough to transcribe another active ribozyme[26]. The emergence of such a ribozyme would have been a critical event in the origin of life, because such polymerase ribozyme could then replicate not only itself, but also a primordial RNA "genome". A polymerase previously developed in the Bartel group[27] was used as a starting point for these efforts, which was in turn based on a ligase ribozyme that was originally developed by David Bartel in the Szostak group[28].

Although the RNA world hypothesis was accepted by some scientists, others have proposed an alternative theory called the "peptide-RNA world". In this theory, simple peptides and RNA molecules co-evolved from the very beginning. Interesting results supporting this hypothesis were obtained by professor Thomas Carell and his team from Ludwig Maximilians University from München. They recently demonstrated that RNA, together with the help of non-canonical nucleosides which are today found in transfer and ribosomal RNAs and which are considered to be relics of the RNA world, was able to generate RNA-peptide chimeras. These molecules raise the possibility of an RNA-peptide world from which ribosomal peptide synthesis could have emerged[29,30]. Moreover, if peptides grow on RNA molecules constructed from D-ribose, this will lead to the self-selection of homo-L-peptides. These results provided a possible explanation for the homo-D-ribose and homo-L-amino acid combination seen in nature and hypothetically explained why the world in which we live is homochiral[31].

Thomas Carell and his team also used data from the satellite Philae about the chemical composition of comet 67P/Churyumov-Gerasimenko. They demonstrated that it is possible to form the purine and pyrimidine nucleosides from organic compounds present on this comet under prebiotically reasonable conditions. These data suggest that prebiotically plausible chemical pathways may have created the key molecules of life on the early Earth[32].

1.8 Give me more, give me more, give me one more round.

It often happens that discoveries are made at different places at roughly the same time. This can happen when knowledge in a specific field has reached the point at which the next step is logical and easily deducible.

This happened in year 1990 when three different research groups reported the discovery of functional RNA motifs using the process of artificial evolution. This technique is also referred to as directed evolution, *in vitro* selection, and SELEX (systematic evolution of ligands by exponential enrichment). This was possible because all of the steps needed to perform artificial evolution in the lab had been worked out. During the 1980s, Thomas Cech and Sidney Altman discovered the catalytic activity of RNA[17,33]. In 1970, Howard Temin and David Baltimore discovered the reverse transcriptase polymerase from retroviruses[34,35]. Efficient solid-phase synthesis of synthetic DNA was

described by Serge Beaucege and Marv Caruthers in 1981[36]. In 1986 Kary Mullis developed a way to amplify DNA called the polymerase chain reaction (PCR)[37] and the production of large amounts of RNA using a synthetic DNA template was possible thanks to the recombinant T7 RNA polymerase, described by John Miligan in 1987[38].

The first of these three initial papers was published by Debra Robertson and Gerald Joyce (Figure I_7a) in March of 1990[39]. They used *in vitro* selection to change the substrate specificity of the self-splicing intron I from *Tetrahymena* described by Thomas Cech and his team[20]. After two rounds of *in vitro* selection, ribozymes were identified that preferentially cleaved a single-stranded DNA compared to the initial ribozyme, which cleaved RNA.

The second paper was published by Craig Tuerk and Larry Gold (Figure I_7b) in August 1990[40]. They used a method they named SELEX to isolate RNA molecules that noncovalently bind to T4 DNA polymerase. To do this, they randomized an 8 nucleotide long loop in the mRNA of T4 DNA polymerase, and selected for RNA sequences that efficiently bind to T4 DNA polymerase. Two different sequences were selected from a pool of 65,536 possible sequences. One was the wild-type sequence and the second was a sequence with four mutations relative to the wild-type sequence. The binding affinities of these two RNA sequences toward T4 DNA polymerase were equivalent.

The third paper was published by Andrew Ellington and Jack Szostak (Figure I_7c) in late August 1990[41]. They used *in vitro* selection to isolate RNA molecules that bind to a variety of organic dyes. They named these RNA molecules aptamers (from Latin "apta", to fit). Ellington and Szostak used a completely random 100 nucleotide-long library flanked by constant primer-binding sites. Their experimental setup was basically *in vitro* selection as it is used today. From the beginning, Ellington and Szostak pushed the limits in terms of library size (10^{15} to 10^{16} molecules is still considered as limit today, as larger libraries are difficult to handle in the laboratory) and library length (it is still generally difficult to synthetize DNA molecules larger than 100 nucleotides in high yields by solid-phase synthesis).
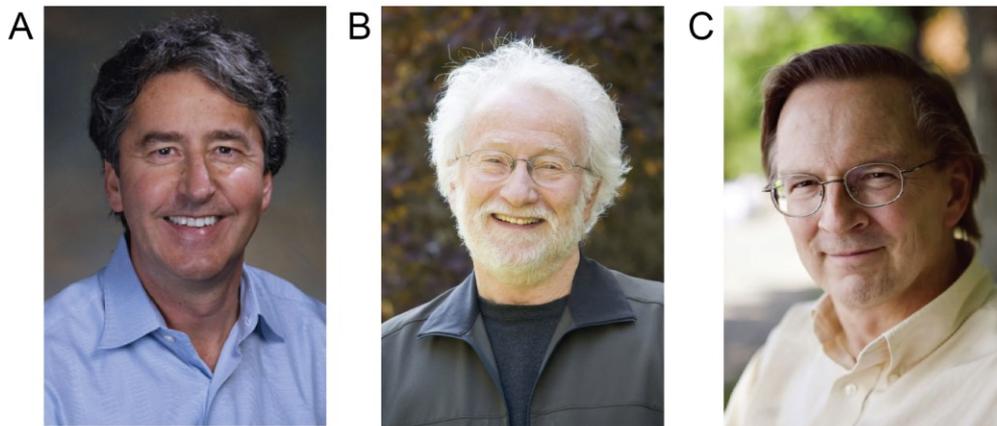
Figure I_7. The holy trinity of *in vitro* selection. (A) Photograph of Gerald Joyce. This photo is owned by BioMedical Graphics. **(B)** Photograph of Larry Gold. This photo is owned by Cold Spring Harbor. **(C)** Photograph of Jack Szostak. This photo is owned by Encyclopedia Britannica.

The method of *in vitro* selection relies on similar processes as Darwinian evolution, and is therefore sometimes also called artificial evolution. RNA or DNA molecules are subject to a selective pressure, and only molecules with a desired function survive. Because DNA can be amplified by PCR, and because it is possible to convert DNA to RNA using RNA polymerase and RNA to DNA using reverse transcriptase, nucleic acid selections are much easier than protein selections. This makes *in vitro* selection such a powerful and useful method.

The first step of an *in vitro* selection experiment is the incubation step. The library is incubated with the substrate or target molecule and sequences are allowed to react. In the next step, the catalytically active sequences are separated from the inactive ones. This isolation process usually requires modification of active sequences, so they can be distinguished from the inactive sequences. One such modificasvehlotion can be the ligation of short oligonucleotide to reacted library members[42,43]. Active sequences are longer and therefore can by isolated by gel electrophoresis. A similar strategy can be used to isolate sequences that cleave RNA or DNA[44-47]. Active sequences are shorter and can also be isolated by gel electrophoresis. Another type of isolation is based on the formation of a complex between an active sequence and a target molecule. If the target molecule is marked with an affinity tag, the complex can be isolated using an appropriately coated matrix such as biotin/streptavidin[48] or thiophosphate/thiopropyl sepharose[49]. Such complexes can also be isolated using native gel electrophoresis or capillary electrophoresis[50].

A significant disadvantage of selection strategies that rely on self-modification is that selected sequences can only catalyse the desired reaction in a single turnover manner. A common strategy to solve this problem is to divide the active sequence into two parts (a substrate strand and an enzyme strand)[51]. Active sequence can then function in a *trans*-acting manner and catalyse multiple turnover reactions. However, this strategy is not applicable to all ribozymes and deoxyribozymes. A better solution would be to directly select for sequences, which are capable of multiple-turnover reactions. However, in this case the active sequences are not modified and it is therefore impossible to separate them from the inactive ones. Possible solutions can be found in microfluidic systems. In this type of experiment, library members are enclosed in droplets in water-in-oil emulsion in such a way that in one droplet contains multiple copies of a single sequence along with an excess of substrate molecules. After incubation, individual droplets can be sorted by fluorescence-activated cell sorting (FACS). If the sequence catalyses a multiple turnover reaction, the signal in a droplet will be stronger than in one containing a molecules than only catalyses a single turnover reaction. This technique was successfully used by the Ellington lab to find multiple turnover ribozymes for RNA ligation[52] and by the Mayer lab to find an RNA motif that enables optical control and light-dependent gene expression in bacteria and mammalian cells[53].

After the selection step, active sequences need to be amplified. Here the *in vitro* selection protocol differs for RNA and DNA molecules. RNA molecules are first reverse transcribed into complementary DNA (cDNA) and then amplified by PCR. However, PCR amplification is not always necessary for RNA molecules, because the cDNA is in the next step of selection (library recovery) transcribed back into RNA by RNA-polymerase and in this step the RNA molecules are also amplified. If the amplification by RNA polymerase is sufficient, the PCR step can be omitted. In contrast, in the case of DNA, PCR is the only method of amplification. Double stranded DNA molecules from PCR can be converted into single stranded ones by degradation of complementary strands by lambda exonuclease (if the reverse primer in PCR was 5' phosphorylated) or by strand separation using streptavidin coated magnetic beads (if the reverse primer in PCR was 5' biotinylated).

Once a signal is detected, the selection is usually stopped after it reaches a plateau. The evolved library is then amplified by PCR and characterised by Sanger sequencing or next generation sequencing.
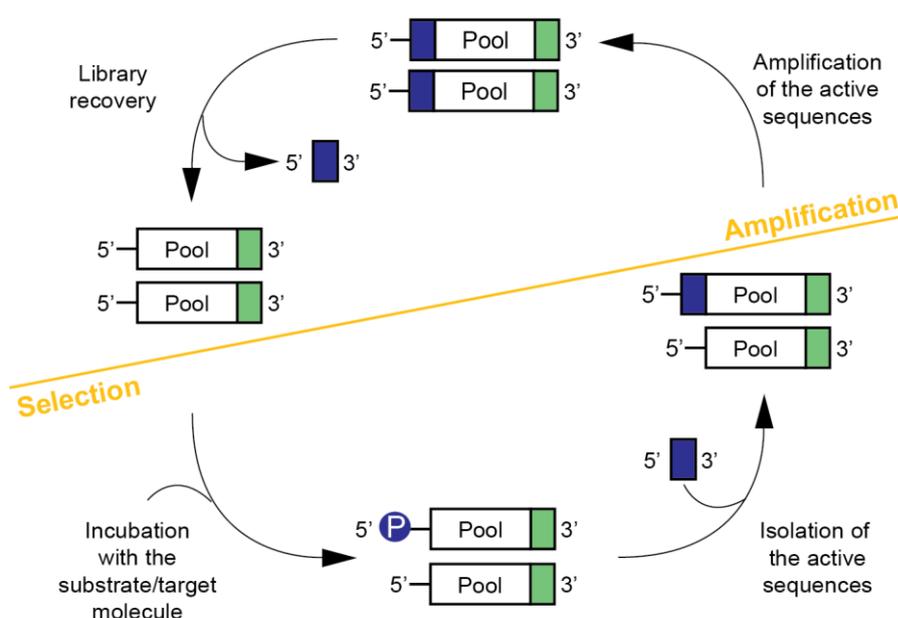
Figure I_8. The general scheme of an *in vitro* selection experiment. The starting library is incubated with the substrate or target molecule. Active sequences are then isolated using various techniques such as gel electrophoresis, capillary electrophoresis, or streptavidin-coated magnetic beads. If the library is RNA, sequences are then transcribed into complementary DNA by reverse transcriptase. In next step, sequences are amplified by PCR using constant regions as primer binding sites. The library is then prepared for the next round by strand separation in the case of DNA or transcription in the case of RNA. This cycle is repeated until the desired activity is observed.

Once a signal is detected, the next step is to sequence the evolved library. Until recently this was done by amplifying the library by PCR, cloning individual sequences into plasmid, transforming the plasmids into *Escherichia coli,* and sequencing by Sanger sequencing[51,54]. Tens to hundreds of sequences were typically obtained using this approach. However, evolved selection libraries typically contain thousands to hundreds of thousands of enriched sequences. Sequencing coverage by Sanger sequencing was therefore insufficient, although it could provide information about the most abundant sequences (and therefore theoretically the most active ones).

More recently, next generation sequencing methods have been developed. These are typically commercially available and have become the preferred method[42,55]. NGS methods have many advantages compared to Sanger sequencing. Several million sequences are typically obtained, which can provide a great deal of additional information relative to Sanger sequencing. Moreover, it is possible to sequence each round of an *in vitro* selection experiment and therefore identify active sequences based on round-to-round enrichments rather than read numbers. This can help to identify improved variants that

arise by mutation during a selection but have low read numbers relative to sequences that were present in the starting library. NGS analysis can identify such sequences and several studies have shown that sequences with the highest read numbers are not always the most active sequences in a library[45,56].

NGS analysis also facillitates secondary structure determination. Secondary structure of aptamers, ribozymes or deoxyribozymes are sometimes determined using programs like Mfold[57] or ViennaRNA[58]. These programs predict the secondary structure based on thermodynamic parameters, but are not able to search for more complex structures like G-quadruplexes, triple helices and other non-Watson-Crick interactions. In addition, they often generate lists of very different structures with similar predicted stabilities, and cannot account for stabilizing interactions formed by interactions between aptamers and ligands. However, thanks to the high sequence numbers in NGS datasets it is possible to use a method called comparative sequence analysis, which does not rely on thermodynamic calculations. Instead, this method uses large datasets of sequences likely to adopt the same secondary structure. Such datasets can be generated by randomly mutagenizing a single variant of a functional motif and performing a second selection to identify the active variants. Identification of pairs of positions that change in a correlated way (called covariations) is a powerful way to identify base pairs in such datasets. These can be further tested by mutagensis. For example, if single mutations at two different positions abolish activity, but combining these mutations does not, it is likely that these positions interact in some way. This method was used to predict the secondary structure of 16S-like ribosomal RNA[59]. Moreover, if the number of sequences with the same secondary structure in the NGS dataset is sufficient, it is possible to generate sequence logos in which the level of conservation of each individual nucleotide can be observed. This can help to identify the minimal sequence needed for function[42].

The final step in an *in vitro* selection experiment is to test individual sequences for activity. However, it is not unusual to obtain inactive sequences even after many rounds of selection. This can be explained because of various types of selection biases such as those related to PCR amplification.

1.9 And what is it good for?

Functional nucleic acids were discovered less than 50 years ago and the method of *in vitro* selection was discovered 34 years ago. The high therapeutic and biotechnological potential of functional nucleic acids led to the development of many different aptamers, ribozymes, and deoxyribozymes. Here are a few examples of molecules, which could be especially useful or are close to being approved for clinical practice.

1.9.1 The good old thrombin aptamer

The thrombin aptamer was isolated in 1992 and was the first DNA aptamer isolated by SELEX[60]. This blocks the protease activity of thrombin, an enzyme in the blood-coagulation cascade, and therefore prevents blood clothing. The minimal core of this aptamer is 15-nucleotides long, and because of its high stability, the crystal structure of the aptamer-thrombin complex was published only a year after its discovery in 1993[61]. This revealed that the thrombin aptamer folds into a stable G-quadruplex structure containing two stacked tetrads.

The thrombin aptamer has high therapeutic potential, because heparin, the currently most used anticoagulant drug, can cause thrombocytopenia (clotting of platelets in the presence of heparin) in some patients. Therefore, it is good to have alternative options. A thrombin aptamer variant named NU172[62] has a higher affinity for thrombin then the original thrombin aptamer. NU172 was submitted for clinical trials by ARCA Biopharma, Inc. (ClinicalTrials.gov identifier NCT00808964) and should be used as a short-acting drug for anticoagulation in surgeries such as coronary artery bypass. NU172 entered Phase II clinical trials. However, despite promising results, it has not been approval for clinical use so far.

A study in Nature Communications published this year by Professor Bruce Sullenger[63] and his team used different approach to improve the affinity of the thrombin aptamer. They found inspiration in hematophagous organisms, which are able to selectively inhibit structurally homologous proteases of coagulation. These organisms achieve this by producing peptide inhibitors which target both the exosites and active sites of specific proteases. For this reason, Sullenger and his team named these aptamers EXACT inhibitors (EXosite-binding aptamers with small molecule ACTive site inhibitors). One of their EXACT inhibitors, HD22-7A-DAB, has extraordinary anticoagulation activity. HD22-7A-DAB is a combination of the thrombin aptamer HD22[64], a 7 nucleotide

long linker and the inhibitor dabigatran (DAB)[65]. The thrombin aptamer HD22 binds to the anion binding exosite 2 of thrombin with high affinity and selectivity, but has weak anticoagulant activity. On the other hand, DAB binds to the thrombin active site, but also has inhibitory activity towards other proteases. DAB has a K_i of 4.5 nM and HD22 is only able to inhibit thrombin activity to a level of 30%. Strikingly, the combination of HD22-7A-DAB has an IC_{50} of 0.95 nM, which is a 500-fold improvement compared to DAB. Moreover, at a 1 nM concentration, thrombin activity is completely blocked. Therefore, simply conjugating a surface binding aptamer with an active site binding inhibitor can have numerous therapeutic applications.

1.9.1 I will isolate hundreds of new aptamers. Is this enough, Larry?

Larry Gold is one of the inventors of *in vitro* selection. Ten years after his initial discovery, he founded a company called SomaLogic, Inc.. The main goal of this company is to develop aptamers against every protein in the human body. The main reason for this decision is that the human genome has around 19,000 genes and thanks to the alternative splicing more then 100,000 human proteins exist, but the majority of publications focus on a few hundred of the most popular and studied proteins. Other interesting and possibly very biologically important proteins can escape our notice just because of missing molecular tools (antibodies) to study them.

Aptamers from SomaLogic are called SOMAmers, because they contain unnatural and modified nucleotides to prevent degradation and to increase chemical variability. Using these SOMAmers, SomaLogic developed the SomaScan platform, which allows the identification and quantification of protein biomarkers for different diseases. In 2010 it was possible using the SomaScan platform to identify and quantify 813 proteins simultaneously[66]. Currently, the commercially available version of the SomaScan platform is able to identify and quantify up to 11,000 proteins.

One example of how SOMAmers can be used comes from the work of Dr. Xiaotang Lu from the Lichtman lab at Harvard University. She and her co-workers have used SOMAmers, which specifically bind to the different cell types in mouse brain. These SOMAmers were fluorescently labelled and enabled detergent-free immunolabelling of brain cells on ultra-structurally reconstructable electron micrographs. Thanks to this labelling, the three-dimensional reconstitution of the brain tissue was possible with all neural circuits between different types of brain cells[67].

1.9.3 Do you want Corn or Broccoli? No thank you, I already have Mango.

In vitro selection was also used to identify RNA and DNA molecules that generate different types of signals. These molecules can be used as signalling components in different biochemical assays.

The most studied group of such signalling sequences are fluorogen-activating RNA aptamers. These are also referred to as Fluorescent Light-Up Aptamers (FLAPs). The first aptamer from this group was named Spinach[68] and mimics the green fluorescent protein (GFP). Spinach binds to the difluorinated analogue of the GFP chromophore, named DFHI, and the emission wavelength and quantum yield resemble enhanced GFP. Many different analogues of Spinach have been isolated, such as Broccoli[69], Corn[70], Chili[71], Pepper[72], Squash[73], and Mango[74]. The mechanisms of these aptamers are similar. Each RNA aptamer binds a small molecule, which is intrinsically non-fluorescent. After binding the aptamer, the conformational flexibility of the ligand is reduced, which enhances its fluorescence.

Fluorogen-activating DNA aptamers, which work using the same type of mechanism, have also been described. These are less common than RNA aptamers. One example binds to a nonfluorescent Hoeschst derivative, which light up after binding[75]. The signal to noise ratio of this aptamer is slightly over 100. Another well known DNA molecule that generates signals is a peroxidase deoxyribozyme isolated by the Sen group[76]. This deoxyribozyme converts the substrate ABTS into a colorimetric product in the presence of hydrogen peroxidase and hemin[76,77]. This deoxyribozyme usually generate a chromogenic product, but a fluorogenic or light signals can also by produced when this reaction is performed with substrate such as tyramine[78] or luminol[79].

Fluorogen-activating RNA aptamers have many practical applications[80–82]. For example, the RNA aptamer Spinach was converted into a sensor for different cellular metabolites. This was achieved by combining a ligand-binding RNA aptamer and the Spinach aptamer. Such sensors were used to detect intracellular levels of adenosine 5-diphosphate or S-adenosylmethionine in *E. coli* cells[83]. Another RNA aptamer, called Mango[74], was isolated by Professor Peter Unrau and his team at Simon Fraser University in Vancouver. Mango binds to derivatives of thiazole orange and it was used for subcellular localization of small non-coding RNAs (5S, U6, and a box C/D scaRNA) in fixed and live mammalian cells[84]. More recently, an improved Mango variant was used to

image RNA molecules (coding β-actin mRNA or long non-coding NEAT1 RNA) in live cells with high contrast at single-molecule density[85]. These examples demonstrate that the Mango aptamer is promising tool for RNA biology.

The fluorogen-activating RNA aptamers are useful for *in vivo* applications. However, for *in vitro* applications such as high-throughput screens or diagnostic applications, where the stability of the signal and overall cost per reaction plays an important role, DNA counterparts will be more useful. Unfortunately, current signalling molecules made of DNA suffer from high background signal and can require harsh reaction conditions which are not compatible with biological samples[76]. Signal to noise ratios are also often low[75]. Recently a light producing deoxyribozyme called Supernova was isolated in our laboratory, providing a useful addition to the toolbox of signalling nucleic acids[42]. However, additional deoxyribozymes that produce other types of signals could also be useful. The development of such deoxyribozymes is the topic of this thesis.

2. Aims

The aims of the PhD project were: to isolate deoxyribozymes which produce fluorogenic or chromogenic signals, to characterize these deoxyribozymes, and to convert them into allosterically regulated sensors.

The PhD project specifically focused on:

1. Using *in vitro* selection to isolate deoxyribozymes, which produce fluorogenic and chromogenic signals.
2. Identifying the minimal catalytic cores of these deoxyribozymes.
3. Elucidating the secondary structures of these deoxyribozymes.
4. Optimizing reaction conditions to achieve the best possible signal-to-noise ratios.
5. Converting these deoxyribozymes into allosterically regulated sensors, which only produce signals in the presence of a target.

3. Results

3.1 Pool design and selection of fluorogenic and chromogenic substrates

One goal in our laboratory is to isolate deoxyribozyme, which can produce different types of signals and be used as signalling components in various types of assays. The first such deoxyribozyme isolated in our laboratory was Hit 1, which was isolated by *in vitro* selection from a library containing a random sequence region of 70 nucleotides (Figure R_1a). Hit 1 was able to produce light in the presence of the 1,2-dioxetane substrate CDP-Star. During this chemiluminescent reaction Hit 1 transfers the phosphate group from the CDP-Star to its own 5'-hydroxyl group. After dephosphorylation, CDP-Star undergoes a spontaneous decomposition, during which light is emitted.

The initial *in vitro* selection experiments that yielded Hit 1 and its improved variant Supernova were done by Kateřina Švehlová (a former PhD student in the Edward Curtis group). These results are summarized in the articles "Supernova: a deoxyribozyme that catalyses a chemiluminescent reaction[42]", "Optimizing the chemiluminescence of a light-producing deoxyribozyme[86]" and in her doctoral thesis "Development and characterization of light-producing deoxyribozymes.[87]". Hit 1 showed weak catalytic activity in the presence of the fluorogenic substrate 4-methylumbelliferyl phosphate (4-MUP) and the chromogenic substrate 4-nitrophenyl phosphate (4-NPP) (Figure R_1b-c). To obtain variants of Hit 1 that react more efficiently with 4-MUP and 4-NPP, we decided to do additional *in vitro* selection experiments. A new library based on Hit 1 was designed in which the sequence of Hit 1 was randomly mutagenized at a rate of 21% per position. A new reverse primer binding site was also added to the 3' end of Hit1 (Figure R_1d, see Pool 1 in Figure 1).

Figure R_1. Pool 1 (used for the initial *in vitro* selection experiments described in this thesis) is based on Hit 1, which showed weak activity in the presence of some fluorogenic and chromogenic substrates. **(A)** Workflow used to identify Hit 1 (a chemiluminescent deoxyribozyme which produces light in the presence of CDP-Star) **(B)** Chemical structures of the chemiluminescent substrate CDP-Star, the fluorogenic substrate 4-methylumbellifery phosphate (4-MUP), and the chromogenic substrates 4-nitrophenol phosphate (4-NPP) and naphthol AS-TR phosphate (NAP) **(C)** Catalytic activity of Hit 1 in the presence of CDP-Star, 4-MUP, 4-NPP, and NAP measured using the ligation assay. NC - negative control, PC - positive control **(D)** Sequence of Hit1.

4-methylumbelliferyl phosphate (4-MUP) is non-fluorogenic substance. After dephosphorylation, 4-methylumbelliferone (4-MU) is formed. 4-MU is a fluorogenic molecule with an excitation wavelength of 358 nm and an emission wavelength of 455 nm (Figure R_2a)[88]. 4-nitrophenyl phosphate (4-NPP) is a colorless substance, which is converted into a yellow substance after dephosphorylation. 4-nitrophenol (4-NP) has a maximal absorption at 405 nm (Figure R_2b)[89]. Both substances are cheap and commercially available and are widely used for detection and biochemical characterization of various protein phosphatases such as alkaline phosphatase (AP)[90] and protein tyrosine phosphatase (PTP)[91] in ELISA assays. They both show good potential for further

applications. The fluorogenic 4-MU has better sensitivity compared to the chromogenic 4-NP. Therefore, it is more suitable for assays in which targets are at low concentrations. However the chromogenic 4-NP can be monitored without specialised equipment, or in the simplest case by eye, and this can be useful for field work in which where specialised equipment is not available.

Phosphoryl transfer reactions catalysed by DNA or RNA molecules were previously described by several different groups[43,92]. Moreover, our starting library (Pool 1) was based on Hit 1, which already catalysed a phosphoryl transfer reaction using CDP-Star and showed weak activity in the presence of fluorogenic and chromogenic substrates (Figure R_1c). These preliminary results greatly increased our chances of successfully isolating deoxyribozymes that generate robust fluorogenic and chromogenic signals (Figure R_2).
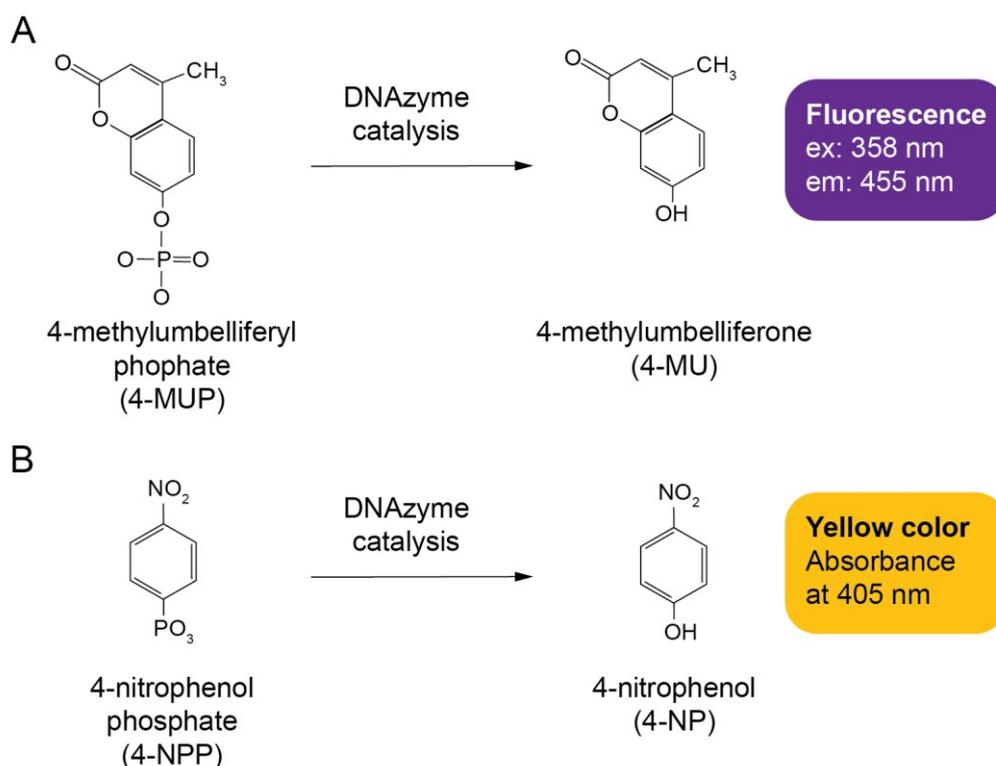
Figure R_2. Reaction scheme for production of fluorogenic and chromogenic products. (A) After dephosphorylation of 4-MUP a fluorogenic product 4-MU is produced with an excitation wavelength of 358 nm and an emission wavelength of 455 nm. **(B)** After dephosphorylation of 4-NPP, the chromogenic yellow product 4-NP with absorbance at 405 nm is formed.

3.2 *In vitro* selection protocol

For isolation of deoxyribozymes that generate fluorogenic and chromogenic signals, we used an indirect selection strategy. It would be desirable to select deoxyribozymes directly based on their ability to generate fluorescence or color, but we are not yet able to do such selections in our laboratory. For this reason, we based our selection strategy on phosphorylation of the 5' end of active DNA molecules (Figure R_3a).

Our starting library Pool 1 was made by 21% random mutagenesis of the 85-nucleotide long Hit 1 (see Table 1). This resulted in a pool of $1,74*10^{14}$ single-stranded DNA molecules. This pool was incubated in selection buffer for 2.4 hours with either 1 mM 4-MUP (in the case of the selection for deoxyribozymes expected to generate fluorogenic products) or 1 mM 4-NPP for deoxyribozymes (in the case of the selection for for deoxyribozymes expected to generate colorimetric products). During the incubation active DNA molecules should transfer the phosphate group from 4-MUP or 4-NPP to their own 5'-hydroxyl group. Active DNA molecules were then ligated to a 20-nucleotide long oligonucleotide, and because this made them 20 nucleotides longer, it was possible to separate them from inactive ones by 6% denaturing polyacrylamide gel electrophoresis (PAGE). DNA molecules which, corresponded to the length of 125 nucleotides (105 nucleotides from the original Pool 1 + 20 nucleotides from the ligated oligonucleotide) were cut out and eluted from the PAGE gel. They were then amplified by PCR. The 20 nucleotide-long oligonucleotide ligated to reacted pool members and the constant 20 nucleotide-long 3'end of Pool 1 served as primer binding sites. Forward primer (FWD1 see Table 1) contained a ribonucleotide at its 3' end and the reverse primer (REV1p see Table 1) was 5' phosphorylated. After PCR the double-stranded 125 nucleotide-long product was treated by λ-exonuclease, which degraded the 5' phosphorylated reverse complement strand. The ribonucleotide at the 3' end of the forward primer made it possible to remove this primer by base hydrolysis and to regenerate the single-stranded 105 nucleotide-long Pool 1 for subsequent rounds of *in vitro* selection (Figure R_3b).

In each round of *in vitro* selection there were two different enrichment steps used to capture catalytically active DNA molecules. Firstly, only 5' phosphorylated DNA molecules will undergo ligation. Secondly, only DNA molecules which contain the 20 nucleotide long donor oligonucleotide ligated to their 5' end in the correct orientation will be amplified by PCR.

Figure R_3. *In vitro* **selection protocol.** **(A)** *In vitro* selection protocol to identify deoxyribozymes that phosphorylate themselves in the presence of 4-MUP or 4-NPP. **(B)** All of the steps that a catalytically active DNA molecule has to undergo during one round of *in vitro* selection protocol.

3.3 Initial *in vitro* selection and reselection

The workflow we used to isolate deoxyribozymes that generate fluorogenic and chromogenic products is represented in Figure R_4a. The initial selection for deoxyribozymes that generate a fluorogenic product (Figure R_4b) was called the Aurora initial selection. After four rounds of selection activity was detected, and after one more round the library was characterized by high-throughput sequencing. If we compare the library before and after initial selection (Figure R_6a) significant enrichment of specific sequences can be observed, which is an indication that the selection was successful. We obtained 3,997,122 reads and 42,556 unique sequences. The most abundant sequence, called Hit 1, had a read number of 219,442. This corresponded to 5.49 % of the total reads (Figure R_6b-c). We tested the top ten sequences (Figure R_6c) from the Aurora initial selection for both self-phosphorylation activity and fluorescence production (Figure R_5a-b). Surprisingly the most active sequence in both assays was Hit 10.

Unfortunately, we were not able to determine the minimal catalytic core and secondary structure of Hit 10 from the Aurora initial selection. We also wanted to search the sequence space surrounding Hit 10 to identify variants with better catalytic properties. Therefore, we decided to do a reselection experiment. The 85 nucleotide-long sequence of Hit 10 was randomly mutagenized at 21% per position, and new reverse primer binding site was design to minimize the possibility of cross contamination with the original Pool 1 (see Pool 2 in Table 1). The incubation time in this reselection was decreased tenfold from 2.4 hours to 0.24 hour. During the Aurora reselection, Pool 2 was subjected to six rounds of selection (Figure R_4c) and then characterized by high-throughput sequencing. Aurora reselection library showed significant enrichment of sequences after reselection (Figure R_6d). We obtained 6,599,688 reads with 849,535 unique sequences. The most abundant sequence, called rHit 1, had a read number of 25,022. This corresponded to 0.38% of the total number of reads (Figure R_6e-f). We tested the top ten sequences (Figure R_6f) from the Aurora reselection for both self-phosphorylation activity and fluorescence production (Figure R_5c-d). The most active sequence in fluorescence production was rHit 5. Further analysis showed that each of the top ten sequences shared the same catalytic core.
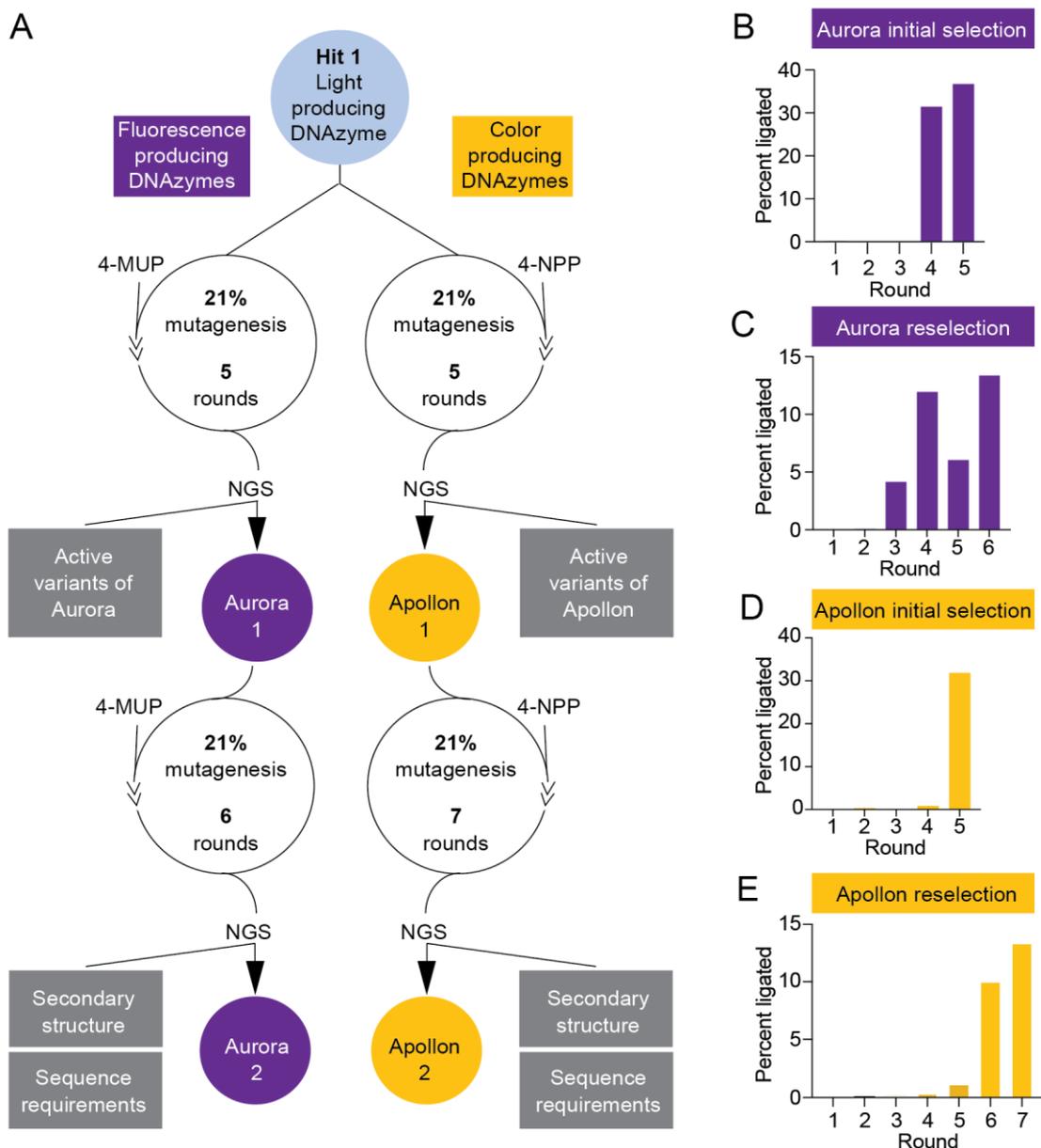
Figure R_4. Discovery of deoxyribozymes that react with the fluorogenic substrate 4-MUP or the colorimetric substrate 4-NPP. (A) Workflow used to identify deoxyribozymes that generate fluorogenic or colorimetric products. **(B-E)** Graph showing percent of ligated molecules during the initial *in vitro* selections and reselection experiments that yielded Aurora and Apollon.

For the Apollon initial selection, Pool 1 had to undergo four rounds of *in vitro* selection before activity was observed, and after one more round the library was sent for high-throughput sequencing. A number of sequences were highly enriched (Figure R_7a, c). For example, the top sequence occupied enormous 78.5 % of the all reads (Figure R_7b). We tested the top five sequences from the Apollon initial selection for both self-phosphorylation activity and color production (Figure R_5e-f). As expected, Hit 1 was the most active sequence in both assays.

For the same reasons that we did an Aurora reselection, we decided to also do an Apollon reselection. Hit 1 from the Apollon initial selection was randomly mutagenized at a rate of 21% per position and new reverse primer binding site was added to minimize the possibility of cross contamination with the original selection library (see Pool 3 in Table 1). The incubation time was also reduced from 2.4 hours to 0.24 hour. Activity was observed after four rounds of selection, and after three more rounds Pool 3 was submitted for high-throughput sequencing (Figure R_4e). A number of sequences were highly enriched (Figure R_7d). The top sequence, named rHit1, had a read number of 3,137,978, which corresponded to 55.5% of all reads (Figure R_7e-f). The top ten sequences from the Apollon reselection were tested for both self-phosphorylation activity and color production. rHit 3 had highest activity in the self-phosphorylation assay (Figure R_5g), while all tested hits showed similar activity for color production (Figure R_5h) and shared the same catalytic core.
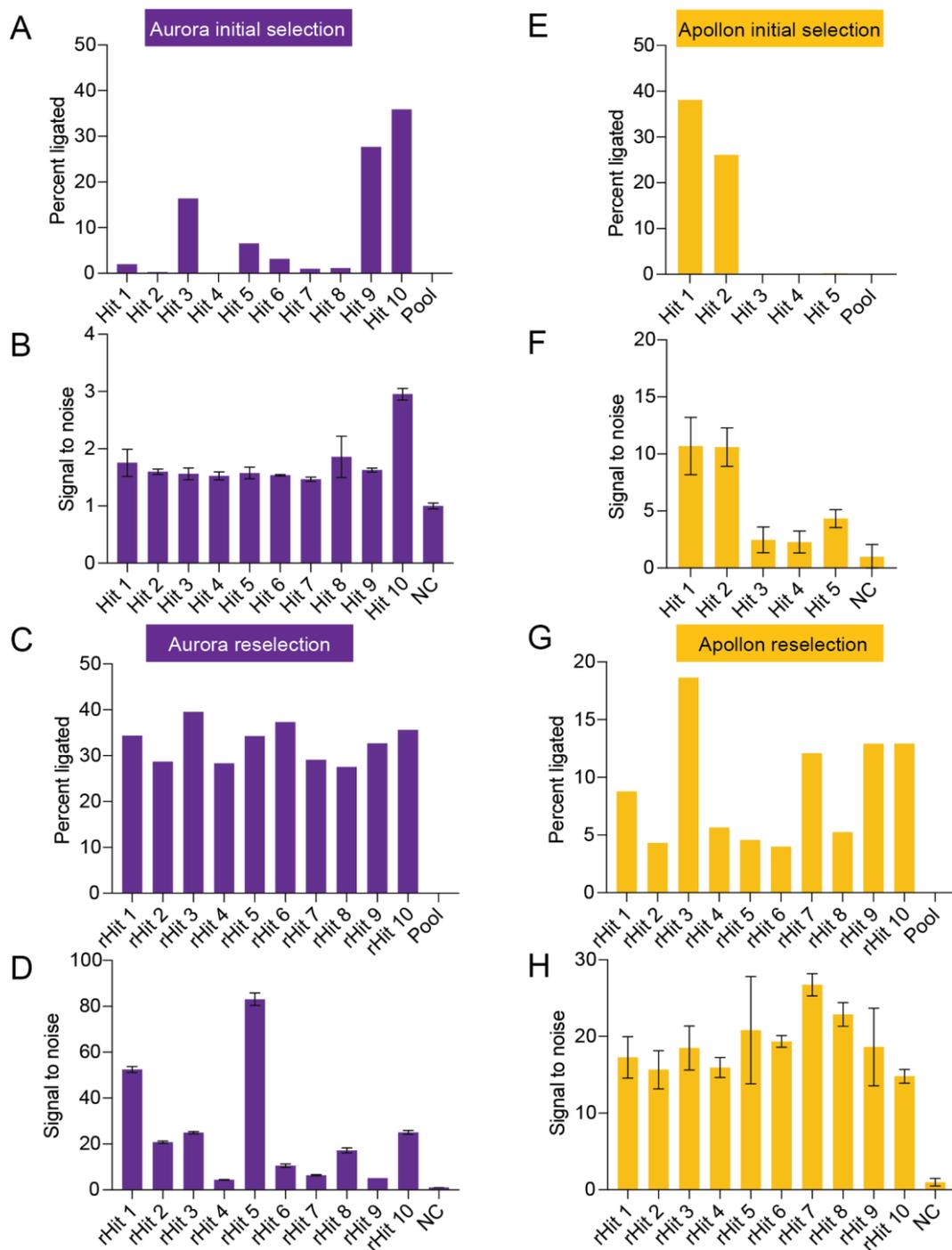
Figure R_5. Catalytic activity of the most abundant Hits from initial selections and reselections. (A, C) Testing of the 10 most abundant Hits from the Aurora initial selection and the Aurora reselection for self-phosphorylation activity. Percent ligated was determined after incubation of 1 µM DNA with 1 mM 4-MUP for 1 hour in 1x Selection buffer **(B, D)** Testing of the same sequences as in panels A and C for fluorescence production. Reactions were performed using 15 µM DNA and 30 µM 4-MUP in 1x Selection buffer. After incubating for 4 hours, 20 µl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(E, G)** Testing the 5 most abundant Hits from the Apollon initial selection and the 10 most abundant Hits from the Apollon reselection for self-phosphorylation activity. Hits were incubated at 1 µM DNA with 1 mM 4-NPP for 1 hour in 1x Selection buffer. **(F, H)** Testing of the same sequences as in panels E and G for color production. Reactions were performed using 30 µM DNA and 100 µM 4-NPP in 1x Selection buffer. After incubating for 4 hours, absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader.

Aurora initial *in vitro* selection

A

B

Hit 1 - 5.49 %
Hit 2 - 3.27 %
Hit 3 - 3.27 %
Hit 4 - 2.77 %
Hit 5 - 2.65 %
Hit 6 - 2.60 %
Hit 7 - 2.54 %
Hit 8 - 2.50 %
Hit 9 - 2.21 %
Hit10 - 2.15 %
Other seq.

C

Name	Read number	Sequence 5'-3'
Hit 1	219,442	GGAAGAGATGGCTACTACCAAGCTACGTGGCCTAATATCATCATGGCCCCGGACCTCAAGAAGTAAGACGTTGGAATGGCGCATG
Hit 2	130,734	GGAGGAGATGAAGACGTCCGCGGCCCAAGCTCGTCTATGCTCAATGCGATCTGCTCCTAGGTCGATTTAGTTGCGACGCGTGCGC
Hit 3	130,647	GGAAGAGAGCGTGCCGTCAGACAGATGAAACTGAAACACCTCAGTGTGAAGTGCTGCGGTGGGCTGTGACCCCCTATGACGGCTC
Hit 4	110,797	GGAAGAGATGTAGACAAGGCAGGCCCCCATTCCTTGCATACTTAGTGCCTAGGACCCCGTGGGGCAGTAACTAGGGATGGAGTTGC
Hit 5	105,729	GGTAGAGAGGACGACGGCCCCGGAAGATTTCTGAATATCCTCCGTGCGCTTTATCGCAGGGGGGATTTAGTTTCGATGGAGCGTC
Hit 6	103,892	GGAAGATAATGCGAGGAGGCAGGTCCCATACCAGAGCGTCTAAGGTCGCACTACCGAATGGTCGTGAGACTCTGGATGGAGTTGC
Hit 7	101,515	GGAAGAGACAGGGAGGAGATAGGCCCCATGCAGTACCTCCCAGATGCGACGATCCACAGGGTGGTACGATATCGCATGGAGCTAT
Hit 8	100,109	GGAAGAGCCGGCGATGAAATAGAGACGAGTTATAATGGTCCAATTCCGCAGGTCCGCCGGGCGGTCTGACTGAGGAGAGCGCATT
Hit 9	88,221	GGAAAAGAGCGCGACAATATTCGTACGATGCATGTACACGCCAGTCCGTAGACCCAAGTGGGGGAGTGTCTTGGGCTGGGGGCAA
Hit 10	85,990	GGAAGAGATGACCAGGGCAGCGGGACGCTGACGAATTTTCTCACTATGTCCGGGACCCGAGGGGCGTGAGGAGTGTTGTGCAATT

Aurora reselection

D

E

rHit 1 - 0.38%
rHit 2 - 0.37%
rHit 3 - 0.20%
rHit 4 - 0.13%
rHit 5 - 0.13%
Other seq.

F

Name	Read number	Sequence 5'-3'
rHit 1	25,022	GGAAGGGATGAGCAGAGTAGCGCGACGATGACGAATTTTGTAACTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGTGCAATT
rHit 2	25,379	GGAAGGGATGAGCAAGGCAGCATGATGCTGACGAATCTAATTACTATGTCCGGGCCCCTATGGGCTTGAGGAGTGTTGTTTAATT
rHit 3	13,326	GGAAGGGATGAACAGGGCAGCAGGATGCTGCAGATTTTTCTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTGTTGTGAAATT
rHit 4	8,839	GGAAGGGATGCGTAAGTAAGCGGGACGCTGACGCATCTTCTTACTATATCCGGGTCCCGAAGGGCGTGAGGAGTGTTGTAAGAGT
rHit 5	8,414	GGAAGGGATGGCAAAGGCAGCGGGACGCTGGCGAAATGTCTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTGTTGTGATAAG

Figure R_6. Statistics from high-throughput sequencing datasets generated in Aurora initial selection and reselection. (A) Distribution of read numbers (expressed in reads per million) in the initial library (Pool1 in Table 1) before and after the Aurora initial selection. **(B)** Chart showing the percent of total reads of the ten most abundant sequences obtained in the Aurora initial selection. **(C)** Name, read number, and sequence of the ten most abundant deoxyribozymes obtained in the Aurora initial selection. **(D)** Distribution of read numbers (expressed in reads per million) in the Aurora reselection library (Pool2 in Table 1) after the Aurora reselection. **(E)** Chart showing the percent of total reads of the five most abundant sequences obtained in the Aurora reselection. **(F)** Name, read number, and sequence of the five most abundant deoxyribozymes obtained in the Aurora reselection.

Apollon initial *in vitro* selection

A

B

0.4%   7.5%

13.6%

78.5%

- Hit 1
- Hit 2
- Hit 3
- Other seq.

C

Name	Read number	Sequence 5'-3'
Hit 1	3,448,950	GGCAGAGACGGCGGCGTCATAGGGACAGTGCTGTACATGCTCACCCTCAGGGGCAATCGGGGGGCGTGAACTTATATGGAGTTAC
Hit 2	598,238	GGAAGATGAGGCTACAACTCAGCGCCCATACGGTATATACTAATTGCTCAAGGTCATATGAGATACTGGCTTCCGCTATCGCATA
Hit 3	19,589	GGTAGACATGGCAACGCCACTGTCACTAAACTGAGCGTGATCAGTAAGCATGCTCATAGGTGAGGGTGTTTTTGGGTGGCTGGTA

Apollon reselection

D

E

26.7%

2.2%

7.1%

8.5%

55.5%

- rHit 1
- rHit 2
- rHit 3
- rHit 4
- Other seq.

F

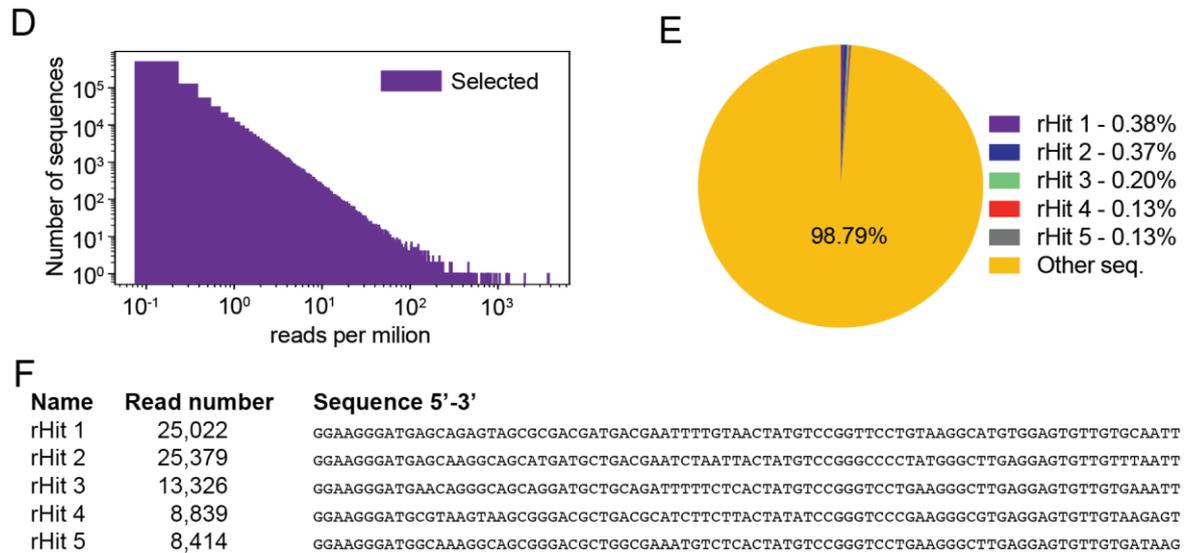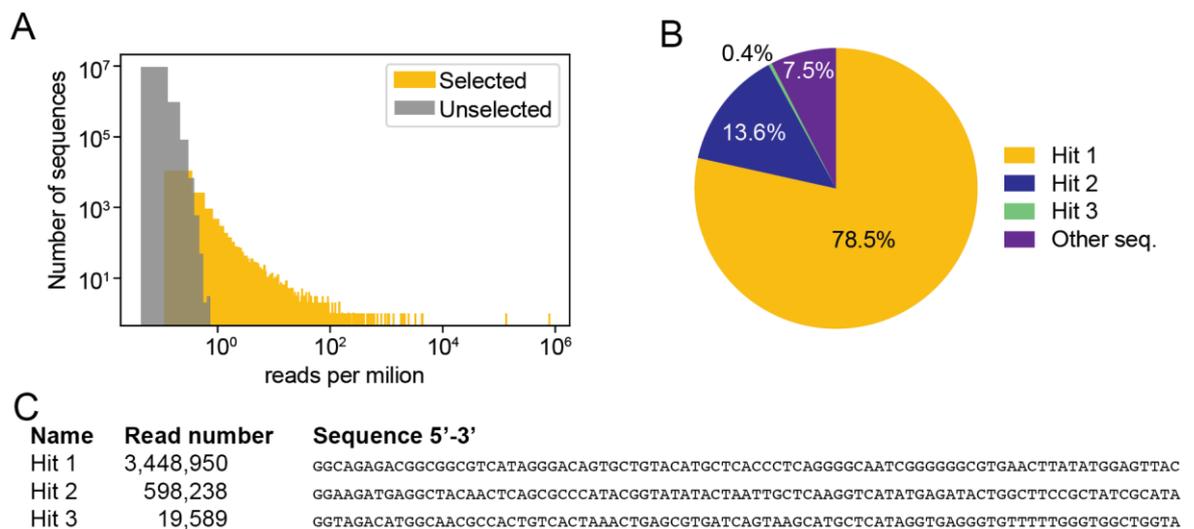| Name	Read number	Sequence 5'-3'
rHit 1	3,137,978	GGCAGAGATGGCGCTGTCATAGGGACAGTGCTACGCATGCTCACCCTCTGGGGCAATCGGGGGGCTTCGACTTGTATAGAGTTAG
rHit 2	480,794	GGCAGAGATGGCGGTGTCATGTTGACAGTGCATTGCATGCTCCCCCTCAGGGGCAATCGGGGGGTGAGAGCTAGTATGGAGTTAA
rHit 3	399,558	GGCAGAGATGGCGGTGTCATAGGTACACTGCTATGCATGCTCACCCTCTGGGGCAATCGGGGGGCATTAACATATATTGATTTAC
rHit 4	126,714	GGCAGAGACGGCGGCGTCCTAGGGACAGTGCTGTACATGCTCCCCCTCAGGGGCAATCGGGGGGTGTAAACTGCGATGTTATTAG

Figure R_7. Statistics from high-throughput sequencing datasets generated in the Apollon initial selection and reselection. (A) Distribution of read numbers (expressed in reads per million) in the initial library (Pool1 in Table 1) before and after the Apollon initial selection. **(B)** Chart showing the percent of total reads of the three most abundant sequences obtained in the Apollon initial selection. **(C)** Name, read number, and sequence of the three most abundant deoxyribozymes obtained in the Apollon initial selection. **(D)** Distribution of read numbers (expressed in reads per million) in the Apollon reselection library (Pool3 in Table 1) after the Apollon reselection. **(E)** Chart showing the percent of total reads of the four most abundant sequences obtained in the Apollon reselection. **(F)** Name, read number, and sequence of the four most abundant deoxyribozymes obtained in the Apollon reselection.

3.4 Aurora and Apollon have different folds than Supernova

Our original expectation was that, if we selected for deoxyribozymes that react with 4-MUP or 4-NPP using a library based on Supernova (Pool 1, Table 1), we would obtain deoxyribozymes with the same secondary structure as Supernova, but with changed substrate binding pockets. However, most of the deoxyribozymes we isolated appeared to be structurally and functionally distinct from Supernova. We initially compared the mutational distances of sequences from all three reselection experiments (Figure R_8a-d). When a selection was previously performed with Pool1 to identify variants that use CDP-Star with improved efficiency[42], the average mutational distance of sequences in the evolved pool from Supernova was around 18 (Figure R_8d, green peak). In contrast, the average mutational distance of variants in reselection libraries that used 4-MUP or 4-NPP compared to Supernova was around 38 (Figure R_8d, yellow and dark blue peaks), suggesting that deoxyribozymes that use CDP-Star, 4-MUP and 4-NPP form different structures. Similar mutational distances were observed when evolved reselection libraries were compared to Apollon or Aurora (Figure R_8b-c).

Analysis of individual sequences (Supernova, Aurora 2 and Apollon 2) from evolved libraries provided additional support for this idea: Aurora 2 and Apollon 2 contained mutations that were not consistent with the sequence requirements of Supernova (Figure R_8e). The substrate specificities of these new deoxyribozymes were also orthogonal to Supernova. For example, the deoxyribozyme Aurora 2 reacted with 4-MUP but not CDP-Star or 4-NPP, whereas Apollon 2 reacted with 4-NPP, but not with CDP-Star or 4-MUP. On the other hand, Supernova reacted with CDP-Star, but not with 4-MUP or 4-NPP under these conditions (Figure R_8f).

These results suggest that, despite being isolated from a library based on Supernova (Pool 1, Table 1), most deoxyribozymes in the initial and reselection libraries that use 4-MUP or 4-NPP as a substrate form structures that are distinct from Supernova.
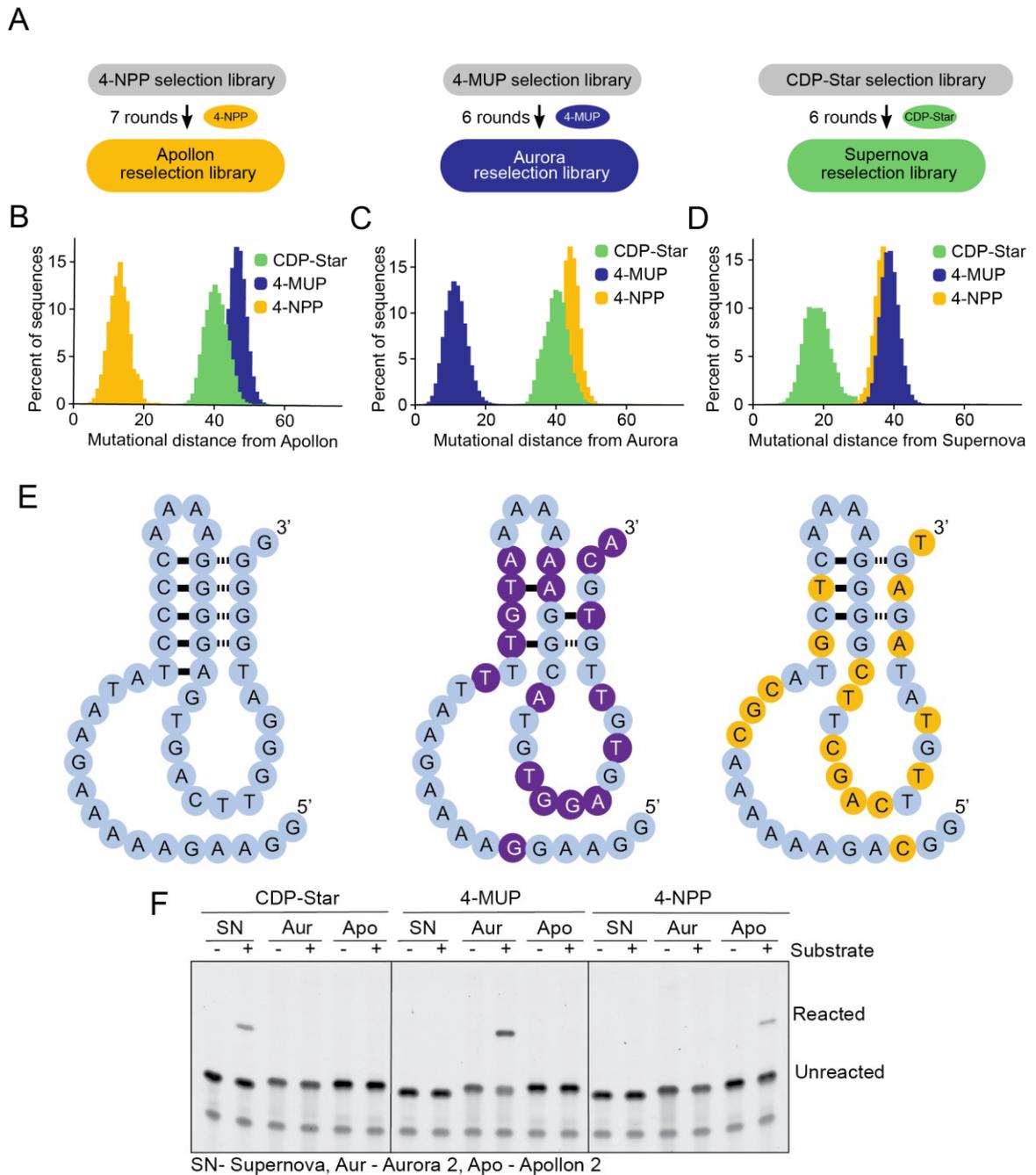
Figure R_8. Aurora and Apollon cannot form the secondary structure of Supernova. (A) Workflow of previous selections in which deoxyribozymes that react with 4-NPP, 4-MUP or CDP-Star were isolated from libraries in which a single variant of Apollon, Aurora or Supernova was randomly mutagenized at a rate of 21% per position. **(B)** Distribution of mutational distances of sequences in a library of Apollon variants relative to Apollon after selection for the ability to react with 4-NPP. **(C)** Distribution of mutational distances of sequences in a library of Aurora variants relative to Aurora after selection for the ability to react with 4-MUP. **(D)** Distribution of mutational distances of sequences in a library of Supernova variants relative to Supernova after selection for the ability to react with CDP-Star. **(E)** Left: secondary structure of the catalytic core of Supernova. Middle: sequence of the homologous positions in Aurora 2 or Right: sequence of the homologous positions in Apollon 2 (corresponding to nucleotides 1-6, 33-42, and 61-82) mapped onto the secondary structure of Supernova. Positions shown in purple or yellow are different from the sequence of Supernova. **(F)** Catalytic activity of Supernova (SN), Aurora 2 (Aur) and Apollon 2 (Apo) in the presence of CDP-Star, 4-MUP, and 4-NPP measured using the ligation assay.

3.5 Determining the minimal catalytic core and secondary structure

3.5.1 The catalytic core of Aurora is a 47-nucleotide long bulged hairpin

Initial analysis of sequences from Aurora reselection high-throughput sequencing dataset revealed two highly conserved regions (corresponding to nucleotides 1-10 and 43-79) separated by 32 less conserved positions (Figure R_9a). The catalytic activity of a 47 nucleotide-long deoxyribozyme in which the nucleotides at positions 11-42, 80-85, and in the 3'binding site were deleted (called Aurora 2 core) was higher than of the full-length sequence (called Aurora 2 full, Figure R_9b-c).
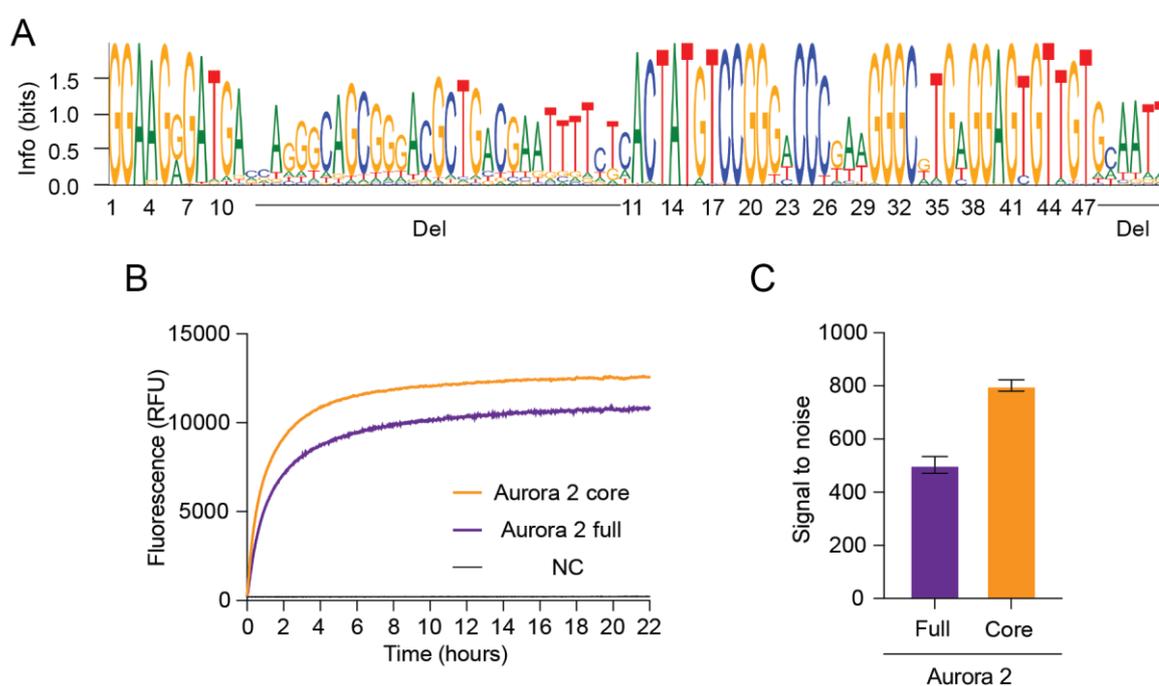
Figure R_9. Sequence requirements and catalytic activity of full-length and minimized Aurora. (A) Sequence logo generated from analysis of variants of Aurora using the high-throughput sequencing dataset from the Aurora reselection. **(B)** Fluorescence production of full length and minimized versions of Aurora 2. Reactions were performed using 15 μM DNA and 30 μM 4-MUP in 1x Aurora buffer containing 200 mM KCl, 1 mM $ZnCl_2$, 5% (v/v) DMSO, and 50 mM HEPES, pH 7.4. Fluorescence was measured for 24 hours using a TECAN Spark plate reader. **(C)** Signal to noise ratio of full-length and minimized versions of Aurora 2. Reactions were performed as in panel B, but after incubation for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader.

The proton NMR spectrum of the 17C40G of Aurora 2 suggests that Aurora forms a structure containing multiple Watson-Crick base pairs (Figure R_10g). Consistent with this observation, comparative sequence analysis[93–95] revealed four pairs of covarying positions in the deoxyribozyme (11-47, 12-46, 17-40, and 26-30) with mutational patterns

that were consistent with base pairs. Mutagenesis experiments in which these putative base pairs were disrupted by point mutations and restored by compensatory mutations provided strong additional support for the proposed interactions (Figure R_10b-f). When we also considered the conserved nucleotides that flank these confirmed base pairs, we concluded that Aurora likely forms 11-base pair hairpin interrupted by an asymmetric bulge (Figure R_10a). This hairpin contains two irregular features: a TT mismatch (positions 13-45, which could reflect noncanonical interaction) and a highly conserved bulged guanine (position 41). The hairpin is capped by a three-nucleotide loop formed by position 27, 28, and 29. Several correlations identified by mutual information analysis including 22-29, 22-27, 23-26, 26-29, and 22-30 (Figure R_10a and Figure R_11) suggest that this loop interacts with the conserved asymmetric bulge formed by positions 20-23 and 33-37 rather than extending into solution. Covariation analysis suggests that nucleotides at the 5' end of Aurora do not form canonical base pairs with one another or with the rest of the deoxyribozyme. However, a network of correlations among positions 4, 6, and 45 (including an AT to GA covariation between positions 4 and 45 which is one of the strongest in the dataset) is consistent with a tertiary interaction that anchors the 5' end of Aurora (which is the reaction site) to the rest of the catalytic core. One possibility is that this interaction helps to position the nucleophilic 5'hydroxyl group in the vicinity of the phosphate group of 4-MUP. Several additional correlations including the 16-34, 6-10, 6-46, 6-12, 6-9, 16-37, and 6-42 (Figure R_11) are consistent with the hypothesis that the 5' end of Aurora is also facing the asymmetric bulge. If this hypothesis is correct, the overall architecture of Aurora is likely a bent hairpin in which nucleotides distant in both the primary sequence and secondary structure converge on this asymmetric bulge.
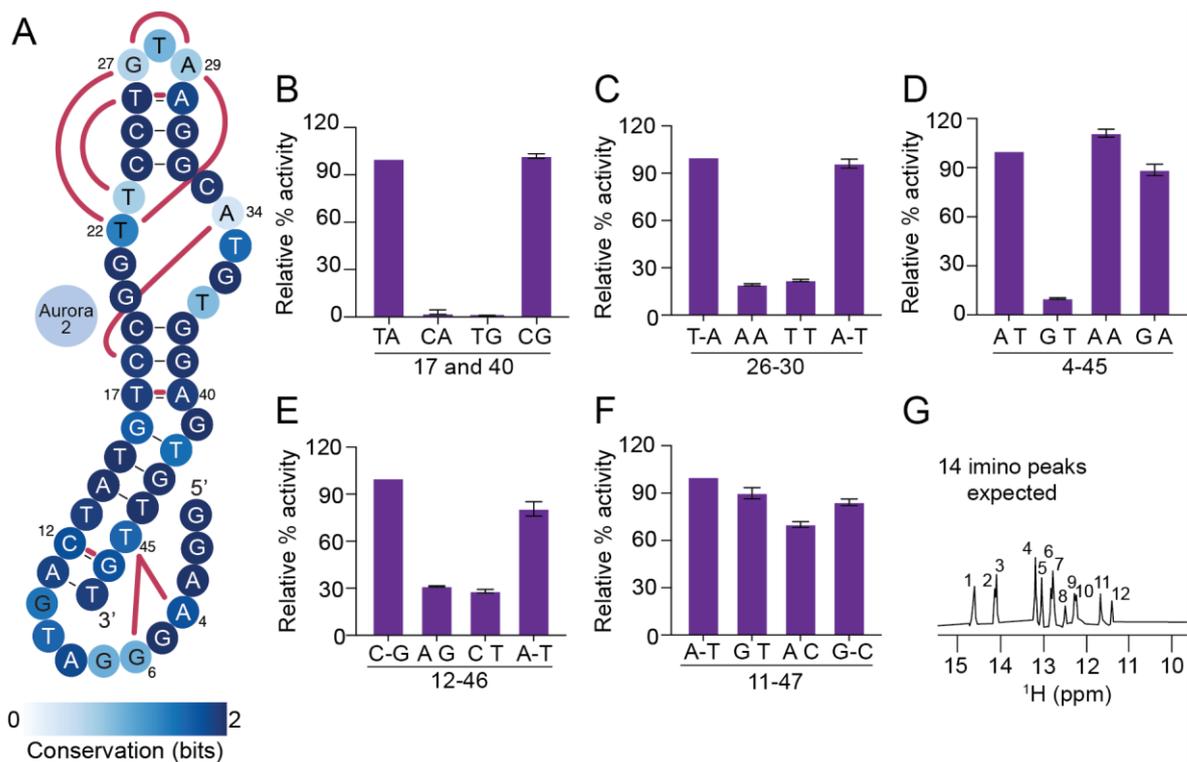
Figure R_10. Testing base pairs and noncanonical interactions in Aurora using double-mutant. cycles. (A) Secondary structure model of Aurora 2. Base pairs are shown using solid black lines, interactions supported by mutual information analysis are shown in maroon, and the degree of conservation at each position is indicated by blue shading. **(B)** Testing the 17-40 base pair. Signal to noise ratios are expressed relative to a variant containing a T-A base pair. **(C)** Testing the 26-30 base pair. Signal to noise ratios are expressed relative to a variant containing a T-A base pair. **(D)** Testing the 4-45 noncanonical pair. Signal to noise ratios are expressed relative to a variant containing an A T pair. **(E)** Testing the 12-46 base pair. Signal to noise ratios are expressed relative to a variant containing a C-G base pair. **(F)** Testing the 11-47 base pair. All signal to noise ratios are expressed relative to a variant containing a A-T base pair. Reactions were performed using 15 μM Aurora and 30 μM 4-MUP in 1x Aurora buffer containing 200 mM KCl, 1 mM ZnCl₂, 5% (v/v) DMSO, and 50 mM HEPES, pH 7.4. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. Experiments were performed using Aurora 2. **(G)** Proton NMR spectrum of the 17C 40G variant of Aurora 2 (Table 1) showing chemical shifts consistent with base pairs.
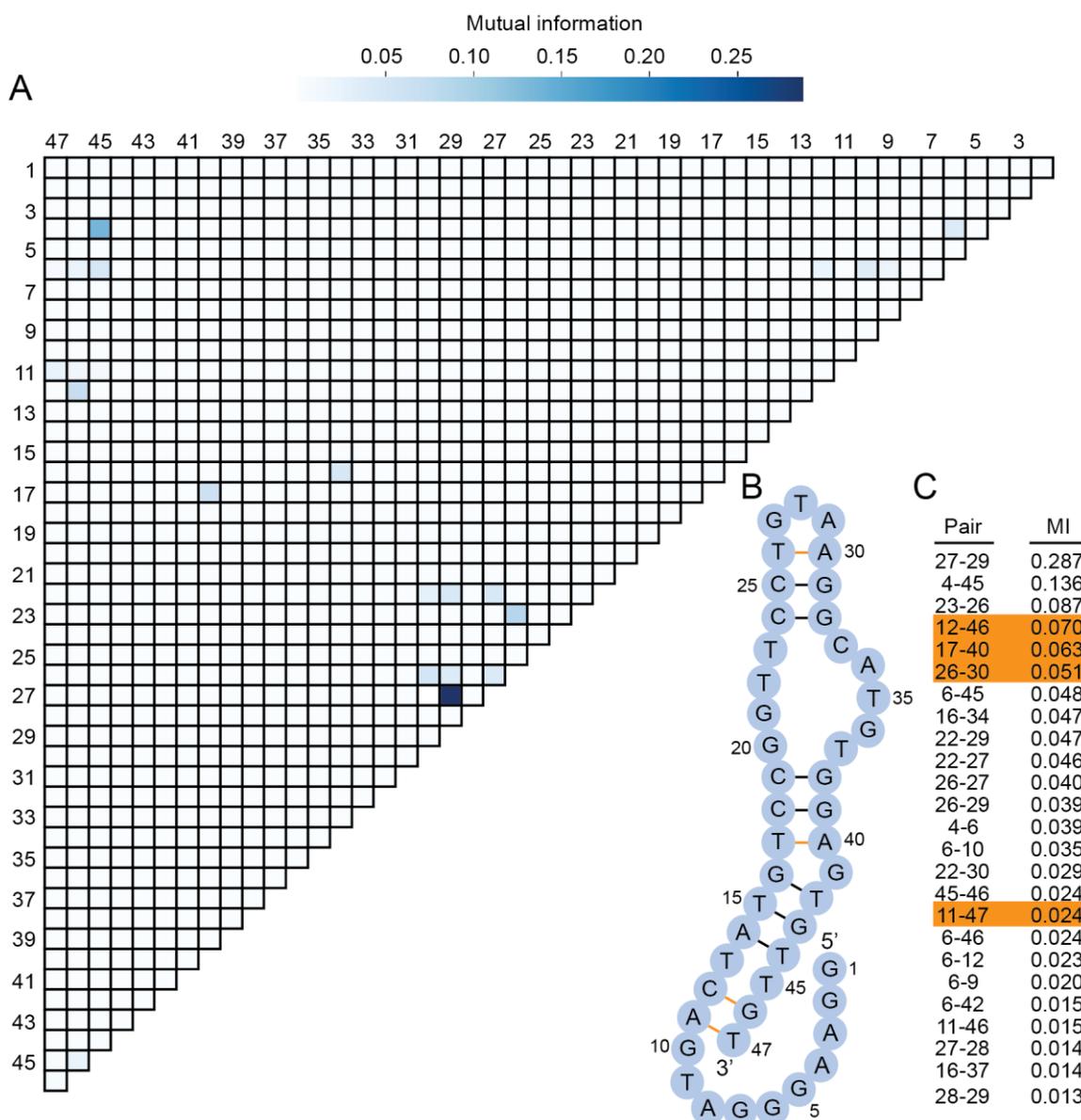
Figure R_11. Identification of correlated pairs of positions in Aurora using mutual information. (A) Heat map showing mutual information correlations between different pairs of positions in the minimized catalytic core of Aurora. **(B)** Secondary structure model of Aurora. Base pairs are shown using solid black lines, and base pairs supported by covariation analysis are shown in orange. **(C)** Mutual information values of the 25 pairs of positions with the highest mutual information values. Pairs of positions that form base pairs are shown in orange.

3.5.2 The catalytic core of Apollon is 50-nucleotides long with three helical regions

The analysis of conservation at each position in the sequences from the Apollon reselection suggested that positions 1-8 and 34-64 likely form the catalytic core of Apollon, whereas other positions might not be necessary for activity (Figure R_12a). The activity of a 39-nucleotide long construct, which only contained positions 1-8 and 34-64 was confirmed

using a spectroscopic color producing assay (Figure R_12c; column marked -4bp). However, additional analysis revealed that the 50-nucleotide long construct in which the region between positions 9 and 33 was extended with a stem had about 20% higher activity than the 39-nucleotide long catalytic core (Figure R_12c; compare the columns marked Apollon and -4bp). For this reason, the 50-nucleotide long variant (which we named Apollon 2) was used for most experiments.

Our next task was to construct a secondary structure model of Apollon. Secondary structure can often be identified by searching for covariations consistent with base pairing in datasets of deoxyribozyme variants[93–95]. However, in the case of Apollon, this approach was limited by the very high levels of conservation at most positions in the catalytic core. Despite this limitation, a secondary structure model was developed by combining the results from comparative analysis with information from double mutant cycles and proton NMR experiments. This model contains three helical regions, called Stem 1, Stem 2 and Stem 3, and several conserved loops and bulges (Figure R_12b). Stem 1 was well-supported by both covariations analysis and double mutant cycles (Figure R_12d). Stem 2 is less important: individual base pairs could be deleted without reducing activity, and a variant in which Stem 2 was deleted completely was only slightly (20%) less active than a variant in which it was present (Figure R_12c). However, mutant cycles in which multiple base pairs in Stem 2 were disrupted: by mutations in the 5' half of the helix, or by mutations in the 3' half of the helix, or restored by mutations in both the 5' and 3' halves of the helix, demonstrated that formation of this stem is important in some sequence backgrounds (Figure R_12e). Stem 3 was harder to prove, because this part of the sequence was almost invariant (Figure R_12a-b) and no covariations were observed. However, several additional lines of evidence supported Stem 3. Firstly, most active variants (and all with a read number of 10,000 or greater) had the potential to form at least seven of eight base pairs in Stem 3, and this degree of pairing was unlikely to occur by chance in randomly chosen subsets of sequences in the starting library (Figure R_12f). Secondly, the proton NMR spectrum of Apollon 2 contains approximately 19 distinct signals consistent with base pairs[96,97] (Figure R_12g). This spectrum is consistent with our suggested secondary structure model, which contains 12 canonical base pairs (each of which would be expected to yield on signal in the 12 to 15 ppm range of the spectrum) and 4 G-T or T-G pairs (each of which would be expected to yield two signals in this part of the spectrum due to the two tautomeric forms of guanine and thymine).
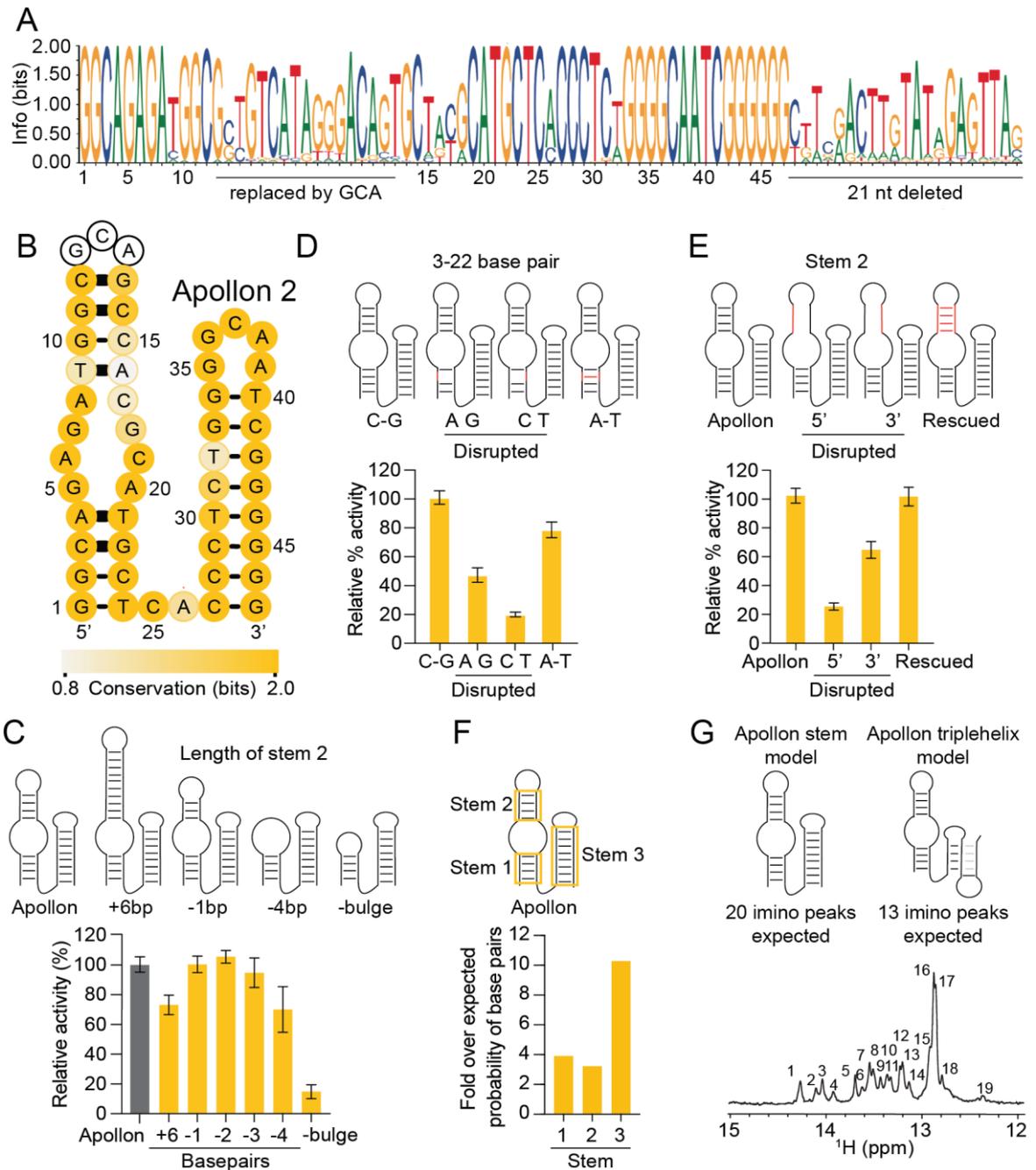
Figure R_12. Sequence requirements and secondary structure of Apollon. (A) Sequence logo generated from analysis of variants of Apollon using high-throughput sequencing data from the Apollon reselection. **(B)** Secondary structure model of Apollon 1. Base pairs are shown using thin black lines, base pairs supported by covariation analysis are shown using thick black lines, and the degree of conservation at each position is indicated by yellow shading. **(C)** Identification of the minimized catalytic core of Apollon. **(D)** Double mutant cycle showing that positions 3 and 22 interact in a manner consistent with base pairing. **(E)** Mutant cycle showing the importance of Stem 2 for Apollon function. **(F)** Graph showing the frequency of variants in the evolved reselection library with all base pairs in Stem 1, Stem 2, or Stem 3 divided by their frequency in the starting reselection library. **(G)** The proton NMR spectrum of Apollon suggests that Stem 3 in Apollon forms a canonical helix rather than a triple helix.

Apollon 2 also contains two stretches of at least three consecutive pyrimidines and three stretches of at least three consecutive purines, and therefore could potentially form triple helical structures like that one proposed to occur in the Supernova deoxyribozyme[42]. However, such triple helical structure is not supported by either comparative sequence analysis, triple-mutant cycles (Figure R_13) or proton NMR experiments (the triple helix model predicted 13 peaks instead of the 19 actually observed).

Taken together, these experiments suggest that Apollon forms a secondary structure consisting of three stems, one with a highly conserved primary sequence.

3.5.3 Testing of the Apollon triple helix model
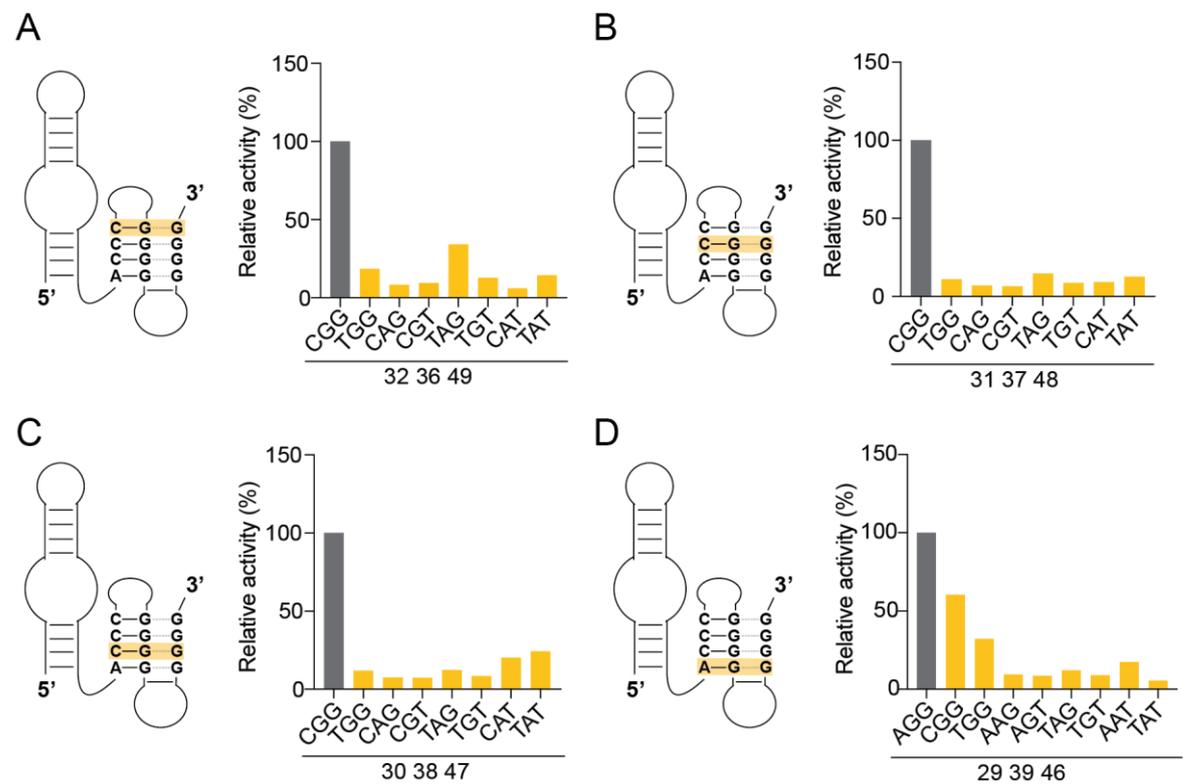
Figure R_13. Mutagenesis experiments do not support the hypothesis that Stem 3 of Apollon contains a purine-motif triple helix. (A) Left: secondary structure model with the putative 32-36-49 base triple highlighted in yellow. Right: triple mutant cycle in which the putative CGG triple is converted to a TAT triple. Activity is normalized to that of Apollon 2. **(B)** Same, but for the 31-37-48 base triple. **(C)** Same, but for the 30-38-47 base triple. **(D)** Same, but for the 29-39-46 triple. Reactions contained 30 μM DNA and 100 μM 4-NPP. The buffer contained 200 mM KCl, 1mM $ZnCl_2$, and 50 mM HEPES pH 7.4. Absorbance at 405 nm was measured after 4 hours using a TECAN Infinite M200 Pro plate reader.

3.6 Aurora generates a robust fluorescent signal and Apollon generates a yellow product that can be seen by eye

To evaluate if the reselection experiments yielded improved variants of Aurora and Apollon, we compared the catalytic activity of the initial isolates (Aurora 1 and Apollon 1; Figure R_14a, b) with the variants with the highest read numbers from the reselection experiments (Aurora 2 and Apollon 2; Figure R_14a, b). Each variant was characterized in the context of the 47-nucleotide long minimal catalytic core of Aurora and 50-nucleotide long minimal catalytic core of Apollon.

Measurements were performed over a range of 4-MUP and 4-NPP concentrations using a ligation assay (which measures the extent of self-phosphorylation). These experiments revealed that the catalytic activity of Aurora 2 was more than 100-fold higher than that of Aurora 1 at some substrate concentrations (Figure R_14c). Most mutations in Aurora 2 occurred in either the asymmetric bulge (positions 20-23 and 33-37) or the loop (position 27-29; Figure R_14a), which highlights the importance of these parts of Aurora for catalytic activity. At saturating substrate concentrations, the maximal rate of Aurora 2 was 0.18 min^{-1} and the concentration of substrate at which the activity of Aurora 2 was half maximal was 30 μM (Figure R_14c).

In the case of Apollon, it was not possible to saturate the binding sites of either Apollon 1 or Apollon 2 even at concentrations of 4-NPP as high as 2 mM, indicating that Apollon has a low affinity for its substrate (Figure R_14e). This is perhaps not surprising given the small size of 4-NPP. However, the maximum k_{obs} was 0.32 min^{-1} which is comparable to both Aurora and Supernova. The catalytic efficiency (k_{cat}/K_M) of Apollon 2 was 10-fold higher than that of Apollon 1 (Figure R_14e). Apollon 2 contained six mutations in the catalytic core relative to Apollon 1. Five of these mutations occurred in or near the bulge between Stem 1 and Stem 2 (Figure R_14b), indicating that this part of the deoxyribozyme is particularly important for catalytic function.

We also investigated the extent to which Aurora is able to enhance fluorescence. When using a continuous experimental setup in which Aurora 2 was mixed with buffer and 4-MUP and fluorescence was continuously monitored using a plate reader (Figure R_15a), signal to noise ratios of more than 100-fold were obtained in hours (Figure R_15b). When using a discontinuous setup in which reactions were quenched with base before measurement, signal to noise ratios were about 6-fold higher, and values over 700-fold

were achieved (Figure R_15c). In both assays, the fluorescent signal gendered by Aurora 2 was at least 10-fold higher than that of Aurora 1 (Figure R_14d and Figure R_15b, c).

Figure R_14. Optimized versions of Aurora and Apollon from reselections produce better fluorescence and chromogenic signals than the versions isolated in initial selections. (A) Sequence and secondary structure model of Aurora 1 and Aurora 2. Positions that differ from Aurora 1 are shown in orange in Aurora 2. **(B)** Sequence and secondary structure model of Apollon 1 and Apollon 2. Positions that differ from Apollon 1 are shown in blue in Apollon 2. **(C)** Catalytic activity of Aurora 1 and 2 over a range of 4-MUP concentrations as measured using a ligation assay. **(D)** Comparison of the fluorogenic reaction of Aurora 1 and Aurora 2 analysed using a plate reader. **(E)** Catalytic activity of Apollon 1 and 2 over a range of 4-NPP concentrations as measured using a ligation assay. **(F)** Comparison of the colorimetric reaction of Apollon 1 and Apollon 2 analysed using a plate reader.
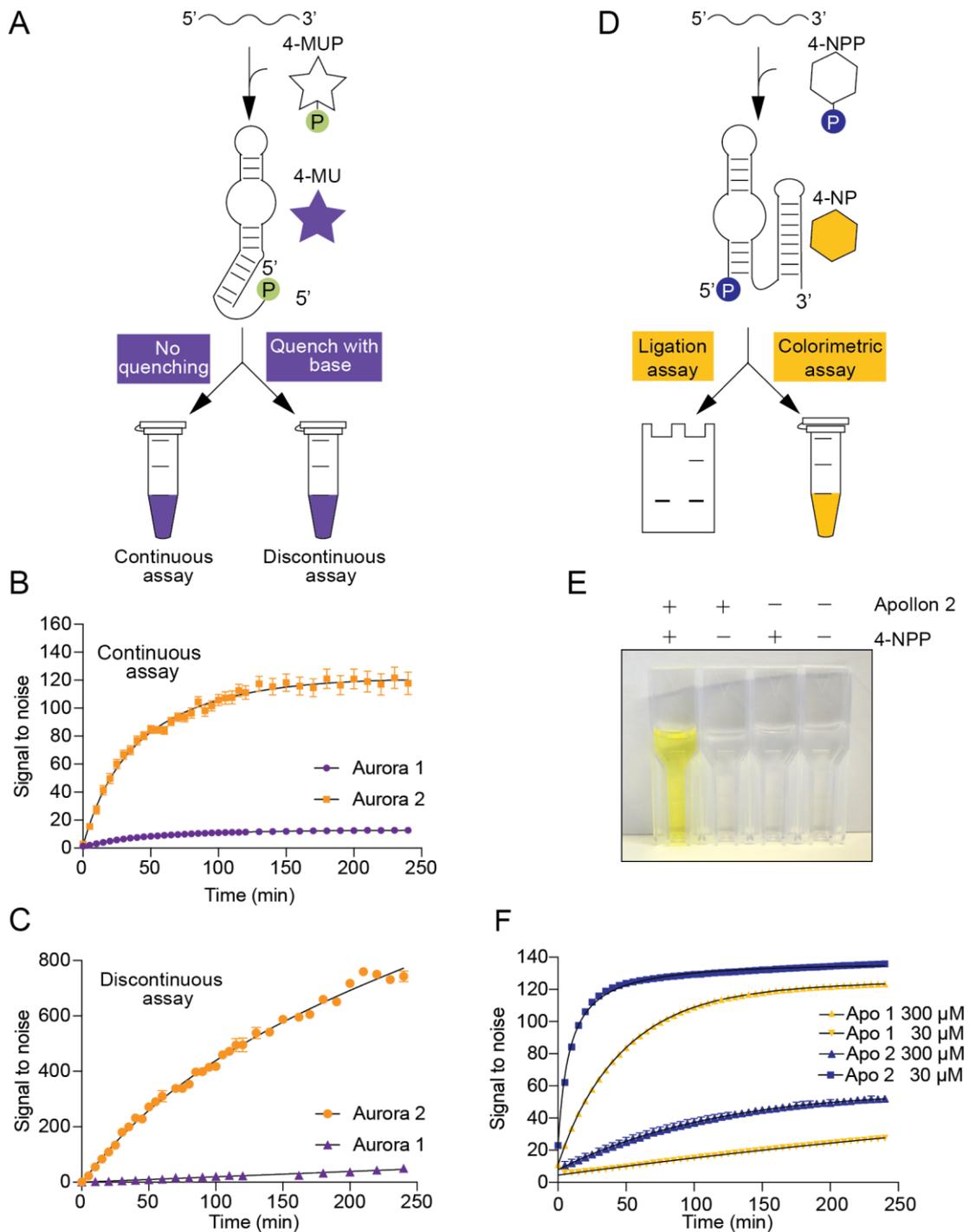
Figure R_15. Aurora generates a robust fluorescent signal and Apollon generates a robust chromogenic signal. (A) Workflow of continuous and discontinuous assays using Aurora. **(B)** Example of a continuous assay for Aurora 1 and Aurora 2 in which the reaction is continually monitored in a plate reader. **(C)** Example of a discontinuous assay for Aurora 1 and Aurora 2 in which time points are quenched with base before measuring fluorescence. **(D)** Assay to measure Apollon self-phosphorylation (ligation assay) and production of a yellow product (colorimetric assay). **(E)** Example of a colorimetric reaction that can be visualized by eye. **(F)** Colorimetric reaction of Apollon 1 and Apollon 2 at different concentrations analysed using a plate reader.

Apollon 2 also generated a stronger colorimetric signal than Apollon 1, with a signal to noise ratio of over 150-fold, but the difference in color production between Apollon 1 and Apollon 2 was not as large as in the case of Aurora 1 and Aurora 2 (Figure R_14f and Figure R_15f). The signal generated by Apollon 2 could also be detected by eye (Figure R_15e). The limit of detection of Aurora 2 was 100 nM and that of Apollon 2 was 1 μM (Figure R_16).

Taken together, these experiments provided information about mutations that enhance the catalytic activity of Aurora and Apollon. They also demonstrate that our selection strategy can be used to identify deoxyribozymes that can generate a robust fluorescent signal or a colorimetric signal that can be seen by eye.
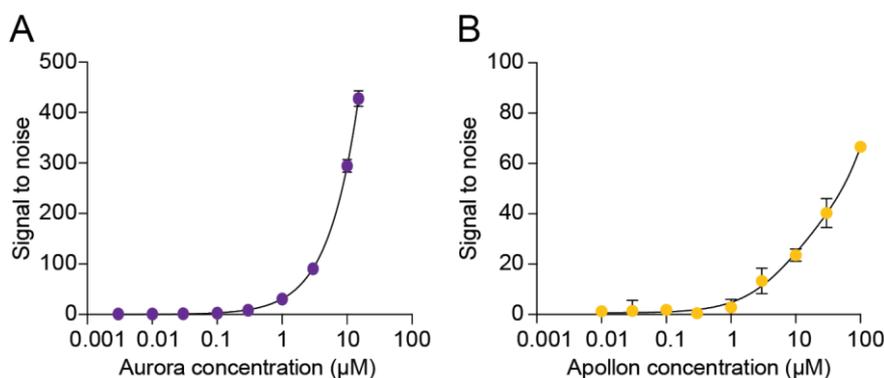

Figure R_16. Detection limit of Aurora and Apollon. (A) Signal to noise ratio of fluorescence production over a range of concentrations of Aurora. Reactions were performed using the indicated concentration of Aurora and 30 μM 4-MUP in a buffer containing 200 mM KCl, 1 mM $ZnCl_2$, 5% (v/v) DMSO, and 50 mM HEPES, pH 7.4. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(B)** Signal to noise ratio of color production over a range of concentrations of Apollon. Reactions were performed using the indicated concentration of Apollon and 100 μM 4-NNP in a buffer containing 200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4. After incubating for 4 hours, absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. Experiments were performed using Aurora 2 and Apollon 2.

3.7 Aurora and Apollon require multiple zinc ions for structure and function

Although these selection experiments provided extensive information about the sequence requirements of Aurora and Apollon, they revealed little about how external factors (such as components of the buffer or temperature) influence reactions. Such factors can significantly affect signal to noise ratios, and can also provide clues about possible catalytic mechanisms of our deoxyribozymes. Therefore, we next extensively characterized

the buffer requirements of Aurora and Apollon using both ligation and spectroscopic assays. The effects of buffer conditions on deoxyribozyme folding were also monitored by proton NMR. We were especially interested in the effects of metal ions on the reactions, because they can play both structural and catalytic roles. Survey revealed both differences and similarities among Aurora, Apollon and Supernova[42,86].

An important difference was that Aurora appears to require monovalent ions for activity (Figure R_17-19) while Apollon and Supernova do not. On the other hand, experiments showed that, of the four ions present in the selection buffer (zinc, lead, cerium, and potassium), lead and cerium could be excluded from the reaction without reducing activity of Aurora, Apollon or Supernova (Figure R_17). Potassium could be replaced by other monovalent metal ions such as lithium, sodium, rubidium, or cesium without affecting catalytic activity (Figure R_18). However, as was previously mentioned, Apollon activity was similar over a range of potassium concentrations while Aurora activity was strongly influenced by potassium concentration (Figure R_19). Activity of Aurora and Apollon did not specifically require HEPES (Figure R_20), but was highly dependent on pH, and activity was only observed between pH 7 and pH 8 (Figure R_21). Apollon is active between 25 and 55 °C, but activity decreased rapidly at higher or lower temperatures (Figure R_22). PEG 200 and DMSO have beneficial effects on Aurora activity, and therefore 5% (v/v) DMSO was used in Aurora reactions, but Apollon activity was inhibited by DMSO (Figure R_23). A particularly interesting parallel is that each of these three deoxyribozymes (Aurora, Apollon and Supernova) requires zinc for activity when in principle they could have required any other combination of the four metal ions present in the selection buffer (Figure R_24). Moreover, titration experiments showed that zinc affects catalytic activity in a highly cooperative way, suggesting that multiple zinc ions are needed for function and for folding of Aurora and Apollon (Figure R_25 and Figure R_26). This is intriguing because zinc ions play catalytic roles in some protein enzymes (such as alkaline phosphatase[98,99]), that catalyse reactions similar to that promoted be Aurora or Apollon. In a more general sense, these results suggest that zinc plays an important role in deoxyribozymes that promote phosphoryl transfer reactions using monophosphorylated substrates, as has been previously noted for deoxyribozymes that promote RNA or DNA cleavage reactions[100–102]. They also highlight the degree to which the presence of specific metal ions during a selection can increase the likelihood of obtaining nucleic acid enzymes with specific functions[103].

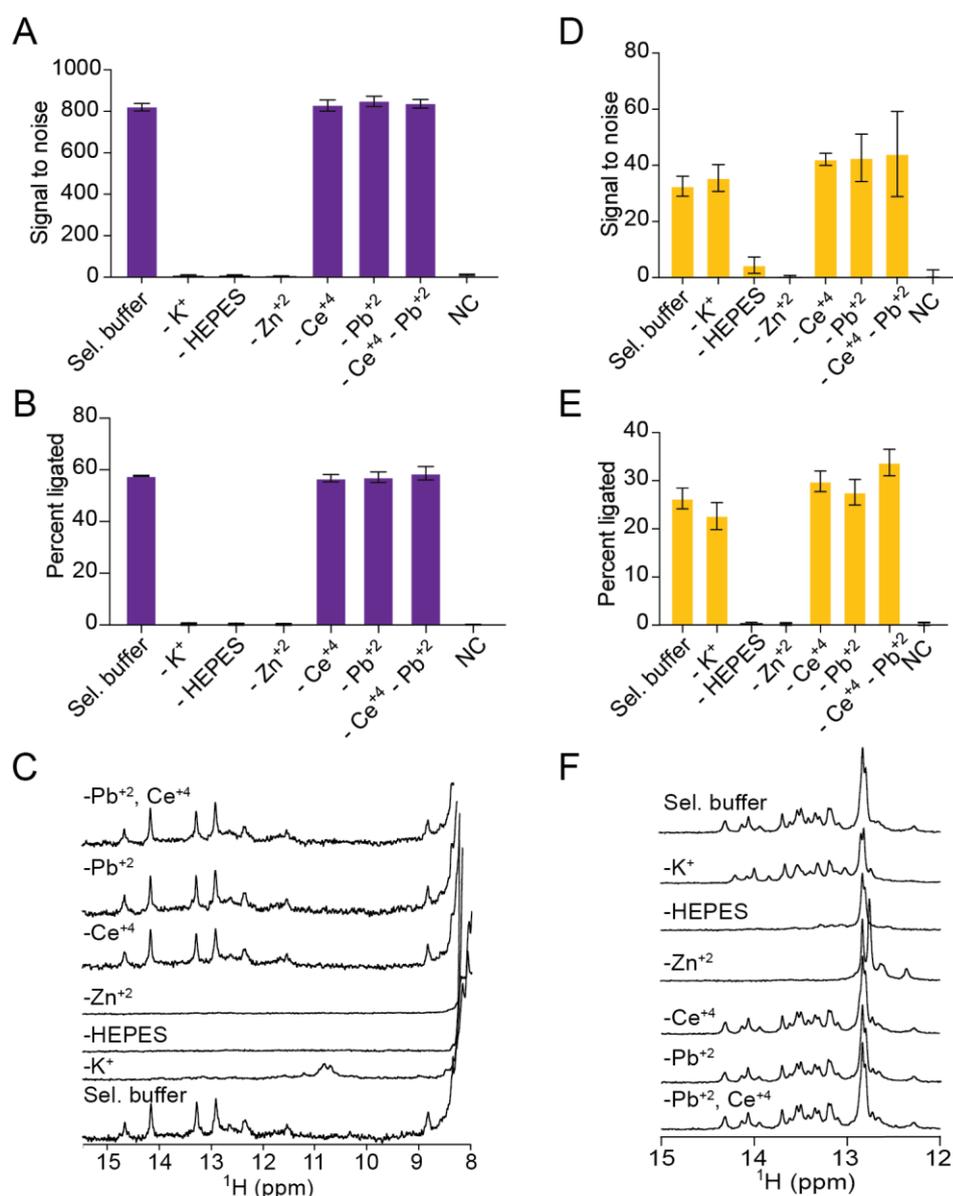## 3.7.1 Simplification of the selection buffer

Figure R_17. Aurora requires potassium, HEPES and zinc for activity and folding, while Apollon only requires HEPES and zinc. (A) Signal to noise ratios of Aurora in a series of buffers in which different components were omitted. The original selection buffer (labelled "Sel. buffer") contained 200 mM KCl, 1 μM Ce(NO$_3$)$_4$, 0.1 μM PbCl$_2$, 1 mM ZnCl$_2$, and 50 mM HEPES pH 7.4. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(D)** Experiment in the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(B, E)** Experiments using the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. **(C, F)** Proton NMR spectra of Aurora and Apollon in the same buffers as used in panel A. Spectra were measured using 300 μM Aurora and 450 μM 4-MUP or 300 μM Apollon and 450 μM 4-NPP. Experiments were performed using Aurora 2 and Apollon 2.
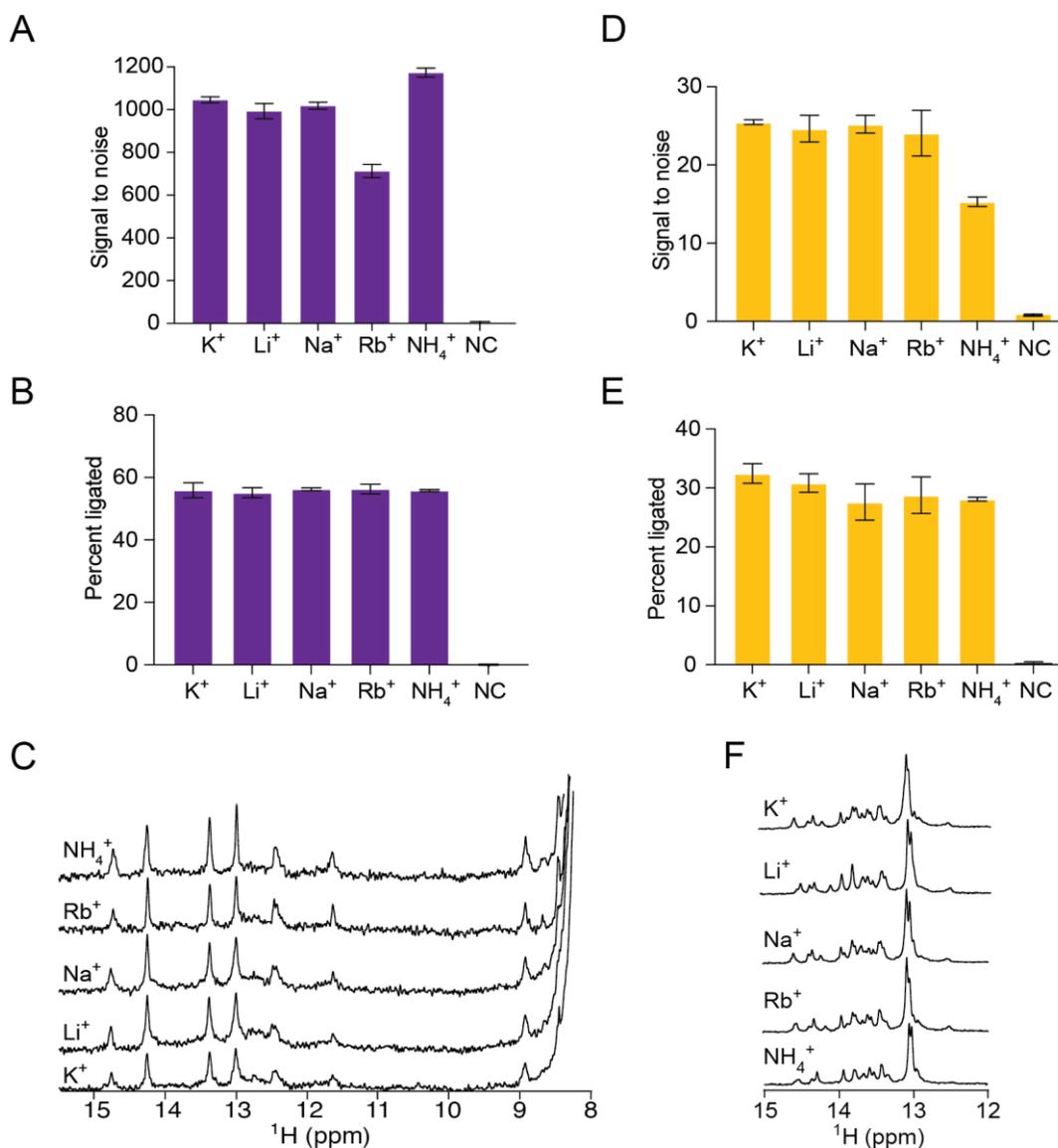
3.7.2 Effect of monovalent ions

Figure R_18. Aurora and Apollon are active in a wide range of monovalent metal ions. (A) Signal to noise ratios of Aurora in a series of buffers in which potassium was replaced with other monovalent metal ions. Buffers contained the indicated monovalent cation at a concentration of 200 mM, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(D)** Experiment in the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(B, E)** Experiment in the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. **(C, F)** Proton NMR spectra of Aurora and Apollon in the same buffers as used in panel A. Spectra were measured using 300 μM Aurora and 450 μM 4-MUP or 300 μM Apollon and 450 μM 4-NPP. Experiments were performed using Aurora 2 and Apollon 2.

3.7.2.1 Effect of potassium concentration
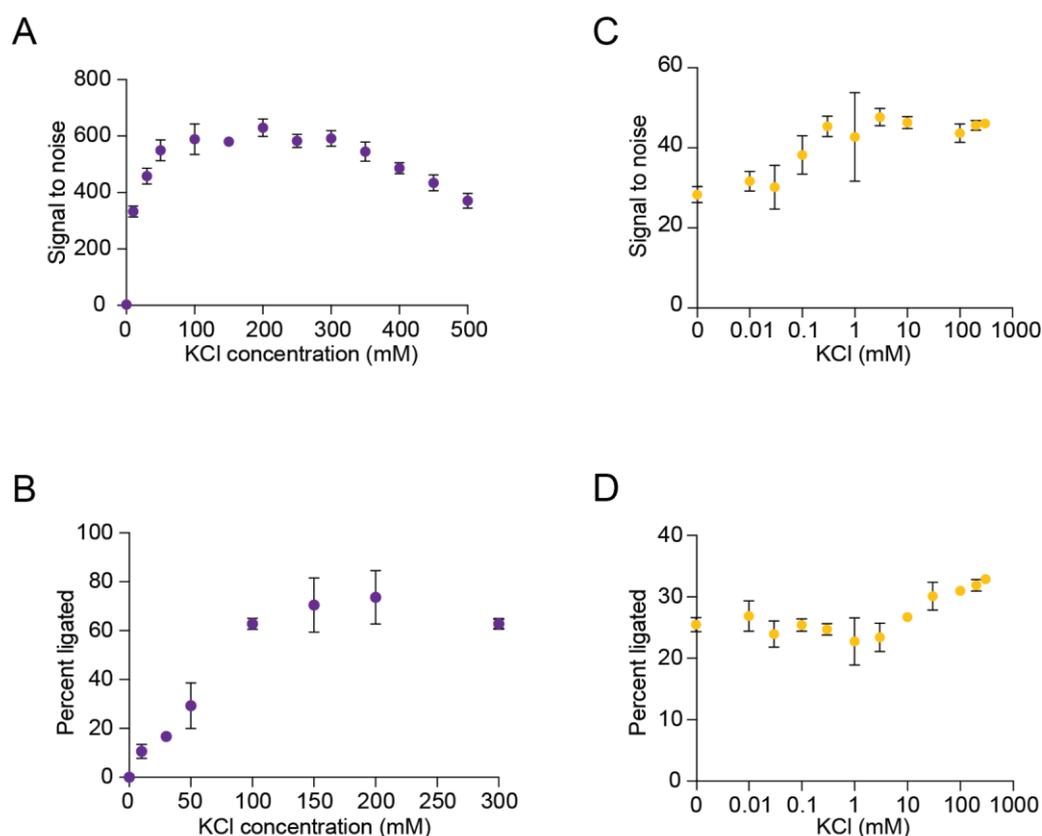
Figure R_19. Aurora activity is dependent on KCl concentration, but Apollon activity is not. (A) Signal to noise ratios of Aurora over a range of potassium concentrations. Buffers contained the indicated concentration of KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(B)** Experiments using the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(C, D)** Experiment using the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. Experiments were performed using Aurora 2 and Apollon 2.

3.7.3 Effect of buffering agents
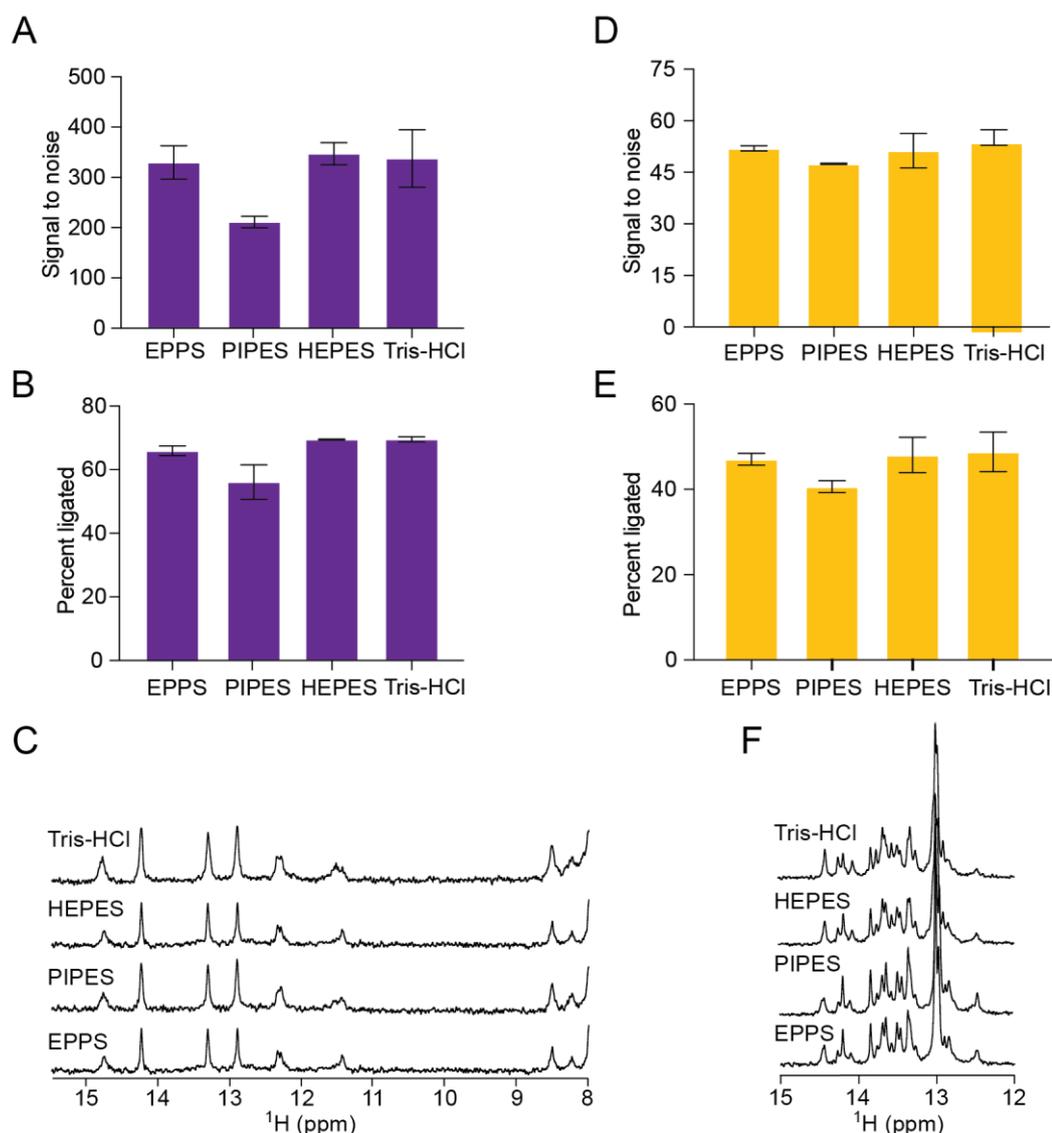
Figure R_20. Aurora and Apollon are active in a wide range of buffering agents. (A) Signal to noise ratio of Aurora in a series of buffers in which HEPES was replaced with another buffering agent. Reactions contained the indicated buffering agent at a concentration of 50 mM at pH 7.4 as well as 200 mM KCl and 1 mM $ZnCl_2$. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(D)** Experiments using the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(B, E)** Experiments using the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. **(C, F)** Proton NMR spectra of Aurora and Apollon in the same buffers as used in panel A. Spectra were measured using 300 μM Aurora and 450 μM 4-MUP or 300 μM Apollon and 450 μM 4-NPP. Experiments were performed using Aurora 2 and Apollon 2.

3.7.4 Effect of pH on fluorescence and color production
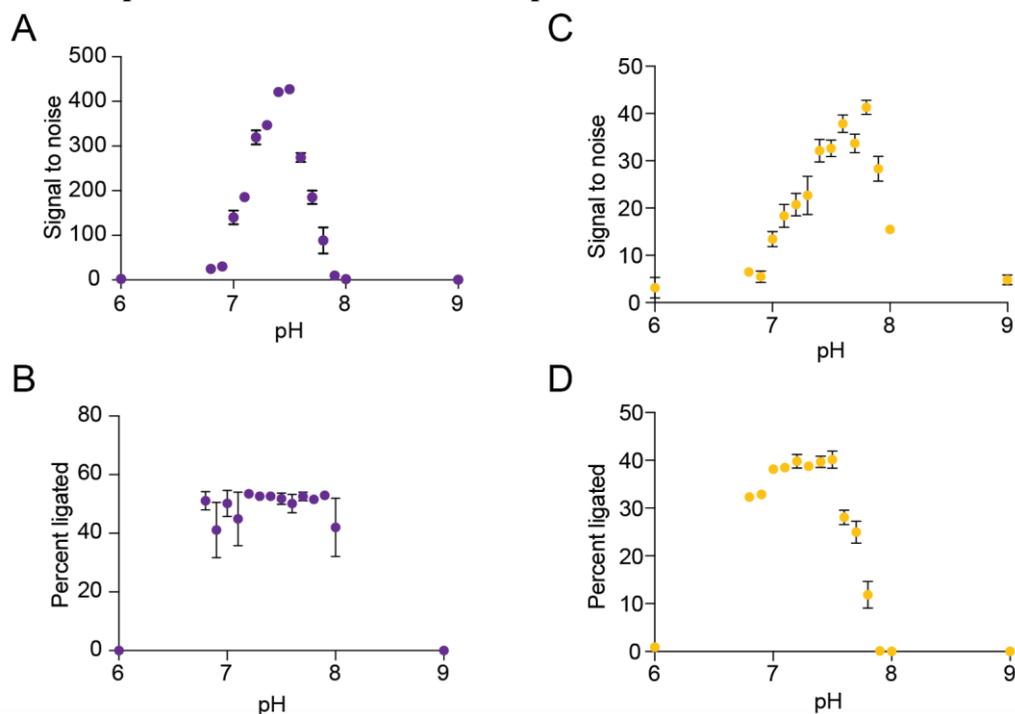

Figure R_21. Effect of pH on Aurora and Apollon catalytic activity. (A) Signal to noise ratio of Aurora as a function of pH. Buffers contained 50 mM HEPES at the indicated pH, 200 mM KCl, and 1 mM $ZnCl_2$. Reactions contained 15 µM Aurora and 30 µM 4-MUP. After incubating for 4 hours, 20 µl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(B)** Experiment in the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 µM Apollon and 100 µM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(C, D)** Experiments in the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 µM Aurora with 1 mM 4-MUP or 1 µM Apollon with 1 mM 4-NPP for 1 hour. Experiments were performed using Aurora 2 and Apollon 2.

3.7.5 Effect of temperature on color production
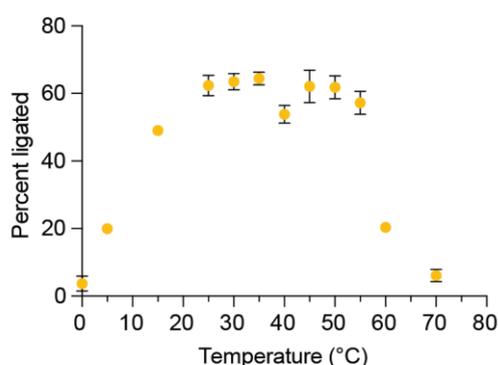

Figure R_22. Temperature dependence of Apollon as measured using the ligation assay. The pH of the buffering agent (HEPES) was adjusted as recommended by the manufacturer so that it would remain at pH 7.4 at different temperatures. Buffers contained 200 mM KCl, 1 mM $ZnCl_2$ and 50 mM HEPES at the indicated pH. The percentage of ligated Apollon was determined after incubating 1 µM Apollon with 1 mM 4-NPP for 1 hour at the indicated temperature. Points show average values from three experiments, and error bars represent one standard deviation. Experiments were performed using Apollon 2.

3.7.6 Effect of organic solvents and crowding agents
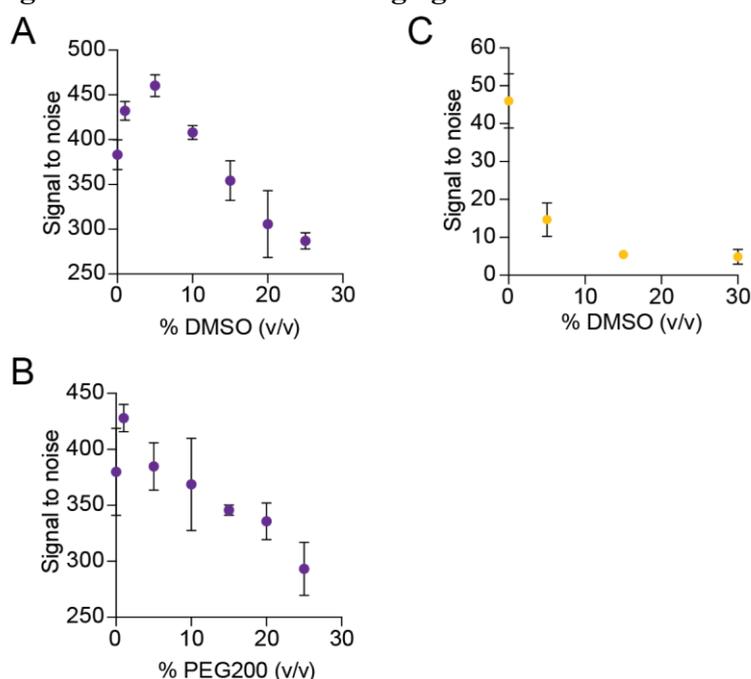
Figure R_23. Effect of molecular crowding agents and organic solvents on the catalytic activity of Aurora and Apollon. (A) Signal to noise ratio of Aurora over a range of DMSO concentrations. Buffers contained the indicated concentration of DMSO, 200 mM KCl, 1 μM Ce(NO$_3$)$_2$, 0.1 μM PbCl$_2$, 1 mM ZnCl$_2$, and 50 mM HEPES, pH 7.4. Reactions were performed using 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(B)** Signal to noise ratio of Aurora over a range of PEG 200 concentrations. Buffers contained the indicated concentration of PEG 200, 200 mM KCl,1 μM Ce(NO$_3$)$_2$, 0.1 μM PbCl$_2$, 1 mM ZnCl$_2$, and 50 mM HEPES, pH 7.4. Reactions were performed using 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(C)** Signal to noise ratio of Apollon over a range of DMSO concentrations. Buffers contained the indicated concentration of DMSO, 200 mM KCl, 1 mM ZnCl$_2$, and 50 mM HEPES, pH 7.4. Reactions were performed using 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. Experiments were performed using Aurora 2 and Apollon 2.

3.7.7 Effect of divalent ions
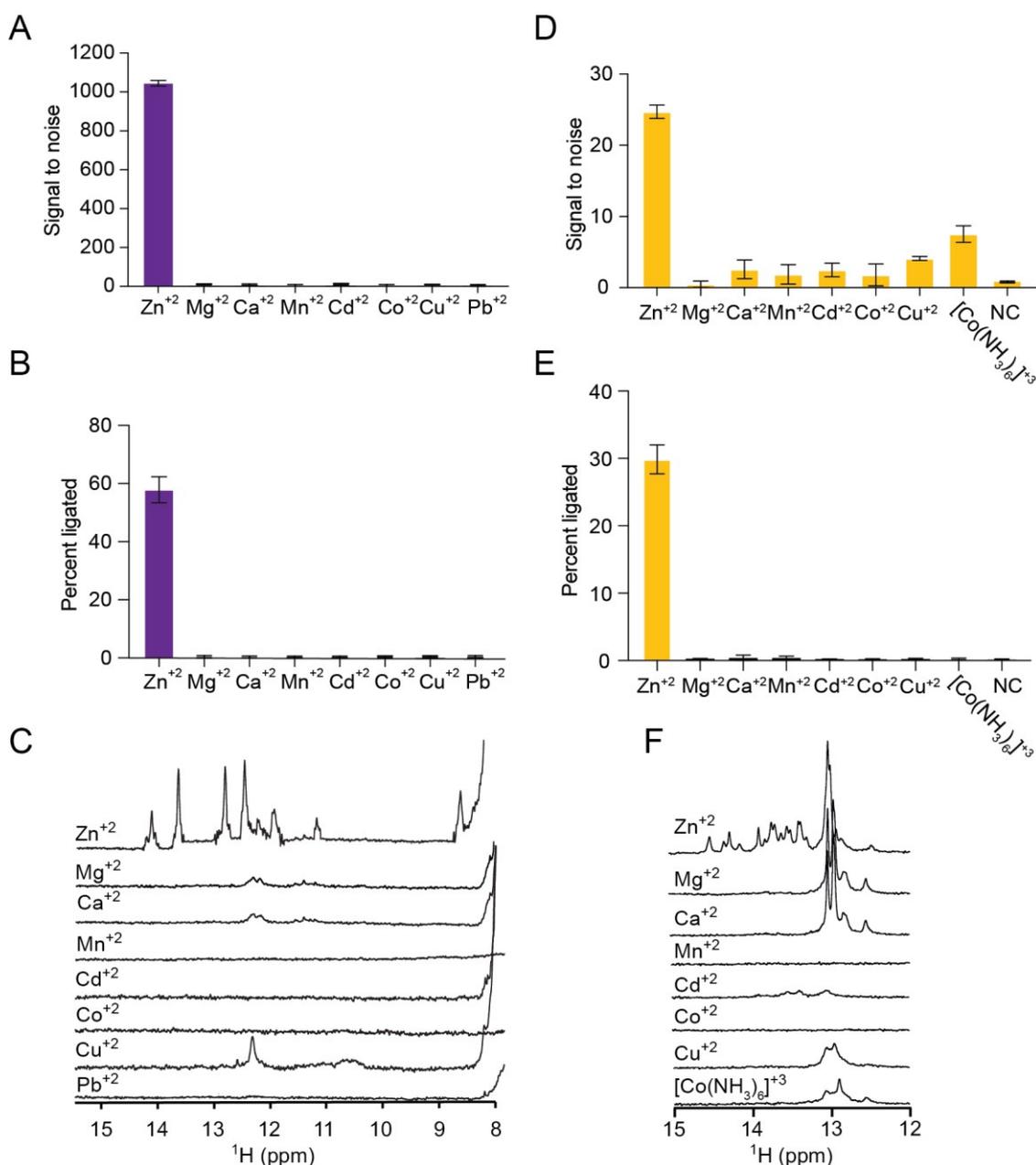
Figure R_24. Aurora and Apollon are active in the presence of zinc, but not in the presence of other divalent metal ions. (A) Signal to noise ratio of Aurora in a series of buffers in which zinc was replaced with other divalent metal ions. Buffers contained the indicated divalent cation at a concentration of 1 mM, 200 mM KCl, and 50 mM HEPES pH 7.4. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(D)** Experiment using the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(B, E)** Experiment using the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. **(C, F)** Proton NMR spectra of Aurora and Apollon in the same buffers as used in panel A. Spectra were measured using 300 μM Aurora and 450 μM 4-MUP or 300 μM Apollon and 450 μM 4-NPP. Experiments were performed using Aurora 2 and Apollon 2.

3.7.7.1 Effect of zinc concentration on fluorescence and color production
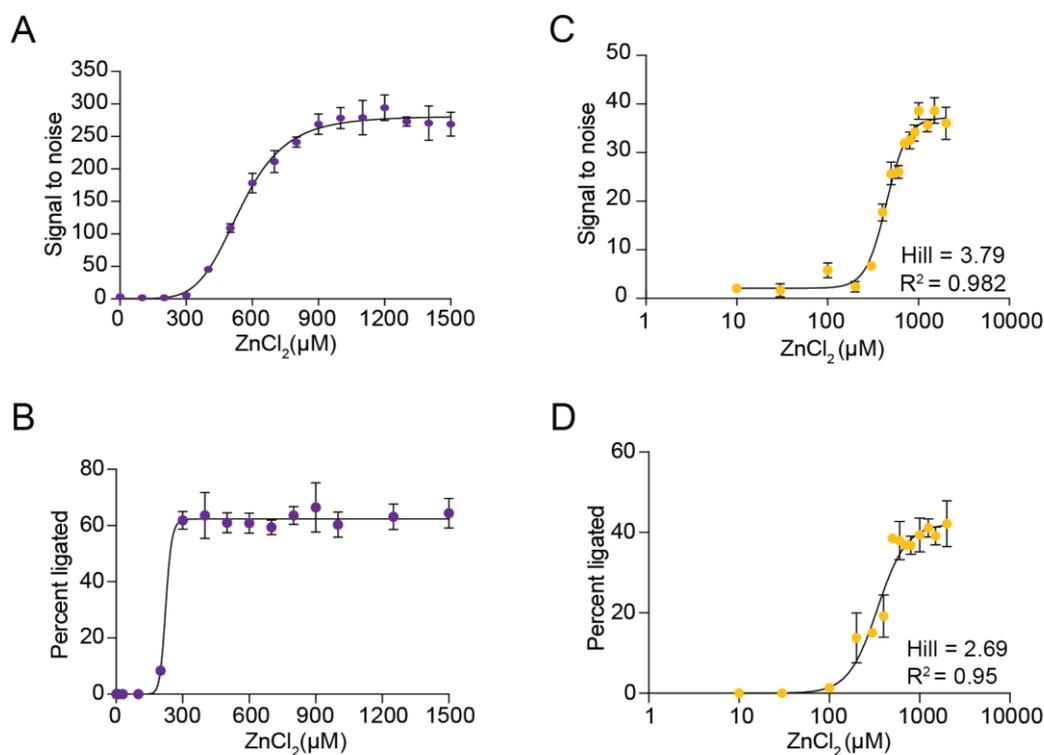
Figure R_25. Zinc has cooperative effect on Aurora and Apollon catalytic activity. (A) Signal to noise ratio of Aurora as a function of zinc concentration. Buffers contained the indicated concentration of $ZnCl_2$, 200 mM KCl, and 50 mM HEPES pH 7.4. Reactions contained 15 μM Aurora and 30 μM 4-MUP. After incubating for 4 hours, 20 μl of 1 M KOH was added and fluorescence was measured using a TECAN Spark plate reader. **(B)** Experiment using the same buffers as in panel A, but showing signal to noise ratios of Apollon. Reactions contained 30 μM Apollon and 100 μM 4-NPP. After incubating for 4 hours, the absorbance at 405 nm was measured using a TECAN Infinite M200 Pro plate reader. **(C, D)** Experiment using the same buffers as in panel A, but showing the percent of ligated Aurora or Apollon measured using the ligation assay. The percent ligated was determined after incubating 1 μM Aurora with 1 mM 4-MUP or 1 μM Apollon with 1 mM 4-NPP for 1 hour. Data were fitted with Hill equation using Prims10 software. Experiments were performed using Aurora 2 and Apollon 2.

3.7.8 Effect of zinc concentration on folding of Aurora and Apollon
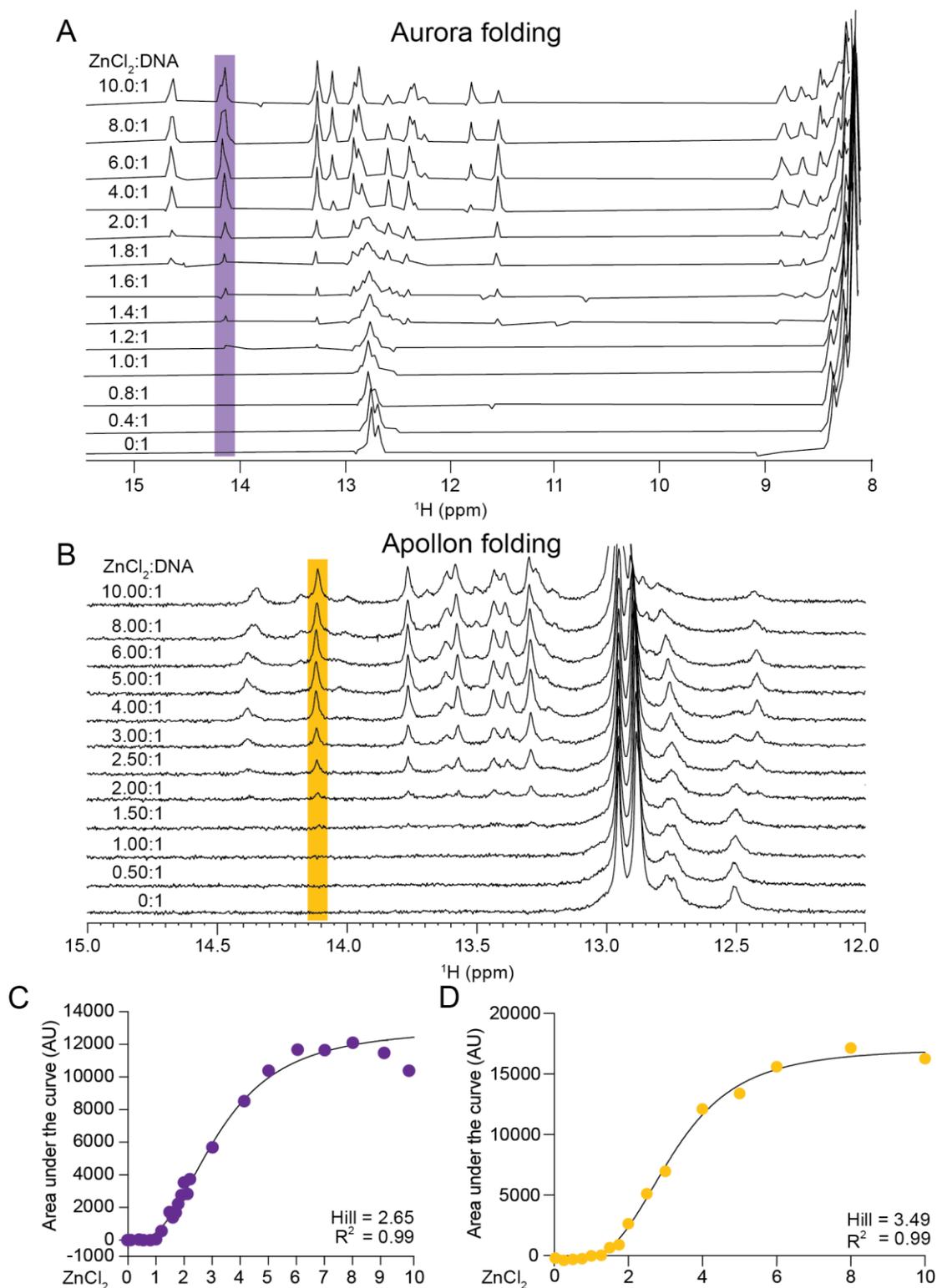
Figure R_26. Cooperative effect of zinc on Aurora and Apollon folding. (A, B) Proton NMR spectra of Aurora and Apollon over a series of concentrations of $ZnCl_2$. Spectra were measured in the presence of 300 µM DNA, 450 µM 4-MUP or 450 µM 4-NPP, 200 mM KCl, and 50 mM HEPES pH 7.4. **(C, D)** Graphs showing the area under the curve of the second peak from the left in panels A or B (indicated with a purple box for Aurora and yellow box for Apollon) Data were fitted with Hill equation using Prims10 software.

3.8 Apollon makes strong contacts with the phosphate group of the substrate

We next investigated how Apollon recognizes its substrate. Isothermal titration calorimetry provided an important clue. No binding between 4-NPP to Apollon was detected when a substrate titration was done using reacted (5' phosphorylated) deoxyribozyme (Figure R_27a). This result suggests that the 5' phosphate remains in the binding site even after the reaction. NMR experiments also showed that folding of Apollon requires the presence of both zinc ion and a phosphate group on either the substrate 4-NPP or on the 5' end of Apollon (Figure R_28b). This is another difference between Aurora and Apollon. Aurora requires 4-MUP and zinc for folding (Figure R_28a).

Additional evidence that Apollon makes strong contacts with the phosphate group came from differential scanning calorimetry. This showed that the melting temperature of the reacted form of the Apollon (which contains a 5' phosphate) is 7°C higher than that of the unreacted form (which contains a 5' hydroxyl group) (Figure R_27b). This difference in melting temperatures was observed only when folding was performed in the presence of zinc, highlighting its important role in deoxyribozyme folding and function. These results are consistent with a model in which Apollon makes important contacts with the phosphate group from 4-NPP. However, other characteristics of the substrate binding site must also be important, since the specificity Apollon is orthogonal to both Aurora and Supernova (Figure R_8f), perhaps due to steric clashes with the bigger substrates used by these deoxyribozymes (Figure R_1b). Furthermore, Apollon does not generate a ligated product when reactions are performed in the absence of 4-NPP, indicating that it cannot phosphorylate itself using the ATP in the ligation buffer (Figure R_8f).

Taken together, these results indicate that Apollon requires a phosphate group on either the 4-NPP substrate or its own 5' hydroxyl group for folding, and has a relatively specific binding site for this substrate. Differential scanning calorimetry provided additional evidence that zinc helps to stabilize the overall fold of the deoxyribozyme, including the structure in the active site.
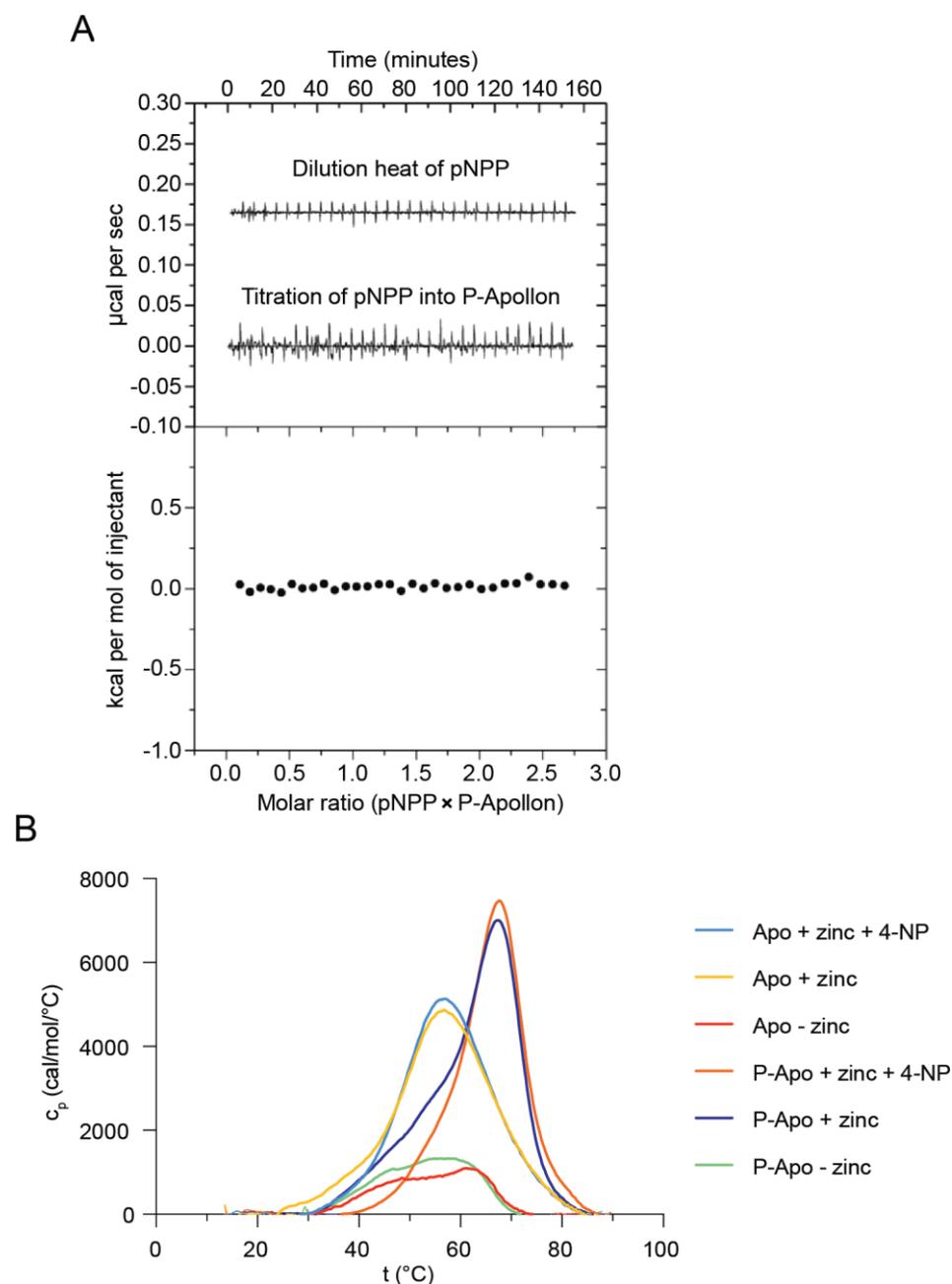
Figure R_27. 5' phosphorylated Apollon does not bind 4-NPP, but zinc and a 5' phosphate help to stabilize the Apollon structure. (A) Isothermal calorimetric titration of 5' phosphorylated Apollon 2 with 4-NPP. Binding of 4-NPP to 5' phosphorylated Apollon 2 was measured using 101.3 μM DNA in 200 mM KCl, 50 mM HEPES pH 7.4, and 1 mM ZnCl$_2$. Stepwise injections of typically 9 μl of 1.3 mM 4-NPP were performed until saturation was achieved. **(B)** Denaturation profiles of Apollon 2 (labelled Apo) and 5'phosphorylated Apollon 2 (labelled P-Apo) with or without 4-NP and with or without zinc. Profiles were measured using 30 μM DNA in 200 mM KCl, 50 mM HEPES pH 7.4, 1 mM ZnCl$_2$ and 60 μM 4-NP if not stated otherwise.
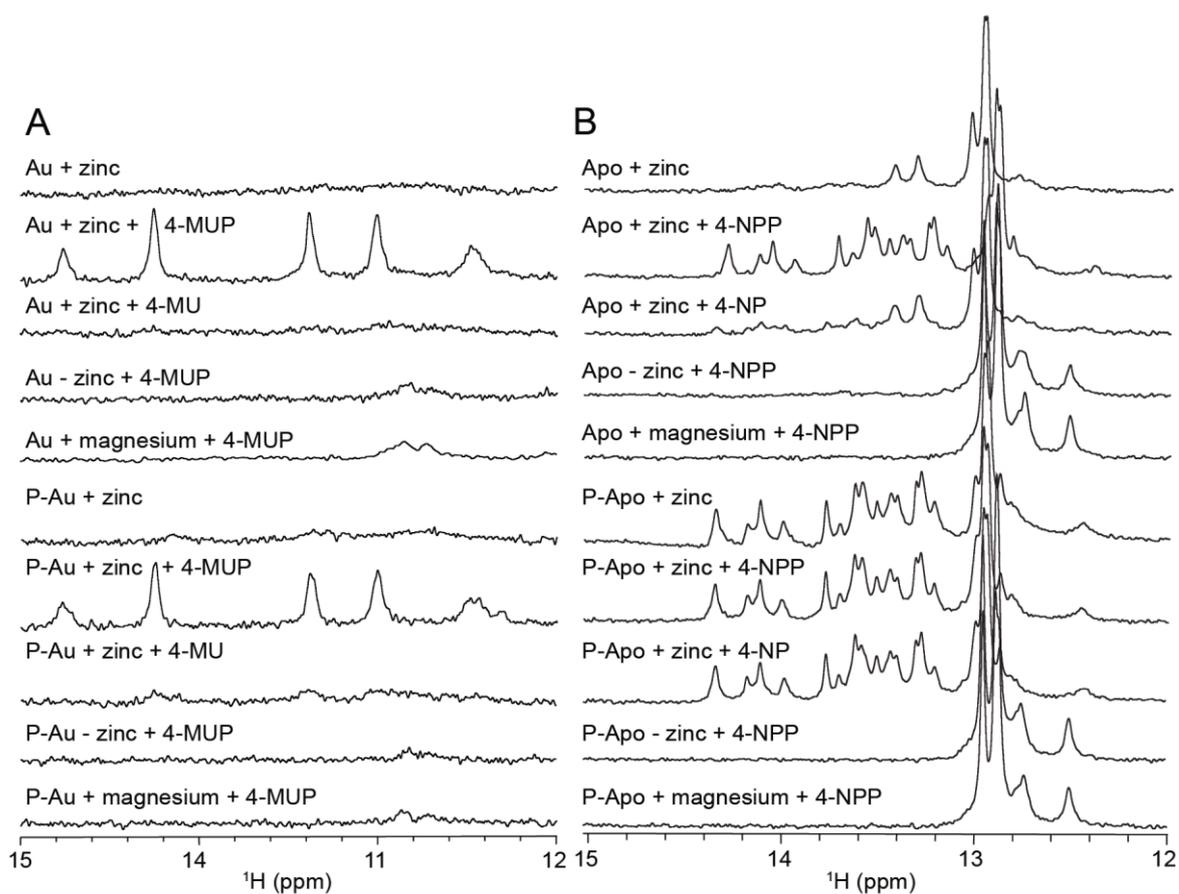
Figure R_28. Aurora requires zinc and 4-MUP to form a stable structure. Apollon requires zinc and a 5' phosphate to form a stable structure. (A) Proton NMR spectra of Aurora 2 (which contains a 5' hydroxyl group and is labelled Au) and 5' phosphorylated Aurora 2 (which contains a 5' phosphate group and is labelled P-Au) with 4-MUP and zinc, with zinc but not 4-MUP, with 4-MU and zinc, with 4-MUP but not zinc, and with 4-MUP and magnesium instead of zinc. **(B)** Proton NMR spectra of Apollon 2 (which contains a 5' hydroxyl group and is labelled Apo) and 5' phosphorylated Apollon 2 (which contains a 5' phosphate group and is labelled P-Apo) with 4-NPP and zinc, with zinc but not 4-NPP, with 4-NP and zinc, with 4-NPP but not zinc, and with 4-NPP and magnesium instead of zinc. Spectra were measured in the presence of 300 μM DNA, 450 μM 4-MUP or 450 μM 4-NPP, 200 mM KCl, 50 mM HEPES pH 7.4, and 1 mM $ZnCl_2$ if not stated otherwise.

3.9 Multiple turnover assay

One limitation of conventional *in vitro* selection experiments is that catalytic molecules must modify themselves in order to survive the selection step. For this reason, many deoxyribozymes and ribozymes identified by conventional selections catalyse single turnover reactions. A common strategy to overcome this limitation is to divide a catalytic motif into a substrate strand (which in our case contains the reaction site) and an enzyme strand (which contains the rest of the motif)[44,92,104]. If a substrate strand is in excess, a single enzyme strand can bind, react with, and release multiple substrate strands (Figure R_29a). For designing such system, the knowledge of the secondary structure of the deoxyribozyme is needed.

In the case of Apollon, we found that deleting stem 2 only reduced activity by 20% (Figure R_12c). This suggested that it might be possible to generate a bimolecular version of Apollon by deleting the loop in the second stem of Apollon. This turned out to be correct, and the catalytic activity of the bimolecular construct (tested using a 1:1 ration of the enzyme strand to the substrate strand) was similar to that of the unimolecular one.

To further investigate the extent to which this variant could promote multiple turnover phosphorylation, a series of bimolecular constructs were generated. The number of base pairs in the helix which is formed between the enzyme strand and the substrate strand were varied (Figure R_29b, Table 1). When incubated in the presence of an excess of the substrate strand, an enzyme strand generated a stronger colorimetric signal than a unimolecular deoxyribozyme incubated at the same concentration (Figure R_29c), suggesting that it promotes multiple turnover catalysis. We also investigated the effect of adding or deleting base pairs between the enzyme strand and substrate strand on the efficiency of the multiple turnover reaction. Constructs with short recognition helices were expected to be limited by substrate strand binding, whereas constructs with long helices were expected to be instead limited by substrate strand release[44]. For Apollon, the optimal length of the helix between the enzyme strand and substrate strand was seven base pairs. This construct catalysed approximately ten turnovers when the reaction was followed for seven days (Figure R_29d). These results show that Apollon can catalyse multiple turnover reaction, and that this can enhance the colorimetric signal.
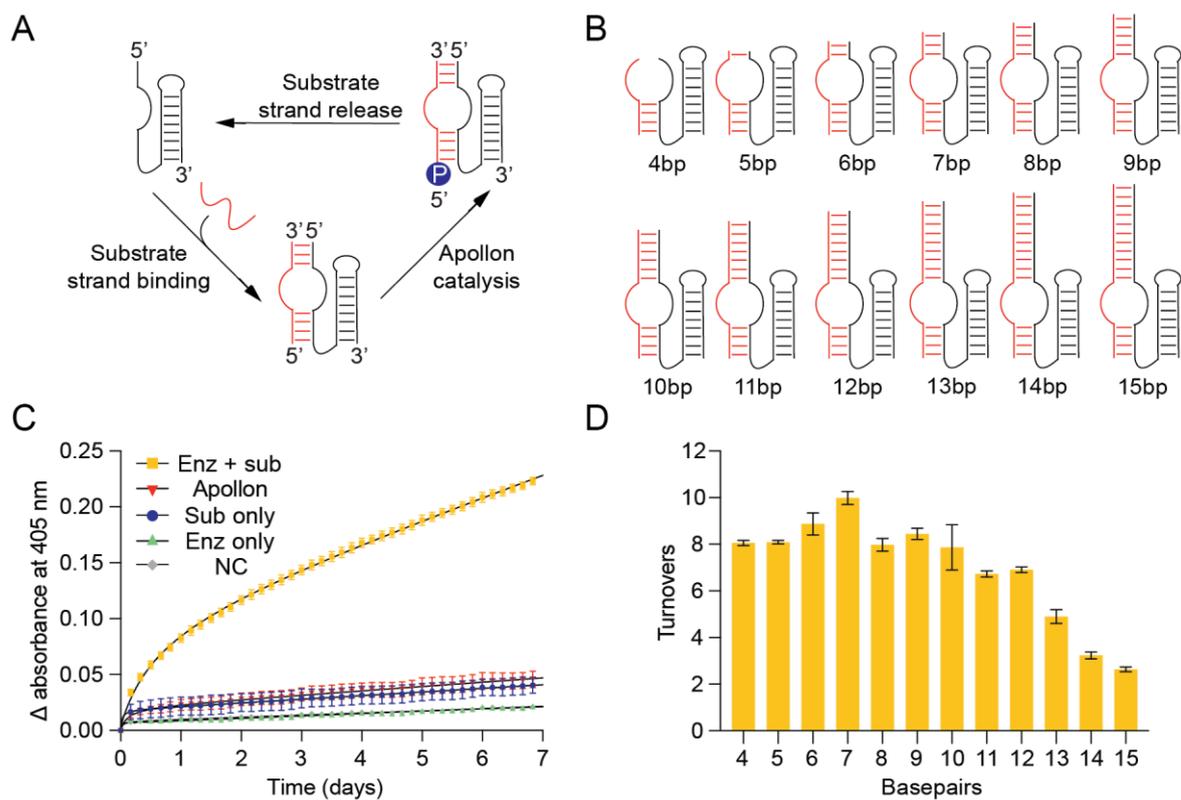
Figure R_29. Multiple turnover reaction of Apollon. (A) Overview of the multiple turnover reaction. **(B)** Schematic representation of constructs with substrate binding helices of different lengths. **(C)** Colorimetric assay comparing the signal generated by a multiple-turnover construct containing a substrate binding helix of seven base pairs with that generated by a unimolecular single-turnover construct. NC shows the signal generated by 4-NPP in buffer alone. **(D)** Graph showing the number of turnovers catalysed by different constructs in seven-day incubations.

3.10 Engineered forms of Aurora and Apollon can detect a range of inputs and enzymes in homogeneous assays

Because the workflow of the reactions catalysed by Aurora and Apollon contains few steps and requires little experimental manipulation, sensors that only generate fluorescence or color in the presence of an input of interest could be useful for applications such as high-throughput screening and diagnostics. Deoxyribozyme sensors that are activated in solution and that can be performed in a single tube without the need for wash steps or additional biochemical purifications would be especially useful. To show that this is feasible, we developed three different sensor architectures based on Aurora and Apollon.

The first type of sensor only generated a signal in the presence of oligonucleotides with specific sequence (Figure R_30). Such sensors could in principle be useful for applications such as diagnostics and molecular computing[105]. To make such sensor, part of sequence of the target oligonucleotide was added into different part of Aurora or Apollon and the full reverse complement of the target was added to the 3' end. Because these two insertions are complementary, they were expected to interact by base pairing and interfere with correct formation of Aurora or Apollon and therefore inhibit deoxyribozyme activity (Figure R_30a, f). Consistent with our model, in the absence of the target oligonucleotide, these sensors were inactive (Figure R_30c, g). However, in the presence of the target oligonucleotide, base paring between the target and the 3' end of the deoxyribozyme prevented this inhibitory interaction from occurring and catalytic activity was restored (Figure R_30c, g). These sensors generated a fluorescent or yellow product only in the presence of the correct target oligonucleotide (Figure R_30b, f) with signal-to-noise ratios (defined here as the signal in the presence of the target oligonucleotide divided by the signal in the absence of the target oligonucleotide) of ~10-fold (Figure R_30c, g). A limitation of this type of senor was low sensitivity, with a limit of detection of approximately 1 μM of target for the Aurora oligo sensor and 20 μM of target for the Apollon oligo sensor (Figure R_30d, h). In addition, some off-target activation was observed (Figure R_30b, f). These observations suggest that our proof-of-principle sensor would benefit from further optimization which could possibly be achieved by reselection[106–109]. The low sensitivity of these sensors is likely related to catalytic turnover because, unlike classical enzymes, a single molecule of the deoxyribozyme sensor can only generate one molecule of fluorescent or colorimetric product.
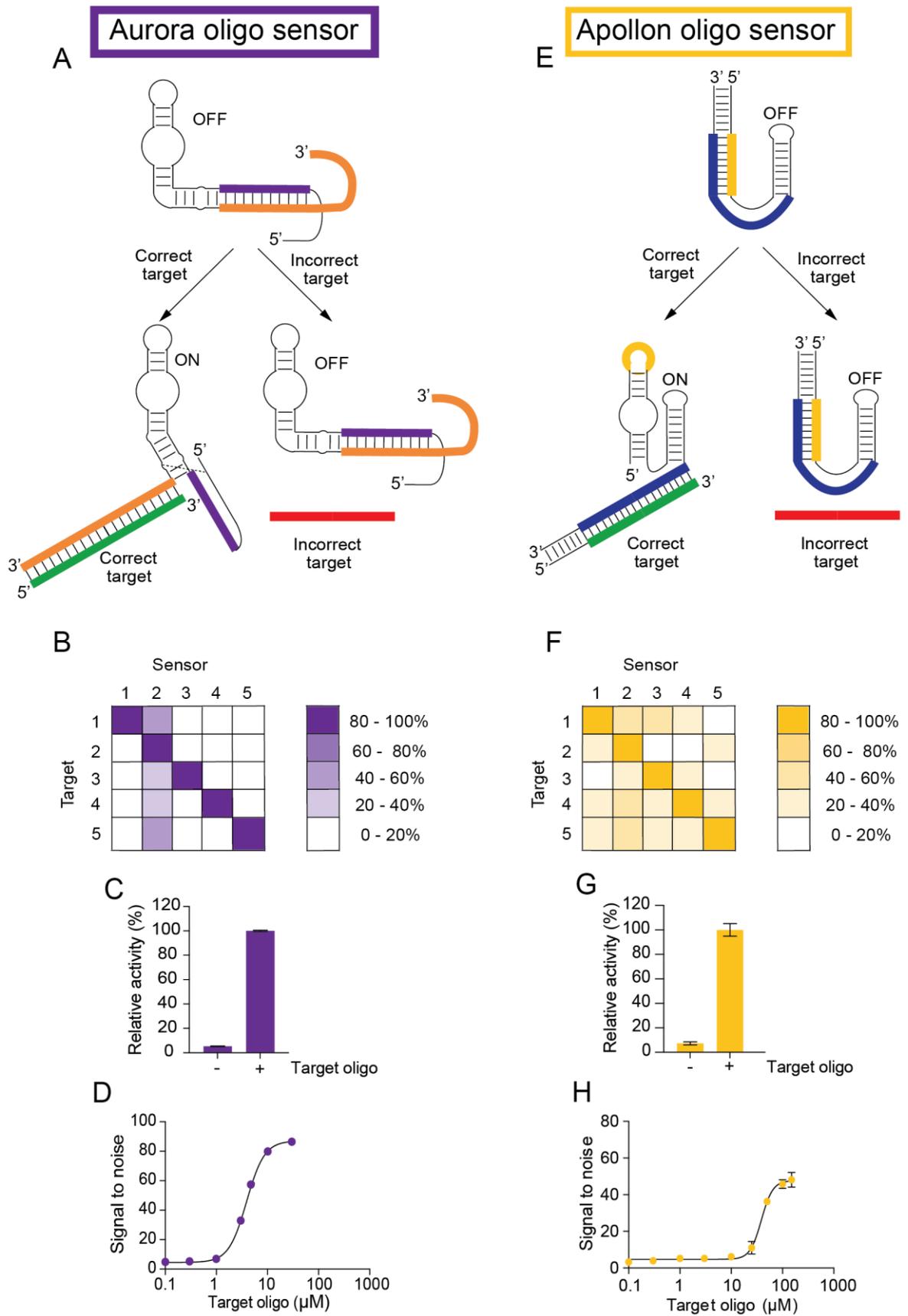
3.10.1 Oligonucleotide sensor

Figure R_30. Programmable oligonucleotide sensors based on Aurora or Apollon. **(A)** Architecture of an Aurora oligonucleotide sensor that detects short oligonucleotides. **(B)** Specificity of the Aurora oligonucleotide sensor. Five different sensors were generated, each of which was designed to detect a different target oligonucleotide. Each square represents a different sensor/target combination. **(C)** Production of fluorescence in the absence (left) and presence (right) of the target oligonucleotide. **(D)** Detection limit of the Aurora oligonucleotide sensor. Reactions were incubated for four hours in the presence of the indicated concentration of the target oligonucleotide, and, after quenching with base, fluorescence was measured using a plate reader. **(E)** Architecture of an Apollon oligonucleotide sensor that detects short oligonucleotides. **(F)** Specificity of the Apollon oligonucleotide sensor. Five different sensors were generated, each of which was designed to detect a different target oligonucleotide. Each square represents a different sensor/target combination. **(G)** Production of color in the absence (left) and presence (right) of the target oligonucleotide. **(H)** Detection limit of the Aurora oligonucleotide sensor. Reactions were incubated for twenty-four hours in the presence of the indicated concentration of the target oligonucleotide, and then absorbance at 405 nm was measured using a plate reader. See Table 1 for the sequences of deoxyribozymes used in the experiments described in this figure.

3.10.2 Ribonuclease sensor

To overcome the sensitivity problem of the oligonucleotide sensor, we investigated whether it was possible to link the single-turnover signal generated by Aurora or Apollon to the catalytic activity of an enzyme that itself catalyses a multiple turnover reaction. Because Aurora and Apollon are made of DNA, the natural solution was to focus on nucleic acids modifying enzymes. With this idea in mind, we developed Aurora and Apollon sensors that are activated by enzymes that cleave RNA. These ribonuclease sensors were constructed by fusing a short DNA oligonucleotide containing a ribonucleotide at its 3' end to the 5' end of Aurora or Apollon (Figure R_31a, d). Aurora and Apollon require a free 5' hydroxyl group for activity, and this modification was therefore expected to abolish catalytic activity. Some ribonucleases cleave RNA at internal sites to generate 3' phosphate (or 2'-3' cyclic phosphate) and 5' hydroxyl termini, and in the presence of such ribonucleases, the RNA linkage in our sensors should be cleaved, the 5' end of Aurora or Apollon should be regenerated and catalytic activity should be restored (Figure R_31b, e). Because protein ribonucleases are capable of multiple turnover catalysis, we also expected that this mechanism would amplify the single-turnover signal generated by Aurora or Apollon.

We tested our ribonuclease sensors using ribonuclease A. Ribonuclease A activated our ribonuclease sensors and signal-to-noise ratios (defined as signal of Aurora RNase sensor or Apollon RNase sensor in the presence of RNase A divided by the signal of Aurora RNase sensor or Apollon RNase sensor in the absence of RNase A) were more then 10-fold for both deoxyribozymes. Furthermore, the detection limits of these sensors were

approximately 10,000-fold lower than for our oligonucleotide sensors (Figure R_31c, f). This increase in sensitivity is likely due to the high turnover number of RNase A. To further probe the mechanism of these sensors, we investigated whether this activation was affected by RiboLock, a commercially available inhibitor of RNase A. RiboLock had no effect on Aurora or Apollon itself, but prevented the Aurora and Apollon ribonuclease sensors from being activated by RNase A (Figure R_31b, e). These results provided additional evidence that our sensors are activated by RNA cleavage. Taken together, these experiments show that assays, which use a covalently blocked form of Aurora or Apollon to detect a multiple-turnover enzyme, can be orders of magnitude more sensitive than those that use unmodified ones.
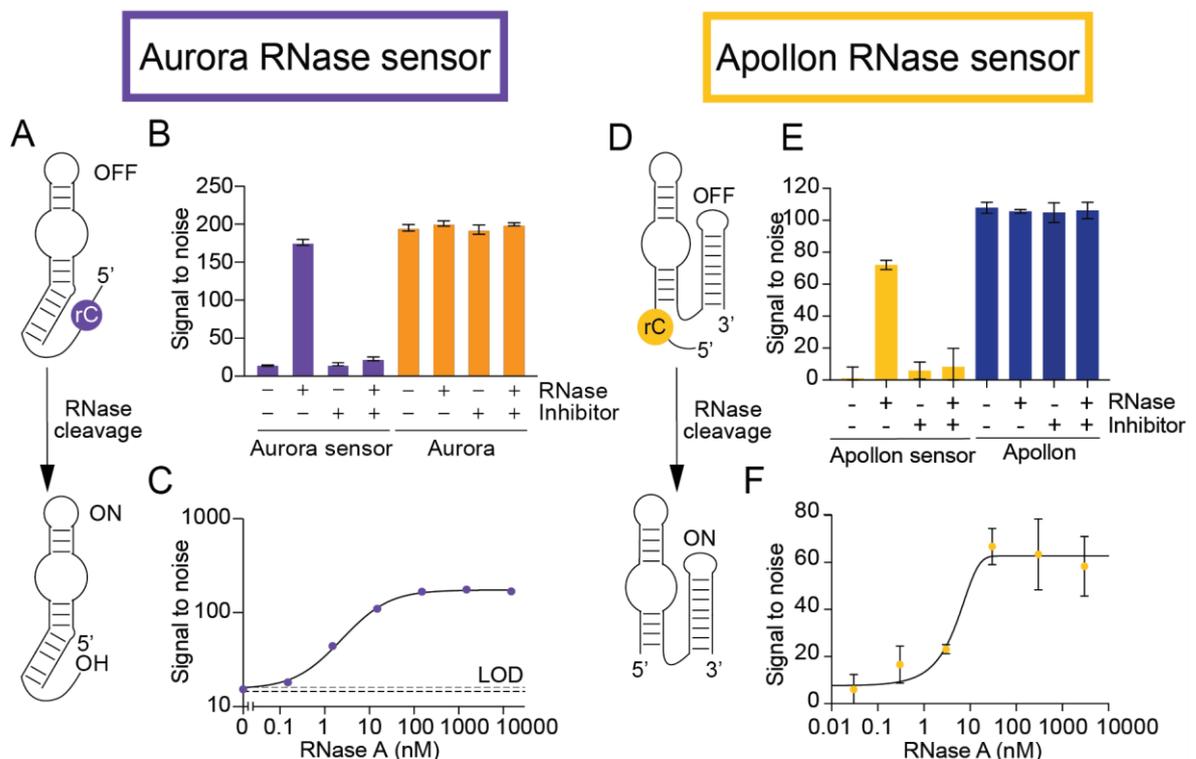
Figure R_31. Ribonuclease sensors based on Aurora and Apollon. (A) Architecture of an Aurora RNase sensor that detects ribonucleases. **(B)** An Aurora RNase sensor is activated by RNase A, but not when a ribonuclease inhibitor is present. In contrast, the catalytic activity of Aurora itself is not affected by either RNase A or this ribonuclease inhibitor. **(C)** The Aurora RNase sensor detects ribonuclease A with a limit of detection of 100 pM. **(D)** Architecture of an Apollon RNase sensor that detects ribonucleases. **(E)** An Apollon RNase sensor is activated by RNase A, but not when a ribonuclease inhibitor is present. The catalytic activity of Apollon itself is not affected by either RNase A or this ribonuclease inhibitor. **(F)** The Apollon RNase sensor detects ribonuclease A with a limit of detection of low nanomolar range.

3.10.3 Nuclease sensor

A third type of sensor used unmodified Apollon to detect nucleases that cleave DNA (Figure R_32a). We note that in principle unmodified Aurora could also be used to construct such a sensor. When Apollon was incubated in the absence of nuclease, Apollon generated a colorimetric signal. However, in the presence of nuclease, Apollon was degraded, which resulted in a reduction of the colorimetric signal. We tested this design using Apollon and the nucleases DNase I and Exonuclease I (Figure R_32b, c). Signal-to-noise ratios (defined as signal of Apollon in the absence of nuclease divided by the signal of Apollon in the presence of nuclease) of more than 20-fold were achieved, with detection limits between 10 and 100 nanomolar (Figure R_32d). This type of sensor can also distinguish between reactions that contain DNase or exonuclease inhibitors and those that do not (Figure R_32b,c), suggesting that these sensors could be used to identify inhibitors of therapeutically important nucleases in high-throughput screens[110–112].
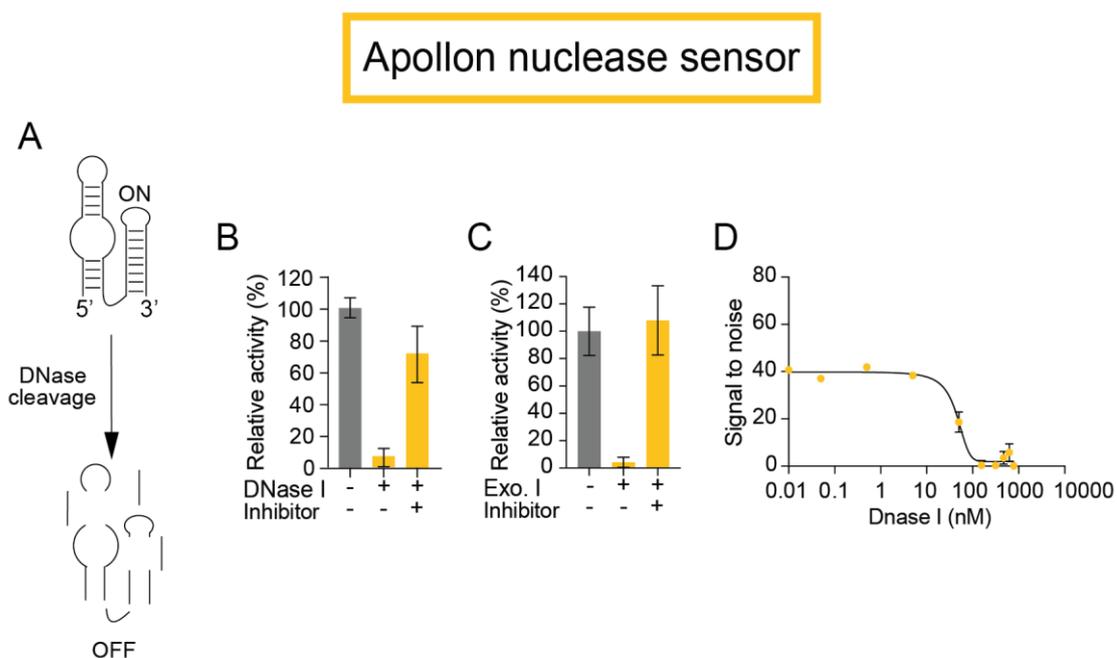
Figure R_32. Apollon nuclease sensor. (A) Architecture of an Apollon sensor that detects nucleases. **(B)** DNase I inactivates the Apollon nuclease sensor, but not when a DNase inhibitor is present. **(C)** Exonuclease I inactivates the Apollon nuclease sensor, but not when an Exonuclease inhibitor is present. **(D)** Sensitivity of the Apollon nuclease sensor for DNase I.

3.10.4 Using Aurora to identify Nsp15 inhibitors in a high-throughput screen

Assays with the Aurora or Apollon ribonuclease sensor can be performed rapidly and inexpensively, and they therefore appear to be well-suited for applications such as high-

throughput screens. To further probe this idea, we investigated whether the Aurora ribonuclease sensor could be used to identify inhibitors of the SARS-CoV-2 endoribonuclease Nsp15. This ribonuclease cleaves 3' of pyrimidines (preferentially uridines) to generated 2'-3' cyclic phosphate and 5' hydroxyl termini[113]. Nsp15 helps to prevent host recognition by degrading double-stranded viral intermediates[113]. Inhibitors of Nsp15 could be potentially used as antiviral agents[114]. Pilot experiments showed that it was possible to construct an Aurora Nsp15 sensor (Table 1) that was activated by Nsp15, and we used this to performed a screen for Nsp15 inhibitors.

A master mix containing Nsp15 and buffer was aliquoted into the wells of 384 well plates, each of which contained a different compound from a 1000-member fragment-based small molecule library (Figure R_33a). A second master mix containing the Aurora Nsp15 sensor was then added to each well. After a short incubation to allow Nsp15 to cleave the RNA linkage and to activate the sensor, zinc and 4-MUP were added to initiate Aurora catalysis. After another incubation, fluorescence was measured in a plate reader (Figure R_33b). In wells containing compounds that do not inhibit Nsp15, the RNA linkage in the Aurora Nsp15 sensor was expected to be cleaved by Nsp15, which should lead to production of a fluorescent signal (Figure R_33b, black points). In contrast, RNA cleavage should not occur and fluorescence should not be produced in wells containing compounds that inhibit either Nsp15 or Aurora (Figure R_33b, orange points). To distinguish compounds that inhibit Nsp15 from those that inhibit Aurora, a counter screen was performed using Aurora instead of the Aurora Nsp15 sensor (Figure R_33c). This counter screen revealed that none of the hits identified in the initial screen inhibited Aurora. This result indicates that these hits are Nsp15 rather than deoxyribozyme inhibitors, and also that they do not quench fluorescence of 4-MU itself.

To confirm our results and to compare them to those obtained using standard methods, the screen was repeated using a FRET assay in which Nsp15 was incubated with library members and a FRET substrate containing a fluorophore at one end and a quencher at the other (5'FAM-AAArUAA-BHQ1-3', Figure R_33d). Cleavage by Nsp15 was expected to result in an increase in fluorescence, while fluorescence in wells containing Nsp15 inhibitors was expected to remain at background levels. The results of this FRET screen were almost identical to those obtained using the Aurora Nsp15 sensor (Figure R_33e). Several selected hits were further characterized as a function of concentration using the Aurora Nsp15 sensor and the FRET substrate. The most potent of these

compounds inhibited Nsp15 with an IC_{50} of 12 μM in an assay that used our Aurora Nsp15 sensor and 11 μM in an assay that used FRET substrate (Figure R_33f). Other tested hits inhibited Nsp15 with IC_{50} values ranging from 7.9 μM to 22 μM (Figure R_34). Taken together, these experiments have demonstrated that Aurora Nps15 sensor can be used in combination with small-molecule libraries to rapidly identify inhibitors in high-throughput screens.
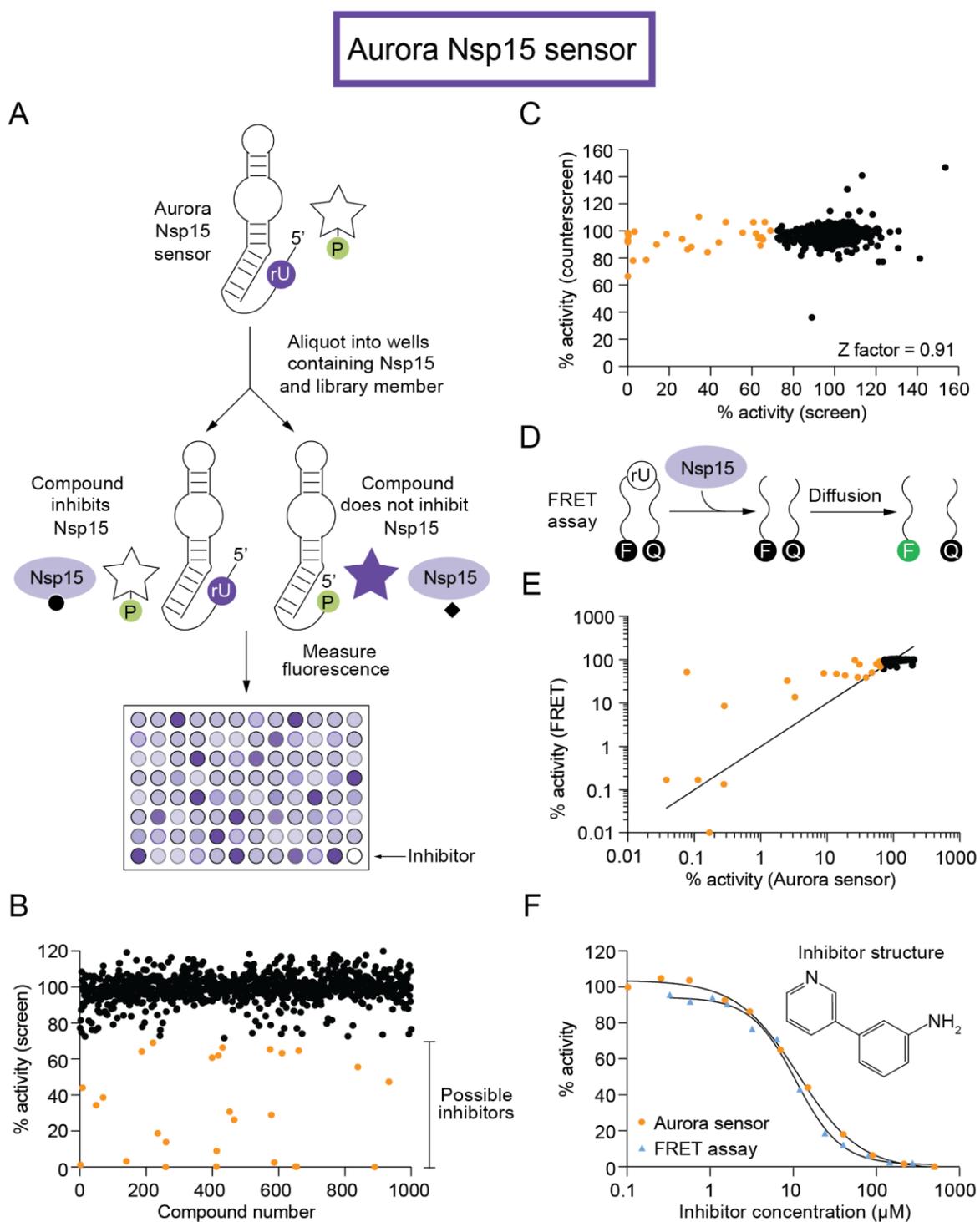
Figure R_33. Identification of small-molecule inhibitors of the SARS-CoV-2 ribonuclease Nsp15 using a fluorescent Aurora Nsp15 sensor. (A) Workflow of high-throughput screen to identify Nsp15 inhibitors. **(B)** Effect of each compound in the 1000-member library on the fluorescence of the Aurora Nsp15 sensor. Potential inhibitors are shown in orange. **(C)** Identification of inhibitors and false positives. The *x*-axis of the graph shows the fluorescent signal generated by the Aurora Nsp15 sensor in the presence of Nsp15 and different compounds in the library, while the *y*-axis shows the fluorescent signal generated by Aurora itself in the presence of Nsp15 and the same compounds. Points with high fluorescence values on both the *x*-axis and the *y*-axis (shown in black) correspond to wells containing compounds that inhibit neither Nsp15 nor Aurora. Points with low fluorescence values on the *x*-axis and a high fluorescence value on the *y*-axis (shown in orange) correspond to wells containing compounds that inhibit Nsp15 but not Aurora. **(D)** Workflow of a FRET assay for ribonuclease activity. **(E)** Comparison of the results of a high-throughput screen for Nsp15 inhibitors using the Aurora Nsp15 sensor (*x*-axis) with a screen of the same library using a FRET assay (*y*-axis). **(F)** Example of an Nsp15 inhibitor identified in the screen. This compound inhibits Nsp15 with an IC50 value of 12 μM when measured using the Aurora Nsp15 sensor and 11 μM when measured using the FRET assay.

3.10.4.1 IC_{50} values for Nsp15 inhibitors measured using the Aurora sensor or FRET assay
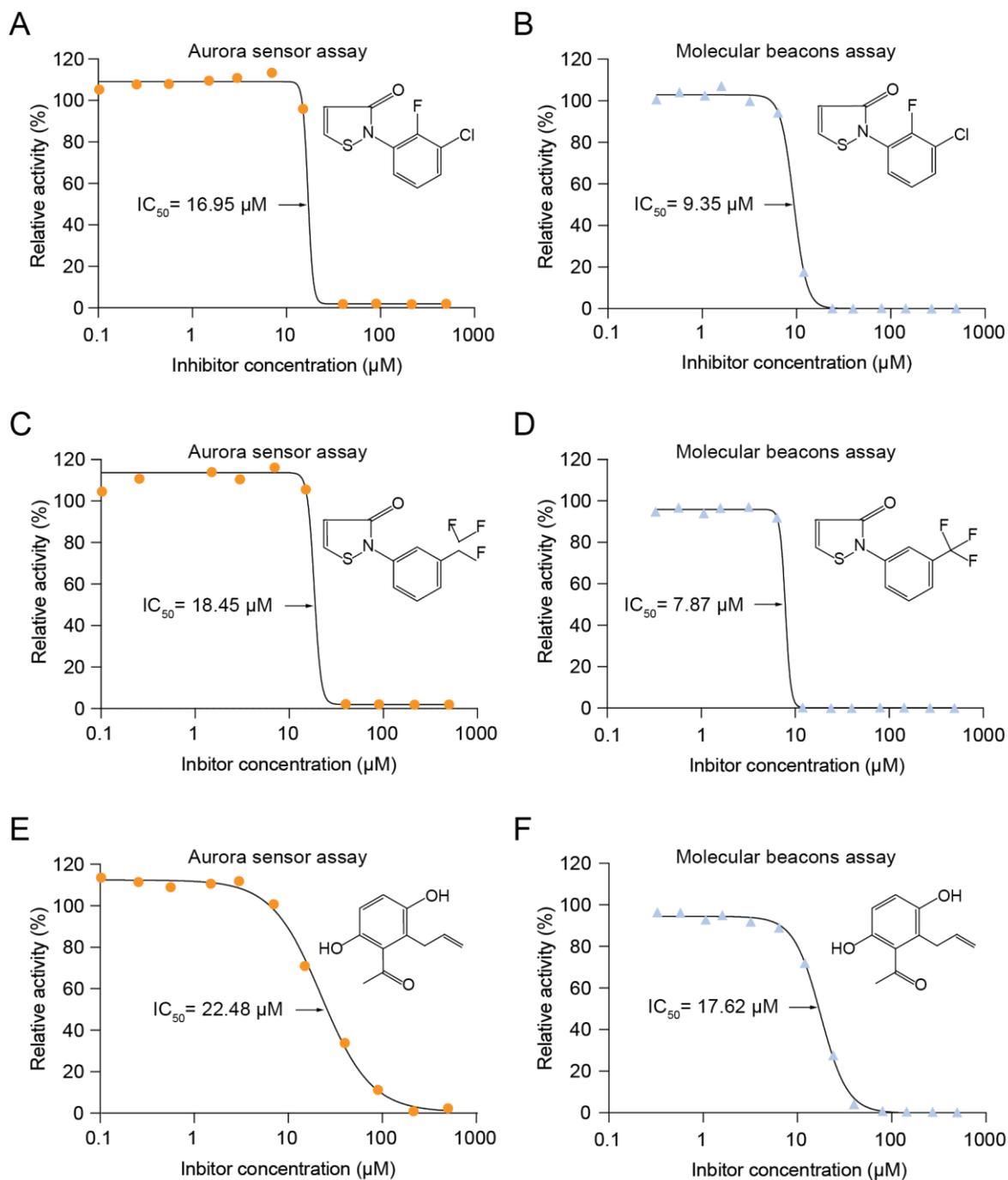
Figure R_34. IC_{50} values for Nsp15 inhibitors isolated in the high-throughput screen. Experiments were performed using an Aurora Nsp15 sensor (left column) and a FRET assay (right column).

4. Discussion

To link the presence of a target molecule with an easily detectable signal is desirable for many types of applications. Since the discovery of catalytic RNA[23,115] and artificial evolution[39–41], a major focus in the field has been to isolate functional nucleic acids that can act as signalling molecules or serve as sensors for detection of various therapeutically important targets[116–118].

Chemiluminescent and fluorescent signals are suitable for certain applications thanks to their high sensitivity, low background and high signal stability, but expensive instrumentation is needed to analyse the results. On the other hands, assays that monitor changes in the color of a sample are less sensitive, but can be analysed without expensive equipment (for example by smartphones or in the simplest case by eye). A light-producing deoxyribozyme called Supernova[42,86] was previously identified in our laboratory, and we therefore focused in this study on the development of fluorescence and color-producing deoxyribozymes.

Currently, the most efficient fluorescence producing systems are based on RNA aptamers. Many fluorogen-activating RNA aptamers with diverse functional properties have been isolated in recent years[119–121,69,68]. Engineered variants of these aptamers were further used to investigate the functions and localisation of cellular RNA molecules, or to monitor metabolite concentrations in real time[83,85]. Although extremely useful for studies of biological systems, such aptamers are less suitable for *in vitro* application such as high-throughput screening, because they are unstable under many conditions due to the ubiquitous presence of ribonucleases[122].

DNA counterparts are more stable and cheaper than fluorogen-activating RNA aptamers. Despite these facts, only few DNA motifs that generate fluorescent signals have been developed[75,123,124]. DNA aptamers that bind and enhance the fluorescence of ligands typically have signal to noise ratios around 100-fold, which is significantly lower than fluorogen-activating RNA aptamers[124]. In addition, in the case of fluorogen-activating DNA or RNA aptamers the fluorophore must remain associated with the aptamer to generate a signal, therefore this approach provides a less permanent and robust readout than a signal generated by a catalyst like a ribozyme or deoxyribozyme.

One type of deoxyribozyme that produces signals was previously discovered by the Sen group[76]. This is a peroxidase deoxyribozyme that converts the substrate ABTS into a colorimetric product in the presence of hydrogen peroxidase and hemin[76,77]. Although this

deoxyribozyme is typically used to generate a chromogenic product, a fluorogenic or light signal can also by generated when this reaction is performed with substrate such as tyramine[78] or luminol[79]. This deoxyribozyme is small, inexpensive to synthetize, and uses a widely available substrate. It is active over range of conditions, including activity at pH 3[125] and activity at 95 °C[126]. Disadvantage of this deoxyribozyme is, that it requires hydrogen peroxide to the reaction, which is not compatible with many assays and has a high background, because the peroxidase reaction is also promoted by hemin itself.

Another approach how to generate fluorescence and chromogenic signals is to use the protein enzyme alkaline phosphatase. A big advantage of alkaline phosphatase is its remarkable catalytic efficiency (k_{cat}/K_M), which approaches the diffusion-controlled limit[90], and its ability to catalyse multiple turnover reactions. On the other hand, alkaline phosphatase is more laborious to manufacture and less stable than deoxyribozymes. Alkaline phosphatase is also nonspecific with regards to its substrate and readily dephosphorylates 4-MUP, 4-NPP, and CDP-Star, which means that it is not suitable for multiplex assays. It is also more difficult to modify proteins by rational design or artificial evolution than DNA or RNA, which is a significant limitation with respect to a sensor development.

These limitations suggested to us that deoxyribozymes, which generate strong fluorogenic and chromogenic signals, could be useful additions to the toolkit of functional DNA parts. Such deoxyribozymes would likely produce a more stable signal than fluorogen-activating RNA/DNA aptamers, have higher signal-to-noise ratios than the existing peroxidase deoxyribozyme, and should also be relatively easy to convert into allosterically modulated sensors. We selected fluorogenic 4-MUP and chromogenic 4-NPP as substrates. Both substrates are cheap, commercially available and produce strong signals after dephosphorylation. Deoxyribozymes that catalyse transfer of a phosphate group from small molecules to their own 5' hydroxyl group have been described previously[42,43,127]. Moreover, the deoxyribozyme Hit 1, which we used as the starting point for our library, shows weak activity with 4-MUP and 4-NPP. These preliminary results suggested that the identification of deoxyribozymes that that efficiently generate strong fluorogenic and chromogenic signals was a feasible goal.

Indeed, such deoxyribozymes were successfully isolated by *in vitro* selection. Under optimal reaction conditions, Aurora enhances fluorescence 700-fold relative to the

reaction in the absence of this deoxyribozyme, and Apollon enhances formation of a yellow product 100-fold relative to the reaction in the absence of this deoxyribozyme.

Knowledge of the secondary structure and minimal catalytic core of a deoxyribozyme is not necessary, but extremely beneficial for sensor development. Since initial *in vitro* selection experiments was done with library based on Hit 1, we expected that new deoxyribozymes would have the same fold as Hit 1, but only differ in their reactivities towards 4-MUP or 4-NPP. However, both Aurora and Apollon have distinct secondary structures. For this reason, additional selections were done to provide information about the neighbouring sequence space of our deoxyribozymes. By analyzing these datasets using mutual information and comparative sequence analysis[93–95] it was possible to determine the minimal catalytic core and secondary structure of our deoxyribozymes. The minimal catalytic core of Aurora is 47 nucleotides long and this variant is even more active than the 85 nucleotide full-length construct. The minimal catalytic core of Apollon is 39 nucleotides long. However, the activity of a 50 nucleotide long variant is about 20% higher, and this construct was therefore used for most of the experiments. Minimization of our deoxyribozymes reduced the cost of synthesis and also improved activity.

The secondary structure of Aurora could not be easily predicted by Mfold[57], possibly because of tertiary interactions among the 5' terminus, loop regions, and the asymmetric bulge. On the other hand, the secondary structure of Apollon was successfully predicted. However, for secondary structure determination we instead used comparative sequence analysis together with NMR spectroscopy and double mutant cycles. Such secondary structure determination is more laborious than simple prediction, but provides more convincing results. Such analysis revealed that the secondary structure of Aurora is an 11-base pair long hairpin interrupted by an asymmetric bulge. Our secondary structure model of Apollon is the same as that predicted by Mfold, and it consists of three stems, one with a highly conserved primary sequence.

The optimization of reaction conditions was extremely useful not only in terms of increasing signal to noise ratios, but also in providing valuable information about limits in which our deoxyribozymes were still active. This information was also useful for sensor development. Signal to noise ratios were increased by increasing the concentration of our deoxyribozymes and by decreasing the concentration of substrates form 1 mM to 30 μM for 4-MUP and from 1 mM to 100 μM for 4-NPP. In the case of Aurora, 5% (v/v) DMSO

also increased the signal and was therefore used in the reaction buffer. An additional improvement came from the observation that the maximum fluorescent yield of 4-MU is at pH 9.4[128], which is not compatible with the pH dependence of Aurora. However, by increasing the pH after the Aurora reaction it was possible to increase the signal-to-noise ratio 6-fold.

Despite the fact that Aurora and Apollon were isolated from the same library, they have different secondary structures and functional properties. For example, Aurora requires potassium, while Apollon does not. However, both Aurora and Apollon need multiple zinc ions for function. Moreover, activity is affected in a highly cooperative manner by zinc concentration, as is activity of the protein enzyme alkaline phosphatase. These findings suggest that zinc plays a crucial role in phosphoryl transfer reactions catalyzed by nucleic acids[100–103]. Our knowledge of the optimal reaction conditions and secondary structures of Aurora and Apollon was particularly useful for sensor development.

As an initial proof of concept, we showed that Aurora and Apollon can be converted by rational design into an oligonucleotide sensor that generates a fluorescent or colorimetric signals only in the presence of a target oligonucleotide. However, the detection limits of these sensors were quite high (in the micromolar range), in part because one molecule of sensor can only produce one molecule of product. These detection limits are far higher than other widely used techniques for oligonucleotide detection such as fluorescent *in situ* hybridization[39], polymerase chain reaction[37,129], and rolling circle amplification[130].

To increase the sensitivity of our deoxyribozyme sensors, we focused on enzymes that modify DNA or RNA. A key point is that one molecule of a protein enzyme can typically activate or deactivate many molecules of the sensor, which leads to dramatic increase in sensitivity. Indeed, our most sensitive sensor could detect ribonucleases with a detection limit of approximately 100 pM. This is comparable with the detection limits of many homogeneous assays that use combinations of aptamers with signalling elements such as dyes, fluorophores, gold nanoparticles or quantum dots[35–38]. We also note that the architecture of our ribonuclease sensor could not be easily adapted to a sensor based on either alkaline phosphatase or the peroxidase deoxyribozyme. In the case of alkaline phosphatase, an RNA oligonucleotide covalently linked to the protein sequence would not be a substrate for most ribonucleases. In the case of the peroxidase deoxyribozyme, an RNA oligonucleotide covalently linked to the 5' hydroxyl group would not be expected to

abolish deoxyribozyme activity, because this deoxyribozyme does not use its 5' hydroxyl group as the reaction site.

After verifying that our ribonuclease sensor worked, we used it to identify inhibitors of the Nsp15 ribonuclease from SARS-CoV-2 in a high-throughput screen. This screen could distinguish between reactions that contained an inhibitor of Nsp15 ribonuclease and those that did not. The screen also provided similar results to those obtained in a parallel screen that used a more standard FRET assay[131,132]. Our deoxyribozyme-based assay is comparable to the FRET assay in terms of simplicity and workflow. However, reagents costs are several fold lower and signal-to-noise ratios are higher for our assay.

A final type of sensor used unmodified Apollon to detect different nucleases. When Apollon was degraded by the nuclease, catalytic activity was lost and no signal was generated. Pilot experiments were done with the nucleases DNase I and Exonuclease I, detection limits were in the nanomolar range, and it was possible to distinguish between reactions that contained nuclease inhibitors and those that did not. A number of nucleases have been described for which finding an inhibitor in a high-throughput screen would be desirable. For example, nucleases such as Sda1[111] from group A *Streptococcus* can play roles in bacterial infections, while nucleases such as DNase γ[110,133] are associated with inflammatory diseases.

In addition to sensors developed in our group, other groups have also started to construct sensors using our deoxyribozyme. For example, the Li group used our light-producing deoxyribozyme Supernova[42,86] as a signalling component in their assay for nucleic acid detection. The CRISPR/Cas12a system was used as a signal amplifier. In the presence of a target viral sequence, Cas12a was activated by a conformation change of the guide RNA. Activated Cas12a acted as exonuclease and degraded Supernova, which resulted in a reduction of light production. The limit of detection of this assay was reported to be in the low picomolar range, and the assay was able to detect a single molecule of target sequence in one microliter of solution[134].

Although these examples show that sensors can be readily constructed by rational design, it is likely that *in vitro* selection could be utilized to improve sensors performance[135–137]. One way to do this is to start with an existing aptamer against a therapeutically important target such as a pathogen[118,138,139], a viral protein[140,141], or a human protein[66,142]. The sequence of the aptamer is inserted into a loop or less conserved

part of a deoxyribozyme and this construct is then randomly mutagenized at a low rate. A conventional *in vitro* selection experiment is then performed using alternating rounds of positive selection (in the presence of the target molecule) and negative selection (in the absence of the target molecule). The final deoxyribozyme sensor should only be active in the presence of the selected target. Combining an existing aptamer with a low rate of mutagenesis makes it possible to explore the neighbouring sequence space of the deoxyribozyme-aptamer fusion for variants with high signal-to-noise ratios. Another strategy, which we are exploring is to insert random sequence regions (rather than aptamers) into loops or less conserved parts of deoxyribozymes. This strategy could in principle make it possible to obtain sensors against targets for which an aptamer is not yet known. If the size of the library is small enough (maximally 10^7 to 10^9 sequences), sensors can be identified in single step selections[45,143,144]. Such selections are much faster than conventional ones. If we can demonstrate that this approach works, it should be possible to rapidly generate sensors to a wide range of targets. However, for now we are exploring this approach using larger libraries and conventional selections. One target is the neuraminidase protein from the H1N1 (2009) pandemic influenza[145]. If sensor development is successful, we plan to develop Aurora sensors for detection of various viral proteins. Such panel of sensors could be used for quick and inexpensive diagnostic tests.

5. Conclusion

In this work, in vitro selection was used to isolate deoxyribozymes capable of generating fluorogenic or chromogenic signals. 4-methylumbelliferyl phosphate (4-MUP) was used as a substrate for fluorescence production, and 4-nitrophenyl phosphate (4-NPP) was used for color production. After dephosphorylation, 4-MUP is converted into the fluorogenic product 4-MU, and colorless 4-NPP is converted into the yellow product 4-NP. Our deoxyribozymes transfer the phosphate group from these substrates to their own 5' hydroxyl group. The deoxyribozyme that dephosphorylates 4-MUP and creates a fluorogenic signal was called Aurora, and the deoxyribozyme that dephosphorylates 4-NPP and creates a chromogenic signal was called Apollon.

After initial *in vitro* selection experiments, the most active deoxyribozymes were randomly mutagenized at a rate of 21%, and additional selections were performed to obtain more efficient variants. This made it possible to construct secondary structure models for two deoxyribozymes. The minimal catalytic core was found to be 47 nucleotides long for Aurora and 39 nucleotides long for Apollon. Kinetic characterization revealed the maximum k_{obs} of 0.18 min^{-1} for Aurora and the maximum k_{obs} of 0.32 min^{-1} for Apollon.

Various reaction conditions were extensively tested to obtain the best possible signal-to-noise ratio and to better understand Aurora and Apollon reaction mechanisms. Surprisingly, cerium and lead could be excluded from the selection buffer without reducing activity of either Aurora or Apollon. However, both deoxyribozymes required 1 mM ZnCl$_2$ for activity, and zinc could not be replaced by any other divalent ion. Aurora required potassium for activity, while Apollon activity was only slightly reduced when potassium was excluded from the reaction. Both were active in various monovalent ions. Aurora and Apollon had a narrow pH optimum around 7.4, which was consistent with the pH used during selection experiments. As long as the pH was 7.4, the specific buffering agent was not important for either Aurora or Apollon. DMSO and PEG200 increased the signal-to-noise ratioof Aurora, and 5% (v/v) DMSO was used in Aurora reactions. On the other hand, DMSO had no effect on Apollon reactions. After optimization, the highest signal-to-noise ratio was over 700-fold for Aurora and over 100-fold for Apollon.

It was also possible to convert both Aurora and Apollon into allosterically regulated sensors. Oligonucleotide sensors were constructed by rational design. These were activated only if a short oligonucleotide of a particular sequence was present in the solution. The

signal-to-noise ratio of the Aurora oligonucleotide sensor relative to the reaction in the absence of oligonucleotide was ~19-fold, and for the Apollon oligonucleotide sensor it was ~14-fold. Additionally, Aurora and Apollon were used as sensors for various nucleases. The Apollon ribonuclease sensor detected RNase A with a limit of detection in the nanomolar range. Apollon was also used as a sensor for the detection of DNase I and Exonuclease I, with a limit of detection in the nanomolar range for both nucleases. Aurora was also used as a sensor to detect RNase A and the Nsp 15 endoribonuclease from SARS-CoV-2. The Aurora ribonuclease sensor had the best sensitivity of all of our sensors, with a limit of detection in the picomolar range. The Aurora Nsp 15 sensor was used in a high throughput screen to identify small molecular inhibitors of the Nsp 15 endoribonuclease from SARS-CoV-2. These inhibitors were in the low micromolar range and were confirmed using an independent FRET assay.

In conclusion, Aurora and Apollon are expected to be valuable additions to the toolbox of functional nucleic acids. They can serve as signalling molecules for many different applications in both basic and applied research. This potential was demonstrated by the identification of inhibitors of Nsp 15 endoribonuclease from SARS-CoV-2 using the Aurora Nsp 15 sensor.

6. Methods

6.1 Design of *in vitro* selection libraries

The library used in initial selections for isolation of fluorogenic or chromogenic deoxyribozymes (see Pool1 in Table 1) was generated by randomly mutagenizing the 85-nucleotide long H1 variant of Supernova (a chemiluminescent deoxyribozyme previously discovered in our group[42]) at a rate of 21% per position. A 20-nucleotide-long primer-binding site was also attached to its 3' end.

The library used for the reselection of fluorogenic deoxyribozyme referred to as Aurora (see Pool2 in Table 1) was based on the sequence of Hit10 from the initial fluorogenic selection. This sequence produced fluorescence with the highest signal-to-noise ratio of any of the tested deoxyribozymes from the initial selection, although it contains the tenth highest read number in the high-throughput sequencing data. The 85-nucleotide long sequence of Hit10 was randomly mutagenized at a rate of 21% per position and a new 20-nucleotide-long primer-binding site was attached to its 3' end.

The library used for the reselection of chromogenic deoxyribozyme referred to as Apollon (see Pool3 in Table 1) was based on the sequence of Hit1 from the initial chromogenic selection. Hit1 produced color with the highest signal-to-noise ratio of any of the tested deoxyribozymes from the initial selection, and it also revealed the highest read number in the high-throughput sequencing data. The 85-nucleotide long sequence of Hit1 was randomly mutagenized at a rate of 21% per position and a new 20-nucleotide-long primer-binding site was attached to its 3' end.

6.2 Initial *in vitro* selection

Pool1 and blocking oligonucleotide (see REV1 in Table 1) were mixed in Milli-Q water, heated at 65°C for 2 minutes, and cooled at room temperature for 5 minutes, 5× selection buffer and disodium salt of the 4-methylumbelliferyl phosphate substrate (4-MUP) or the disodium salt of 4-nitrophenyl phosphate (4-NPP) were then added. Final concentrations were 1 μM Pool1, 1.5 μM REV1, 1× selection buffer (200 mM KCl, 1 mM $ZnCl_2$, 1 μM $Ce(NO_3)_4$, 0.1 μM $PbCl_2$, and 50 mM HEPES pH 7.4) and 1 mM 4-MUP or 1 mM 4-NPP. After incubation for 2.4 hours, DNA was purified by ethanol precipitation.

A 20-nucleotide-long oligonucleotide (see FWD1 in Table 1) was then ligated to Pool1 library members containing a 5' phosphate. To ensure effective ligation, the reaction was performed with a 35-nucleotide-long splint oligonucleotide (see Splint1 in Table 1) which was partially complementary to both FWD1 and the 5' end of Pool1. FWD1, Pool1, and Splint1 were mixed with Milli-Q water, heated at 65 °C for 2 minutes, and cooled at room temperature for 5 minutes. 10× T4 DNA ligase buffer and T4 DNA ligase were added. Final concentrations were 2.5 μM Pool1, 2.5 μM FWD1, 2.5 μM Splint1, 1× T4 DNA ligase buffer, and 0.5 Weiss units of T4 DNA ligase per 1.0 μg of Pool1. Ligation mixture was then incubated for 5 minutes at 37 °C. Ligated DNA molecules were then separated from un-ligated ones by 6% Urea-PAGE. DNA molecules that co-migrated with a 125-nucleotide-long marker were cut from the gel. Then they were eluted into 0.3 M NaCl solution, mixed with 5 μg of yeast tRNA (served as a co-precipitant), and concentrated by ethanol precipitation.

DNA was then amplified by PCR using Q5 HotStart DNA Polymerase and the FWD1r and REV1p primers (see in Table 1). Final concentrations were 500-times diluted DNA molecules (corresponding to reacted library members from Pool1 which had undergone ligation), 0.5 μM FWD1r, 0.5 REV1p, 1× Q5 reaction buffer, 1× Q5 high GC enhancer, 0.2 mM dNTPs, and 0.02 U Q5 HotStart DNA polymerase per 1.0 μl of the PCR reaction mixture. Double-stranded PCR products were isolated using a Macherey-Nagel PCR Clean-up kit.

Because the strand complementary to Pool1 was amplified using REV1p, which contains a 5' phosphate, it was possible to regenerate the single-stranded DNA Pool1 by digesting this strand with λ-exonuclease. This was achieved by incubating each 5 μg of double-stranded PCR product with Milli-Q water, 1× Lambda exonuclease reaction buffer, and 1 μl (5 U) of Lambda exonuclease in a volume of 50 μl. This reaction was incubated at 37°C for 60 minutes. The resulting 125-nucleotide-long single-stranded DNA molecules were purified using a Macherey-Nagel PCR Clean-up kit.

The last step was to regenerate the 5' end of the Pool1, so it could undergo another round of artificial evolution. This was possible because the FWD1r primer used in the PCR contained a single RNA linkage at its 3' end and the FWD1r primer could therefore be removed by base hydrolysis. This was accomplished by heating the 125-nucleotide long single-stranded DNA molecules in Milli-Q water at 65 °C for 2 minutes, cooling at room temperature for 5 minutes, and adding 10× base hydrolysis buffer (10× base hydrolysis

buffer contained: 200 mM Trizma base, 4 M KOH, and 40 mM EDTA). The reaction mixture was incubated at 90°C for 10 minutes. The resulting 105-nucleotide-long DNA molecules each containing a 5' hydroxyl group) were isolated by 6% Urea-PAGE, eluted into 0.3 M NaCl, and concentrated by ethanol precipitation.

For fluorogenic initial *in vitro* selection with 4-MUP and for chromogenic initial *in vitro* selection with 4-NPP, five rounds of this protocol were performed. The evolved library was then amplified by PCR, purified using a Macherey-Nagel PCR Clean-up kit, and sequenced by Eurofins Genomics using an amplicon paired-end sequencing run.

6.3 Reselection

Reselection conditions were the same as those used in the initial selections except for the following differences. For fluorogenic reselection with 4-MUP differences were: First, Pool2 was used instead of Pool1. Second, the library was incubated with 4-MUP for 14.4 minutes rather than 2.4 hours. Third, a new blocking oligonucleotide and a reverse primer (REV2/REV2p) were used (see in Table 1). This library was sequenced after the sixth round by Eurofins Genomics using an amplicon paired-end sequencing run. For chromogenic reselection with 4-NPP differences were: First, Pool3 was used instead of Pool1. Second, Pool3 was incubated with 4-NPP for 14.4 minutes rather than 2.4 hours. Third, different reverse primers (REV3/REV3p) and blocking oligonucleotide were used (see in Table 1). The Pool3 was sequenced after seven rounds of selection by Eurofins Genomics using an amplicon paired-end sequencing run.

6.4 Analysis of fluorescence and color production

Fluorescence production was measured as follows: oligonucleotides corresponding to individual sequences from evolved libraries were resuspended in Milli-Q water, heated at 65 °C for 2 minutes, and cooled at room temperature for 5 minutes. After adding 5× selection buffer or 5× Aurora buffer, samples were transferred to a white half-area 96-well plate (Corning). 4-MUP was then added. In continuous assays, fluorescence was measured for 4 hours using a Tecan Spark plate reader (Tecan Group). In discontinuous assays, after incubating for a specific time, samples were quenched with 20 μl of 1 M KOH and fluorescence was measured using a plate reader. In a typical experiment final concentrations were 15 μM of the tested oligonucleotide and either 1× selection buffer

(200 mM KCl, 1 mM $ZnCl_2$, 1 μM $Ce(NO_3)_4$, 0.1 μM $PbCl_2$, 50 mM HEPES pH 7.4) or 1×
Aurora buffer (200 mM KCl, 1 mM $ZnCl_2$, 50 mM HEPES pH 7.4, 5% (v/v) DMSO), and
30 μM 4-MUP. Fluorescence was measured with the following settings: excitation 358
(±5) nm, emission 455 (±5) nm, optimal gain, 30 flashes, Z position calculated from one
well in the plate.

Color production was measured as follows: oligonucleotides corresponding to individual
sequences from evolved libraries were resuspended in Milli-Q water, heated at 65°C for 2
minutes, and cooled at room temperature for 5 minutes. After adding 5× Apollon buffer,
samples were transferred to a clear half-area 96-well plates (Corning). 4-NPP was added,
and absorbance at 405 nm was measured for 4 hours using a TECAN Infinite M200 PRO
plate reader (TECAN Group) if not stated otherwise. Final concentrations in a typical
experiment were 30 μM of the tested oligonucleotide and 1× Apollon buffer (200 mM
KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), and 100 μM 4-NPP. Color production was
measured with the following settings: absorbance at 405 (±5) nm, 25 flashes, Z position
was calculated from one well in the plate.

6.5 Analysis of phosphorylation using a ligation assay

To measure the extent of 5' self-phosphorylation, oligonucleotides corresponding to
individual sequences from evolved libraries were resuspended in Milli-Q water, heated at
65°C for 2 minutes, and cooled at room temperature for 5 minutes. 5x Aurora buffer and 4-
MUP or 5× Apollon buffer and 4-NPP were then added. Final concentrations in a typical
phosphorylation reaction were 1 μM of the tested oligonucleotide, 1× Aurora buffer
(200 mM KCl, 1 mM $ZnCl_2$, 5% (v/v) DMSO and 50 mM HEPES pH 7.4), and 1 mM 4-
MUP or 1× Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), and
1 mM 4-NPP unless stated otherwise. Reactions were incubated for various times at room
temperature and stopped by the addition of EDTA to a concentration of 25 mM.
Oligonucleotides were then purified by ethanol precipitation, and reacted oligonucleotides
(now containing a 5' phosphate) were ligated to a short oligonucleotide as described in the
section "6.2 Initial *in vitro* selection". Reacted and unreacted molecules were separated by
6% Urea-PAGE. DNA was visualized by staining with GelRed using the protocol
recommended by the manufacturer. Gels were scanned using a Typhoon laser scanner and
the percentage of reacted and unreacted molecules was quantified using ImageQuant TL

software.

6.6 Calculation of signal-to-noise ratios

For fluorescence production signal to noise ratios were defined as the fluorescence of a sample in the presence of deoxyribozyme divided by the fluorescence of the sample in the absence of the deoxyribozyme. The background signal was defined as the fluorescence of $1\times$ Aurora buffer (200 mM KCl, 50 mM HEPES, pH 7.4, 1 mM $ZnCl_2$ and 5% (v/v) DMSO) and was subtracted before calculating signal to noise ratios.

For color production signal-to-noise ratios were defined as the absorbance of a sample at 405 nm in the presence of deoxyribozyme divided by the absorbance at 405 nm in the absence of the deoxyribozyme. The background signal was defined as the absorbance at 405 nm of $1\times$ Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES, pH 7.4), and was subtracted before calculating signal-to-noise ratios.

6.7 Optimization of reaction conditions

To maximize fluorescence or color production, we performed reactions over a range of conditions using Aurora 2 and Apollon 2 (see in Table 1). Optimal KCl, $ZnCl_2$, and HEPES concentrations, the optimal pH of the buffer and optimal temperature was determined by titration. We also tested the effects of different monovalent metal ions, divalent metal ions, buffering agents, organic solvents and molecular crowding agents on deoxyribozymes activity. Activity was measured by analysis of fluorescence or color production (using a plate reader assay) and self-phosphorylation (using a ligation assay).

6.8 Kinetic measurements and analysis

Kinetic measurements were performed using Aurora 1, Aurora 2, Apollon 1 and Apollon 2 (see in Table 1), and reactions were analysed using a ligation assay. Deoxyribozyme was mixed with Milli-Q water, heated at 65 °C for 2 minutes, and cooled at room temperature for 5 minutes. Aurora 1 or Aurora 2 was mixed with 5x Aurora buffer and 4-MUP. Apollon 1 or Apollon 2 was mixed with $5\times$ Apollon buffer and 4-NPP. Final concentrations were 1 μM deoxyribozyme, 1x Aurora buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4, 5% (v/v) DMSO) or $1\times$ Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), and from 1 μM to 1.5 mM 4-MUP or 4-NPP. Reactions were incubated for specific times at room temperature and stopped by adding

EDTA to a concentration of 25 mM. Reactions were stopped at time points that corresponded to the linear phase of the reaction. After ethanol precipitation, reacted deoxyribozymes (containing a 5' phosphate) were ligated to a short oligonucleotide as described in the section "6.2 Initial *in vitro* selection". Reacted (ligated) molecules were separated from unreacted (un-ligated) molecules by 6% Urea-PAGE. DNA was visualized by staining with GelRed using the protocol recommended by the manufacturer and gels were scanned using a Typhoon laser scanner. The percentage of reacted and unreacted deoxyribozyme was quantified using ImageQuant TL software. k_{cat} and K_m values were obtained using Prism 9 software. Curves were fitted using the equation (1) to obtain K_m and with the equation (2) to obtain k_{cat}.

(1)
$$V_0 = \frac{V_{max}[S]}{K_m + [S]}$$

(2)
$$k_{cat} = \frac{V_{max}}{[E]}$$

Where V_0 is the rate of reaction at particular substrate concentration, V_{max} is the maximum rate of reaction, K_m is the substrate concentration at which the reaction rate is 50% of the V_{max}, k_{cat} is the number of maximum substrate molecules converted by the deoxyribozyme into a product per unit time at the saturation of substrate on the deoxyribozyme, [S] is the substrate concentration, and [E] is the deoxyribozyme concentration.

6.9 Next generation sequencing and data analysis

All libraries were sequenced by Eurofins Genomics using amplicon paired-end sequencing runs. Raw reads were processed using a pipeline consisting of adaptor trimming (cutadapt v1.18), read merging (fastq-join v1.3.1), unifying of read orientation (fastx barcode splitter), primer clipping (cutadapt v1.18), length filtering (cutadapt v1.18), and counting of unique sequences (bash). All further analysis was performed using an in-house python script available at https://github.com/Jardic/aurora_selection_analysis or https://github.com/Jardic/apollon_selection_analysis.

6.10 NMR experiments

Oligonucleotides representing individual sequences from evolved libraries were purchased from GENERI BIOTECH s.r.o with HPLC-purification. DNA was resuspended in Milli-Q water, heated at 65°C for 2 minutes, cooled at room temperature for 5 minutes, and 5× Aurora/Apollon buffer was added (In NMR experiments DMSO was excluded from Aurora reaction buffer, so the reactions buffers were identical.). Concentrations at this point were 15 μM DNA, 200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4. Samples were concentrated to 500 μM DNA using Ultra-Amicon Centrifugal Filter Units (cut-off 3 kDa), and a 1.5 molar excess of 4-MUP or 4-NPP, D_2O and DSS were added. Final concentrations were 500 μM DNA, 200 mM KCl, 1 mM $ZnCl_2$, 50 mM HEPES pH 7.4, 750 μM 4-MUP or 4-NPP, 10% (v/v) D_2O, and a trace amount of DSS if not stated otherwise. NMR experiments were performed on a Bruker Advance III HD 850 MHz system equipped with an inverse triple-resonance cryo-probe. Spectral analyses were performed using TOPSPIN (Bruker)[146] and MestReNova software.

6.11 Isothermal titration calorimetry

The thermodynamics of 4-NPP binding to 5' phosphorylated Apollon 2 were monitored at 25°C using a VP-ITC microcalorimeter (MicroCal Inc./Malvern Instruments Ltd., UK). Reactant solutions were prepared in a 1x Apollon buffer containing 200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4. The exact concentration of the HPLC-purified DNA used in these experiments was determined using a NanoDrop100 spectrophotometer. Typically, 9 μl of 1.3 mM 4-NPP were injected stepwise into a sample cell containing 1.43 ml of 101.3 μM Apollon 2 until saturation was achieved. Experimental titrations were accompanied by corresponding control experiments in which 4-NPP was injected into 1x Apollon buffer that did not contain Apollon 2. Thermodynamic parameters were determined by MicroCal software implemented in Origin 7.0 (MicroCal Inc./Malvern Instruments Ltd., UK).

6.12 Differential scanning calorimetry

Thermal denaturation experiments of Apollon 2 and 5' phosphorylated Apollon 2 in the presence of 4-nitrophenyl (4-NP) or zinc were performed using a high precision VP-DSC differential scanning calorimeter (MicroCal, GE Healthcare, Northampton, MA, USA). HPLC-purified Apollon 2 or 5' phosphorylated Apollon 2 was annealed in 1× Apollon

buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), in same samples zinc was omitted from the buffer. The exact concentration of deoxyribozyme was determined using a NanoDrop100 spectrophotometer and adjusted to 30 μM, while the concentration of 4-NP was 60 μM. The DNA sample and the buffer reference solutions were degassed by stirring under vacuum and carefully loaded into the calorimeter cells in order to avoid bubble formation. Solutions were heated from 10°C to 90°C at a rate of 1°C/min. Denaturation profiles were superimposed after baseline correction using MicroCal Origin software. The reversibility of heat denaturation was monitored by repeating experiments with already heated samples cooled to 10°C, with additional scans performed after 15 minutes of thermosetting.

6.13 Oligonucleotide detection using engineered versions of Aurora or Apollon

Five different Aurora oligonucleotide sensors and five different Apollon oligonucleotide sensor were designed (ten in total). Each sensor detected a different target oligonucleotide (see in Table 1). To detect the presence of a specific target oligonucleotide, the sensor was mixed with the target in water, heated at 98°C for 2 minutes, and immediately cooled on ice for 5 minutes. In the case of Aurora oligonucleotide sensor, 5× Aurora buffer and DMSO were then added. Samples were transferred to a white half-area 96-well plate (Corning), 4-MUP was added, and the reaction mixture was incubated for 4 hours at room temperature. Final concentrations were 5 μM of the oligonucleotide sensor, 10 μM of the target oligonucleotide, 1× Aurora buffer (200 mM KCl, 50 mM HEPES pH 7.4, 1 mM $ZnCl_2$ and 5% (v/v) DMSO), and 30 μM 4-MUP. After 4 hours the reaction was stopped by adding 20 μl of 1 M KOH, and fluorescence was then measured using a Tecan Spark plate reader. In the case of Apollon oligonucleotide sensor, 5× Apollon buffer was then added. Samples were transferred to a clear half-area 96-well plate (Corning), and 4-NPP was added. Final concentrations were 50 μM of the oligonucleotide sensor, 100 μM of the target oligonucleotide, 1× Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), and 100 μM 4-NPP. After incubating for 24 hours at room temperature, the absorbance at 405 nm was measured using a TECAN Infinite M200 PRO plate reader. Analysis of fluorescence and color production was performed as described in the section "Calculation of signal-to-noise ratios."

6.14 DNase I and Exonuclease I inhibitor detection using Apollon

Apollon 2 was mixed with Milli-Q water, 10× DNase I buffer or 10× Exonuclease I buffer, and either DNase I or Exonuclease I in a volume of 20 μl. Final concentrations were 150 μM Apollon 2.5 μM DNase I, and 1× DNase I buffer (10 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, and 0.1 mM $CaCl_2$) for detection of DNase I, and 150 μM Apollon 2, 60 U Exonuclease I, and 1× Exonuclease I buffer (67 mM glycine-KOH pH 9.5, 6.7 mM $MgCl_2$, and 1 mM DTT) for detection of Exonuclease I. After an incubation at 37°C for 30 minutes to allow DNase I or Exonuclease I to cleave and inactivate Apollon 2, 80 μl of 1× Apollon reaction mixture (200 mM KCl, 1.25 mM $ZnCl_2$, 50 mM HEPES pH 7.4, and 125 μM 4-NPP) was added. Samples were transferred to clear half-area 96-well plate (Corning). The reaction mixture was incubated at room temperature for 4 hours, and the absorbance at 405 nm was then measured using a TECAN Infinite M200 PRO plate reader. Analysis of color production was performed as described in the section "6.6 Calculation of signal-to-noise ratios." 1 mM $ZnCl_2$ was used as an inhibitor of DNase I and Exonuclease I, and was mixed with these enzymes before they were added to the premix reaction.

6.15 Multiple turnover reaction of Apollon

The Apollon constructs used for multiple turnover reactions (see in Table 1) were mixed with Milli-Q water, heated at 65°C for 2 minutes, and cooled at room temperature for 5 minutes. 5× Apollon buffer (1 M KCl, 5 mM $ZnCl_2$, and 250 mM HEPES pH 7.4) was then added. Samples were transferred to a clear half-area 96-well plate (Corning), and 4-NPP was added. Final conditions were 250 μM of the 5' (substrate) strand, 5 μM of the 3' (enzyme) strand, 1× Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, 50 mM HEPES pH 7.4), and 1 mM 4-NPP in a volume of 100 μl. The reaction mixture was incubated for 7 days at room temperature. Every 4 hours the absorbance at 405 nm was measured using a TECAN Infinite M200 PRO plate reader. Analysis of color production was performed as described in the section "6.6 Calculation of signal-to-noise ratios." Turnovers were calculated relative to the absorbance produced by 5 μM Apollon 2 over 7 days.

6.15 Ribonuclease detection using covalently locked versions of Aurora or Apollon

The Aurora ribonuclease sensor or the Apollon ribonuclease sensor was mixed with MiliQ

water heated at 65°C for 2 minutes, and cooled at room temperature for 5 minutes. Then for Aurora ribonuclease sensor 5× Aurora buffer and DMSO were added. Samples were transferred to a white half-area 96-well plate (Corning), and 4-MUP and either RNase A (Thermo Fisher Scientific) alone or RNase A and RiboLock (Thermo Fisher Scientific) were added. The reaction mixture was incubated for 4 hours at room temperature. Final concentrations were 5 μM of the RNase A sensor, 500 nM RNase A or 500 nM RNase A plus 500 nM RiboLock, 1× Aurora buffer (200 mM KCl, 50 mM HEPES pH 7.4, 1 mM $ZnCl_2$ and 5% (v/v) DMSO), and 30 μM 4-MUP if not stated otherwise. After 4 hours the reaction was stopped by adding 20 μl of 1 M KOH to the reaction mixture. Fluorescence was then measured using a Tecan Spark plate reader. For Apollon ribonuclease sensor 5× Apollon buffer (1 M KCl, 250 mM HEPES pH 7.4, 5 mM $ZnCl_2$) was added. Samples were transferred to a clear half-area 96-well plate (Corning), and 4-NPP and either RNase A (Thermo Fisher Scientific) alone or RNAse A and RiboLock (Thermo Fisher Scientific) were added. The reaction mixture was incubated for 4 hours at room temperature. Final conditions were 30 μM of the ribonuclease sensor, 300 nM RNase A, 1× Apollon buffer (200 mM KCl, 1 mM $ZnCl_2$, and 50 mM HEPES pH 7.4), and 100 μM 4-NPP in a final volume of 100 μl if not stated otherwise. After 4 hours the absorbance at 405 nm was measured using a TECAN Infinite M200 PRO plate reader. Analysis of fluorescence and color production was performed as described in the section "6.6 Calculation of signal-to-noise ratios."

6.16 Plasmid construction, expression, and purification of Nsp15 from SARS-CoV-2

Nsp15 cloning, expression and purification were performed as described[147] with minor modifications. A synthetic DNA sequence encoding an *E. coli* codon optimized version of Nsp15 was cloned into a pMCSG7 vector using Gibson assembly. Cloning was confirmed by Sanger sequencing. The final pSARS-CoV-2-Nsp15_6×His vector encoded the full-length Nsp15 protein fused to an N-terminal hexahistidine tag via a TEV protease cleavage site. *E. coli* NiCo21(DE3) cells (New England Biolabs) were transformed with this plasmid. For large-scale expression and purification, a 3 litres culture of LB medium was grown at 37°C in a LEX bioreactor (Epiphyte3) in the presence of 100 μg/ml ampicillin. Once the culture reached OD_{600} ~ 1.0, flasks were moved to an 18°C bioreactor bath and supplemented with 0.1% glucose and 40 mM K_2HPO_4 (final concentration). Protein

expression was induced by the addition of 0.2 mM IPTG for 16 hours at 18°C. Bacterial cells were harvested by centrifugation at 7000g and cell pellets were resuspended in 40 ml lysis buffer (50 mM HEPES, 500 mM NaCl, 5% [v/v] glycerol, 20 mM imidazole, 10 mM β-mercaptoethanol, pH 8.0) per litter of culture and lysed using a CF1 high-pressure homogenizer. Cellular debris was removed by centrifugation at 25000g for 40 min at 4°C. The supernatant was filtered through a 0.45 µm filter, mixed with 2 ml of Ni^{2+} Sepharose equilibrated with lysis buffer, and the suspension was added to a gravity-flow column. Unbound proteins were removed by washing with 40 ml of lysis buffer. Bound proteins were eluted with 10 ml of lysis buffer supplemented with 500 mM imidazole pH 8.0. A final purification was performed using a Superdex 200 column equilibrated in lysis buffer in which 10 mM β-mercaptoethanol was replaced by 1 mM TCEP. Fractions containing Nsp15 were collected. Lysis buffer was replaced with storage buffer (150 mM NaCl, 20 mM HEPES pH 7.5, 1 mM TCEP) via repeated concentration and dilution using a 30 kDa MWCO filter (Amicon-Millipore). The final protein sample was concentrated to 1 mg/ml, aliquoted, snap frozen in liquid nitrogen and stored at -80°C until further use.

6.17 High-throughput screen for Nsp15 inhibitors using an Aurora Nsp15 sensor

Small molecules from a 1000-member fragment screen library (Maybridge) were transferred to the wells of 384-well plates using an Echo 550 liquid handler. Nsp15 protein in 1× Nsp15 buffer (50 mM KCl, 20 mM HEPES pH 7.4, 5 mM $MnCl_2$, 0.003% (v/v) Tween20) was then added using a CERTUS Flex liquid handler. After mixing, the Aurora Nsp15 sensor (in 1× Nsp15 buffer, see in Table 1) was added using a CERTUS Flex liquid handler. Reactions were mixed again. Final concentrations were 25 µM Aurora Nsp15 sensor, 400 nM Nsp15 protein, 1× Nsp15 buffer (50 mM KCl, 20 mM HEPES pH 7.4, and 5 mM $MnCl_2$), 0.003% (v/v) Tween20, and 200 µM small molecule from the fragment screen library in a volume of 20 µl. After incubating at room temperature for 1 hour to allow Nsp15 to cleave and activate the Aurora Nsp15 sensor, 80 µl of 1× Aurora reaction mixture (50 mM KCl, 20 mM HEPES pH 7.4, 1.25 mM $ZnCl_2$, 6.25% (v/v) DMSO, and 18.75 µM 4-MUP) was added using a CERTUS Flex liquid handler. The $ZnCl_2$ in this buffer inhibited the Nsp15 protein while activating Aurora for catalysis. Final concentrations were 5 µM Aurora Nsp15 sensor, 80 nM Nsp15 protein, 40 µM small molecule from the fragment screen library, 1× Aurora/Nsp15 buffer (50 mM KCl, 20 mM

HEPES pH 7.4, 1 mM $ZnCl_2$ 1 mM $MnCl_2$), 0.0006% Tween20, 5% (v/v) DMSO, and 15 μM 4-MUP in a volume of 100 μl. The reaction mixture was incubated at room temperature for 4 hours, and fluorescence was measured using a Tecan Spark plate reader. Fluorescence was measured in a black flat bottom 384-well plate (Corning). Analysis of fluorescence production was performed as described in the section "6.6 Calculation of signal-to-noise ratios." A counter screen was also performed to confirm that the inhibitors identified in the initial screen inhibit the Nsp15 protein rather than Aurora. The counter screen was performed as described above, but Aurora 2 was used instead of the Aurora Nsp15 sensor.

6.18 High-throughput screen for Nsp15 inhibitors using a FRET assay

Small molecules from a 1000-member fragment screen library (Maybridge) were transferred to the wells of 384-well plates using an Echo 550 liquid handler. Nsp15 protein in 1× Nsp15 buffer (50 mM KCl, 20 mM HEPES pH 7.4, 5 mM $MnCl_2$, 0.003% (v/v) Tween20) was then added using a CERTUS Flex liquid handler. After mixing, the FRET substrate (5'-FAM-AAArUAA-BHQ1-3') in 1× Nsp15 buffer was added using a CERTUS Flex liquid handler. Reactions were mixed again. Final concentrations were 25 μM FRET substrate, 400 nM Nsp15 protein, 1× Nsp15 buffer (50 mM KCl, 20 mM HEPES pH 7.4, and 5 mM $MnCl_2$), 0.003% (v/v) Tween20, and 200 μM small molecule from the fragment screen library in a volume of 20 μl. Fluorescence was measured every 20 minutes for 1 hour in a black flat bottom 384-well plate (Corning) using a Tecan Spark plate reader with the following settings: excitation 485 (±5) nm, emission 527 (±5) nm, gain 134, 30 flashes, Z position calculated from the well.

6.19 Analysis of data from high-throughput screens

Wells containing 1 mM $ZnCl_2$ (which inhibited Nsp15 at this concentration) served as negative controls, and were used to determine background levels of fluorescence. The average value of this background was subtracted from the fluorescence values obtained from all other wells. Wells containing aliquots of DMSO alone rather than DMSO plus small molecule were used as positive controls. After subtraction of the background, the average value of these positive controls was defined as 100% Nsp15 activity. Activity of Nsp15 in the presence of small molecules from the fragment screen library was calculated

relative to this positive control value. Z-factors[148] were calculated for each 384-well plate to determine the quality of the screen. Z-factors were calculated using the equation Z-factor = 1 - [3(σ_p + σ_n)] / (μ_p - μ_n) where μ_p is the mean fluorescence of the positive control, μ_n is the mean fluorescence of the negative control, σ_p is the standard deviation of the mean fluorescence of the positive control, and σ_n is the standard deviation of the mean fluorescence of the negative control.

6.20 Calculation of IC$_{50}$ values

IC$_{50}$ values were measured for small molecules that strongly inhibited Nsp15 in high-throughput screens. Solutions containing different concentrations of these inhibitors were transferred to the wells of 384-well plates using an Echo 550 liquid handler. To obtain the same volume in each well, the drops containing small molecules were backfilled with DMSO to 200 nl. For each inhibitor characterized, IC$_{50}$ values were measured using both the Aurora Nsp15 sensor and using the FRET assay. After determining the relative activity of Nsp15 at each concentration of inhibitor, IC$_{50}$ values were calculated using Prism 9 software (GraphPad).

7. Material

7.1 List of chemicals used

1000-member Fragment screen library	Maybridge Chemical Company Ltd., UK
4-methylumbelliferyl phosphate	Sigma-Aldrich, USA
4-methylumbelliferone	Sigma-Aldrich, USA
4-nitrophenyl phosphate	Sigma-Aldrich, USA
4-nitrophenol	Sigma-Aldrich, USA
6xLoading dye	ThermoFisher Scientific, USA
Amicons Ultra Centrifugal Filter, 3 kDa MWCO	Sigma-Aldrich, USA
Amicons Ultra Centrifugal Filter, 30 kDa MWCO	Sigma-Aldrich, USA
Acrylamide/Bis-acrylamide, 40% solution, BioReagent	Bio-Rad, USA
Agarose, BioReagent, for molecular biology, low EEO	Sigma-Aldrich, USA
Ammonium chloride for molecular biology, suitable for cell culture, ≥99.5%	Sigma-Aldrich, USA
Boric acid, BioReagent, for molecular biology	Sigma-Aldrich, USA
Cadmium chloride, 99.99% trace metals basis	Sigma-Aldrich, USA
Calcium chloride dihydrate, ReagentPlus®, ≥99.0%	Sigma-Aldrich, USA
Cerium(IV) oxide, 99%	Sigma-Aldrich, USA
Cobalt(II) chloride hexahydrate, ACS reagent, 98%	Sigma-Aldrich, USA
Copper(II) chloride, 99.999% trace metals basis	Sigma-Aldrich, USA
dNTP (each 25 mM)	Jena Bioscience GmbH, DE
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, USA
Ethanol 96% for UV	Penta s.r.o., CZ
Ethylenediaminetetraacetic acid (EDTA), BioUltra, anhydrous	Sigma-Aldrich, USA
GelRed, Nucleic Acid Gel Stain 10 000x in water	VWR International, USA
GeneRuler TM 100 bp Plus DNA Ladder	ThermoFisher Scientific, USA
Gibson Assembly Cloning Kit	New England BioLabs (NEB), USA
HEPES ≥ 99.5% (titration)	Sigma-Aldrich, USA

IPTG	Sigma-Aldrich, USA
LB Medium, Powder, 20 g/L	Serva GmbH, DE
Lead(II) chloride, powder, 98%	Sigma-Aldrich, USA
Lithium chloride, powder, 98%	Sigma-Aldrich, USA
Magnesium chloride, BioReagent, $\geq$ 97.0%	Sigma-Aldrich, USA
Manganese(II) chloride tetrahydrate, ReagentPlus®, $\geq$99%	Sigma-Aldrich, USA
NucleoSpin Gel and PCR Clean-up for QIAcube	Macherey-Nagel GmbH, DE
Poly(ethylene glycol), BioUltra, wt 200	Sigma-Aldrich, USA
Potassium chloride, BioXtra, $\geq$ 99.0%	Sigma-Aldrich, USA
Potassium hydroxide, anhydrous, $\geq$99.95% trace metals basis	Sigma-Aldrich, USA
RiboLock RNase Inhibitor (40 U/µL)	ThermoFisher Scientific, USA
Rubidium chloride, *ReagentPlus*®, $\geq$99.0% trace metals basis	Sigma-Aldrich, USA
Sodium chloride, BioXtra, $\geq$ 99.5% (AT)	Sigma-Aldrich, USA
Trizma base, BioPerformance Certified	Sigma-Aldrich, USA
Urea, powder, Bioreagent, for molecular biology	Sigma-Aldrich, USA
Zinc chloride, reagent grade, $\geq$ 98.0%	Sigma-Aldrich, USA

7.2 List of enzymes used

COVID-19 recombinant protein: Nsp15	in house production (IOCB), CZE
DNase I, RNase-free (2 U/µl)	New England BioLabs (NEB), USA
Exonuclease I (20 U/µl)	New England BioLabs (NEB), USA
Hot Start Q5 polymerase (2 U/µl)	New England BioLabs (NEB), USA
Lambda Exonuclease (5 U/µl)	New England BioLabs (NEB), USA
RNase A, DNase and protease-free (10 mg/mL)	ThermoFisher Scientific, USA
T4 DNA ligase (2,5 WU/µl)	Jena Bioscience GmbH, DE
T4 PNK (10 U/µl)	New England BioLabs (NEB), USA

7.3 List of buffers used

1x TBE buffer (Tris-Borate-EDTA buffer):
89 mM Trizma base, 2 mM EDTA, 89 mM Boric accid, pH 8.3

1x T4 DNA ligase buffer (Jena Bioscience GmbH, DE):
50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, 2.5 μg/ml BSA, pH 7.8

1x Lambda exonuclease reaction buffer (NEB, USA):
67 mM Glycine-KOH, 2.5 mM $MgCl_2$, 50 μg/ml BSA, pH 9.4 at 25°C

1x Q5 polymerase buffer (NEB, USA):
25 mM TAPS-HCl, 50 mM KCl, 2 mM $MgCl_2$, 1 mM β-mercapto ethanol, pH 9.3

1x Base hydrolysis buffer:
20 mM Trizma base, 400 mM KOH, 4 mM EDTA, pH ≈ 14

1x Cell lysis buffer:
50 mM HEPES, 500 mM NaCl, 5% [v/v] glycerol, 20 mM imidazole, 10 mM β-mercapto ethanol, pH 8.0

1x DNase I buffer (NEB, USA):
10 mM Tris-HCl, 2.5 mM $MgCl_2$, and 0.1 mM $CaCl_2$, pH 7.5

1x Exonuclease I buffer (NEB, USA):
67 mM glycine-KOH, 6.7 mM $MgCl_2$, and 1 mM DTT, pH 9.5

1x Nsp15 storage buffer:
150 mM NaCl, 20 mM HEPES pH 7.5, 1 mM TCEP

1x Nsp15 reaction buffer:
50 mM KCl, 20 mM HEPES, 5 mM $MnCl_2$, 0.003% (v/v) Tween20, pH 7.4

1x Selection buffer:
50 mM HEPES, 200 mM KCl, 1 μM $Ce(NO_3)_4$, 0.1 μM $PbCl_2$, 1 mM $ZnCl_2$, pH 7.4

1x Aurora buffer:
50 mM HEPES, 200 mM KCl, 5% (v/v) DMSO, 1 mM $ZnCl_2$, pH 7.4

1x Apollon buffer:

50 mM HEPES, 200 mM KCl, 1 mM ZnCl$_2$, pH 7.4

7.4 List of oligonucleotides used

Name:	Nucleotide sequence (from 5' end to 3' end):
Supernova	GGAAGAAAAGAATATCCCCAAAAGGGGAGTGACTTGGGATGGGGG
Pool1	GGAAGAGATGGCGACGACACAGGGACGATGCCGAATATCCTCAGTGCGCAGGGCCGCAGGGGGGAGTGACTTGGGATGGGGGGTCCACTAATGATCTGCCCGATG (underlined NTs were mutagenized at the rate of 21%)
FWD1	ACCGCTCAGGTGTAGTATCA
REV1	CATCGGGCAGATCATTAGTG
Splint1	GTCGCCATCTCTTCCTGATACTACACCTGAGCGGT
FWD1r	ACCGCTCAGGTGTAGTATCrA
REV1p	pCATCGGGCAGATCATTAGTG
Pool2	GGAAGAGATGACCAGGGCAGCGGGACGCTGACGAATTTTCTCACTATGTCCGGGACCCGAGGGGCGTGAGGAGTGTTGTGCAATTCTCATGAACTATCCGCTGGA (underlined NTs were mutagenized at the rate of 21%)
REV2	TCCAGCGGATAGTTCATGAG
REV2p	pTCCAGCGGATAGTTCATGAG
Pool3	GGCAGAGACGGCGGCGTCATAGGGACAGTGCTGTACATGCTCACCCTCAGGGGCAATCGGGGGGGCGTGAACTTATATGGAGTTACTAACACTCTTGACGAGCCGT (underlined NTs were mutagenized at the rate of 21%)
REV3	ACGGCTCGTCAAGAGTGTTA
REV3p	pACGGCTCGTCAAGAGTGTTA
Aurora 1 - full length 85 nt	GGAAGAGATGACCAGGGCAGCGGGACGCTGACGAATTTTCTCACTATGTCCGGGACCCGAGGGGCGTGAGGAGTGTTGTGCAATT
Aurora 1 - core 47 nt	GGAAGAGATGACTATGTCCGGGACCCGAGGGGCGTGAGGAGTGTTGT
Aurora 2 - full length 85 nt	GGAAGGGATGGCAAAGGCAGCGGGACGCTGGCGAAATGTCTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTGTTGTGATAAG
Aurora 2 - core 47 nt	GGAAGGGATGACTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGT
Aurora Oligo sensor 1	GGAAGGGATGGAAGGTCAATACTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGTATTGACCTTCATTGACCTTC

Aurora Oligo sensor 2	GGAAGGGATGGGGCACTGATACTATGTCCGGTTCCTGTA AGGCATGTGGAGTGTTGTATCAGTGCCCAGTGGGCACG
Aurora Oligo sensor 3	GGAAGGGATGGATGATCGGAACTATGTCCGGTTCCTGT AAGGCATGTGGAGTGTTGTTCCGATCATCCGAAGATCAG
Aurora Oligo sensor 4	GGAAGGGATGAAGTAATAGCACTATGTCCGGTTCCTGT AAGGCATGTGGAGTGTTGTGCTATTACTTATCTTTCCGA
Aurora Oligo sensor 5	GGAAGGGATGTAAAAGATAAACTATGTCCGGTTCCTGT AAGGCATGTGGAGTGTTGTTTATCTTTTATTCGTGTGTA
Aurora Target oligo 1	GAAGGTCAATGAAGGTCAAT
Aurora Target oligo 2	CGTGCCCACTGGGCACTGAT
Aurora Target oligo 3	CTGATCTTCGGATGATCGGA
Aurora Target oligo 4	TCGGAAAGATAAGTAATAGC
Aurora Target oligo 5	TACACACGAATAAAAGATAA
Aurora RNase A sensor	AAAAA(rC)GGAAGGGATGAATATGTCCGGTTCCAGTATG GCGTGTGGAGTGTTTT
Aurora Nsp15 sensor	AAAA(rU)GGAAGGGATGAATATGTCCGGTTCCTTTTAGG CGTGTGGAGTGTTTT
Aurora 2 base pairs tested by double-mutant cycles	
Aurora 2 17C40A	GGAAGGGATGACTATGCCCGGTTCCTGTAAGGCATGTG GAGTGTTGT
Aurora 2 17T40G	GGAAGGGATGACTATGTCCGGTTCCTGTAAGGCATGTG GGGTGTTGT
Aurora 2 17C40G	GGAAGGGATGACTATGCCCGGTTCCTGTAAGGCATGTG GGGTGTTGT
Aurora 2 26A30A	GGAAGGGATGACTATGTCCGGTTCCAGTAAGGCATGTG GAGTGTTGT
Aurora 2 26T30T	GGAAGGGATGACTATGTCCGGTTCCTGTATGGCATGTG GAGTGTTGT
Aurora 2 26A30T	GGAAGGGATGACTATGTCCGGTTCCAGTATGGCATGTG GAGTGTTGT
Aurora 2 4G45T	GGAGGGGATGACTATGTCCGGTTCCTGTAAGGCATGTG GAGTGTTGT
Aurora 2 4A45A	GGAAGGGATGACTATGTCCGGTTCCTGTAAGGCATGTG GAGTGTAGT
Aurora 2 4G45A	GGAGGGGATGACTATGTCCGGTTCCTGTAAGGCATGTG GAGTGTAGT
Aurora 2 12A46G	GGAAGGGATGAATATGTCCGGTTCCTGTAAGGCATGTG GAGTGTTGT
Aurora 2 12C46T	GGAAGGGATGACTATGTCCGGTTCCTGTAAGGCATGTG GAGTGTTTT

Aurora 2 12A46T	GGAAGGGATGAATATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTTT
Aurora 2 11G47T	GGAAGGGATGGCTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGT
Aurora 2 11A47C	GGAAGGGATGACTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGC
Aurora 2 11G47C	GGAAGGGATGGCTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGC
Aurora Hit 1	GGAAGAGATGGCTACTACCAAGCTACGTGGCCTAATATCATCATGGCCCCGGACCTCAAGAAGTAAGACGTTGGAATGGCGCATG
Aurora Hit 2	GGAGGAGATGAAGACGTCCGCGGCCCAAGCTCGTCTATGCTCAATGCGATCTGCTCCTAGGTCGATTTAGTTGCGACGCGTGCGC
Aurora Hit 3	GGAAGAGAGCGTGCCGTCAGACAGATGAAACTGAAACACCTCAGTGTGAAGTGCTGCGGTGGGCTGTGACCCCCTATGACGGCTC
Aurora Hit 4	GGAAGAGATGTAGACAAGGCAGGCCCCATTCCTTGCATACTTAGTGCCTAGGACCCCGTGGGGCAGTAACTAGGGATGGAGTTGC
Aurora Hit 5	GGTAGAGAGGACGACGGCCCCGGAAGATTTCTGAATATCCTCCGTGCGCTTTATCGCAGGGGGGATTTAGTTTCGATGGAGCGTC
Aurora Hit 6	GGAAGATAATGCGAGGAGGCAGGTCCCATACCAGAGCGTCTAAGGTCGCACTACCGAATGGTCGTGAGACTCTGGATGGAGTTGC
Aurora Hit 7	GGAAGAGACAGGGAGGAGATAGGCCCCATGCAGTACCTCCCAGATGCGACGATCCACAGGGTGGTACGATATCGCATGGAGCTAT
Aurora Hit 8	GGAAGAGCCGGCGATGAAATAGAGACGAGTTATAATGGTCCAATTCCGCAGGTCCGCCGGGCGGTCTGACTGAGGAGAGCGCATT
Aurora Hit 9	GGAAAAGAGCGCGACAATATTCGTACGATGCATGTACACGCCAGTCCGTAGACCCAAGTGGGGGAGTGTCTTGGGCTGGGGGCAA
Aurora Hit 10	GGAAGAGATGACCAGGGCAGCGGGACGCTGACGAATTTTCTCACTATGTCCGGGACCCGAGGGGCGTGAGGAGTGTTGTGCAATT
Aurora rHit 1	GGAAGGGATGAGCAGAGTAGCGCGACGATGACGAATTTTGTAACTATGTCCGGTTCCTGTAAGGCATGTGGAGTGTTGTGCAATT
Aurora rHit 2	GGAAGGGATGAGCAAGGCAGCATGATGCTGACGAATCTAATTACTATGTCCGGGCCCCTATGGGCTTGAGGAGTGTTGTTTAATT

Aurora rHit 3	GGAAGGGATGAACAGGGCAGCAGGATGCTGCAGATTT TTCTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTG TTGTGAAATT
Aurora rHit 4	GGAAGGGATGCGTAAGTAAGCGGGACGCTGACGCATC TTCTTACTATATCCGGGTCCCGAAGGGCGTGAGGAGTG TTGTAAGAGT
Aurora rHit 5	GGAAGGGATGGCAAAGGCAGCGGGACGCTGGCGAAAT GTCTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTG TTGTGATAAG
Aurora rHit 6	GGAAGGGATGAGCAAGGCAGCGGGACGCTGACTTAGT TTCTCACTATGTCCGGTACCCGAAGGGCATGAGGAGTG TTGTGAAATA
Aurora rHit 7	GGAAGAGATATCGAGGTCAGCCGGAAGCTGACGAGTT TTTTCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTG TTGTGAAACT
Aurora rHit 8	GGAGGAGATGACAAGGGCAGCGGAACGCTGCCGAAGA CTCTAACTATGTCCGGTTCCTGAAAGGCATGAGGAGTG TAGTGCAATT
Aurora rHit 9	GGAAGAGATAAAATGGGCAGCGCGTGGCTGCCGAATT TTCGCACTATATCCGGGACCCTCTGGGCGTGCGGAGTG TTGTGCAATT
Aurora rHit 10	GGAAGGGATGACGAGGGCAGCGAGACGCTGACGAGTT TTCGCACTATGTCCGGGTCCTGAAGGGCTTGAGGAGTG TTGTGCAATT
Apollon 1-full length 85nt	GGCAGAGACGGCGGCGTCATAGGGACAGTGCTGTACA TGCTCACCCAGGGGCAATCGGGGGGCGTGAACTTATAT GGAGTTAC
Apollon 1-core 50nt	GGCAGAGACGGCGCAGCTGTACATGCTCACCCTCAGG GGCAATCGGGGGG
Apollon 2-full length 85nt	GGCAGAGATGGCGCTGTCATAGGGACAGTGCTACGCA TGCTCACCCTCTGGGGCAATCGGGGGGCTTCGACTTGT ATAGAGTTAG
Apollon 2-core 64nt	GGCAGAGATGGCGCTGTCATAGGGACAGTGCTACGCA TGCTCACCCTCTGGGGCAATCGGGGGG
Apollon 2-core 50nt	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGGGGG
5'P Apollon 2	[Pho]GGCAGAGATGGCGCAGCCACGCATGCTCACCCTC TGGGGCAATCGGGGGG
Apollon 2-core 39nt	GGCAGAGACGCATGCTCACCCTCTGGGGCAATCGGGG GG
Apollon 2 3A25G	GGAAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGGGGG
Apollon 2 3C25T	GGCAGAGATGGCGCAGCCACGCATCCTCACCCTCTGGG GCAATCGGGGGG

Apollon 2 3A25T	GGTAGAGATGGCGCAGCCACGCATACTCACCCTCTGGG GCAATCGGGGGG
Apollon 2 - 5'disrupted	GGCAGAGAACTGGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGGGGG
Apollon 2 - 3'disrupted	GGCAGAGATGGCGCACAGTCGCATGCTCACCCTCTGGG GCAATCGGGGGG
Apollon 2 - rescue	GGCAGAGAACTGGCACAGTCGCATGCTCACCCTCTGGG GCAATCGGGGGG
Apollon 2 Stem 2 - 1 bp	GGCAGAGAGGCGCAGCCCGCATGCTCACCCTCTGGGG CAATCGGGGGG
Apollon 2 Stem 2 - 2 bp	GGCAGAGAGCGCAGCCGCATGCTCACCCTCTGGGGCA ATCGGGGGG
Apollon 2 Stem 2 - 3 bp	GGCAGAGACGCAGCGCATGCTCACCCTCTGGGGCAAT CGGGGGG
Apollon 2 - bulge	GGCAGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon 2 triple helix model testing	
Apollon 2 32T36G49G	GGCAGAGATGGCGCAGCCACGCATGCTCACCTTCTGGG GCAATCGGGGGG
Apollon 2 32C36A49G	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTAG GGCAATCGGGGGG
Apollon 2 32C36G49T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGGGTG
Apollon 2 32T36A49G	GGCAGAGATGGCGCAGCCACGCATGCTCACCTTCTAGG GCAATCGGGGGG
Apollon 2 32T36G49T	GGCAGAGATGGCGCAGCCACGCATGCTCACCTTCTGGG GCAATCGGGGTG
Apollon 2 32C36A49T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTAG GGCAATCGGGGTG
Apollon 2 32T36A49T	GGCAGAGATGGCGCAGCCACGCATGCTCACCTTCTAGG GCAATCGGGGTG
Apollon 2 31C37A48G	GGCAGAGATGGCGCAGCCACGCATGCTCACTCTCTGGG GCAATCGGGGGG
Apollon 2 31T37G48G	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGA GGCAATCGGGGGG
Apollon 2 31C37G48T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGGTGG
Apollon 2 31T37A48G	GGCAGAGATGGCGCAGCCACGCATGCTCACTCTCTGAG GCAATCGGGGGG
Apollon 2 31T37G48T	GGCAGAGATGGCGCAGCCACGCATGCTCACTCTCTGGG GCAATCGGGTGG
Apollon 2 31C37A48T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGA GGCAATCGGGTGG
Apollon 2 31T37A48T	GGCAGAGATGGCGCAGCCACGCATGCTCACTCTCTGAG GCAATCGGGTGG
Apollon 2 30T38G47G	GGCAGAGATGGCGCAGCCACGCATGCTCATCCTCTGGG GCAATCGGGGGG
Apollon 2 30C38A47G	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG

	AGCAATCGGGGGGG
Apollon 2 30C38G47T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGGTGGG
Apollon 2 30T38A47G	GGCAGAGATGGCGCAGCCACGCATGCTCATCCTCTGGA GCAATCGGGGGG
Apollon 2 30T38G47T	GGCAGAGATGGCGCAGCCACGCATGCTCATCCTCTGGG GCAATCGGTGGG
Apollon 2 30C38A47T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG AGCAATCGGTGGG
Apollon 2 30T38A47T	GGCAGAGATGGCGCAGCCACGCATGCTCATCCTCTGGA GCAATCGGTGGG
Apollon 2 29C39G46G	GGCAGAGATGGCGCAGCCACGCATGCTCCCCCTCTGGG GCAATCGGGGGG
Apollon 2 29T39G46G	GGCAGAGATGGCGCAGCCACGCATGCTCTCCCTCTGGG GCAATCGGGGGG
Apollon 2 29A39A46G	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG DACAATCGGGGGG
Apollon 2 29A39G46T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG GGCAATCGTGGGG
Apollon 2 29T39A46G	GGCAGAGATGGCGCAGCCACGCATGCTCTCCCTCTGGG ACAATCGGGGGG
Apollon 2 29T39G46T	GGCAGAGATGGCGCAGCCACGCATGCTCTCCCTCTGGG GCAATCGTGGGG
Apollon 2 29A39A46T	GGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGG DACAATCGTGGGG
Apollon 2 29T39A46T	GGCAGAGATGGCGCAGCCACGCATGCTCTCCCTCTGGG ACAATCGTGGGG
Apollon Oligo sensor 1	GGCAGAGATGGCGAAGGTCAATGCCACGCATGCTCAC CCTCTGGGGCAATCGGGGGGGATTGACCTTCATTGACCT TCGCCATCTCTGCC
Apollon Oligo sensor 2	GGCAGAGATGGCGGGCACTGATGCAGCCACGCATGCT CACCCTCTGGGGCAATCGGGGGGGATCAGTGCCCAGTG GGCACGGCCATCTCTGCC
Apollon Oligo sensor 3	GGCAGAGATGGCGATGATCGGAGCAGCCACGCATGCT CACCCTCTGGGGCAATCGGGGGGGTCCGATCATCCGAAG ATCAGGCCATCTCTGCC
Apollon Oligo sensor 4	GGCAGAGATGGCAAGTAATAGCGCAGCCACGCATGCT CACCCTCTGGGGCAATCGGGGGGGGCTATTACTTATCTT TCCGAGCCATCTCTGCC
Apollon Oligo sensor 5	GGCAGAGATGGCTAAAAGATAAGCAGCCACGCATGCT CACCCTCTGGGGCAATCGGGGGGGTTATCTTTTATTCGT GTGTAGCCATCTCTGCC
Apollon Target Oligo 1	GGCAGAGATGGCGAAGGTCAATGAAGGTCAA
Apollon Target Oligo 2	GGCAGAGATGGCCGTGCCCACTGGGCACTGAT
Apollon Target Oligo 3	GGCAGAGATGGCCTGATCTTCGGATGATCGGA
Apollon Target Oligo 4	GGCAGAGATGGCTCGGAAAGATAAGTAATAGC
Apollon Target Oligo 5	GGCAGAGATGGCTACACACGAATAAAAGATAA

Apollon- RNase A sensor	AAAArCGGCAGAGATGGCGCAGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_4bp	GGCAGAGA
Apollon MT_3'part_4bp	CGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_5bp	GGCAGAGAT
Apollon MT_3'part_5bp	ACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_6bp	GGCAGAGATG
Apollon MT_3'part_6bp	CACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_7bp	GGCAGAGATGG
Apollon MT_3'part_7bp	CCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_8bp	GGCAGAGATGGCG
Apollon MT_3'part_8bp	CGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_9bp	GGCAGAGATGGCG
Apollon MT_3'part_9bp	CGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_10bp	GGCAGAGATGGCGC
Apollon MT_3'part_10bp	GCGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_11bp	GGCAGAGATGGCGCT
Apollon MT_3'part_11bp	AGTGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_12bp	GGCAGAGATGGCGCTG
Apollon MT_3'part_12bp	CAGTGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_13bp	GGCAGAGATGGCGCTGT
Apollon MT_3'part_13bp	ACAGTGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_14bp	GGCAGAGATGGCGCTGTC
Apollon MT_3'part_14bp	GACAGTGCCACGCATGCTCACCCTCTGGGGCAATCGGGGGG
Apollon MT_5'part_15bp	GGCAGAGATGGCGCTGTCA

Apollon MT_3'part_15bp	TGACAGTGCCACGCATGCTCACCCTCTGGGGCAATCGG GGGG
Apollon Hit 1	GGCAGAGACGGCGGCGTCATAGGGACAGTGCTGTACA TGCTCACCCTCAGGGGCAATCGGGGGGCGTGAACTTAT ATGGAGTTAC
Apollon Hit 2	GGAAGATGAGGCTACAACTCAGCGCCCATACGGTATA TACTAATTGCTCAAGGTCATATGAGATACTGGCTTCCG CTATCGCATA
Apollon Hit 3	GGTAGACATGGCAACGCCACTGTCACTAAACTGAGCGT GATCAGTAAGCATGCTCATAGGTGAGGGTGTTTTTGGG TGGCTGGTA
Apollon Hit 4	GGTAGGGATGAGGATGAGAATGGGCGGACGCGGAATA ACCTCATGGCGCATCGACGAGGGGGGGAGGCACTAGG GAATAACAGTC
Apollon Hit 5	GGCAGAGAAGGCGGCGTAATATGGACAGAGCTGTACA TGCTCCCCCTCAGGGGCAATCGGGGGGGTGAAAAATTAT ATCGAGTTCG
Apollon rHit 1	GGCAGAGATGGCGCTGTCATAGGGACAGTGCTACGCA TGCTCACCCTCTGGGGCAATCGGGGGGGCTTCGACTTGT ATAGAGTTAG
Apollon rHit 2	GGCAGAGATGGCGGTGTCATGTTGACAGTGCATTGCAT GCTCCCCCTCAGGGGCAATCGGGGGGGTGAGAGCTAGT ATGGAGTTAA
Apollon rHit 3	GGCAGAGATGGCGGTGTCATAGGTACACTGCTATGCAT GCTCACCCTCTGGGGCAATCGGGGGGGCATTAACATATA TTGATTTAC
Apollon rHit 4	GGCAGAGACGGCGGCGTCCTAGGGACAGTGCTGTACA TGCTCCCCCTCAGGGGCAATCGGGGGGGTGTAAACTGCG ATGTTATTAG
Apollon rHit 5	GGCAGAGACAGCGACGGGATATAAACAGTTCTGTACA TGCTCACCCTCTGGGGCAATCGGGGGGGCTTAAAGGCTT AGGGAGTTAG
Apollon rHit 6	GGCAGAGATGGCGGCGTCATATGGAGGCCGCTATGCA TGCTCCCCCTCAGGGGCAATCGGGGGGGTGTAAACATAT AAGGAGTTTG
Apollon rHit 7	GGCAGAGATGGCGGTGGCATAGTGACAGTGCCGCGCA TGCTCACCCTCTGGGGCAATCGGGGGGGCGATAACTTAT AAGGCGTTAC
Apollon rHit 8	GGCAGAGACGACAGCGTCATACAGACACTGGTGTGCA TGCTCACCCTGGGGGGCAATCGGGGGGGCGACAATTTAT ATAGGCTTAC
Apollon rHit 9	GGCAGAGATGGCAGCCTCATAGGGATGGTGCCATACA TGCTCACCCGGAGGGGCAATCGGGGGGGCGATAAAGTA TATGGAATTAC

Apollon rHit 10	GGCAGAGAAGGCGGCGACTTAGCGACCATGCTTCGCA TGCTCCCCCTCAGGGGCAATCGGGGGGTGTGAATTTAA AAAGAGATT

7.5 List of devices used

ÄKTA™	GE HealthCare, USA
analytic weights RS232	VWR International, USA
analog Heat Block	VWR International, USA
analog Vortex Mixer	VWR International, USA
automatic pipets Pipetman	GILSON, FR
Bruker Avance III HD 850 MHz NMR spectrometr	Bruker, DE
centrifuge Eppendorf 5424 R	Eppendorf, DE
CERTUS FLEX liquid dispenser	Trajan Scientific and Medical, AU
CF1 High pressure cell disruptor	Constant Systems Ltd., UK
compact UV lamp UVP UVGL-15	Thermo Fisher Scientific, USA
Echo 550 Liquid Handler	Beckman Coulter, USA
electrophoretic aparatur for agarose gels	Biorad, USA
electorphoretic aparatur for polyacrylamid gels Biometr Model V15.17	An Analytic Jena Company, DE
gel scanner Quantum	Vilber, DE
gel scanner Thyphoon 9000	GE HealthCare, USA
LEX Bioreactor	Epiphyte Three Inc., CA
magnetic hotplate stirrer MS-H280-Pro	DLAB Scientific Inc., USA
microcentrifuge MiniStar silverline	VWR International, USA
microplate reader STECAN infinite M200Pro	Scheoller instruments, DE
microplate reader TECAN Spark	Scheoller instruments, DE
microwave oven SVERIN 900W, Grill	P-Lab a.s., CZ
multicanal automatic pipets Eppendorf	Eppendorf, DE
Orbi ShakerTM	Benchmark, USA
PCR Thermal cycler T100	Biorad, USA
pHenomenal pH 1100L	VWR International, USA
Power sourse Power pac basic HC	Biorad, USA

250V/3,0A/300W

Power sourse 300V	VWR International, USA
Precision balances, Science Education	VWR International, USA
refrigerated centrifuge Hettich Micro 220R	Scheoller instruments, DE
spektrofotometr NanoDrop 1000	Thermo Fisher Scientific, USA
Superdex 200 columm	Cytiva, USA
Tube rotator 444-0500	VWR International, USA
VP-ITC microcalorimeter	MicroCal Inc./Malvern Instruments Ltd., UK
VP-DSC differential scanning calorimeter	(MicroCal, GE Healthcare, Northampton, MA, USA

7.6 List of cell cultures used

E. coli NiCo21 (DE3) competent cells	New England BioLabs (NEB), USA

9. Author's contribution to the following publications:

This thesis is based on the work, which is described in two manuscripts published in Nucleic Acids Research. These manuscripts are listed below:

Volek M., Kurfürst J., Drexler M., Svoboda M., Srb P., Veverka V., and Curtis E. A. (2024). Aurora: a fluorescent deoxyribozyme for high-throughput screening. *Nucleic Acids Researc*h, 10.1093/nar/gkae467.

Volek M., Kurfürst J., Kožíšek M., Srb P., Veverka V., and Curtis E. A. (2024). Apollon: a deoxyribozyme that generates a yellow product. *Nucleic Acids Research*, 10.1093/nar/gkae490.

Martin Volek was the first author of both of these papers. He did all the experiments described in these manuscripts except below listed exceptions: **Jaroslav Kurfürst** did the bioinformatics analysis. **Michal Svoboda** was responsible for cloning and purification of endoribonuclease Nsp15 from SARS-CoV2 and helped with high-throughput screening experiments. **Matúš Drexler** operated the CERTUS Flex liquid handler and Echo 550 liquid handler and helped with high-throughput screening experiments. **Milan Kožíšek** measured the isothermal titration calorimetric data and the differential scanning calorimetric data. **Pavel Srb** helped to measure NMR data and was helpful with the NMR data interpretation. **Václav Veverka** provided valuable comments during the manuscript's preparation and during the revision process. **Edward Curtis** oversaw the whole project, previously developed the *in vitro* selection protocol together with Kateřina Švehlová, and he wrote the first version of both manuscripts.

10. References

1. Mieschr, J. F. Ueber die chemische Zusammensetzung der Eiterzellen. (1871 (4: pp.441-460)).

2. Dahm, R. Friedrich Miescher and the discovery of DNA. *Dev Biol* **278**, 274–288 (2005).

3. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum Genet* **122**, 565–581 (2008).

4. Miescher, J. F. *Die Spermatozoen einiger Wirbelthiere: ein Beitrag zur Histochemie.* (1874).

5. Griffith, Fred. The Significance of Pneumococcal Types. *J Hyg (Lond)* **27**, 113–159 (1928).

6. Avery, O. T., MacLeod, C. M. & McCarty, M. STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III. *Journal of Experimental Medicine* **79**, 137–158 (1944).

7. Witkin, E. M. Remembering Rollin Hotchkiss (1911–2004). *Genetics* **170**, 1443–1447 (2005).

8. Hershey, A. D. & Chase, M. INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE. *Journal of General Physiology* **36**, 39–56 (1952).

9. Olby, R. C. The Path to the Double Helix: The Discovery of DNA. (1974)

10. Chargaff, Erwin., Vischer, Ernst., Doniger, Ruth., Green, Charlotte. & Misani, Fernanda. THE COMPOSITION OF THE DESOXYPENTOSE NUCLEIC ACIDS OF THYMUS AND SPLEEN. *Journal of Biological Chemistry* **177**, 405–416 (1949).

11. Chargaff, E. Structure and function of nucleic acids as cell constituents. *Fed Proc* **10**, 654–659 (1951).

12. Watson, J. D. & Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

13. Franklin, R. E. & Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **171**, 740–741 (1953).

14. Watson, J. D. & Crick, F. H. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**, 964–967 (1953).

15. Meselson, M. & Stahl, F. W. THE REPLICATION OF DNA IN ESCHERICHIA COLI. *Proceedings of the National Academy of Sciences of the United States of America* **44**, 671 (1958).

16. Delbrück, M. On the replication of desoxyribonucleic acid (dna). *Proceedings of the National Academy of Sciences* **40**, 783–788 (1954).

17. Cech, T. R. Nobel lecture. Self-splicing and enzymatic activity of an intervening sequence RNA from Tetrahymena. *Biosci Rep* **10**, 239–261 (1990).

18. Yao, M. C., Kimmel, A. R. & Gorovsky, M. A. A small number of cistrons for ribosomal RNA in the germinal nucleus of a eukaryote, Tetrahymena pyriformis. *Proc Natl Acad Sci U S A* **71**, 3082–3086 (1974).

19. Wild, M. A. & Gall, J. G. An intervening sequence in the gene coding for 25S Ribosomal RNA of Tetrahymena pigmentosa. *Cell* **16**, 565–573 (1979).

20. Zaug, A. J. & Cech, T. R. In vitro splicing of the ribosomal RNA precursor in nuclei of Tetrahymena. *Cell* **19**, 331–338 (1980).

21. Carin, M. *et al.* In vitro splicing of the ribosomal RNA precursor in isolated nucleoli from Tetrahymena. *Nucleic Acids Res* **8**, 5551–5566 (1980).

22. Cech, T. R., Zaug, A. J. & Grabowski, P. J. In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell* **27**, 487–496 (1981).

23. Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147–157 (1982).

24. Leseigneur, G. *et al.* ESA's Cometary Mission Rosetta—Re-Characterization of the COSAC Mass Spectrometry Results. *Angewandte Chemie International Edition* **61**, e202201925 (2022).

25. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618–618 (1986).

26. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-Catalyzed Transcription of an Active Ribozyme. *Science* **332**, 209–212 (2011).

27. LAWRENCE, M. S. & BARTEL, D. P. New ligase-derived RNA polymerase ribozymes. *RNA* **11**, 1173–1180 (2005).

28. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, 1411–1418 (1993).

29. Müller, F. *et al.* A prebiotically plausible scenario of an RNA–peptide world. *Nature* **605**, 279–284 (2022).

30. Singer, J. N. *et al.* Loading of Amino Acids onto RNA in a Putative RNA-Peptide World. *Angewandte Chemie International Edition* **62**, e202302360 (2023).

31. Węgrzyn, E. *et al.* RNA-Templated Peptide Bond Formation Promotes L-Homochirality. *Angewandte Chemie International Edition* **63**, e202319235 (2024).

32. Feldmann, J. *et al.* A Unifying Concept for the Prebiotic Formation of RNA Pyrimidine Nucleosides. *ChemistryEurope* **1**, e202300013 (2023).

33. Altman, S. Enzymatic Cleavage of RNA by RNA (Nobel Lecture). *Angewandte Chemie International Edition in English* **29**, 749–758 (1990).

34. Baltimore, D. Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature* **226**, 1209–1211 (1970).

35. Temin, H. M. & Mizutani, S. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226**, 1211–1213 (1970).

36. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* **22**, 1859–1862 (1981).

37. Mullis, K. *et al.* Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* **51 Pt 1**, 263–273 (1986).

38. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783–8798 (1987).

39. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).

40. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

41. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

42. Svehlova, K., Lukšan, O., Jakubec, M. & Curtis, E. A. Supernova: A Deoxyribozyme that Catalyzes a Chemiluminescent Reaction. *Angewandte Chemie International Edition* **61**, e202109347 (2022).

43. Li, Y. & Breaker, R. R. Phosphorylating DNA with DNA. *Proceedings of the National Academy of Sciences* **96**, 2746–2751 (1999).

44. Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc Natl Acad Sci U S A* **94**, 4262–4266 (1997).

45. Streckerová, T., Kurfürst, J. & Curtis, E. A. Single-round deoxyribozyme discovery. *Nucleic Acids Research* **49**, 6971–6981 (2021).

46. Scheitl, C. P. M., Lange, S. & Höbartner, C. New Deoxyribozymes for the Native Ligation of RNA. *Molecules* **25**, 3650 (2020).

47. Liaqat, A., Stiller, C., Michel, M., Sednev, M. V. & Höbartner, C. N6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angewandte Chemie International Edition* **59**, 18627–18631 (2020).

48. Song, Z. *et al.* Development of a CD63 Aptamer for Efficient Cancer Immunochemistry and Immunoaffinity-Based Exosome Isolation. *Molecules* **25**, 5585 (2020).

49. Murphy, M. B., Fuller, S. T., Richardson, P. M. & Doyle, S. A. An improved method for the in vitro evolution of aptamers and applications in protein detection and purification. *Nucleic Acids Research* **31**, e110 (2003).

50. Mosing, R. K. & Bowser, M. T. Isolating aptamers using capillary electrophoresis-SELEX (CE-SELEX). *Methods Mol Biol* **535**, 33–43 (2009).

51. Breaker, R. R. & Joyce, G. F. A DNA enzyme that cleaves RNA. *Chem Biol* **1**, 223–229 (1994).

52. LEVY, M., GRISWOLD, K. E. & ELLINGTON, A. D. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* **11**, 1555–1562 (2005).

53. Pietruschka, G. *et al.* An RNA Motif That Enables Optozyme Control and Light-Dependent Gene Expression in Bacteria and Mammalian Cells. *Advanced Science* **11**, 2304519 (2024).

54. Curtis, E. A. & Bartel, D. P. New catalytic structures from an existing ribozyme. *Nat Struct Mol Biol* **12**, 994–1000 (2005).

55. Dupont, D. M., Larsen, N., Jensen, Jan. K., Andreasen, P. A. & Kjems, J. Characterisation of aptamer–target interactions by branched selection and high-throughput sequencing of SELEX pools. *Nucleic Acids Res* **43**, e139 (2015).

56. Cho, M. *et al.* Quantitative selection of DNA aptamers through microfluidic selection and high-throughput sequencing. *Proc Natl Acad Sci U S A* **107**, 15373–15378 (2010).

57. Zuker, M. On finding all suboptimal foldings of an RNA molecule. *Science* **244**, 48–52 (1989).

58. Gruber, A. R., Bernhart, S. H. & Lorenz, R. The ViennaRNA web services. *Methods Mol Biol* **1269**, 307–326 (2015).

59. Woese, C. R., Gutell, R., Gupta, R. & Noller, H. F. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. *Microbiol Rev* **47**, 621–669 (1983).

60. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).

61. Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. & Tulinsky, A. The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J Biol Chem* **268**, 17651–17654 (1993).

62. Troisi, R., Napolitano, V., Spiridonova, V., Russo Krauss, I. & Sica, F. Several structural motifs cooperate in determining the highly effective anti-thrombin activity of NU172 aptamer. *Nucleic Acids Res* **46**, 12177–12185 (2018).

63. Yu, H. *et al.* Aptameric hirudins as selective and reversible EXosite-ACTive site (EXACT) inhibitors. *Nat Commun* **15**, 3977 (2024).

64. Tasset, D. M., Kubik, M. F. & Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J Mol Biol* **272**, 688–698 (1997).

65. Hauel, N. H. *et al.* Structure-based design of novel potent nonpeptide thrombin inhibitors. *J Med Chem* **45**, 1757–1766 (2002).

66. Gold, L. *et al.* Aptamer-Based Multiplexed Proteomic Technology for Biomarker Discovery. *PLOS ONE* **5**, e15004 (2010).

67. Lu, X. *et al.* Probing Molecular Diversity and Ultrastructure of Brain Cells with Fluorescent Aptamers. *bioRxiv* doi:10.1101/2023.09.18.558240.

68. Pothoulakis, G., Ceroni, F., Reeve, B. & Ellis, T. The Spinach RNA Aptamer as a Characterization Tool for Synthetic Biology. *ACS Synth. Biol.* **3**, 182–187 (2014).

69. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J. Am. Chem. Soc.* **136**, 16299–16308 (2014).

70. H, K. & Sr, J. A Fluorogenic RNA-Based Sensor Activated by Metabolite-Induced RNA Dimerization. *Cell chemical biology* **26**, (2019).

71. C, S., N, P., Kr, G. & C, H. A Multicolor Large Stokes Shift Fluorogen-Activating RNA Aptamer with Cationic Chromophores. *Chemistry (Weinheim an der Bergstrasse, Germany)* **25**, (2019).

72. X, C. *et al.* Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. *Nature biotechnology* **37**, (2019).

73. Sk, D. *et al.* Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag. *Nature chemical biology* **18**, (2022).

74. Ev, D. *et al.* RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS chemical biology* **9**, (2014).

75. Sando, S., Narita, A. & Aoyama, Y. Light-up Hoechst-DNA aptamer pair: generation of an aptamer-selective fluorophore from a conventional DNA-staining dye. *Chembiochem* **8**, 1795–1803 (2007).

76. Sen, D. & Poon, L. C. H. RNA and DNA complexes with hemin [Fe(III) heme] are efficient peroxidases and peroxygenases: how do they do it and what does it mean? *Crit Rev Biochem Mol Biol* **46**, 478–492 (2011).

77. Travascio, P., Li, Y. & Sen, D. DNA-enhanced peroxidase activity of a DNA aptamer-hemin complex. *Chemistry & Biology* **5**, 505–517 (1998).

78. Nakayama, S. & Sintim, H. O. Biomolecule detection with peroxidase-mimicking DNAzymes; expanding detection modality with fluorogenic compounds. *Mol. BioSyst.* **6**, 95–97 (2009).

79. Xiao, Y., Pavlov, V., Gill, R., Bourenko, T. & Willner, I. Lighting Up Biochemiluminescence by the Surface Self-Assembly of DNA–Hemin Complexes. *ChemBioChem* **5**, 374–379 (2004).

80. Trachman, R. J. & Ferré-D'Amaré, A. R. Tracking RNA with light: selection, structure, and design of fluorescence turn-on RNA aptamers. *Q Rev Biophys* **52**, e8 (2019).

81. Neubacher, S. & Hennig, S. RNA Structure and Cellular Applications of Fluorescent Light-Up Aptamers. *Angew Chem Int Ed Engl* **58**, 1266–1279 (2019).

82. Bouhedda, F., Autour, A. & Ryckelynck, M. Light-Up RNA Aptamers and Their Cognate Fluorogens: From Their Development to Their Applications. *Int J Mol Sci* **19**, 44 (2017).

83. Paige, J. S., Duc, T. N., Song, W. & Jaffrey, S. R. Fluorescence imaging of cellular metabolites with RNA. *Science* **335**, 1194 (2012).

84. Autour, A. *et al.* Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat Commun* **9**, 656 (2018).

85. Cawte, A. D., Unrau, P. J. & Rueda, D. S. Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nat Commun* **11**, 1283 (2020).

86. Jakubec, M., Pšenáková, K., Svehlova, K. & Curtis, E. A. Optimizing the Chemiluminescence of a Light-Producing Deoxyribozyme. *Chembiochem* **23**, e202200026 (2022).

87. Švehlová, K. Development and characterization of light-producing deoxyribozymes. (2022).

88. Robinson, D. & Willcox, P. 4-Methylumbelliferyl phosphate as a substrate for lysosomal acid phosphatase. *Biochimica et Biophysica Acta (BBA) - Enzymology* **191**, 183–186 (1969).

89. Tabatabai, M. A. & Bremner, J. M. Use of p-nitrophenyl phosphate for assay of soil phosphatase activity. *Soil Biology and Biochemistry* **1**, 301–307 (1969).

90. Simopoulos, T. T. & Jencks, W. P. Alkaline phosphatase is an almost perfect enzyme. *Biochemistry* **33**, 10375–10380 (1994).

91. Mercan, F. & Bennett, A. M. Analysis of Protein Tyrosine Phosphatases and Substrates. *Curr Protoc Mol Biol* **0 18**, Unit-18.16 (2010).

92. Lorsch, J. R. & Szostak, J. W. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36 (1994).

93. Ekland, E. H. & Bartel, D. P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res* **23**, 3231–3238 (1995).

94. Gutell, R. R., Power, A., Hertz, G. Z., Putz, E. J. & Stormo, G. D. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res* **20**, 5785–5795 (1992).

95. Gutell, R. R. Ten lessons with Carl Woese about RNA and comparative analysis. *RNA Biol* **11**, 254–272 (2014).

96. Plavec, J. Dna. in *NMR of Biomolecules* 96–116 (John Wiley & Sons, Ltd, 2012). doi:10.1002/9783527644506.ch5.

97. Churcher, Z. & Johnson, P. NMR for non-experts; a practical guide for applying NMR methods in studies of aptamer-ligand interactions. **4**, 3–9 (2020).

98. Holtz, K. M. & Kantrowitz, E. R. The mechanism of the alkaline phosphatase reaction: insights from NMR, crystallography and site-specific mutagenesis. *FEBS Lett* **462**, 7–11 (1999).

99. McCall, K. A., Huang, C. & Fierke, C. A. Function and mechanism of zinc metalloenzymes. *J Nutr* **130**, 1437S–46S (2000).

100. Chandra, M., Sachdeva, A. & Silverman, S. K. DNA-catalyzed sequence-specific hydrolysis of DNA. *Nat Chem Biol* **5**, 718–720 (2009).

101. Gu, H., Furukawa, K., Weinberg, Z., Berenson, D. F. & Breaker, R. R. Small, Highly Active DNAs That Hydrolyze DNA. *J. Am. Chem. Soc.* **135**, 9121–9129 (2013).

102. Zhang, C. *et al.* New DNA-hydrolyzing DNAs isolated from an ssDNA library carrying a terminal hybridization stem. *Nucleic Acids Research* **49**, 6364–6374 (2021).

103. Yarus, M. How many catalytic RNAs? Ions and the Cheshire cat conjecture. *FASEB J* **7**, 31–39 (1993).

104. Uhlenbeck, O. C. A small catalytic oligoribonucleotide. *Nature* **328**, 596–600 (1987).

105. Furukawa, K. & Minakawa, N. Allosteric control of a DNA-hydrolyzing deoxyribozyme with short oligonucleotides and its application in DNA logic gates. *Org. Biomol. Chem.* **12**, 3344–3348 (2014).

106. Koizumi, M., Soukup, G. A., Kerr, J. N. Q. & Breaker, R. R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Mol Biol* **6**, 1062–1071 (1999).

107. Levy, M. & Ellington, A. D. ATP-dependent allosteric DNA enzymes. *Chem Biol* **9**, 417–426 (2002).

108. Kertsburg, A. & Soukup, G. A. A versatile communication module for controlling RNA folding and catalysis. *Nucleic Acids Res* **30**, 4599–4606 (2002).

109. Sgallová, R. & Curtis, E. A. Secondary Structure Libraries for Artificial Evolution Experiments. *Molecules* **26**, 1671 (2021).

110. Kolarevic, A., Yancheva, D., Kocic, G. & Smelcerovic, A. Deoxyribonuclease inhibitors. *Eur J Med Chem* **88**, 101–111 (2014).

111. Walker, M. J. *et al.* DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nat Med* **13**, 981–985 (2007).

112. Larsen, B. D. & Sørensen, C. S. The caspase-activated DNase: apoptosis and beyond. *FEBS J* **284**, 1160–1170 (2017).

113. Deng, X. & Baker, S. C. An 'Old' protein with a new story: Coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology* **517**, 157–163 (2018).

114. Faheem, null *et al.* Druggable targets of SARS-CoV-2 and treatment opportunities for COVID-19. *Bioorg Chem* **104**, 104269 (2020).

115. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).

116. Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem Sci* **41**, 595–609 (2016).

117. Micura, R. & Höbartner, C. Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **49**, 7331–7353 (2020).

118. Ali, M. M., Aguirre, S. D., Lazim, H. & Li, Y. Fluorogenic DNAzyme probes as bacterial indicators. *Angew Chem Int Ed Engl* **50**, 3751–3754 (2011).

119. Lu, X., Kong, K. Y. S. & Unrau, P. J. Harmonizing the growing fluorogenic RNA aptamer toolbox for RNA detection and imaging. *Chem. Soc. Rev.* **52**, 4071–4098 (2023).

120. Jaffrey, S. R. RNA-Based Fluorescent Biosensors for Detecting Metabolites in vitro and in Living Cells. *Adv Pharmacol* **82**, 187–203 (2018).

121. Autour, A. *et al.* Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nat Commun* **9**, 656 (2018).

122. Zhou, J. & Rossi, J. Aptamers as targeted therapeutics: current potential and challenges. *Nat Rev Drug Discov* **16**, 181–202 (2017).

123. Kato, T., Shimada, I., Kimura, R. & Hyuga, M. Light-up fluorophore--DNA aptamer pair for label-free turn-on aptamer sensors. *Chem Commun (Camb)* **52**, 4041–4044 (2016).

124. Kolpashchikov, D. M. & Spelkov, A. A. Binary (Split) Light-up Aptameric Sensors. *Angew Chem Int Ed Engl* **60**, 4988–4999 (2021).

125. Chen, L. *et al.* A Glucose-Powered Activatable Nanozyme Breaking pH and H2O2 Limitations for Treating Diabetic Infections. *Angewandte Chemie International Edition* **60**, 23534–23539 (2021).

126. Guo, Y. *et al.* A Thermophilic Tetramolecular G-Quadruplex/Hemin DNAzyme. *Angew Chem Int Ed Engl* **56**, 16636–16640 (2017).

127. Camden, A. J., Walsh, S. M., Suk, S. H. & Silverman, S. K. DNA Oligonucleotide 3′-Phosphorylation by a DNA Enzyme. *Biochemistry* **55**, 2671–2676 (2016).

128. Lavis, L. D. & Raines, R. T. Bright Ideas for Chemical Biology. *ACS Chem. Biol.* **3**, 142–155 (2008).

129. Cui, C., Shu, W. & Li, P. Fluorescence In situ Hybridization: Cell-Based Genetic Diagnostic and Research Applications. *Front Cell Dev Biol* **4**, 89 (2016).

130. Christian, A., Pattee, M., Attix, C. & Tucker, J. Rolling circle amplification detection of RNA and DNA. (2002).

131. Canal, B. *et al.* Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of nsp15 endoribonuclease. *Biochem J* **478**, 2465–2479 (2021).

132. Choi, R. *et al.* High-throughput screening of the ReFRAME, Pandemic Box, and COVID Box drug repurposing libraries against SARS-CoV-2 nsp15 endoribonuclease to identify small-molecule inhibitors of viral activity. *PLOS ONE* **16**, e0250019 (2021).

133. Yamada, Y. *et al.* DR396, an apoptotic DNase γ inhibitor, attenuates high mobility group box 1 release from apoptotic cells. *Bioorganic & medicinal chemistry* **19**, 168–171 (2011).

134. Qin, C., Bian, X., Lv, W. & Li, B. Label-free chemiluminescent CRISPR/Cas12a biosensing strategy for the detection of nucleic acids and non-nucleic acids utilizing DNAzyme Supernova. *Sensors and Actuators B: Chemical* **402**, 135143 (2024).

135. Curtis, E. A. Pushing the Limits of Nucleic Acid Function. *Chemistry – A European Journal* **28**, e202201737 (2022).

136. Robertson, M. P. & Ellington, A. D. In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol* **17**, 62–66 (1999).

137. Piganeau, N., Jenne, A., Thuillier, V. & Famulok, M. An Allosteric Ribozyme Regulated by Doxycyline. *Angew Chem Int Ed Engl* **39**, 4369–4373 (2000).

138. Ali, M. M. *et al.* A Rapid Sputum-based Lateral Flow Assay for Airway Eosinophilia using an RNA-cleaving DNAzyme Selected for Eosinophil Peroxidase. *Angewandte Chemie International Edition* **62**, e202307451 (2023).

139. Mirzakhani, K., Gargari, S. L. M., Rasooli, I. & Rasoulinejad, S. Development of a DNA Aptamer for Screening Neisseria meningitidis Serogroup B by Cell SELEX. *Iran Biomed J* **22**, 193–201 (2018).

140. Schmitz, A. *et al.* A SARS-CoV-2 Spike Binding DNA Aptamer that Inhibits Pseudovirus Infection by an RBD-Independent Mechanism**. *Angewandte Chemie International Edition* **60**, 10279–10285 (2021).

141. Zhang, Z. *et al.* High-Affinity Dimeric Aptamers Enable the Rapid Electrochemical Detection of Wild-Type and B.1.1.7 SARS-CoV-2 in Unprocessed Saliva. *Angew Chem Int Ed Engl* **60**, 24266–24274 (2021).

142. Kraemer, S. *et al.* From SOMAmer-Based Biomarker Discovery to Diagnostic and Clinical Applications: A SOMAmer-Based, Streamlined Multiplex Proteomic Assay. *PLoS One* **6**, e26332 (2011).

143. Sednev, M. V., Liaqat, A. & Höbartner, C. High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq). *J. Am. Chem. Soc.* **144**, 2090–2094 (2022).

144. Singh, N. K. *et al.* High-affinity one-step aptamer selection using a non-fouling porous hydrogel. *Nat Biotechnol* 1–8 (2023) doi:10.1038/s41587-023-01973-8.

145. Albiñana, C. B. *et al.* Kinetic, thermodynamic and structural analysis of tamiphosphor binding to neuraminidase of H1N1 (2009) pandemic influenza. *European Journal of Medicinal Chemistry* **121**, 100–109 (2016).

146. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

147. Kim, Y. *et al.* Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci* **29**, 1596–1605 (2020).

148. Zhang, J.-H., Chung, T. D. Y. & Oldenburg, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *SLAS Discovery* **4**, 67–73 (1999).

Svoluji k zapůjčení této práce pro studijní účely a prosím, aby byla řádně vedena evidence vypůjčovatelů.

Jméno a příjmení S adresou	Číslo OP	Datum vypůjčení	Poznámka