

Optimisation by selection for oligonucleotide-based synthetic molecular machinery



Qian Zhang

St. Hugh's College

University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity Term 2024

Abstract

Nucleic acid nanotechnology has emerged as a promising frontier for artificial molecular machines, owing to the intrinsic programmability of nucleic acids. However, its potential is constrained by the slow kinetics of strand displacement reactions, which typically occur over minutes. In contrast, ATP hydrolysis-powered biological motors operate on millisecond timescales. Inspired by the efficient chemomechanical cycles of nature, this thesis aims to develop nucleic acid-based motors that allosterically bind to their fuel, catalyse its chemical conversion, and autonomously repeat this cycle until fuel depletion.

The primary strategy used was *in vitro* selection to develop nucleic acid-based mechanocatalysts that couple the chemical reactions of small-molecule fuels with the mechanical cycles of nucleic acid nanostructures. GTP was chosen as the fuel and covalently attached to a library containing an integrated triphosphate-specific aptamer. After eight rounds of selection and high-throughput sequencing, ten enriched sequences were identified. The most active sequence exhibited activity comparable to that of RNA 5' pyrophosphohydrolase, which converts the 5' triphosphate of RNA into a 5' monophosphate. Truncation experiments identified a minimal 77-nucleotide ribozyme that retained catalytic activity. Kinetic analysis revealed an observed rate constant of $k_{\text{obs}} = 0.08 \text{ min}^{-1}$. However, under multiple-turnover conditions (with free GTP), no significant catalytic activity was observed. Another strategy attempted to replace the adenine-specific aptamer in a catalyst-aptamer conjugate with a triphosphate-recognising aptamer. However, this design was hindered by difficulties in synthesising the organic catalytic centre.

This work explores a new approach to enhancing nucleic acid-based molecular machines. Although sustained catalytic activity was not achieved under multiple-turnover conditions, this thesis provides insight into the feasibility and challenges of developing enzyme-like autonomous molecular machines and informs future developments.

Acknowledgements

I would like to thank the Marie Skłodowska-Curie project ArtMoMa (Artificial Molecular Machines) under the European Union's Horizon 2020 research and innovation programme and the Oxford Physics Endowment for Graduates (OXPEG) for their financial support and the extensive training and resources provided through this programme.

I would like to express my deepest gratitude to my supervisor, Professor Andrew Turberfield, for giving me this opportunity and for his guidance, ideas, patience, and support throughout my DPhil. I am grateful to Dr Jon Bath for teaching me all the laboratory techniques essential for this work. I extend my sincere thanks to Dr Phil Holliger for his valuable advice on the design of the *in vitro* selection experiments and for the opportunity to spend two months in his lab at the MRC-LMB in Cambridge. I also thank all the members of his group, particularly Dr Edoardo Gianni and Samantha Kwok, for their help with high-throughput sequencing. I am grateful to Dr Rana Abdul Razzak, who supervised me throughout the organic synthesis, and I greatly appreciate the generous advice from Professor Tom Brown and Dr Asha Brown on the proposed synthetic protocols. I am also thankful to Dr Ashley Marshall of the Snaith group for organising shared lab space during COVID.

I feel very lucky to have been part of the Turberfield group. Thanks to all the members for making it such a friendly and supportive place. Special thanks to Jon for being a great mentor in both science and life, always generous with advice, support, and encouragement. I am also grateful to him and Claudia for hosting so many memorable gatherings. Heartfelt thanks to Rafa and Seham for their constant encouragement. Rafa, your uncle's wisdom and calm presence were often exactly what I needed. Seham, thank you for always being there with kindness and support. To Catherine and Alma, thank you for the

many delicious cakes that brightened up our days in the lab. Rana, thank you for your company during those weekends in the lab. Singming, it has been a pleasure to share the ArtMoMa experience with you. And to Erik, Joakim, Behnam, Sonji, and Eryk, thank you for creating such a warm and welcoming environment.

I would like to thank all the ArtMoMa ESRs (in alphabetical order): Alessandro, Brian, Dania, Henry, Loli, Lorenzo, Maria, Martin, Max, Nina, Philippe, Rachele, and Roza. Thank you for making ArtMoMa such a memorable and enjoyable experience. I sometimes regret being too sleepy to fully join in the late-night fun at the annual ArtMoMa schools, but I was always impressed by how you managed to party into the night and still turn up on time the following morning. I would also like to thank my friends at St Hugh's College, Michelle, Julia, Erik, Marcello, and Sharan, for sharing so many memorable moments. A special thank you to Amartya for always being there for me, offering continuous mental and emotional support. Thanks to Tina, Dora, Yifeng, Qin, and Lijin for the many happy and delicious Saturday dinners. Last but not least, I would like to thank my family for their love, support, and encouragement, which have been invaluable in helping me complete this journey.

It has been a great journey. Working on an interdisciplinary project in such a dynamic environment has been an incredibly rewarding experience. I have learned a lot from different fields and have developed a deep interest in molecular biology, especially RNA molecules. I have had the privilege of meeting and learning from so many inspiring individuals along the way. Oxford has been a wonderful place and these years will remain unforgettable.

Contents

List of Abbreviations	vii
1 Background	1
1.1 Introduction and motivation	1
1.2 Artificial molecular machines	2
1.2.1 Energy ratchets	4
1.2.2 Information ratchets	6
1.3 DNA nanotechnology	10
1.3.1 Nucleic acids as building materials	10
1.3.2 Oligonucleotide-based molecular machines	17
1.4 Functional nucleic acids	25
1.4.1 Hierarchical folding of functional nucleic acids	26
1.4.2 Aptamers	29
1.4.3 Catalytic nucleic acids	34
1.4.4 Applications of functional nucleic acids in dynamic oligonucleotide-based systems	42
1.5 Creating functional nucleic acids by selection and engineering	45
1.5.1 In vitro selection of aptamers	46
1.5.2 In vitro selection of catalytic nucleic acids	48
1.5.3 Modular nucleic acid catalysts	52
2 A proposed optimised mechanism for autonomous molecular machines made from nucleic acids	54
2.1 An optimised mechanism	55

2.2	Two strategies to create the mechanocatalyst	58
3	Tethered catalyst approach	62
3.1	Introduction	62
3.2	Materials and Methods	63
3.3	Synthesis outcomes and discussion	65
4	In vitro selection approach	69
4.1	Experimental design	70
4.1.1	Library design	70
4.1.2	Selection design	71
4.2	Selection procedure	73
4.2.1	Library preparation	73
4.2.2	In vitro selection	74
4.3	Evaluation of selection progression	77
4.3.1	Input-to-Output ratio analysis across selection rounds	77
4.3.2	Bulk ligation activity of sub-libraries	78
4.3.3	Melting curve analysis of sublibraries	80
4.4	Sanger sequencing	83
4.4.1	Cloning and sequencing	83
4.4.2	Activity analysis	85
4.5	High-throughput sequencing	87
4.5.1	Preparation of sequencing templates	87
4.5.2	Sequencing data analysis	88
4.5.3	Activity analysis of the top 10 sequences	91
5	Post-selection engineering and analysis	99
5.1	Sequence engineering	99
5.2	3D structure simulation	106
5.3	Kinetic analysis	107
5.3.1	Single-turnover kinetics	107
5.3.2	Multiple-turnover kinetics	110

6	Conclusions and future directions	116
6.1	Discussion and conclusions	116
6.2	Future directions	118
A	Appendix	122
A.1	Lists of oligonucleotides	122
A.2	Gel Electrophoresis	133
A.3	Supplementary Figures	134
A.4	High-throughput sequencing data analysis script	139
A.5	Single-turnover kinetics analysis script	144

List of Abbreviations

AMP	Adenosine monophosphate
ATP	Adenosine triphosphates
cDNA	Complementary DNA
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
GMP	Guanosine monophosphate
GTP	Guanosine triphosphates
HTS	High Throughput Sequencing
IDT	Integrated DNA Technologies
IVT	In vitro transcription
NEB	New England Biolabs
nt	Nucleotides
NTP	Nucleoside triphosphate
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PPi	Pyrophosphate
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
ssDNA	Single-stranded DNA

Chapter 1

Background

1.1 Introduction and motivation

All living organisms are sustained by nanoscale molecular machines, with adenosine triphosphate (ATP) being the most prevalent energy source in biological systems. ATP is generated through oxidative phosphorylation during respiration or photosynthesis, and its hydrolysis serves as a source of energy to various molecular machines, including biosynthetic machinery, transmembrane pumps, and intracellular transport systems. The concept of artificial molecular machines dates back more than 150 years to Maxwell's thought experiments ^[1]. Objects on the nanometre scale experience constant Brownian motion, and energy is required to rectify these motions rather than to overcome inertia to initiate movement ^[2]. Research in artificial molecular machines has flourished, leading to the award of the Nobel Prize in Chemistry in 2016 ^[3–5].

It has been widely acknowledged that biological molecular machines operate as mechanocatalysts whose mechanical cycles are coupled to the hydrolysis of ATP through Brownian ratchets ^[6, 7]. Thus, the development of kinetic asymmetric chemomechanical cycles driven by the catalysis of exergonic chemical reactions has become a crucial step in the creation of autonomous molecular machines ^[8].

The Leigh group reported the first autonomous chemically fuelled synthetic motor in 2016 ^[9]. Meanwhile, eight years earlier, the Turberfield group introduced an autonomous bipedal DNA walker that operates by catalysing the hybridisation of inert hairpin fuels ^[10]. As information-coded molecules, nucleic acids provide unparalleled levels of control,

precision, and modularity in nanotechnology. However, this DNA walker operates at very slow speeds due to slow strand displacement reactions. Furthermore, protein motors exhibit various motility behaviours, but their mechanical cycles are universally coupled to ATP hydrolysis and share a similar ATP binding pocket ^[11–14].

Inspired by this, we hypothesised that molecular machines made from oligonucleotides might also be able to operate in a similar manner by coupling programmable nucleic acid reactions with energy dissipating processes of small-molecule fuels ^[15]. These machines could be driven by a universal small-molecule fuel and would possess a common fuel-interacting module. Nucleic acids exhibit properties similar to those of proteins, including specific binding, catalytic activity, and allosteric regulation ^[16–18]. Hence, in theory, it should be possible to create an oligonucleotide-based mechanocatalyst in which its mechanical state is coupled to a catalysed chemical reaction of a small-molecule fuel. The reconfiguration resulting from the binding and release of small molecules following the catalysed reaction is faster than strand displacement reactions between large molecules (oligonucleotides) by branch migration. Small-molecule fuels could also be used at much higher concentrations (millimolar) compared to the typical concentrations used for DNA fuels (nano- to micromolar), which generally yields faster kinetics ^[19].

In this chapter, I will begin by exploring the Brownian ratchet mechanisms in molecular machines. Following this, I will discuss the use of nucleic acids as a construction material and provide a brief review of DNA-based molecular machines, highlighting the underlying Brownian ratchet mechanisms. The feasibility of the hypothesised concept will then be discussed through a brief review of the structural, chemical, and kinetic properties of functional nucleic acids. Finally, I will introduce methods for developing nucleic acids with desired functions, including in vitro selection and engineering techniques. Our proposed optimised mechanism, which involves the development of a nucleic acid-based mechanocatalyst powered by a small-molecule fuel will be presented in the next chapter.

1.2 Artificial molecular machines

Since the famous "Maxwell's demon" proposed by James Clerk Maxwell over one and a half centuries ago, artificial molecular machines have been one of the end goals in the

field of nanotechnology ^[20, 21]. Motions are governed by different physical principles at different length scales. In contrast to the macroscopic world where energy is required to overcome inertia to make objects move, molecular-scale objects constantly experience random Brownian motion and operate under low Reynolds number conditions ^[2]. In order to make directional movements, they must either "swim in molasses and walk in a hurricane" ^[22] or find ways to harness random Brownian motion ^[23]. It is widely accepted that the latter is a more plausible scenario, and strategies to bias Brownian motion are called Brownian ratchets ^[24–26].

In Maxwell's thought experiment, the demon would extract directed motion from random Brownian motion by controlling a tiny door between two chambers containing gas ^[1]. It would allow fast-moving (high-energy) molecules to go to one chamber and slow-moving (low-energy) molecules to go to the other, thereby decreasing the total entropy of the system at no energy cost. However, a real-life Maxwell's demon must adhere to the second law of thermodynamics. Consequently, it expends more energy gauging the speed of each molecule approaching the door than it gains from the resulting net decrease in entropy. Feynman's ratchet and pawl device also illustrated that energy is required to harness Brownian motion ^[27]. The device consists of two main parts mechanically connected by a rod: a paddle wheel in one chamber and a ratchet with a pawl in another chamber, each filled with gas at different temperatures, T_1 and T_2 , respectively. Feynman demonstrated that the ratchet would rotate continuously in one direction only when the whole system is out of equilibrium ($T_1 \neq T_2$). This is because the pawl itself undergoes Brownian motion and when $T_1 = T_2$, the failure rate equals the forward ratcheting rate, and thus the device would not produce any net work.

Nature has evolved a wide range of biological molecular machines that sustain the life of organisms ^[28]. For example, intracellular vesicle transport and muscle contraction are performed by linear motor proteins walking along polar and periodic tracks (such as actin or microtubule filaments) ^[29–31], and the conversion of energy stored in the proton gradient into ATP is carried out by the rotary motor F_0F_1 -ATP synthase, which operates across the membrane by making 360° rotations ^[32]. Despite differing in structure and function, they all consume energy (most commonly from ATP hydrolysis, ion gradients, or redox

reactions) and make repetitive and directional movements [33]. Over the past three decades, research into molecular machines has flourished, attracting widespread interest from chemists, physicists, and biologists alike. This growing field explores how external energy drives controllable mechanical motion, as well as the underlying principles for designing artificial molecular machines [8]. Several Brownian ratchet mechanisms have been proposed, generally categorised into two groups: energy ratchets based on Feynman’s ratchet and pawl, and information ratchets based on “Maxwell’s demon” [6, 7].

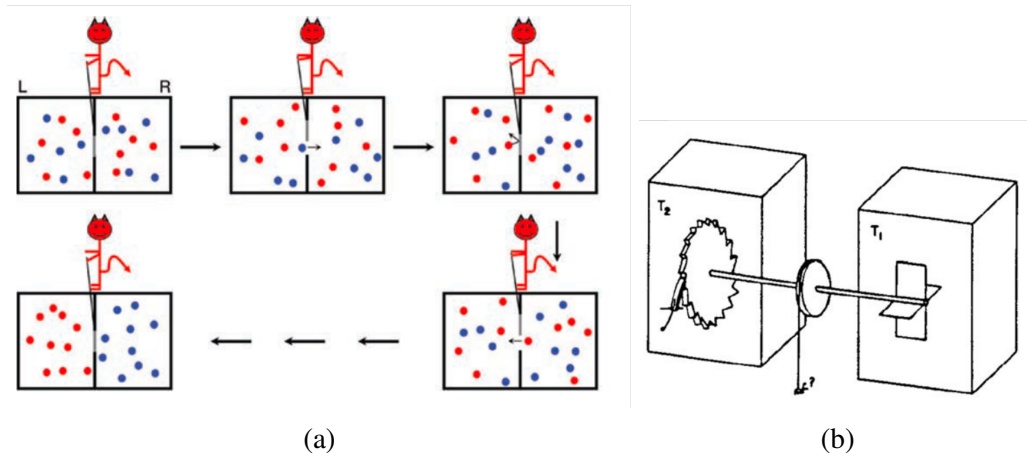


Figure 1.1: (a) Maxwell’s demon. The demon standing on the chambers sorts the high-energy molecules (red dots) from the low-energy molecules (blue dots) by selectively opening the door. Image adapted from ref. [7]. (b) Feynman’s ratchet and pawl device. Image adapted from ref. [34].

1.2.1 Energy ratchets

A Brownian ratchet system consists of three components [24]: a Brownian object, an anisotropic medium, and an energy input that drives the system out of equilibrium. In energy ratchets, input energy is provided through changes in external conditions, which in turn alter the relative depths of the energy wells and barriers on the potential energy surface [7]. This leads to a net directional bias in molecular motion. A classic energy profile for an energy ratchet is shown in Figure 1.2a. A system in which motion occurs has a periodic series of two different minima and two different maxima on the potential energy surface. Initially, the system is in equilibrium, and the Brownian object is in one of the lowest-energy minima. An energy input changes the conditions, disrupting the balance, raising the object’s energy, and lowering the energy barrier for a forward

movement. The object then undergoes directional relaxation (power stroke) to the new energy minimum. A second energy input returns the energy profile to its original state and repeats this cycle, driving the object to move directionally along the track.

The molecular walker that can walk directionally along a track shown in Figure 1.2b is an example of an energy ratchet. This walker features two dynamic orthogonal feet: a hydrazone foot, which hydrolyses under acidic conditions and locks under basic conditions, and a disulfide foot, which locks under acidic conditions and hydrolyses under basic conditions [35]. The direction in which the walker travels depends on a sequence of four stimuli [35]: basic or acidic conditions to selectively dissociate or associate the feet, and 500 nm ($h\nu^1$) or 365 nm ($h\nu^2$) light to induce or release ring strain between the walker and the track through an E-Z isomerisation of a stilbene unit located between the aldehyde and disulfide footholds on the track. In Figure 1.2b, the walker travels from left to right. Initially, the rear foot (disulfide) breaks under the basic condition. Then, the application of $h\nu^1$ disrupts the balance of the system by driving the isomerisation of stilbene from E to Z. This creates an energetic preference for forming the 2,3-macrocycle over the 1,2-macrocycle in a neutral pH medium, resulting in a forward step through a power stroke. The following application of $h\nu^2$ restores the ring strain between the walker and the track. Under acidic conditions, the rear foot (hydrazone) breaks, preparing the walker for a second forward step.

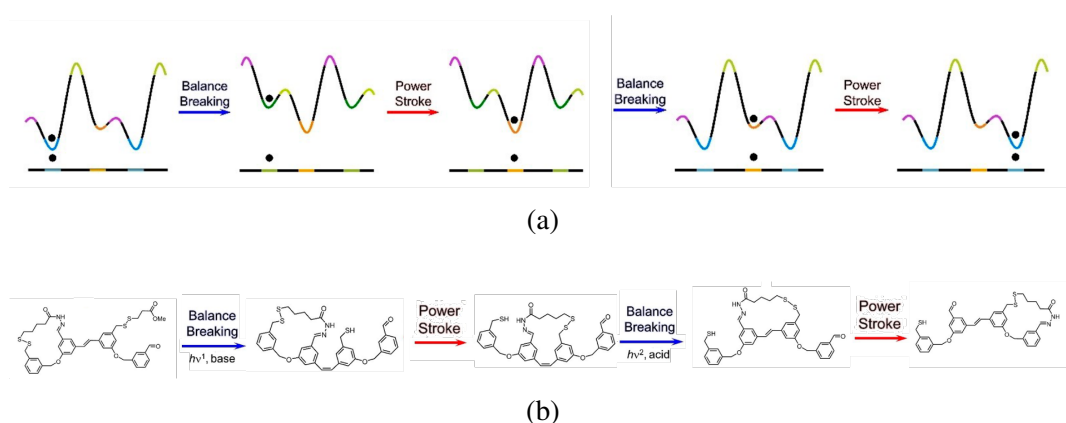


Figure 1.2: (a) A typical energy profile for an energy ratchet. For a detailed explanation of its operation, refer to the text. (b) A light and pH driven molecular walker. For a detailed explanation of its operation, refer to the text. Images and captions adapted from ref. [8].

Numerous energy ratchets have been recorded in the literature. Despite their varied designs and the different fluctuations in external conditions that drive them - such as

changes in pH ^[36, 37], temperature ^[38, 39], or electrochemical potential ^[40, 41] - all rely on switching energy profiles in response to fluctuations in external conditions to achieve directional movement ^[8]. However, this mechanism is fundamentally different from the Brownian ratchets of biological molecular motors. Biomolecular motors are driven by chemical fuels and operate autonomously until the fuel is depleted ^[33].

1.2.2 Information ratchets

ATP is the most common energy source for biomolecular motors. For example, the motor protein kinesin consumes one ATP molecule for each 8-nm walking step along the microtubule, producing around 40 pN · nm work (around $10k_B T$) which is roughly half of the free energy released from the hydrolysis of each ATP molecule ^[42]. Stepping occurs on the microsecond to millisecond timescale, whereas the local temperature gradient generated by the thermal energy released during ATP hydrolysis dissipates within picoseconds to nanoseconds over the nanometre-scale stepping distance. Thus, the energy ratchet model does not adequately explain the mechanism of operation, as the energy from ATP hydrolysis is not used to alter the overall energy profile or directly drive mechanical motion ^[43]. Instead, it is widely recognised that they operate through information ratchet mechanisms, where the Brownian object acts as a catalyst for the chemical fuel, and its mechanical cycle is coupled to the chemical fuel-to-waste conversion ^[6, 7, 44].

A typical chemomechanical cycle ^[45] of a chemically fuelled molecular information ratchet operates as illustrated in Figure 1.3a. The Brownian object (the catalyst) interconverts between mechanical states A and B. Both states A and B can bind to the fuel, forming intermediate complexes A' and B', respectively, which also interconvert dynamically. The relative rate constants for the fuel-catalyst association and the waste-forming reaction depend on the mechanical state of the catalyst, creating a chemical gating mechanism. Simultaneously, the relative rates of mechanical transitions between A and A' and between B and B' depend on the chemical state of the fuel, creating a mechanical gating mechanism. The combination of chemical and mechanical gating results in kinetic asymmetry, which biases the direction of movement ^[8].

Amano and colleagues ^[46] highlighted the close relationship between kinetic asym-

metry in Brownian information ratchets and the well-established Curtin–Hammett principle in organic chemistry. The Curtin–Hammett principle states that for two interconverting substrates reacting with the same reagent to produce two different products (kinetic resolution), the product ratio is determined by the equilibrium constant between the interconverting starting materials and the relative activation energies of the reactions [47, 48]. The authors pointed out that the main difference between a traditional kinetic resolution and a Brownian ratchet is that the latter operates through a cyclic process. The two paths (fuel-catalyst association and waste formation reaction) that make up the information ratchet can be presented separately using the Curtin–Hammett principle (Figure 1.3b). The difference in the transition state energies ($\Delta\Delta G^\ddagger$) creates the kinetic asymmetry in each step.

The overall directionality is determined by the Curtin–Hammett asymmetry factor (F_{C-H}):

$$F_{C-H} = \frac{\text{Forward coupled cycles}}{\text{Backward coupled cycles}} = e^{\left(\frac{\Delta\Delta G_{\text{fuelling}}^\ddagger + \Delta\Delta G_{\text{waste forming}}^\ddagger}{RT}\right)} \quad (1.1)$$

or alternatively by the ratcheting constant K_r [49]:

$$K_r = \frac{F_{C-H} \cdot e^{\left(-\frac{\Delta\mu}{RT}\right)} + 1 + \gamma}{F_{C-H} + e^{\left(-\frac{\Delta\mu}{RT}\right)} + \gamma} \quad (1.2)$$

where R is the gas constant and T is the absolute temperature, and $\Delta\mu$ is the difference in the chemical potential between the fuel and the waste and γ is an uncoupling constant that describes the relative likelihood of slip cycles:

$$\gamma = \frac{\text{Slip cycles}}{\text{Backward coupled cycles}} \quad (1.3)$$

For an out-of-equilibrium system ($\Delta\mu > 0$), the system has a forward direction when $F_{C-H} > 1$ or $K_r > 1$, a backward direction when $F_{C-H} < 1$ or $K_r < 1$, and no net directionality when $F_{C-H} = 1$ or $K_r = 1$.

The Curtin–Hammett principle can also be applied to reactions associated with biomolecular motors to visualise the underlying mechanism of biased motion [46]. For example, consider the kinesin discussed above. Kinesin is considered to have two identical

feet that carry cargo and walk along a microtubule ^[50]. It is powered by catalysing the conversion of ATP to ADP and inorganic phosphate (Pi). As shown in Figure 1.3d, each foot undergoes three different reactions, resulting in three distinct mechanical states with different standard chemical potentials: ATP-bound (T), ADP-bound (D), and empty (E). When ATP is bound, the foot catalyses its hydrolysis, and then the ADP-bound head detaches from the microtubule track. The empty foot binds to the track again once the bound ADP is released, making it available for another ATP. The ADP release reaction involves two interconverting mechanical states with an equilibrium constant K_s : the detached foot can be trailing or leading. The overall Curtin–Hammett asymmetry factor F_{kinesin} is ^[46]:

$$F_{\text{kinesin}} = \frac{k_{+T}^F k_{+h}^B k_{-D}^F}{k_{+T}^B k_{+h}^F k_{-D}^B} K_s \quad (1.4)$$

There is a continuing debate about whether the directionality of biomolecular motors originates from biased diffusion (Brownian ratchet) or a mechanical mechanism through a spring-like element (power stroke) ^[42, 51, 52]. Nevertheless, it is reasonable to consider the Brownian ratchet mechanism as the fundamental principle underlying chemically driven molecular motors, with the power stroke being one of the elements that contribute to the directionality of Brownian ratchets ^[46]. When the two interconverting mechanical states have different standard chemical potentials, a conformational change (power stroke) occurs along that downhill free energy gradient. In this scenario, $K_s \neq 1$, contributing to the overall directionality as shown in equation 1.4. This concept is illustrated intuitively in Figure 1.3d, where the standard chemical potential difference, denoted as ΔG_s° , is highlighted in the kinetic resolution of the ADP release step.

The first autonomous chemically-fuelled motor (Figure 1.3c) designed based on the information ratchet mechanism was reported by the Leigh group in 2016 ^[9]. This motor system features a small molecular ring (macrocycle) that continuously rotates directionally around a cyclic molecular track, driven by the chemical fuel 9-fluorenylmethoxycarbonyl chloride (Fmoc-Cl). There are two reaction sites for the fuel on the cyclic molecular track. A pyridine-based catalyst promotes carbonate-forming reactions between Fmoc-Cl and the track, blocking the passage of the macrocycle. Fmoc groups are then cleaved by triethylamine (NEt₃) independently of the location of the macrocycle. The key to kin-

etic asymmetry in the system is that the Fmoc attachment rate varies depending on the macrocycle's position. Due to steric interactions between the macrocycle and the bulky catalyst, the attachment is faster when the macrocycle is far from the reactive site than when it is near. This motor achieved an approximately 4:1 ratio of forward to backward directionality.

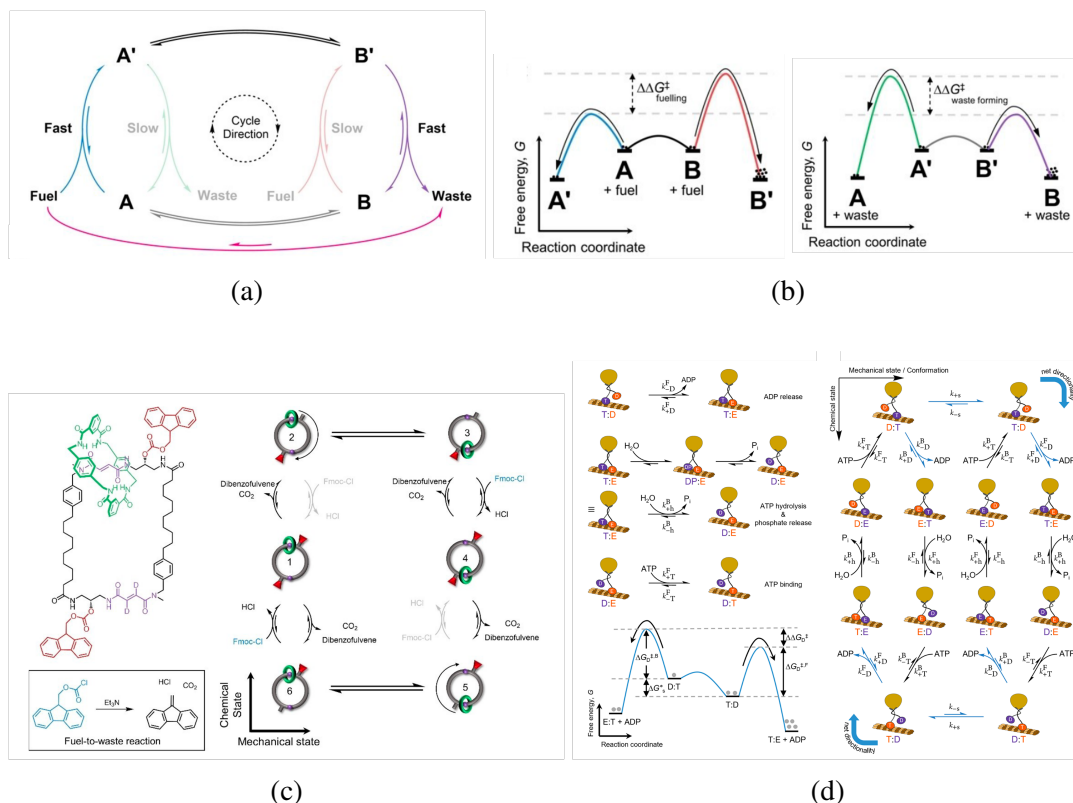


Figure 1.3: (a) Reaction diagram of a chemically fuelled molecular information ratchet. State A and State B interconvert and both react with the fuel, with State A reacting more rapidly. Similarly, State A' and State B' interconvert, with State B' undergoing a faster waste release reaction. This dynamic asymmetry drives a clockwise cycle progression. Image and caption adapted from ref. [8] (b) Curtin-Hammett diagram showing the fuelling reaction (left) and the waste-forming reaction (right) within the chemomechanical cycle. The direction of mechanical motion is biased by differences in the energy of transition states. Image and caption adapted from ref. [8] (c) An autonomous Fmoc-Cl fuelled motor. For a detailed explanation of its operation, refer to the text. Image and caption adapted from ref. [8] (d) The three different reactions that kinesin feet undergo are shown on the top left. The bottom left presents a Curtin-Hammett diagram of the ADP release. The chemomechanical cycle for kinesin is illustrated on the right. For a detailed explanation of its operation, refer to the text. Image and caption adapted from ref. [46]

More autonomous chemically-fuelled molecular motors have been documented in recent years. These motors exhibit various designs and modes of motion, such as pumping [53] and rotation [54], and are driven by catalysing different chemical reactions [44]. While these small-molecule systems are impressive, they need to be built from scratch,

often requiring complex chemistry and custom synthetic routes ^[55]. Meanwhile, in the interdisciplinary field of DNA nanotechnology ^[56], the concept of oligonucleotide-based molecular machines has pioneered new frontiers in the development of artificial molecular machines, providing unparalleled levels of control, precision, and modularity. In the next section, I will introduce nucleic acids as a building material for complex structures and molecular machines, and discuss the underlying Brownian ratchets in their operation mechanisms.

1.3 DNA nanotechnology

1.3.1 Nucleic acids as building materials

Nucleic acids, including ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), are essential molecules in biology. These biomolecules are linear polymers composed of nucleotide (nt) monomers (ribonucleotide for RNA and deoxyribonucleotide for DNA). Each monomer consists of a 5-carbon sugar (ribose in RNA and deoxyribose in DNA), a phosphate group, and a nitrogenous base (Figure 1.4a). The sugar in DNA lacks a hydroxyl group at the 2' carbon. Nucleotides are joined through phosphodiester bonds formed between the hydroxyl group at the 3' carbon of the sugar ring and the phosphate group of the next nucleotide, resulting in a sugar-phosphate backbone (Figure 1.4c). There are five types of bases: adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U). A, G, and C are common to both RNA and DNA, while T is found only in DNA, and U is exclusive to RNA. A and G are purines; C, T, and U are pyrimidines. Nucleic acids are known as carriers of genetic information, which is encoded in the sequence of the nitrogenous base attached to the sugar-phosphate backbone. The sequence is denoted by single-letter base abbreviations and is read from the 5' to the 3' end, corresponding to the direction of strand growth during biological synthesis.

Nucleotides with different bases can form specific base pairs through hydrogen bonds following the rules proposed by James Watson and Francis Crick ^[57]: a purine pairs with a pyrimidine; A pairs with T (or U in RNA); C pairs with G (Figure 1.4b). Through hydrogen bonds, a nucleic acid strand can adhere to another strand with a reverse com-

plementary sequence to form a double helix. This process is called hybridisation. In biology, DNA is typically found as an antiparallel double-stranded (dsDNA) helix with a complementary strand. The most common type of dsDNA is the B-form double helix (B-DNA) (Figure 1.4d). B-DNA is a right-handed double helix with two antiparallel ribose-phosphate backbones on the outside and planar base pairs (bp) in the middle [58]. The aromatic bases stack closely on top of each other with no space for the solvent [59]. Base stacking and backbone constraints give rise to a helical structure. The combination of hydrogen bonding and stacking makes the double helix a stable and stiff structure. Under physiological conditions, the persistence length, which quantifies the bending stiffness, is roughly 50 nm for dsDNA [60]. An RNA molecule is usually single-stranded in nature. When in its double-stranded form (dsRNA), it is usually an A-form double helix (A-RNA) with a persistence length of around 62 nm [61]. In contrast to B-DNA, which features a broad major groove and a narrow minor groove with 10 base pairs per turn and an average helical rise of 3.4 Å, A-RNA has a narrow and deep major groove and a wide and shallow minor groove with 11 base pairs per turn and a typical rise of 2.5 Å [62]. Although persistence length is influenced by factors such as sequence, ionic strength, and temperature, single-stranded nucleic acids are generally much more flexible compared to their double-stranded versions, with significantly shorter persistence lengths, approximately 2 nm for single-stranded DNA (ssDNA) and 1 nm for single-stranded RNA (ssRNA) [63–65].

The melting temperature T_M is defined as the temperature at which half of the duplexes are denatured. It can be calculated using the following equation:

$$T_M = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln C_T} \quad (1.5)$$

where ΔS° is the standard entropy change, ΔH° is the standard enthalpy change, R is the gas constant, and C_T is the total concentration of oligonucleotide strands.

Pioneered by Tinoco [66] and Zimm [67] more than 60 years ago, the Nearest-Neighbor (NN) model has been the most widely used method for T_M prediction [68]. In the NN model, the thermodynamic parameters ΔH° , ΔS° , and ΔG° for a base pair in a duplex depend on the identities of its neighbouring base pairs. The overall ΔH° , ΔS° , and ΔG°

for the duplex are the sums of the corresponding parameters for each pair of neighbouring base pairs within the duplex.

For example, for a duplex with sequence 5'CCGTA/3'GGCAT:

$$\begin{aligned}\Delta G^\circ(\text{predicted}) &= \Delta G^\circ(\text{CC/GG}) + \Delta G^\circ(\text{CG/GC}) \\ &\quad + \Delta G^\circ(\text{GT/CA}) + \Delta G^\circ(\text{TA/AT}) \\ &\quad + \Delta G^\circ(\text{Init}) + \Delta G^\circ(\text{Sym})\end{aligned}\tag{1.6}$$

where ΔG° denotes the standard free-energy change, $\Delta G^\circ(\text{Sym})$ is an additional entropic penalty when the duplex is self-complementary, and $\Delta G^\circ(\text{Init})$ is the terminal base pair-dependent initiation parameter ^[69, 70].

ΔG° can also be calculated from the NN parameters ΔH° and ΔS° :

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ\tag{1.7}$$

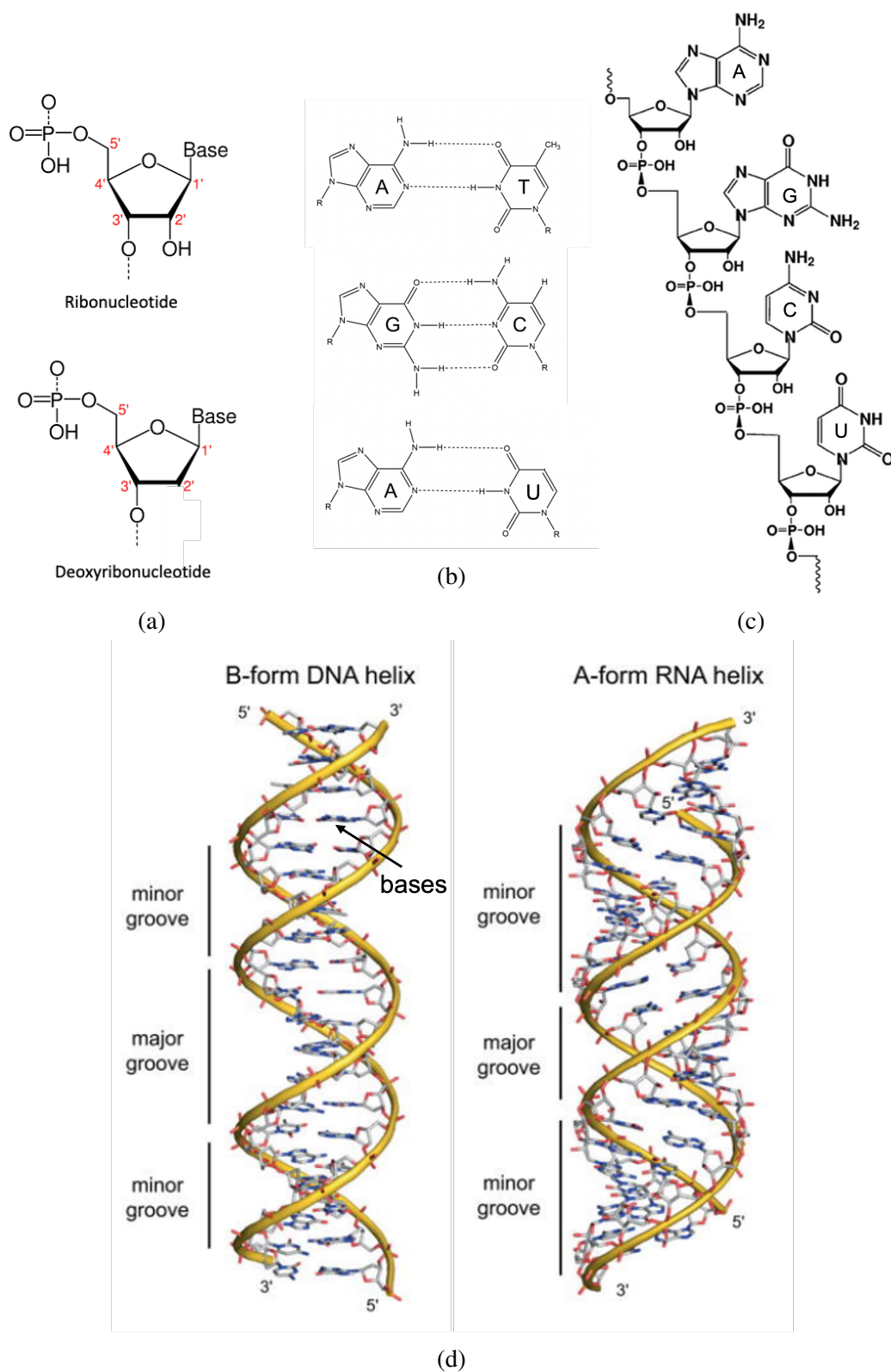


Figure 1.4: (a) Molecular structures of ribonucleotide and deoxyribonucleotide. (b) Watson-Crick base pairings. (c) The connection of nucleotides in a single-stranded RNA. (d) B-form DNA helix and A-form RNA helix. Image and caption adapted from ref. [71].

The programmable base-pairing interactions, together with their controllable flexibility and rigidity, make nucleic acids an ideal material for building nanostructures with precise control. Between the two types of nucleic acids, DNA has been studied more intensively than RNA. This is mainly because RNA has more complex molecular interactions and is more prone to hydrolysis due to the presence of the 2'-hydroxyl group in ribose sugar [72, 73], and also because DNA production is more cost-effective [74]. In 1982, Nadrian Seeman's landmark paper laid the theoretical foundation for the field of DNA nanotechnology by proposing the design of specific oligonucleotide sequences to form stable junctions instead of the usual linear duplexes [75].

Since then, numerous intricate 2D and 3D self-assembled DNA nanostructures have been created [56, 76]. These nanostructures can be roughly classified into two categories: unscaffolded and scaffolded structures [77]. Unscaffolded structures can be constructed by organising rigid duplexes and flexible branched junctions through the annealing of multiple oligonucleotides [78]. This method has been used to generate a diverse range of DNA polyhedra with varying sizes and geometries [79–81]. The DNA tetrahedron depicted in Figure 1.5a is an example [80]. Another strategy for creating unscaffolded structures is using DNA tiles [82–85]. DNA tiles are simple and symmetric motifs assembled from DNA strands with designed sequences. Figure 1.5b shows a three-arm-junction DNA tile and a variety of polyhedra assembled by hybridising their sticky ends [86]. LEGO®-like DNA bricks have been developed to construct larger and more intricate 3D structures (Figure 1.5c) [87, 88]. These modular units, made of a single strand, form the rigid elements of the desired structure through hybridisation of their four interaction domains during simple one-step annealing reactions. Consequently, brick-based assembly does not require the precise adjustments to strand stoichiometry that are often required in hierarchical tile-based assembly [76]. Scaffolded structures are formed using a technique called DNA origami, where a long ssDNA strand serves as a scaffold, and hundreds of short staple strands are annealed to specific regions of the scaffold strand to fold it into the desired 2D or 3D shape [89]. The M13 virus genome (a 7249-nt-long circular ssDNA) is the most common choice for scaffolding due to its well-defined sequence and widespread commercial availability [90]. Different structures can be created using the same scaffold

and different staple strands. Each scaffold recruits a single complete set of staples, without the need to control the relative stoichiometry of hundreds of staple strands. The smiley face and the nanoscale box with a lid shown in Figure 1.5d are early examples of DNA origami^[89, 91].

During the past decade, advances in understanding the intermolecular and intramolecular interactions of RNA molecules, coupled with the decreasing cost of RNA production through biological and chemical synthesis, have propelled rapid progress in the field of RNA nanotechnology^[92, 93]. Its increasing popularity is mainly due to the demonstrated advantages in nanomedicine, notably its minimal genotoxic effects within the body compared to DNA^[94–96]. Although RNA and DNA exhibit slightly different folding properties, the basic principles of DNA nanotechnology are readily applicable to their RNA counterparts^[97, 98].

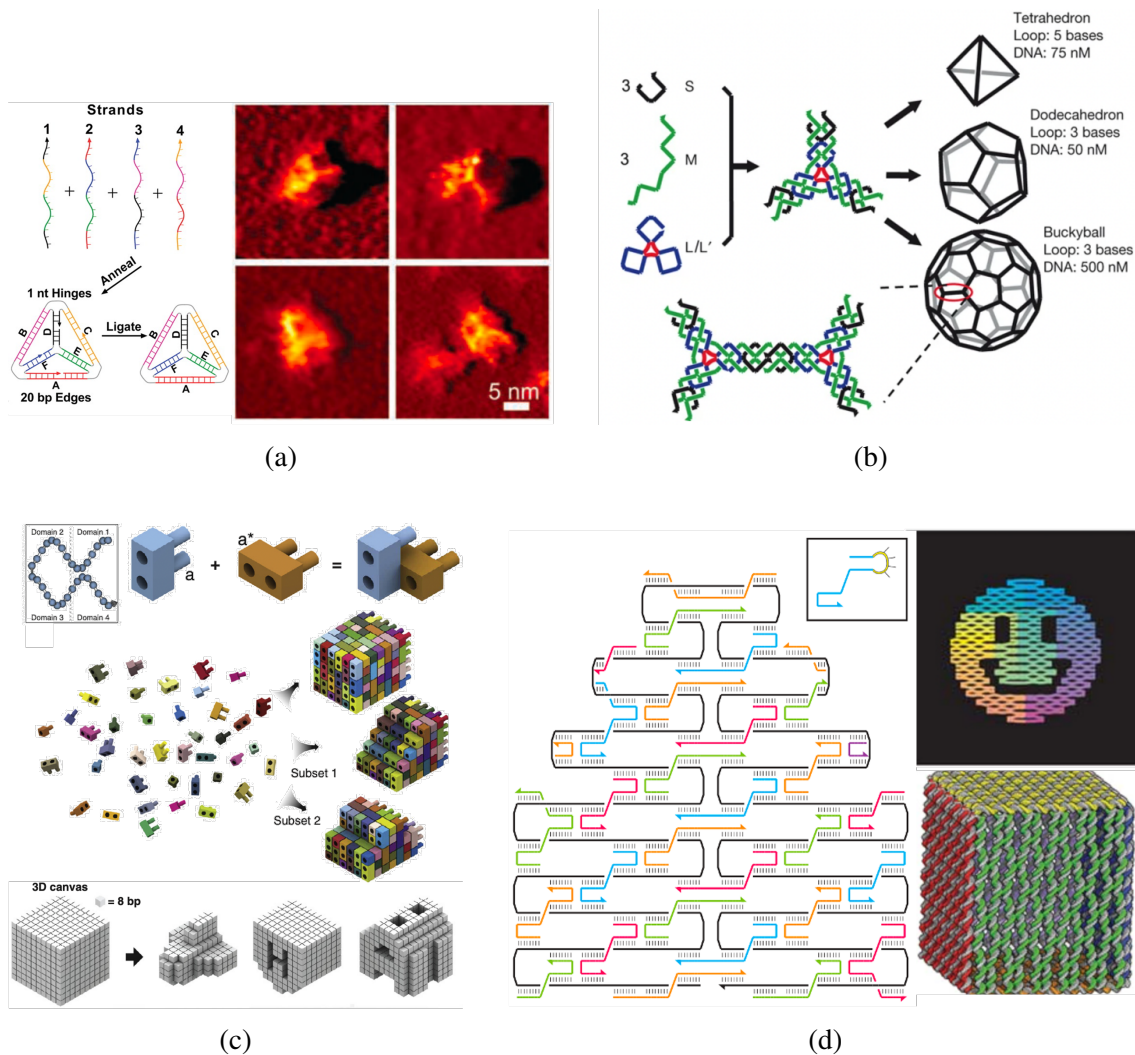


Figure 1.5: Self-assembly of 2D and 3D nanoscale architectures from oligonucleotides. (a) Design and atomic force microscope images of the self-assembly of a DNA tetrahedron through annealing four oligonucleotides. Regions with the same colour complement each other. Image and caption adapted from ref. [80]. (b) A three-arm-junction DNA tile was used to assemble a tetrahedron, a dodecahedron, and a buckyball. This assembly was achieved by controlling both the strand stoichiometry and the length of the flexible linkage (highlighted in red) within the DNA tiles. Image and caption adapted from ref. [86, 99]. (c) 3D structures assembled using DNA bricks. Each DNA brick contains four 8-nt domains and is able to connect four neighbouring bricks. The bricks resemble LEGO[®]-like blocks, and form 90° dihedral angles with neighbouring bricks. This geometry arises from the approximately 3/4 right-handed helical twist of 8 base pairs of DNA. Intricate 3D shapes are designed using a "3D canvas", with each voxel representing 8 base pairs (2.5 nm x 2.5 nm x 2.7 nm). Image and caption adapted from ref. [87]. (d) A DNA origami smiley face and a single-layer DNA origami box created using a biological m13 scaffold strand and hundreds of synthetic staple strands. Image and caption adapted from ref. [89, 91]

1.3.2 Oligonucleotide-based molecular machines

Toehold-mediated strand displacement reactions

A strand in a duplex can be displaced by another homologous strand, resulting in the formation of a new duplex. This spontaneous process can occur via two pathways depending on the temperature ^[100]. When the temperature is near T_M , the first pathway dominates. In this pathway, the original duplex completely dissociates before the invading strand hybridises. At lower temperatures, the second pathway is more likely. Here, the invading strand starts hybridising from one end as the original duplex frays (the ends of the duplex partially unzip). This hybridisation then displaces the old strand through a random walk of the branch point ^[101]. The random walk is commonly referred to as the three-way branch migration. Both of the two pathways rely on complete or partial dissociation of the duplex. Morrison and Stols ^[102] used fluorescence measurement to study the thermodynamic and kinetic behaviours of DNA hybridisation in solution. Under typical reaction conditions, the dissociation rate constant for a 10-bp dsDNA was measured at approximately $1 \times 10^{-2} \text{ s}^{-1}$, and the rate for a 20-bp dsDNA was around $1 \times 10^{-10} \text{ s}^{-1}$. This implies that spontaneous dissociation is negligible on the common length scale (a few tens of bp) used in DNA nanotechnology. Nevertheless, it was noted in the paper that a nucleation site for the invading strand, such as a mismatch or a single-stranded region on the duplex, could promote the rate of strand displacement ^[102].

In 2000, Yurke and coworkers ^[103] first used this strategy to speed up strand displacement reactions by introducing a single-stranded overhang to a duplex, now known as a "toehold". They constructed a dynamic pair of DNA tweezers (Figure 1.6a). The tweezers have two stiff double-stranded arms connected by a flexible 4-nt single-stranded hinge. Each arm has a single-stranded sticky end (toehold). In the open state, the two arms move freely around the hinge with a mean distance between the ends of around 6 nm. The addition of a single-stranded DNA "fuel" (F) pulls the arms together, bringing the tweezers to the closed state through hybridisation with the two toeholds. The new sticky end of F then serves as another toehold for the "antifuel" ($\bar{F}$) to open the tweezers by toehold-mediated strand displacement (TMSD), generating a double-stranded waste product. Since then, TMSD has been commonly used as a powerful mechanism in dy-

namic DNA nanotechnology ^[104].

Figure 1.6b illustrates a typical TMSD process. First, the invading strand (F) binds to the substrate through hybridisation with the toehold (T). This docking step is reversible due to the potential fraying of the toehold. Once the toehold is bound, the invading strand competes with the incumbent strand (I) for binding to the substrate. There is no significant thermodynamic bias for the branch point to move in either direction since the free energy difference for both processes is roughly zero.

The TMSD reaction shown in Figure 1.6b can be simply represented using the following equation:



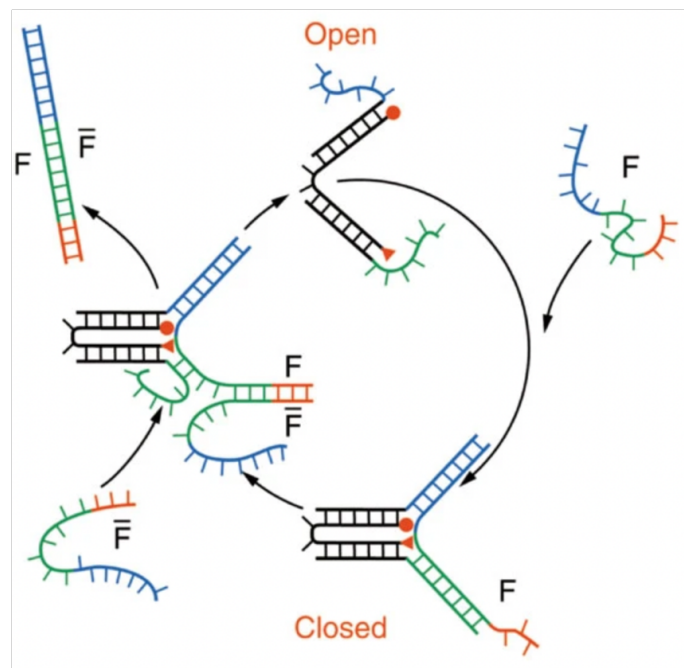
where k_1 and k_2 are forward and reverse rate constants.

In stoichiometric conditions, the equilibrium concentration ratio of $[D1]$ (the incumbent strand paired with the substrate) to $[D2]$ (the substrate strand paired with the fuel) is [105]:

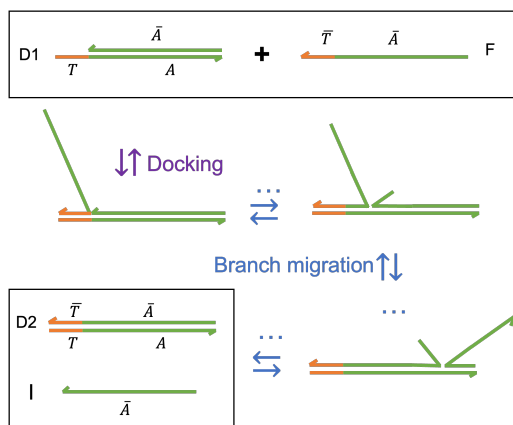
$$\frac{[D1]}{[D2]} = \sqrt{\frac{k_2}{k_1}} = \sqrt{e^{\Delta G/kT}} \quad (1.9)$$

where ΔG is the Gibbs free energy change due to the duplex formation of the toehold, k is Boltzmann constant, and T is absolute temperature measured in Kelvin.

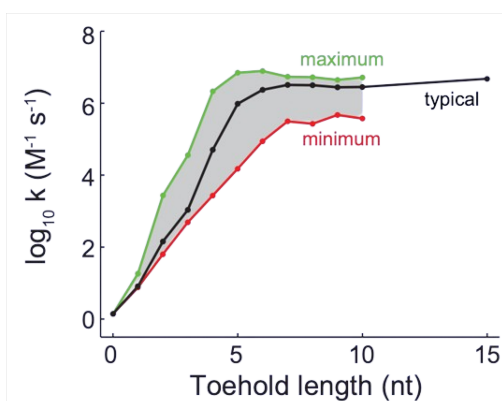
The formation of each base pair releases an average free energy of approximately $1.4 \text{ kcal} \cdot \text{mol}^{-1}$ ^[68], which indicates that a 3-nt toehold can drive the reaction to near 100% completion. Zhang and Winfree ^[106], observed that the rate of strand displacement reactions increases exponentially with the toehold length and plateaus around 6 nucleotides, and it can be controlled by varying the toehold length and sequence composition. Figure 1.6c shows that these kinetics can be modeled based on the characteristics of the toehold, with the rate constant of strand invasion at room temperature ranging from $1 \text{ M}^{-1} \text{ s}^{-1}$ to $6 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$.



(a)



(b)



(c)

Figure 1.6: (a) Construction and operation of the molecular tweezers. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[103]. (b) Toehold-mediated strand displacement reaction. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[107]. (c) Dependence of reaction rate on toehold length and composition. The green trace represents the kinetics of a toehold made only of G/C nucleotides, the red trace represents a toehold made only of A/T nucleotides, and the black traces show a toehold with a mix of all four nucleotides. The gray area between the green and red traces indicates the range of possible kinetics based on the toehold length. Image and caption adapted from ref. ^[108].

Non-autonomous DNA-based molecular machines

In 2004, Sherman and Seeman^[109] reported the first DNA walker driven by TMSD (Figure 1.7a). The walker has two double-stranded legs and two single-stranded feet, named A and B. These legs are connected by three flexible linker strands. The rigid triple crossover (TX) track features three single-stranded footholds, numbered 1, 2, and 3 from left to right. Initially, the walker attaches to footholds 1 and 2 through hybridisation with two anchor strands, 1A and 2B. The 1A strand is complementary to both foot A and foothold 1, while the 2B strand is complementary to both foot B and foothold 2. Fuel strand 2B competes with foot B and foothold 2 for binding to anchor strand 2B. This competition eventually lifts foot B from the track through toehold (purple)-mediated strand displacement. Next, the addition of a new anchor strand, 3B (complementary to both foot B and foothold 3), allows foot B to step forward to foothold 3. In the same way, fuel strand 1A lifts foot A, while anchor strand 2A attaches foot A to foothold 2. The walker moves one step to the right after each cycle of adding four sequential fuel and anchor strands.

Using a similar mechanism of sequential addition of fuel and anchor strands, the Seeman group built a programmable DNA nanoscale assembly line (Figure 1.7c) in 2010^[110]. The system operates on an approximately 300 nm-wide 2D DNA origami with 18 protruding single-stranded footholds and three DNA cargo delivery machines carrying different DNA-bound gold nanoparticles (C1, C2, and C3). The worker on the assembly line is a DNA walker with four feet (F) and three hands (H). Sequential additions of fuel and anchor strands control the movement of the walker on the DNA origami track. When the walker is in proximity to a cargo delivery machine, the gold cargo handover happens through competitive hybridisation between the hand and the cargo delivery machines to capture the strand attached to the gold particle. The handover of cargo can be programmed to be either "on" or "off", and there are 8 different combinations in total. The behaviour of the system was investigated by using polyacrylamide gel electrophoresis (PAGE), atomic force microscopy (AFM), and scanning electron microscopy (SEM). The authors demonstrated that the expected products of the walker-cargo complex were formed in yields greater than 75% in each of the 8 scenarios.

Both of the above systems work non-autonomously through Brownian energy ratchet

mechanisms, as shown in Figure 1.7b. The introduction of fuel strands changes the whole energy landscape of the system, causing the walker to move towards new energy minima by diffusion. Each step involves external intervention, and the source of energy is the free energy from the formation of extra base pairs in the stable duplex waste.

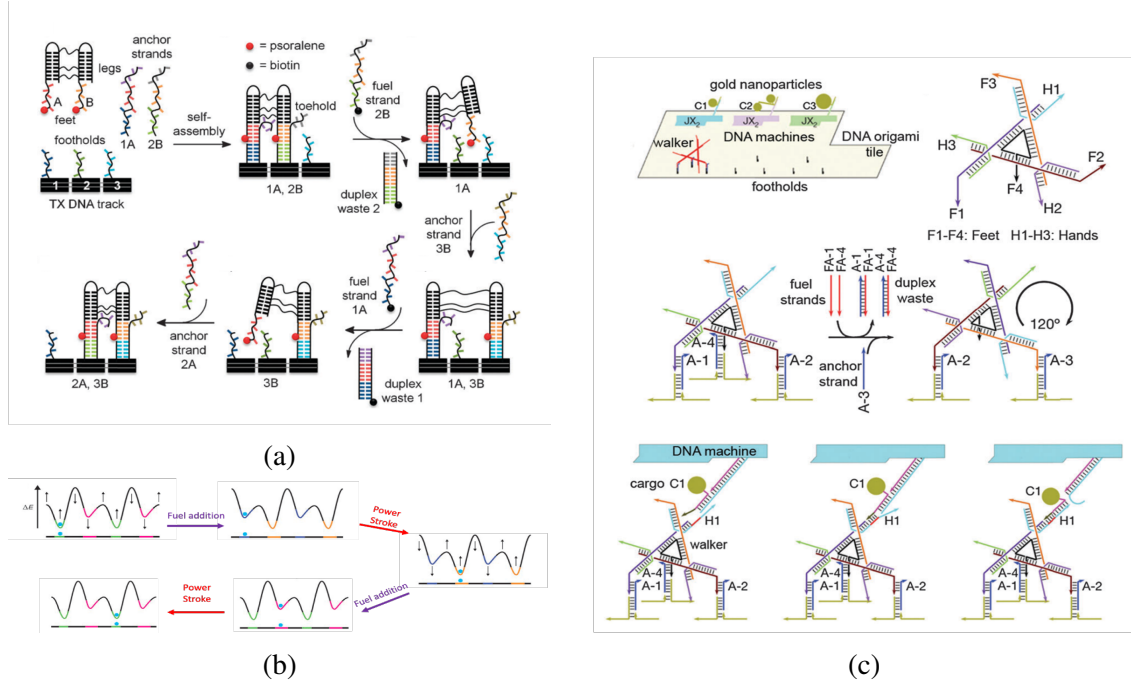


Figure 1.7: (a) The first DNA walker that is driven by TMSD ^[109]. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[55]. (b) A typical energy profile for non-autonomous DNA-based machines driven by TMSD. (c) The construction of the DNA nanoscale assembly line reported by the Seeman group. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[110].

Autonomous DNA-based molecular machines

Shortly after the report of the first stepwise DNA walker in 2004, an autonomous DNA walker that moves unidirectionally along a double-stranded track was published by the Reif group and the Turberfield group ^[111]. In this paper, directed movement is coupled with enzymatic reactions that involve three enzymes. In this process, first, the walker covalently binds to the first foothold. Then, the walker attaches to the next foothold through toehold hybridisation. Next, a ligase (T4 ligase) ligates the nicks between the walker and the new foothold, creating a recognition site for one of two restriction enzymes (PflIM I and BstAP), which then perform a selective cleavage reaction. In 2005, Bath and colleagues introduced a different system, reducing the number of enzymes required from

three to one ^[112]. The linear motor operates via a “burnt bridges” mechanism, using a nicking enzyme that cuts the single-stranded stator on the track where the single-stranded motor is bound. The short strand cut from the stator then dissociates from the motor due to its reduced melting temperature, revealing a toehold region. The motor can relocate to the next neighbouring single-stranded stator via TMSD. The motor is prevented from moving backwards because the toehold region of the previous stator is removed by the nicking enzyme. In the same year, Tian and co-workers reported an autonomous DNA nanodevice that operates without protein enzymes ^[113]. They adopted a mechanism similar to that of Bath’s motor, with the key difference being that the walker itself is a catalytic DNA (DNAzyme) that cleaves the track. Observing the autonomous movement of early DNA nanodevices is challenging due to the short length of the tracks ^[114]. To address this, Wickham and coworkers ^[115] assembled a 100-nm-long DNA track on a 2D DNA origami and used real-time AFM to directly observe the movements of a motor driven by a nicking enzyme ^[112] (Figure 1.8a). Cha and coworkers published a DNAzyme-based walker that transports a gold nanoparticle over 3 μm on a carbon nanotube track (Figure 1.8b) in 2014 ^[116]. The motor’s movement and precise positioning of the motor relative to the nanotube track were tracked in real time with fluorescence microscopy and AFM.

In recent years, more enzymes that act on nucleic acids, such as RNase H ^[117] and RNase A ^[118], have been introduced into autonomous nucleic-acid-based systems using mechanisms similar to those discussed above. Although these systems demonstrate remarkable autonomy and controllability ^[115, 119] and high processivity over micrometer distances ^[116], enzyme-driven burnt bridge mechanisms have limitations. The biochemical activity of protein enzymes requires more delicate environmental conditions, and most tracks can be used only once by a single motor ^[120].

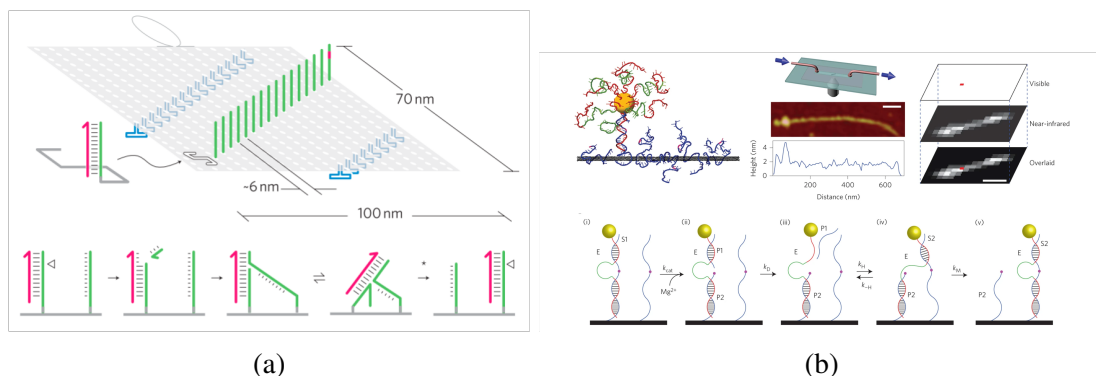


Figure 1.8: (a) An autonomous DNA motor driven by a nicking enzyme on DNA origami tile. The DNA origami tile (grey) contains the single-stranded stators of the track (green). A nicking enzyme cuts the stator bound to the motor, exposing a toehold region on the motor (magenta). This toehold allows the motor to move to the adjacent intact stator through TMSD. Image and caption adapted from ref. ^[115]. (b) A DNAzyme motor that move along carbon nanotubes. The DNAzyme motor has a catalytic region (green) flanked by two recognition arms (red) that are complementary to the DNA stator strands (blue) decorated on the surface of carbon nanotubes (black). Each gold nanoparticle cargo is decorated with multiple DNAzyme motors. Image and caption adapted from ref. ^[116].

A catalysis-driven DNA walker

In 2008, the Turberfield group developed a clever design of a purely DNA-based autonomous motor with a reusable track ^[10]. The system consists of a single-stranded track with three binding regions (1, 2, and 3) and a walker with two single-stranded legs connected through hybridization at their 5' ends. Each leg has an identical foot (foot A and foot B) at its 3' end (Figure 1.9a). The walker is driven by two hairpin fuels (H1 and H2). The autonomous movement is based on the concept of using DNA loops as fuel, a mechanism proposed by Turberfield and colleagues in 2003 ^[121]. As depicted in Figure 1.9b, spontaneous hybridisation between two complementary strands is inhibited when the ends of each strand are complementary and hybridised to form a hairpin structure. This inhibition occurs in two ways ^[121]: first, the stem of the hairpin makes the single-stranded loop less flexible and prevents the formation of a helix if the loop is small and the stem is stable enough; second, to form a full duplex, the base pairs in the two stems must be disrupted to open the loops. This hybridization process can be made over 1000 times faster by using a catalyst strand that hybridises with the stem region and opens the loop through strand displacement ^[122].

As illustrated in Figure 1.9a, in the initial state, the walker is attached to the track through partial hybridisation with the foot-binding regions 1 and 2. There is an overlap-

ping binding site on the track indicated by a green arrow (i). Foot B acts as a catalyst to open the loop of foot A, revealing a toehold for hairpin fuel H1 (ii). H1 then lifts foot A from the track (iii), and foot A acts as a catalyst to open the loop of H1 (iv), revealing a toehold that allows H1 to be removed from foot A by H2 through strand displacement (v). Once released (vi), foot A can return to position 1 or move forward to position 3 (vii). Although the rebinding step is not directionally biased, the lifting of the left foot is designed to be kinetically preferred.

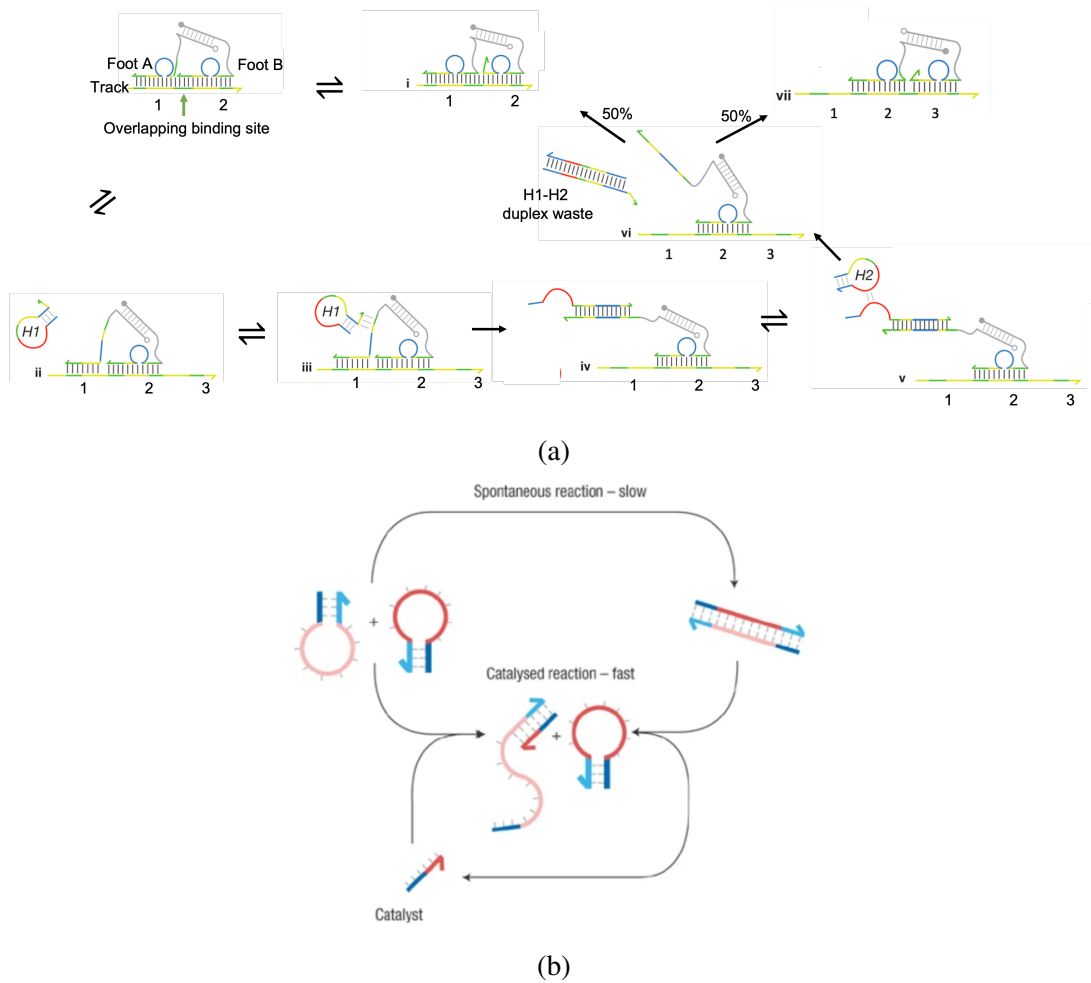


Figure 1.9: (a) Design of the catalysis-driven DNA walker. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[10]. (b) Hairpin loops as a fuel. Image adapted from ref. ^[123].

This motor system can be considered as a catalysis-driven information ratchet. The mechanical cycle of the motor is coupled with the hybridisation reaction of the hairpin fuels H1 and H2, and the walker is the catalyst for the reaction. The overall directionality is biased by the left and right feet having different rate constants for binding to H1. The energy input that drives the system out of equilibrium is the free energy released from the

hybridisation of H1 and H2.

However, this autonomous walker suffers from low speed. Its stepping rate is limited by the rate of hybridisation of opened hairpin fuel H1 and hairpin fuel H2, which occurs at a typical rate of $10^4 \text{ M}^{-1}\text{s}^{-1}$ [122]. In the presence of 100 nM H2, the expected mean velocity is 0.006 nm/s with a step size of 6 nm, which is remarkably slow compared to protein motors, which exhibit speeds ranging from 0.02 to 60 $\mu\text{m/s}$ [124]. Nevertheless, this speed is similar to the speeds observed in other TMSD-based autonomous DNA walkers [115, 116], with the exception of one that achieved speeds up to nanometers per second, driven by a burnt-bridges mechanism using RNase H [117].

In addition to TMSD reactions, stimuli such as pH change [125], photonic signals [126], and oscillating electric current [127] have been used in dynamic DNA nanotechnology. However, these systems generally operate in stepwise mode through energy ratchet mechanisms.

1.4 Functional nucleic acids

Compared to DNA, RNA is a more versatile molecule in nature. There are three main types of RNA in the cell: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). mRNAs carry the genetic code transcribed from DNA and instruct ribosomes to assemble amino acids into the primary structures of the encoded proteins. In the non-coding regions of mRNA, riboswitches regulate gene expression by binding specific metabolites [128]. These regulatory elements consist of an aptamer domain and an expression platform, connected by a switching sequence. Aptamers are oligonucleotides that bind ligands with high affinities and specificities. When a ligand binds to the aptamer, the switching sequence undergoes a conformational change, altering the structure of the downstream expression platform and regulating the expression of the associated gene [129, 130]. tRNA decodes the mRNA sequence and transfers the appropriate amino acid to the growing amino acid chain. The ribosome, the protein-synthesis factory, is composed of rRNA and its bound proteins. It is believed that the ribosome is essentially a ribozyme (ribonucleic acid enzyme), where the specific catalytic activity is provided solely by RNA molecules [131–133].

The reservoir of functional nucleic acids has expanded significantly since the advent of in vitro selection, or systematic evolution of ligands by exponential enrichment (SELEX) in 1990 ^[134–136]. This method has proven that DNA molecules can have functions and led to the discovery of DNA aptamers and DNAzymes ^[137, 138]. Aptamers against a broad spectrum of targets, ranging from small metal ions ^[139] to micrometre-sized cells, have been identified by in vitro selection, together with a number of catalytic nucleic acids for different types of reactions ^[140, 141].

1.4.1 Hierarchical folding of functional nucleic acids

In 1948, Pauling ^[142] proposed that the specific binding and catalysis observed in antigen-antibody and enzyme-substrate reactions result from structural complementarity, which maximises interactions between the folded polypeptide chain and the ligand or substrate. Similarly, as nucleic acid counterparts to protein antibodies and enzymes, the various functions and specific activities of aptamers and catalytic nucleic acids are largely dependent on their ability to fold into complex 3D structures ^[141].

The organisation of nucleic acid is classified into three levels: primary, secondary, and tertiary ^[143]. The primary structure refers to the sequence of bases along the linear polynucleotide chain. The secondary structure refers to the two-dimensional configuration that emerges when the strand folds onto itself by base pairing and forms double-stranded regions joined by unpaired single-stranded region. As depicted in Figure 1.10a, complementary regions within a strand form double helices through hybridisation, with the connecting single-stranded regions appearing as hairpins, loops, bulges, internal loops, and three-way junctions ^[143].

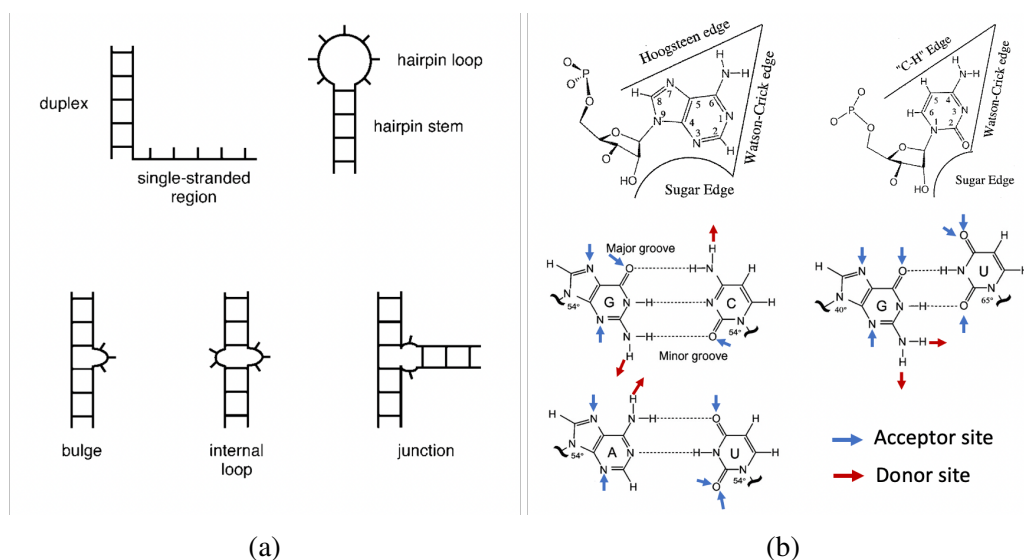


Figure 1.10: (a) Common 2D structural elements. Image and caption adapted from Ref. [143]. (b) Upper panel: edges available for Hydrogen bonding. Lower panel: possible interaction sites around the standard Watson–Crick pairs and the wobble G–U pair. Images and caption adapted from ref. [144, 145]

In addition to conventional base pairs, non-Watson–Crick (nWC) pairs are widespread in RNA structures, playing a crucial role in RNA folding and specific interactions. In fact, nWC pairs account for about 30% of all base pairs in structured RNAs [146]. As illustrated in Figure 1.10b, each base has three edges. One edge of one base can form at least two hydrogen bonds with any of the three edges of another base, either in the cis or trans orientation relative to the glycosidic bonds. This leads to twelve basic geometric families of base pairs, with sheared G–A, G–A imino, A–U reverse Hoogsteen, A–C wobble, and G–U wobble being the five most prevalent nWC pairs [144, 147]. The geometry of nWC pairs differs from that of WC pairs. For example, in regular A-form RNA, all base pairs are strongly isosteric and share a similar glycosidic angle ($\sim 54^\circ$) with respect to the phosphodiester backbone [145]. However, in a G–U pair, the U slightly shifts towards the deep groove, altering the local helical twist angles and disrupting the alignment of the structure [148]. G–U base pairs also have three hydrophilic acceptor sites in the deep major groove but do not have any hydrophilic donor site (Figure 1.10b) [149]. The nWC pairs are important in the formation of 3D structures [150], and the distinctive structural characteristics of the nWC pairs are potential targets for specific interactions [151].

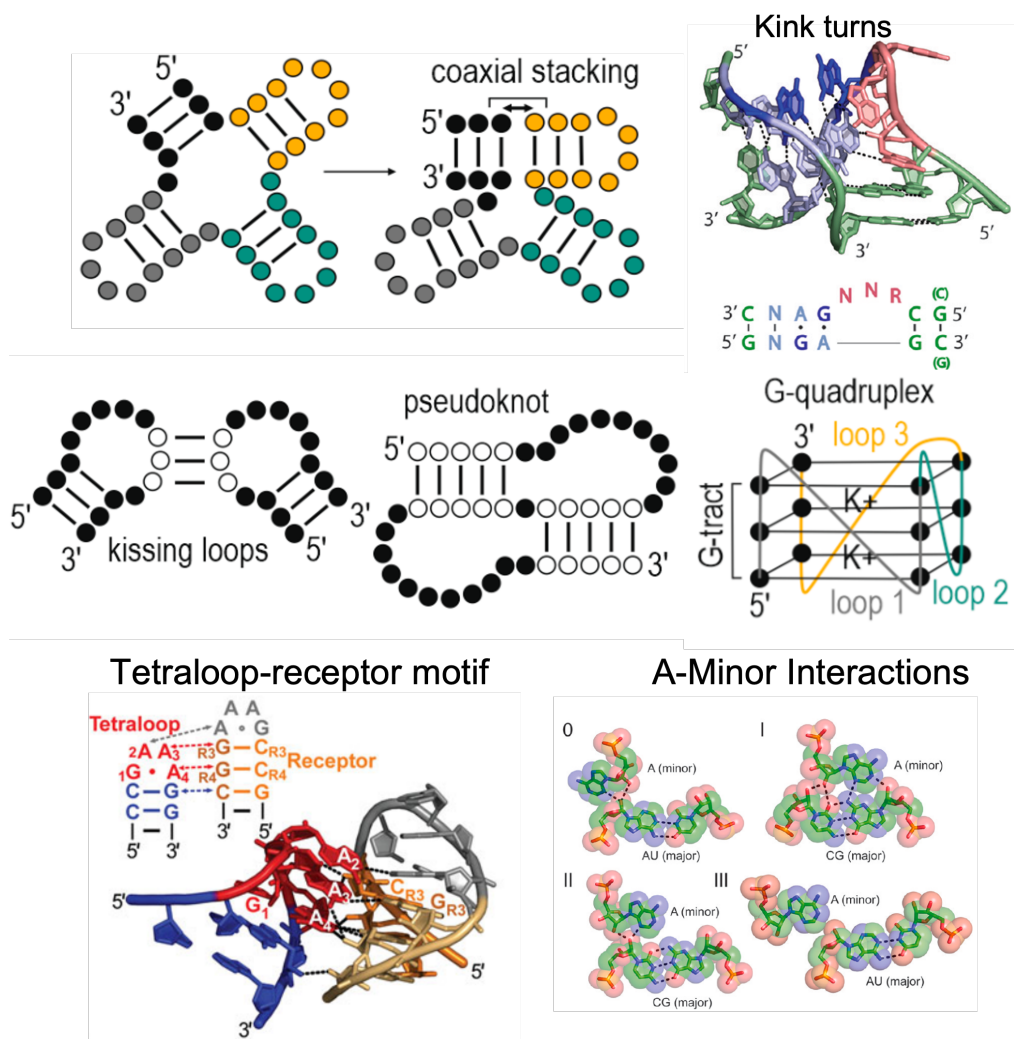


Figure 1.11: Common 3D structural elements. Image and caption adapted from ref.^[152].

Higher-order organisation in nucleic acids is achieved through interactions among secondary structure elements, as shown in Figure 1.11. Doudna and colleagues ^[143] divided the common tertiary structural motifs into three categories: interactions between two double-stranded regions, between two single-stranded regions, and between a double-stranded region and a single-stranded region. One key motif is the coaxial stacking of two duplexes, which form a single pseudocontinuous helix ^[153]. Another common conformation is parallel packing, typically induced by the tetraloop-receptor motif. This involves a terminal tetraloop that interacts with a distant helix, often featuring GNRA sequences (N is any nucleotide, R a purine), and receptors usually are minor grooves or internal loops within helices ^[154]. The kink-turn (k-turn) motif introduces a sharp kink into a helix and is a common structural element. A canonical k-turn has a 3-nt purine-rich bulge on its 3' side flanked by the A-G and G-A base pairs ^[155]. Both the k-turn and the tetraloop-

receptor motifs involve minor groove triples, where an unpaired nucleotide and a base pair in a helix form hydrogen bonds in the minor groove ^[153]. The unpaired nucleotide is usually adenine (A-minor interactions), with G-C being the preferred base pair ^[156]. Various A-minor motifs, distinguished by different hydrogen bonding schemes, are shown in Figure 1.11 ^[157]. The loop-loop "kissing" complex is generated by the hybridising the loops of two hairpins. The pseudoknot motif results from long-distance base-pairing between a hairpin loop and an extended single-stranded region ^[158]. In addition, guanine-rich sequences can form four-stranded 3D structures stabilised by potassium ions, called G-quadruplexes ^[159].

Metal ions play a crucial role in the folding of nucleic acids. They facilitate and stabilise complex structural arrangements by shielding electrostatic repulsion between phosphate groups on the backbone ^[160, 161]. Nucleic acids interact with metal cations in two modes: diffusive binding and site-specific binding ^[162]. In aqueous solutions, most metal ions bind diffusively. They remain hydrated and accumulate near the negatively charged backbone due to nonspecific electrostatic attraction ^[163]. In the site-specific binding mode, metal ions are dehydrated or partially dehydrated and directly interact with the nucleic acids, leading to stronger interactions and increased compactness of the folded structures. This specific interaction is selective and depends on the ions' charge density, volume exclusion, and hydration status ^[164]. The folding process is rapid. The accumulation of metal ions, which occurs in microseconds, reduces electrostatic barriers to folding and drives the formation of tertiary structures in milliseconds. Site-specific binding triggers further structural rearrangement, resulting in a stable tertiary structure ^[165]. Divalent cations are much more effective in site-specific binding-induced structural rearrangement than monovalent cations, since they have an entropic advantage in stabilising compact structures ^[166].

1.4.2 Aptamers

Thermodynamics and kinetics of aptamer–ligand binding

The interactions between an aptamer (A) and its ligand (L) can be simply described as follows:



where AL represents the aptamer–ligand complex, k_{on} is the association rate constant that describes the rate at which the aptamer and the ligand bind to form the complex, and k_{off} (often written as k_d) is the dissociation rate constant that describes the rate at which the bound complex dissociates into its free components.

The affinity between an aptamer and its target can be characterised by the equilibrium dissociation constant K_d

$$K_d = \frac{[A][L]}{[AL]} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (1.11)$$

where $[A]$, $[L]$, and $[AL]$ represent the equilibrium concentrations of the aptamer, ligand, and associated complex (in units M), respectively. K_d is the concentration of the ligand at which half of the binding sites on the aptamer are occupied. A lower K_d indicates a higher affinity between the aptamer and the ligand, meaning that there is a higher chance that the two molecules are bound at any given moment and vice versa.

Within an aptamer–ligand–solvent system, complex interactions and heat exchanges occur among the components, including aptamers, ligands, water, and buffer ions ^[167]. Like any spontaneous process, aptamer–ligand binding happens only when it results in a negative change in the Gibbs free energy (ΔG) of the system. The relationship between ΔG and K_d is expressed as follows:

$$\Delta G = RT \ln K_d \quad (1.12)$$

where R is the gas constant, and T is the temperature in Kelvin.

ΔG can be decomposed into its enthalpic and entropic contributions using the following equation:

$$\Delta G = \Delta H - T\Delta S \quad (1.13)$$

where ΔH is the change in enthalpy and ΔS is the change in entropy of the system upon ligand binding.

Binding models

The investigation of protein–ligand binding mechanisms began long before the discovery of aptamers. The lock-and-key model proposed by Emil Fischer in 1894 is one of the first models for studying protein–ligand interactions ^[168]. This model (Figure 1.12a) assumes that both receptors and ligands are rigid structures, where only a ligand of the correct size and shape (the key) can fit into the active site of the receptor (the lock). Experimental and theoretical studies on geometrically complementary colloidal particles have shown that lock-and-key binding is predominantly entropically driven ^[169–173]. The effect of exclusion volume or the maximisation of solvent entropy is concluded to be the governing influence on the forces that act between the colloidal particles ^[167]. This process is highly dependent on the size, shape, and surface roughness of the particles but is independent of the composition or surface chemistry. The complete displacement of the constrained solvent molecules during the binding process significantly increases solvent entropy, which makes entropic contributions dominant in the lock-and-key binding mechanism.

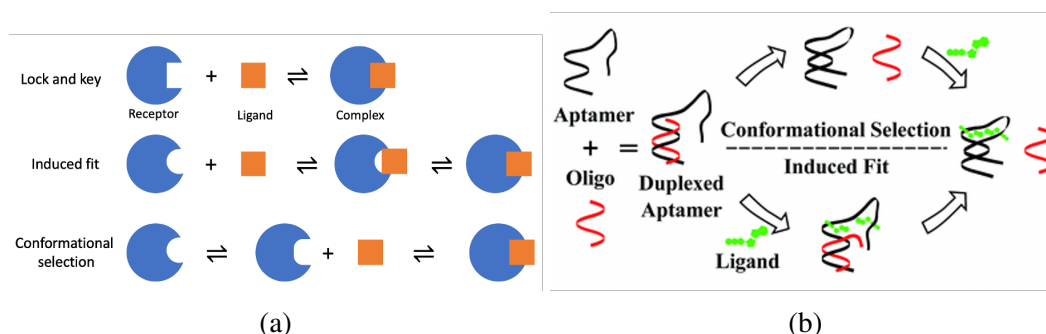


Figure 1.12: Schematic illustrations of ligand binding models. (a) The three proposed protein-ligand binding pathways. (b) Ligand binding through conformational selection or induced fit in duplexed aptamers. Images adapted from ref. ^[174].

However, it cannot explain all scenarios, particularly in which proteins bind to ligands without an exact initial shape match. To overcome this, Daniel Koshland proposed the induced fit theory (Figure 1.12a) in 1958 ^[175]. This model takes into account the flexibility of receptors, suggesting that the binding sites change their conformation upon interaction with the ligand. In the induced fit model, initial contacts between the ligand and the protein, driven by collisions, must be strong enough and have a sufficient lifetime to drive the conformational changes needed to form a fully bound complex ^[176]. There is

an entropy gain from the release of solvent molecules, but the large negative change in the enthalpy of newly formed noncovalent interactions dominates, making induced fit binding an enthalpy-dominated process ^[177]. The conformational selection model (Figure 1.12a) proposed in 1965 integrates features of both the lock-and-key and induced fit models ^[178]. In this model, the receptor protein adopts various interconvertible conformations in its unbound state, and the ligand selectively binds to the pre-existing conformation that best matches the ligand, which is followed by conformational adjustments in the protein to form strong noncovalent interactions. The binding is driven by both the gain in the solvent entropy and the decrease in the enthalpy of the system, and they dominate sequentially ^[167]. Regarding which is the governing mechanism, studies have shown that it depends on the concentration of ligands and the timescale of receptor dynamics ^[179–181].

Due to the limited nucleoside diversity and charged backbone of nucleic acids, fitting ligands into the binding sites of aptamers generally displays suboptimal shape complementarity. Therefore, aptamers commonly undergo ligand-induced adaptive folding around the ligand ^[182]. High binding affinities (nM to μ M) and specificities can be achieved through hydrogen bonding, van der Waals forces, stacking interactions, and electrostatic interactions between the ligand and the aptamer ^[182, 183]. The adaptive binding nature of aptamers has been used for ligand-responsive biosensing, often in the form of a duplexed aptamer that is partially hybridised to a complementary oligo ^[184]. The binding of aptamer and ligand in these biosensors operates through two mechanisms (Figure 1.12b): conformational selection ^[185] and induced fit ^[174]. In the former mechanism, the duplex dissociates before the aptamer binds to the ligand, and the ligand-induced conformational change is localised to the binding pocket. Ligand binding disrupts the equilibrium between the duplexed and dissociated aptamer, producing a concentration-dependent signal for ligand sensing. In the latter mechanism, the duplexed aptamer can bind to a ligand while in the duplex state. The ligand induces a conformational change in the global architecture, releasing the complementary strand.

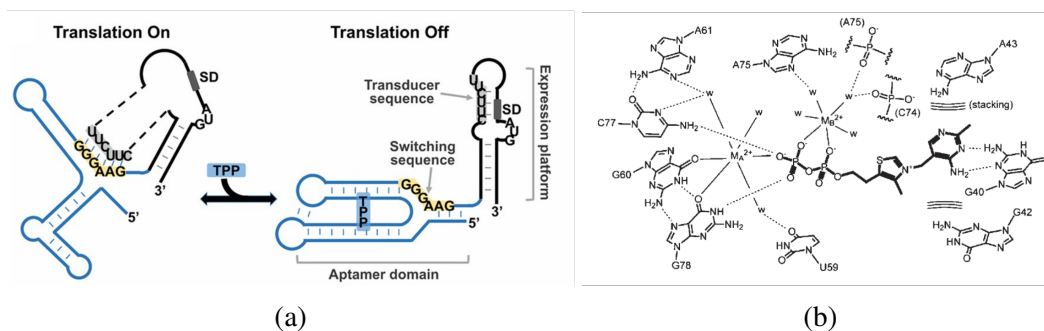


Figure 1.13: (a) The switching mechanism of a TPP-dependent riboswitch. The riboswitch comprises an aptamer domain (blue), a switching sequence (highlighted in yellow), and an expression platform (black). In the absence of TPP, the transducer sequence (grey) interacts with the switching sequence within the aptamer domain to keep the expression platform (SD sequence) exposed. When TPP binds, it induces a conformational change in the aptamer domain, disrupting strand hybridisation. As a result, the transducer sequence detaches from the switching sequence and turns off the translation through hybridisation with the SD sequence. Image and caption adapted from ref. [186]. (b) TPP recognition by the TPP riboswitch. TPP is recognised through base stacking, base pairing, and metal ion-mediated binding. Image and caption adapted from ref. [187].

As highlighted earlier, riboswitches serve as prime examples in demonstrating the concept of allosteric aptamers. Thiamine pyrophosphate (TPP) riboswitches (Figure 1.13a) are among the most studied natural riboswitch classes. They are widespread and are found in all three domains of life [188]. The conformational change induced by TPP binding in the aptamer region impacts the downstream structure, regulating gene expression [189]. As depicted in Figure 1.13b, the TPP is recognised by the aptamer region through base stacking, hydrogen bonding, and metal ion-mediated binding. The pyrimidine ring of TPP is recognised through stacking and a pseudo-base pair interaction with the sugar edge of G40. The central thiazole ring of TPP interacts with the RNA via van der Waals forces. The pyrophosphate is indirectly recognised by forming a chelate complex with two partially hydrated divalent cations [190].

However, not all aptamers undergo significant conformational changes upon ligand binding. To investigate factors and characteristics that allow an aptamer to act as an effective regulatory element within a riboswitch, Kaiser and colleagues conducted a comparative analysis of the ligand-binding kinetics of three different ciprofloxacin-binding aptamers (A, prePS, and RS), only two (preRS and RS) of which exhibit significant structural switching [191]. Despite the differences in both the sequence and the secondary structure, the three aptamers have very similar K_d values [192]. Time-resolved stopped-flow

fluorescence experiments showed that all three aptamers underwent a two-step binding process. In the first step, driven by nonspecific electrostatic interactions, the aptamer and the ligand form an intermediate bound complex. In the second step, the aptamer undergoes a conformational change to create a tightly binding pocket. Compared to aptamer A, the other two aptamers exhibited lower k_1 (association rate in step 1), higher k_{-1} (dissociation rate in step 1), lower k_2 (forward rate in step 2) and significantly lower k_{-2} (reverse rate in step 2), resulting in a virtually irreversible binding step. Therefore, the authors suggested that a strongly preferred structural adaptation of the RNA to the ligand during the final binding step with low dissociation rates is crucial for the allostery of an aptamer. Nevertheless, using aptamer-complementary element scanning, Munzar and coworkers generated comprehensive binding kinetic landscapes of duplexed aptamers, revealing a widespread induced-fit mechanism [193].

1.4.3 Catalytic nucleic acids

Nucleic acids also leverage their intricate structures for chemical catalysis. According to the transition state theory, the observed rate constant k_{obs} agrees with the following equation 1.14:

$$k_{\text{obs}} = \left(\frac{k_B T}{h} \right) \kappa e^{-\frac{\Delta G^\ddagger}{RT}} \quad (1.14)$$

where the prefactor $\left(\frac{k_B T}{h} \right) \kappa$ includes k_B (Boltzmann constant), T (temperature), h (Planck's constant), and κ (transmission coefficient). The exponential term $e^{-\frac{\Delta G^\ddagger}{RT}}$ represents the fraction of molecules that have enough energy to overcome the activation barrier $\Delta G^\ddagger$, where R is the gas constant.

Catalytic nucleic acids reduce the activation barrier and enhance reaction rates through various strategies, such as substrate binding, which facilitates the proper orientation and proximity of the reactant, thereby stabilising the transition state. Metal ions, whether diffused or specifically bound to the negatively charged polyelectrolyte, are essential for both structure formation and catalysis. Additionally, the aromatic heterocyclic bases exhibit varying levels of basicity and acidity, allowing them to participate directly in catalytic reactions [194].

With the exception of rRNA, which catalyses peptidyl transfer, all known natural ribozymes catalyse cleavage or ligation of nucleic acids. Despite its chemical instability, RNA can exhibit a half-life exceeding one year under near-physiological conditions, with a rate constant of about 10^{-7} min^{-1} [195]. Ribozymes can accelerate reactions by at least a million-fold, reaching rates of about 1 min^{-1} . They typically use general acid-base catalysis (Figure 1.14a) or metal ion catalysis (Figure 1.14b), or a combination of both [196].

Nucleolytic ribozymes, a family of nine small ribozymes (70-150 nt), catalyse site-specific cleavage through the nucleophilic attack of the 2'-oxyanion on the neighbouring 3'-phosphorus centre, forming a cyclic 2',3'-phosphate and a 5'-hydroxyl group [197]. Some can also catalyse reverse ligation if the cleaved products are aligned [194]. Of the eight nucleolytic ribozymes with known structures, six use guanine as a general base [197]. The N1 of guanine deprotonates the 2'-hydroxyl, generating a 2'-oxyanion, a superior nucleophile compared to hydroxyl, with an enhancement factor up to 10^{10} [73]. Guanine's pKa (approximately 9.2 at N1) can shift towards neutrality through electrostatic stabilisation and hydrogen bonding in the microenvironment of folded nucleic acid [198, 199]. For the other two nucleolytic ribozymes, the general base is a deprotonated inner-sphere water molecule associated with a hydrated metal ion [200, 201]. Compared with the general base, the general acid varies more. In hairpin and VS ribozymes, adenine N1 protonates O5', improving its leaving group ability [202, 203]. In twister ribozyme, adenine's N3 donates the proton [204]. In HDV and TS ribozymes, a cytosine nucleobase likely acts as the general acid [205, 206]. The pKa values of adenine and cytosine can shift towards neutrality in the electronegative environment of folded structures [207]. The 2'-OH group of a guanine nucleoside, possibly activated by a metal ion, functions as a general acid in the hammerhead ribozyme, while the pistol ribozyme uses a hydrated metal ion [208].

RNase P ribozymes and self-splicing introns represent the other two groups of natural ribozymes that catalyse P-O bond chemistry in RNA backbones. Unlike nucleolytic ribozymes, which use adjacent 2' -OH and produce a cyclic 2',3'-phosphate, these ribozymes use an external nucleophile and direct metal coordination, forming a 3'-OH terminus [209]. RNase P ribozymes remove the 5' leader sequence of tRNA precursors through phosphodiester bond hydrolysis via a two-metal ion mechanism [210]. Here, a

metal-bound hydroxide acts as the nucleophile, while another metal ion coordinates a water molecule to protonate the 3'-oxyanion leaving group ^[211]. Self-splicing introns remove noncoding regions from precursor RNAs through a two-step transesterification mechanism ^[209]. The 5'-splice site is attacked by the 3'-OH of an external guanosine for the Group I introns, while the 2'-OH of an adenosine within the intron is the nucleophile for Group II introns. Divalent metal ions coordinate the phosphoryl oxygen, making the phosphorus centre more susceptible to nucleophilic attack and stabilising the negative charge of the leaving group ^[212, 213]. Atomic mutagenesis suggests that a third metal ion in Group I introns may activate the nucleophile by interacting with the 3'-OH of the attacking guanosine ^[214]. Figure 1.14b depicts the proposed mechanism of three metal ions. The second exon ligation and intron release step is completed by the 3'-OH of the 5'-exon attacking the 3'-splice site.

The rRNA facilitates the nucleophilic attack of an amine in a carbonyl centre, and the catalysis does not depend on nucleobases or metal ions ^[215]. Instead, rRNA contributes to the 10⁷-fold acceleration of peptide bond formation by providing correct geometry and a desolvated environment, facilitating substrate positioning and alignment, and reducing activation entropy ^[216, 217]. The A76 2' -OH of the tRNA plays an important role in a proton shuttle mechanism (Figure 1.14c), where a proton shuttles between the attacking amine and the carbonyl centre through the 2' -O of the terminal A76 ^[218].

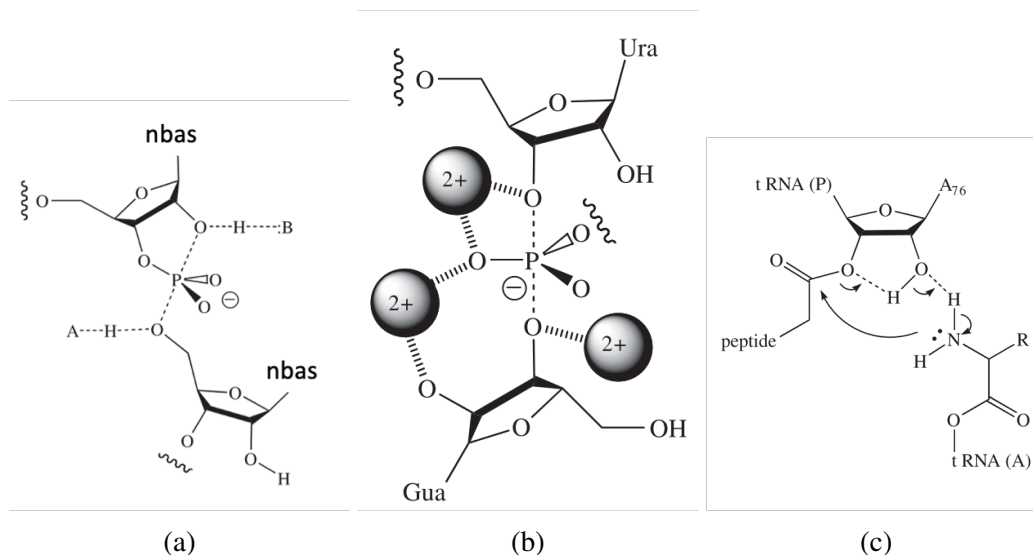


Figure 1.14: Proposed mechanisms of chemical catalysis in natural ribozymes. (a) General acid–base catalysis in self-cleavage reactions, with B serving as a general base and A as a general acid. (b) RNA metalloenzymes. The transition state of the group I intron stabilised by three metal ions. (c) A possible mechanism of peptidyl transfer. The A-site tRNA’s terminal amino group, aided by the 2’-oxyanion group, attacks the carbonyl centre of the P-site tRNA, resulting in the formation of a new peptide bond. Images and caption adapted from ref. [215].

In addition to P-O bond (phosphoryl transfer reactions) and P-N bond (peptide bond formation) chemistries, in vitro selection has led to the development of artificial catalytic nucleic acids capable of catalysing a wider array of chemical reactions in a laboratory setting [219–223]. Examples of other chemistries catalysed by in vitro-selected ribozymes are listed in Figure 1.15.

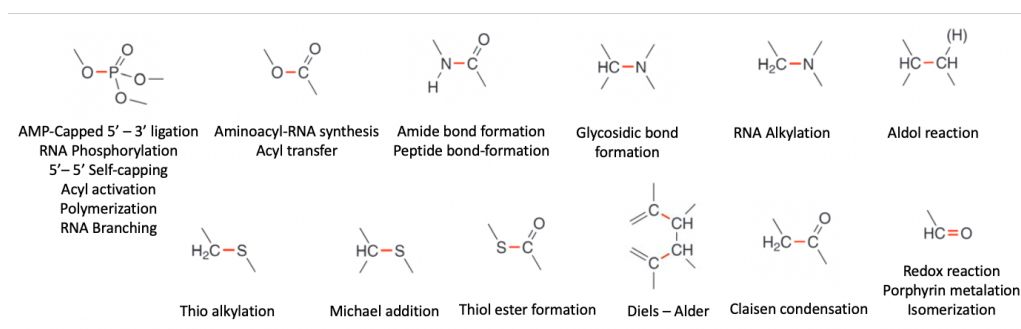


Figure 1.15: Examples of chemistries catalysed by in vitro-selected ribozymes. The relevant chemical bonds are in red. Image and caption adapted from ref. [224]

Catalytic capabilities of nucleic acids

Figure 1.16 is a simple kinetic model for a typical enzyme-catalysed reaction. The reaction started from the formation of the enzyme-substrate complex (ES), followed by the

rapid transformation to the transition state (E'S') and the formation of the enzyme-product complex (EP), and finally the release of the product (P) and regeneration of the enzyme (E).

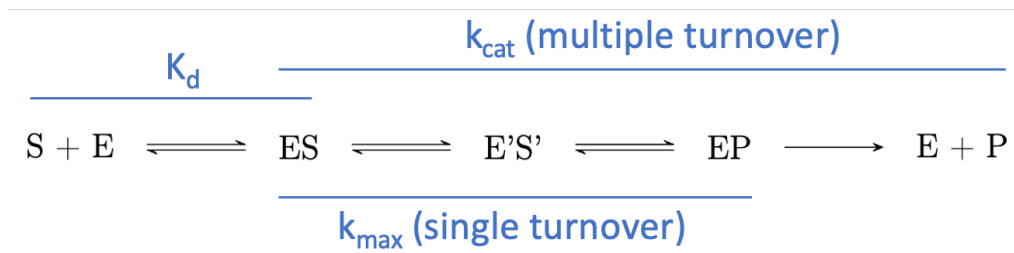


Figure 1.16: A simple kinetic model for a typical enzyme-catalysed reaction. Image and caption adapted from ref. [225].

For kinetic studies of ribozymes and DNAzymes, rate constants are usually obtained from assays under single- and multiple-turnover conditions [225]. Single-turnover kinetics is also referred to as pre-steady-state kinetics. The experiments are set with an excess of the enzyme (E) relative to the substrate (S), and the observed rate constant k_{obs} is measured. Under optimal conditions, k_{max} (the maximum k_{obs}) reflects the chemical conversion of the catalysis ((ES) $\rightleftharpoons$ EP). The rate constant for a specific reaction is determined by plotting the fraction product formed versus time, using Equation 1.15.

$$\frac{[P]_t}{[P]_t + [S]_t} = A(1 - e^{-kt}) \quad (1.15)$$

where A is the maximum fraction, k is the rate constant, and $[P]_t$ and $[S]_t$ are the concentrations of each at time t. k_{max} is obtained by plotting k_{obs} versus enzyme concentration or other conditions, such as pH and metal ion concentration.

Multiple-turnover assays are performed under saturating substrate conditions ($[S] \gg [E]$), which are often called steady-state conditions. The rate of the reaction is determined by the Michaelis–Menten Equation 1.16:

$$v = \frac{d[P]}{dt} = \frac{V_{\text{max}}[S]}{K_M + [S]} = \frac{k_{\text{cat}}[E][S]}{K_M + [S]} \quad (1.16)$$

where v is the initial rate measured as the slope of the linear region of the product versus time plot; V_{max} represents the maximum rate under a specific enzyme concentration; K_{cat} , also called turnover number, reports on all steps including substrate release along the

pathway after $[ES]$; K_M is the Michaelis–Menten constant and it is equal to the $[S]$ required to give a $1/2V_{\max}$. K_M reflects the association of the enzyme with the substrate, but it is different from K_d .

To investigate the catalytic power of nucleic acids, Breaker and coworkers ^[196] studied the kinetics of 14 in vitro selected artificial ribozymes and DNAzymes that catalyse site-specific cleavage of RNA. All the assays were conducted under single-turnover conditions, ignoring potential rate-limiting steps of substrate binding and product dissociation. The authors proposed four possible strategies (named α , β , γ , and θ) that could be used individually or together to accelerate the internal phosphorus transfer reaction. They found half of the ribozymes and DNAzymes fell within the $\alpha\gamma$ speed limit of $\sim 2 \text{ min}^{-1}$, while it is proposed that ribonuclease A uses all four strategies and has a k_{obs} of $\sim 8000 \text{ min}^{-1}$. The authors concluded that, given the ability of some ribozymes and DNAzymes to exceed this limit, nucleic acids could potentially exhibit catalytic efficiency comparable to their protein counterparts, provided that similar strategies are employed and this can be biased during in vitro evolution. Furthermore, better catalytic power can be achieved by engineering the available sequences and optimising reaction conditions to increase the fraction of active species. For example, a VS ribozyme with a k_{obs} of approximately 600 min^{-1} has been identified from a pool of mutants ^[226]; a k_{obs} of 870 min^{-1} of a natural hammerhead ribozyme was reported under optimal pH and Mg^{2+} concentration ^[227], and a tertiary-stabilized hammerhead ribozyme was measured to have a cleavage rate of 3000 min^{-1} in 1 mM Mn^{2+} ^[228]. Therefore, nucleic acid enzymes are not intrinsically less capable than protein enzymes in terms of chemical rate enhancement.

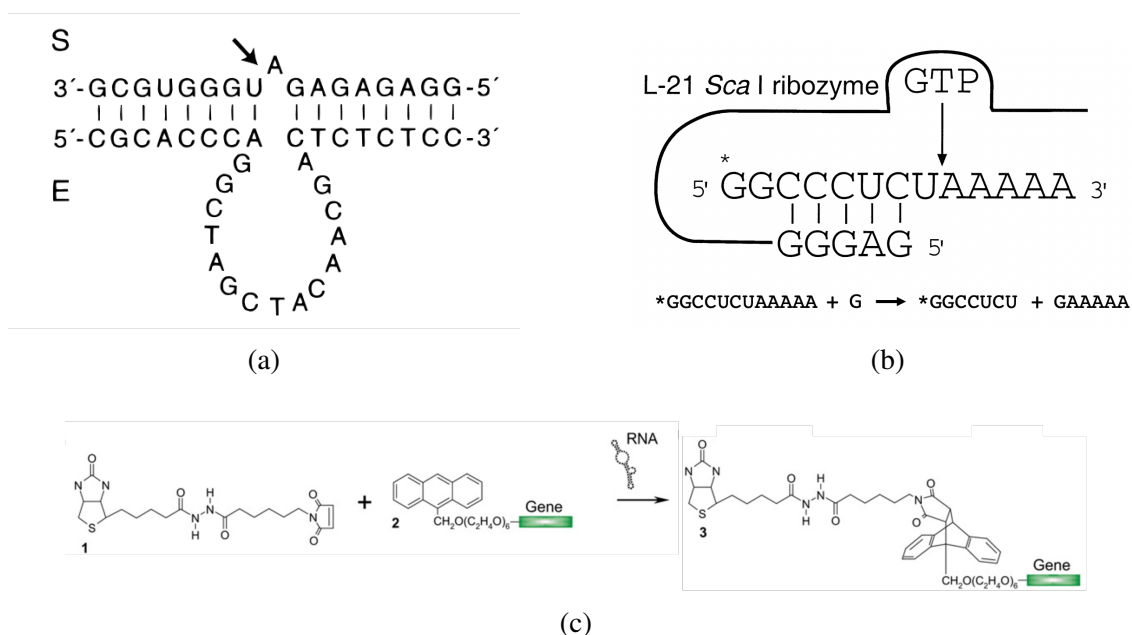


Figure 1.17: (a) The 10-23 DNA enzyme (E) and its RNA substrate (S). Image and caption adapted from ref. [229]. (b) Pathway for an RNA cleavage catalysed by the L-21 Sca I ribozyme. Image and caption adapted from ref. [230]. (c) Ribozyme catalyzed Diels-Alder cycloaddition of biotin-maleimide (1) and 9-anthracenylmethyl hexaethylene glycol (AHEG) (2) covalently coupled to the gene. Image and caption are adapted from ref. [231].

The catalytic efficiency is determined by $k_{\text{cat}}/K_{\text{M}}$. It represents how effectively the enzyme captures the substrate for catalytic turnover. It also reflects the step to the highest transition state along the reaction pathway in the reaction free energy diagram [232]. When the ratio $k_{\text{cat}}/K_{\text{M}}$ reaches the diffusion-controlled limit of $10^8\text{--}10^9\text{ M}^{-1}\text{ s}^{-1}$, indicating that every collision between the enzyme and the substrate leads to a catalysed reaction and the chemical step is faster than diffusion, the enzyme is called a "perfect enzyme" [233]. For oligonucleotide substrates, most collisions between the substrate and the ribozyme or DNAzyme during diffusion are ineffective. This is because the formation of the substrate-enzyme complex requires the initiation of two or three base pairs so that the association occurs faster than the dissociation [234]. Therefore, the $k_{\text{cat}}/K_{\text{M}}$ value is determined mostly by the kinetics of the association between the enzyme and the nucleic acid substrate ($10^6\text{--}10^7\text{ M}^{-1}\text{ s}^{-1}$) [234, 235]. The 10-23 DNAzyme is an RNA-cleaving enzyme with a 15-nt catalytic domain, flanked by two 8-nt substrate-recognition domains (as shown in Figure 1.17a) [236]. The catalytic efficiency of the 10-23 DNAzyme is approximately the same (around $10^9\text{ M}^{-1}\text{ min}^{-1}$) under both multiple-turnover and single-turnover conditions, indicating that substrate binding (hybridisation) rather than product release is the

rate-limiting step for turnover. The L-21 scal ribozyme is an engineered version of the group I intron of *Tetrahymena thermophila* [237]. It binds to its RNA substrate at rates of 10^6 – $10^7 \text{ M}^{-1} \text{ s}^{-1}$ and recruits an external GTP to act as a nucleophile, as illustrated in Figure 1.17b. Studies of its kinetics reveal that the binding of GTP and the oligonucleotide substrate are mutually independent. In the presence of 10 mM Mg^{2+} , the value $k_{\text{cat}}/K_{\text{M}}$ is $9 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$. In the absence of GTP, the ribozyme still exhibits activity if the 5' end is a guanine nucleotide. Although the chemical catalysis rate is significantly slower under these conditions, the $k_{\text{cat}}/K_{\text{M}}$ value was only slightly reduced, measured at $6 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$. This finding suggests that hybridisation with the substrate oligonucleotide, rather than the binding of GTP, is the rate-limiting step [230].

A true catalyst must be able to act in trans and catalyse successive reactions without itself being changed. RNase P is the only natural ribozyme identified so far that turns over in vivo [238]. Nucleic acid enzymes that act on themselves can usually be converted to work in trans and are capable of multiple turnovers, typically by separating the catalytic and substrate domains [239]. For nucleic acid substrates, stable Watson–Crick base pairing between the enzyme and the substrate makes catalytic turnover inherently difficult for ribozymes and DNAzymes [240]. Nevertheless, the 10–23 DNAzyme can reach k_{cat} of 3.4 min^{-1} , depending on the substrate sequence and the buffer conditions [236]. This DNAzyme can be tailored to target and cleave nearly any RNA molecule containing a purine–pyrimidine cleavage site and the product-release step is not rate-limiting until the substrate-recognition domain exceeds 10 nt with a high GC content [229].

For small-molecule substrates, trans-acting ribozymes and DNAzymes can be evolved directly through in vitro selection under multi-turnover conditions [241]. Sometimes, enzymes selected under single-turnover conditions are also capable of turnover, but this occurs fortuitously [242]. Agresti and coworkers developed ribozymes that catalyse Diels–Alder cycloadditions in trans with multiple turnovers (Figure 1.17c) [231]. The selection process began with five rounds of low-barrier selection for species catalysing intramolecular single-turnover reaction, which enriched the library to a manageable size for subsequent rounds of in vitro compartmentalization selection for multiple-turnover species. Although selected under single-turnover conditions, the library showed detectable trans-

acting activity after the first 5 rounds of selection. The most active ribozyme from the final round exhibited multiple-turnover kinetics with k_{cat} of $\sim 6 \text{ min}^{-1}$. While the value is not particularly impressive, they are comparable to those of their natural protein counterparts. The rates are limited by the slow product release, which is proposed to be due to its similarity to the transition state of the substrate complex.

1.4.4 Applications of functional nucleic acids in dynamic oligonucleotide-based systems

Aptamers have been used as reconfigurable elements in dynamic nucleic acid nanostructures, commonly through overhang hybridisation, sequence extension, or modular sequence integration [243, 244]. Figure 1.18a illustrates an RNA-based Förster resonance energy transfer (FRET) system [245]. This system consists of two fluorescent RNA aptamers and an allosteric aptamer specific to S-adenosylmethionine (SAM) between them on an RNA origami scaffold. Upon binding of SAM, a new fluorophore is detected because it brings the two fluorescent aptamers into close proximity, which enables FRET. In addition to fluorescence, ligand binding can also induce other signal changes, such as plasmonic circular dichroism (CD) signals in the biosensor depicted in Figure 1.18c. This device consists of two 10-helix bundles, each carrying a gold nanorod and decorated with two different thrombin-binding aptamers. Thrombin binding alters the chiral state of the device, leading to changes in the CD spectra. Aptamers can be used to create smart DNA nanostructures that respond to chemical cues. An example is the multilayered DNA nanocage shown in Figure 1.18b [246]. These multilayered DNA tetrahedra self-assemble through the hybridisation of ATP aptamers and complementary strands. Upon ATP binding, the structure-switching of the aptamer induces dehybridisation and disassembly of the nanocage. Muscat and coworkers [247] leveraged the structure-switching properties of aptamers to direct the route of a molecular cargo. Figure 1.18d illustrates a junction on the DNA track where the cargo chooses directions. This junction is regulated by reciprocally acting aptamers (red) that bind adenosine, alternately obstructing and exposing the toehold binding region (black) on different branches of the track.

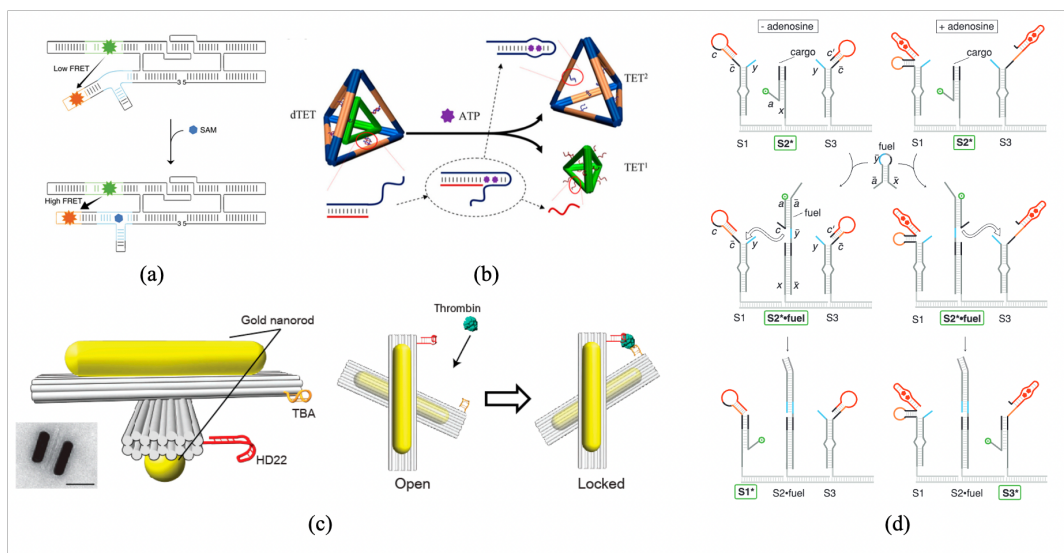


Figure 1.18: Applications of aptamers in dynamic nucleic acid structures. (a) SAM-induced FRET between fluorescent aptamers on an RNA origami scaffold. Image and caption adapted from ref. [245]. (b) ATP-activated layer separation of a multilayered DNA tetrahedrons. Image and caption adapted from ref. [246]. (c) A thrombin sensor with the plasmonic metamolecule. Image and caption adapted from ref. [248]. (d) The sensor-controlled junction. Image and caption adapted from ref. [247].

The off-rate constants of aptamer-bound small molecules are typically in the range of 10^{-2} – 10^{-5} s⁻¹, corresponding to dwell times from minutes to hours [249]. Therefore, aptamers in DNA nanotechnology function more like single-use "molecular fuses" [250]. Various DNAzyme- or ribozyme-aptamer conjugates have been reported, in which the aptamer domain regulates the activity of the catalytic domain. These artificial riboswitches are called aptazymes [251, 252]. Figure 1.19a depicts an aptazyme system which consists of an aptamer domain (red), a peroxidase-mimicking DNAzyme (blue), and a blocking strand (green) [253]. In the absence of ligands, the DNAzyme is partially blocked and inactive. The binding of the ligand (AMP or lysozyme) to its aptamer leads to the detachment of the aptamer region from the blocking strand, resulting in the thermal melting of the duplexed DNAzyme. The free DNAzyme quickly forms the active G-quadruplex structure with a hemin-docking site, which serves as a cofactor in catalysing the oxidation of ABTS⁻ to the coloured ABTS⁻ by H₂O₂.

The applications of catalytic nucleic acids in dynamic oligonucleotide-based systems have mostly focused on oligonucleotide-cleaving DNAzymes and ribozymes. They have been used to power the mechanic motions of DNA devices (Figure 1.19b), such as the open-close motors and burnt-bridges-type walkers discussed above [254]. They can be

used for amplified biosensing under non-PCR conditions ^[255–258]. Figure 1.19c illustrates a designed target-assisted self-cleavage (TASC) probe, consisting of a target-binding site (red), a DNAzyme (blue), and a substrate oligo with a linker (black) ^[259]. The probe undergoes TASC when activated via hybridisation with a target strand, separating a FRET pair positioned on either side of the cleavage site. Unlike molecular beacons that produce a stoichiometric signal from the target, this system employs the target strand as a catalyst for the probe's cleavage, allowing it to be reused after the cleaved products dissociate. They can also be used to construct logic gates for molecular computing ^[260–262]. Figure 1.19d shows an Or and an And gates using an RNA-cleaving DNAzyme and TMSD ^[263].

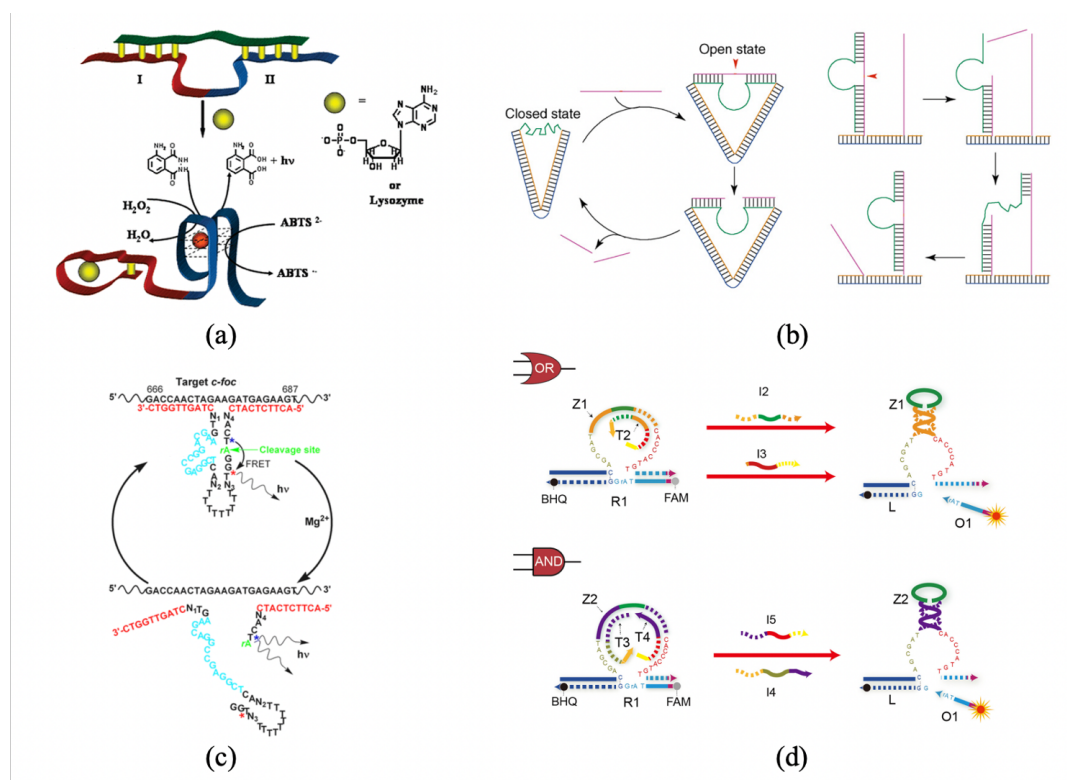


Figure 1.19: Applications of catalytic nucleic acids. (a) An AMP/lysozyme-dependent aptazyme that produces colorimetric or chemiluminescence readout signals. Image and caption adapted from ref. ^[253] (b) DNAzyme (green)-based open-close motor and walker. Image and caption adapted from ref. ^[254]. (c) DNAzyme for amplified target oligonucleotide sensing. Image and caption adapted from ref. ^[255]. (d) DNAzyme-based logic gates. Image and caption adapted from ref. ^[263].

The turnover rates of such systems are generally slow because they rely on the hybridisation and dehybridisation of oligonucleotide strands. For example, the target-assisted self-cleavage probe, which theoretically produces amplified signals through catalytic turnover,

exhibited only four turnovers within 90 minutes, even with a 10-fold excess of the probe [255]. For catalytic nucleic acids using small-molecule substrates, the predominant application has been as the catalytic component of biosensors, such as the aptzyme shown in Figure 1.19a [253, 264].

Most of the research on small-molecule ribozymes has focused on reactions of metabolic relevance, supporting the RNA world theory [224]. This widely accepted theory of the origin of life suggests that an RNA-based metabolism may have existed before the DNA-RNA-protein system seen in modern biology [265, 266].

1.5 Creating functional nucleic acids by selection and engineering

In 1990, three groups independently reported on the technique of in vitro selection to develop nucleic acids with novel functions. The Szostak lab coined the word "aptamer" (from Latin "aptus" and Greek "meros", which means "to fit" and "region", respectively) to refer to nucleic acids that specifically bind to targets [134]. They selected RNA aptamers against several organic dyes by in vitro selection. The Gold lab coined the word SELEX (Systematic Evolution of Ligands by Exponential Enrichment) and selected RNA aptamers against T4 DNA polymerase [135]. In the same year, the Joyce lab reported a ribozyme that cleaves ssDNA [136]. These three papers marked the beginning of the field of artificial functional nucleic acids. Today, in vitro selection or SELEX remains the most common method to produce new functional nucleic acids.

Figure 1.20 illustrates the general procedure for in vitro selection. The process begins with a pool of chemically synthesised DNA molecules, each containing a T7 promoter for transcription, two constant regions as primer binding sites for amplification, and a random region in the middle. The initial RNA library is usually biologically synthesised from the DNA library by in vitro transcription. Lorsch and Szostak performed experiments to estimate the likelihood of finding a functional RNA within a random pool. They found that around 1 in 10^{13} RNA could catalyse phosphoryl transfer reactions, and about 1 in 10^{10} RNA could bind specifically to a particular organic dye [267]. Typically, the starting lib-

rary has about 10^{15} different species. The molecules in the library are then partitioned according to whether they exhibit the desired property. The concentration of functionally active molecules increases by 100- to 1000-fold after each cycle of selection and partitioning due to nonspecific selection [268]. The selected species are then reverse transcribed and amplified using polymerase chain reaction (PCR) to yield a practical amount of material. These PCR products are then transcribed back to RNA for the next round of selection. This selection cycle is typically repeated 5-15 times until a pool predominantly containing species with the desired functions is generated. The selection process can be monitored either by assessing the bulk activity of each sub-library or by assessing enrichment through the complexity of each sub-library [269].

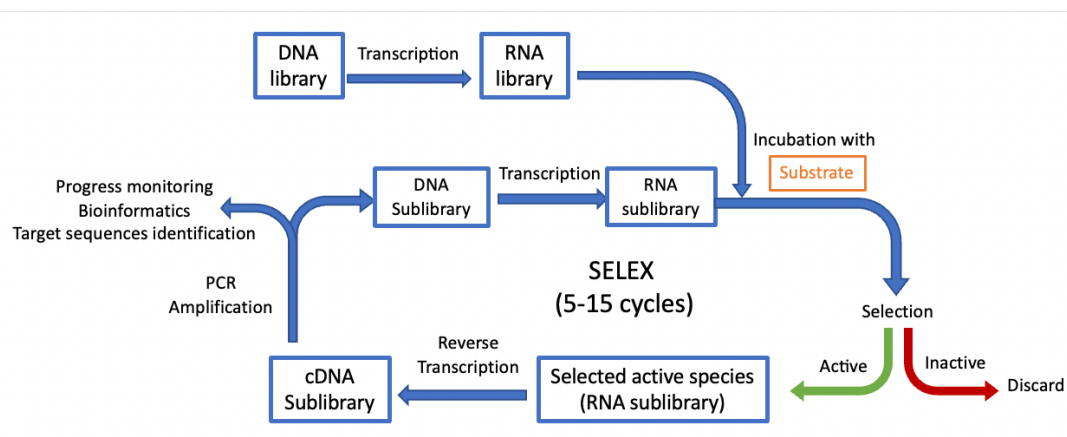


Figure 1.20: Illustration of a typical in vitro selection process for functional RNA molecules. For DNA-based selections, the transcription and reverse transcription steps are omitted, and an additional step for generating ssDNA is included after PCR.

1.5.1 In vitro selection of aptamers

Aptamers are typically selected by affinity chromatography. As the library flows through the affinity column with covalently immobilised ligands, species that bind to the ligand are retained in the column, while the flowthrough containing non-binding species is discarded. The bound sequences are then eluted using buffers with high concentrations of ligands. Figure 1.21a illustrates the selection process of an aptamer that recognises the α and γ phosphate groups of ATP [270]. Phosphate discrimination is achieved by incorporating a negative selection step. In this step, a column with immobilised AMP is used to remove species that bind to the adenosine moiety. Without a negative selection step,

all previously evolved ATP aptamers have been found to recognise the adenosine moiety. The authors suggested that this is because aptamers that recognise the adenosine moiety tend to have simpler secondary structures and are therefore more abundant in random sequence pools.

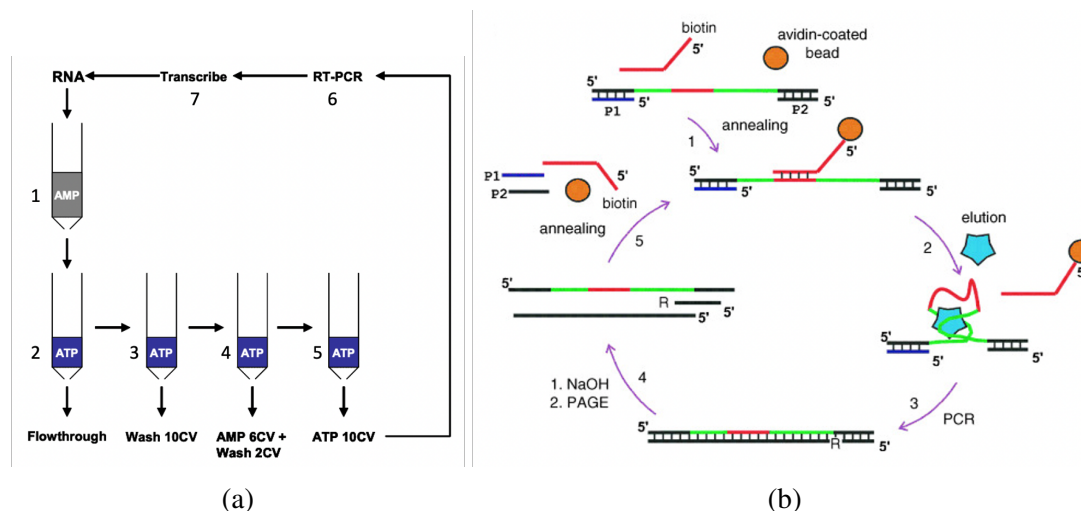


Figure 1.21: Illustrations of the SELEX procedure for aptamers. (a) Selection process of an aptamer that recognises the triphosphate moiety of ATP. Image adapted from ref. [270]. (b) Selection process of structure-switching aptamers. Image adapted from ref. [271].

Aptamers that undergo significant structural changes upon binding to ligands can be generated through "capture-SELEX", a method in which the ligands are typically not immobilised [271–273]. As illustrated in Figure 1.21b [271], the library design consists of a docking site (red) flanked by two random regions (green) and two primer binding sites (black). The docking site is complementary to a biotinylated strand, which can be captured using avidin-coated magnetic beads. When the target molecule is present, the allosteric binders detach from the capture strand, allowing separation from the inactive ones using a magnet. However, this selection strategy inevitably favours mutated docking sites that occurred during PCR amplification, as they tend to detach more easily from the capture strand [271].

Many aptamers that were selected without specific emphasis on allostery have also demonstrated the ability to undergo structural switching upon binding to their target, producing signals such as FRET, fluorescence, and colorimetry [184, 274, 275]. In fact, most switching aptamers were created through post-selection engineering [276, 277]. Figure 1.22a [184] illustrates two common engineering strategies for creating a ligand-specific

biosensor. One strategy involves splitting the aptamer into two oligos that reassemble only in the presence of the ligand. The other strategy involves making the aptamer partially duplex with a complementary strand, which dehybridises from the aptamer upon the addition of the ligand. Two examples of engineering an ATP aptamer based on these two strategies are shown in the upper panel of Figure 1.22b. The secondary structures of several other aptamers that are commonly engineered to be biosensors are shown in the lower panel of Figure 1.22b.

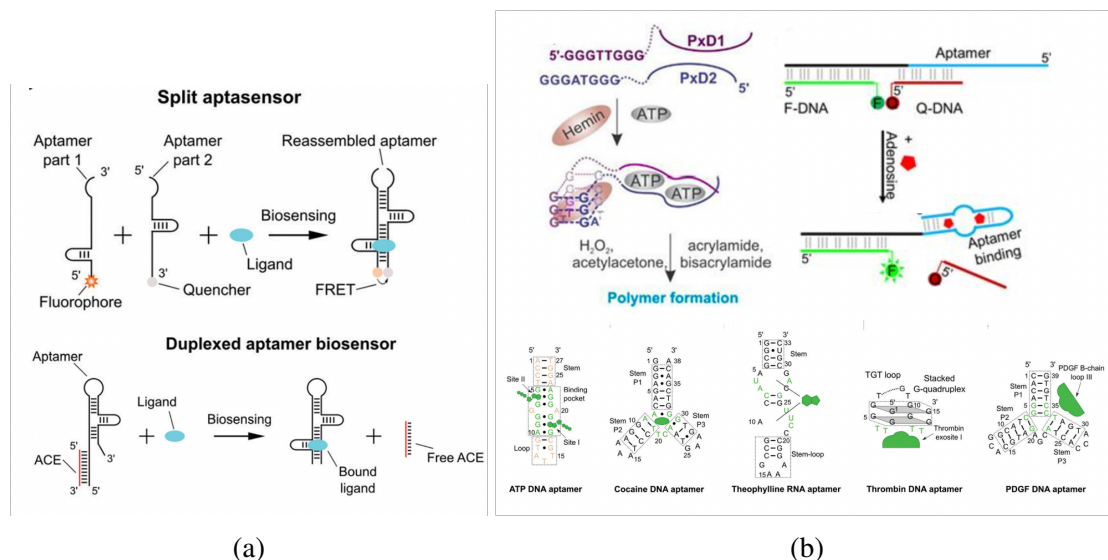


Figure 1.22: Engineering aptamers to create biosensors (a) Common strategies for engineering aptamers. Images adapted from ref. [184]. (b) Upper panel: Two examples of using an ATP aptamer engineered as biosensors; lower panel: Secondary structures of several other aptamers commonly engineered to be biosensors. Images adapted from ref. [184, 278, 279].

1.5.2 In vitro selection of catalytic nucleic acids

Catalytic nucleic acids are considered more complex than aptamers, because while both can bind to targets, catalytic nucleic acids also facilitate catalysis [140]. For catalytic nucleic acids that have nucleic acids as substrates, both the binding and selection steps are straightforward: recognition is achieved through base pairing, and separation can be accomplished by PAGE purification or attachment to beads modified with sequence tags. Figure 1.23a illustrates the selection and separation steps involved in selecting the first RNA ligase ribozymes [280]. One of the nucleic acid substrates is covalently attached to the 220-nt random library oligonucleotide, while the other substrate has a tag and a primer

binding site. Active species ligate the substrate with a tag, and then they can be captured using modified agarose beads by hybridisation.

For small-molecule substrates, one way to produce catalytic nucleic acids is to select aptamers that bind to a transition state analogue of a reaction and then screen these aptamers for catalytic activity ^[219]. However, this approach for catalytic nucleic acids has been less successful compared to catalytic antibodies and is constrained by the availability of transition state analogues ^[281]. Another common method involves attaching the substrate molecule to the library via a linker ^[282], as illustrated in Figure 1.23a. Catalytically active species modify themselves and can then be isolated through various chemistries, depending on the desired reaction.

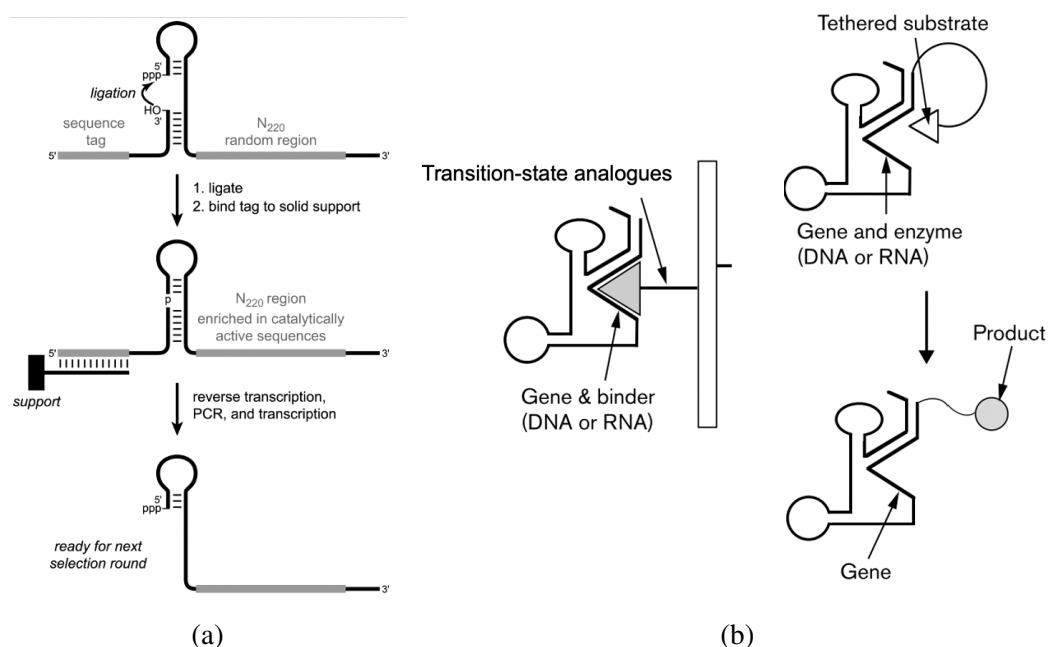
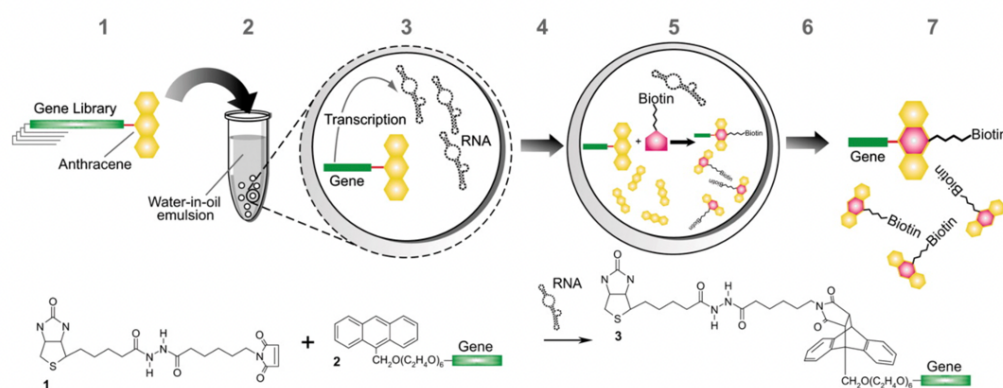


Figure 1.23: Methods to identify catalytically active nucleic acids. (a) Separation of active ribozymes from oligonucleotide substrates through hybridisation. For a detailed explanation of the procedure, refer to the text. Image adapted from ref. ^[280]. (b) Two methods for selecting catalytic nucleic acids for small-molecule substrates: (left) selection of aptamers that bind to a transition state analogue of the reaction; (right) tethering the substrate to the library and fishing out the species tethered with the product. Images adapted from ref. ^[281].

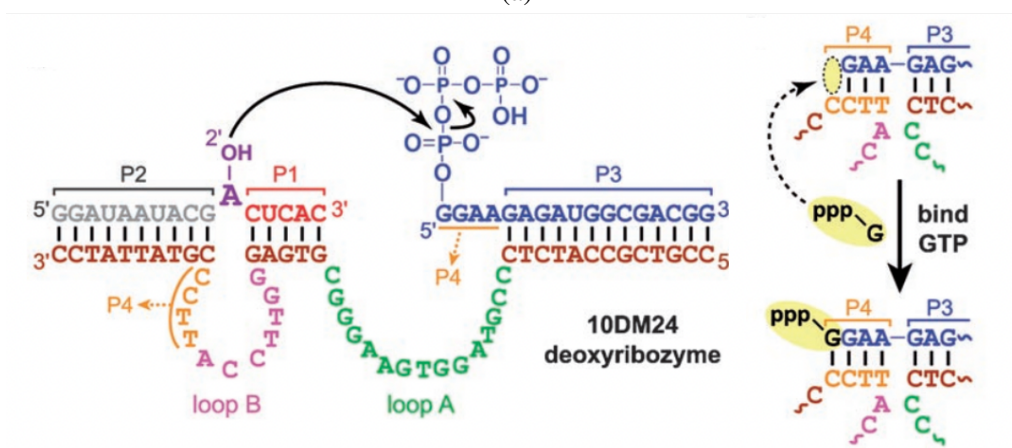
Due to the lack of selection pressure for substrate binding and product release, ribozymes and DNAzymes evolved using tethered substrates typically exhibit limited or no turnover abilities ^[281]. Griffiths and Tawfik developed a method called *in vitro* compartmentalisation (IVC) that enables selection for all enzymatic properties: substrate binding, catalysis, and turnover ^[283]. Originally developed to generate "artificial cells" for se-

lecting protein enzymes, this method creates small reaction compartments for individual genes using a water-in-oil (w/o) emulsion with aqueous droplets. Each droplet contains the components necessary for transcription, translation, and fewer than one gene with tethered substrates. After the reaction, the genes encoding the active species, now tethered with products, are isolated and amplified for further rounds of selection ^[241]. This method has been used to select catalytic nucleic acids, resulting in the production of several ribozymes ^[33-36]. A schematic diagram for the selection of multiple-turnover ($k_{\text{cat}} \approx 0.3/\text{s}$) Diels–Alderase ribozymes using IVC is illustrated in Figure 1.24a ^[231]. First, attach one of the substrates to a DNA library (1). Then, create a water-in-oil emulsion to compartmentalise DNA molecules in aqueous droplets with components for transcription (2). After the RNA library is generated in each droplet through in vitro transcription (3), add the other substrate labelled with biotin to the emulsion (4) and allow it to diffuse into the compartments (5). Finally, break the emulsion (6) and recover the active ribozymes by fishing out their corresponding DNA molecules with the biotin tag (7).

However, a major disadvantage of IVC is the restriction on the number of species in the starting pool, limited to 10^8 to 10^{11} , since there are approximately 10^{10} aqueous droplets per ml of emulsion ^[241]. In addition, both the transcription and the catalysed reactions occur in the same environment, which may not always be compatible. Creating water-in-oil emulsions with uniform compartmentalisation and maintaining stable emulsions throughout the selection process can be technically challenging. There is also a risk of cross-contamination between compartments, especially during emulsion formation and breaking.



(a)



(b)

Figure 1.24: Creating multiple-turnover catalytic nucleic acids. (a) Schematic diagram of the selection procedure of multiple-turnover Diels–Alderase ribozymes by using IVC. For a detailed explanation of its operation, refer to the text. Image and caption adapted from ref. [231]. (b) Left: Secondary structure showing the interaction between 10DM24 and the oligonucleotide substrates P1–P2 and P3. Right: Engineering a GTP binding site into 10DM24. Image and caption adapted from ref. [284].

Hobartner and Silverman reported a multiple-turnover DNAzyme by engineering a small-molecule substrate binding site into a previously selected single-turnover DNAzyme 10DM24 [284]. The original 10DM24 DNAzyme catalyzes the nucleophilic attack of an internal 2'-hydroxy group of an RNA oligonucleotide (P1–P2) at the 5'-triphosphate of another RNA (P3), resulting in the production of 2', 5'-branched RNA [285]. As illustrated in Figure 1.24b, 10DM24 binds to the two substrates through base pairing. Like most other catalytic nucleic acids that act on oligonucleotide substrates, 10DM24 exhibits single-turnover ligation activity due to product inhibition [286]. To reduce product inhibition, the authors engineered the substrate by removing the terminal GTP from P3 and using free GTP as the substrate. Despite the low turnover rate caused by extensive Watson–Crick

interactions between P1-P2 and 10DM24, the authors observed five turnovers over five hours of incubation with excess GTP and the other oligonucleotide substrate P1-P2 ^[284].

1.5.3 Modular nucleic acid catalysts

Inspired by the fact that many natural protein enzymes have discrete substrate recognition and catalyst domains ^[287], Dokukin and Silverman ^[288] explored a modular approach to evolve new catalytic nucleic acids for small-molecule substrates. As illustrated in Figure 1.25a, they integrated a predetermined ATP aptamer module adjacent to a random region to identify new DNAzymes for tyrosine side chain phosphorylation of a CAAYAA hexapeptide substrate tethered to the anchor DNA, using ATP as a phosphoryl donor through in vitro selection. Active species are then captured using 8VP1 ^[289], a DNAzyme that catalyses covalent attachment of an RNA tag (capture oligo) to a phosphotyrosine. After 14 rounds of selection, they identified an active sequence named 14JS101 from the library containing 40 random nucleotides. To determine whether the ATP aptamer contributed to the activity, they conducted a mutational analysis. The results indicated that 14JS101 is indeed a modular tyrosine kinase DNAzyme, where the ATP aptamer domain binds ATP and the domain evolved from the initially random region is responsible for catalysis.

The Willner group introduced a concept of the construction of modular nucleic acid-based catalysts for small-molecule substrates using existing components ^[290]. This concept, termed "nucleoapzyme", involves the conjugation of an aptamer to a catalytic centre, which could be a ribozyme, a DNAzyme, or a metal-ligand complex with corresponding catalytic activity. The idea is that the aptamer domain brings the substrate in close proximity to the catalytic centre, increasing the local concentration of the substrate and thereby enhancing the catalytic activity ^[291]. Hemin/G-quadruplex (hGQ) is a DNAzyme that catalyses H₂O₂-stimulated oxidation of dopamine to aminochrome ^[292]. Golub and coworkers constructed a series of hGQ and dopamine aptamer conjugates with different configurations (Figure 1.25b, upper panel). They found that linking the dopamine aptamer to the 3' end of hGQ through a TATA spacer led to a 20-fold increase in activity compared to the separated hGQ and aptamer domains ^[290]. Similarly, Biniuri and col-

leagues developed ATPase-mimicking nucleozymes by linking an ATP aptamer with a bis-Zn²⁺-pyridyl-salen-type catalyst (Figure 1.25b, lower panel) [293]. While the individual ATP aptamer and bis-Zn²⁺-pyridyl-salen-type catalyst were not catalytically active, all conjugates demonstrated catalytic activity. The configuration with the complex attached to the 3' end of the aptamer via a TT linker achieved a k_{cat} of 6.88 min⁻¹.

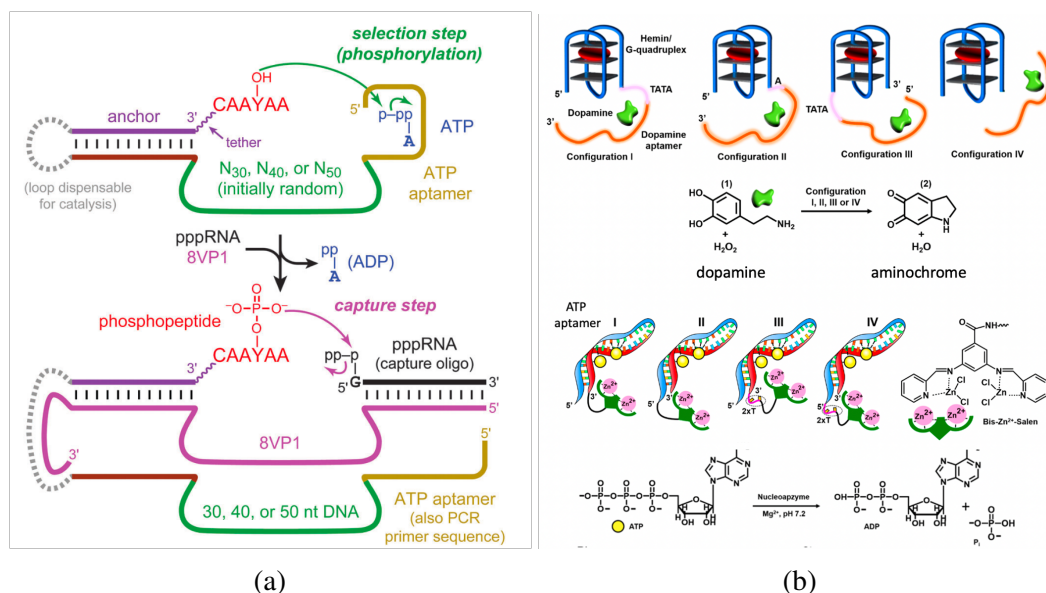


Figure 1.25: (a) Design of the in vitro selection process for modular tyrosine kinase DNAzyme. For a detailed explanation of the procedure, refer to the text. Image and caption adapted from ref. [288]. (b) Upper panel: schematic representation of a nucleozyme consisting of a dopamine aptamer and hemin/G-quadruplex for H₂O₂-stimulated oxidation of dopamine to aminochrome. Lower panel: schematic representation of a nucleozyme consisting of an ATP aptamer and a bis-Zn²⁺-pyridyl-salen-type catalyst for the hydrolysis of ATP to ADP. Images and captions adapted from ref. [294].

Chapter 2

A proposed optimised mechanism for autonomous molecular machines made from nucleic acids

Despite being chemically less diverse, nucleic acids are biomolecules comparable to functional proteins found in nature when it comes to ligand binding and chemical catalysis. Additionally, nucleic acids offer several advantages, including better stability, ease of production, programmability and modularity. In vitro evolution and engineering techniques enable the development of new functional species with virtually any desired function. Furthermore, with negatively charged sugar-phosphate backbones and aromatic bases, nucleic acids can easily be adsorbed onto various materials, such as gold nanoparticles and graphene oxide, through interactions such as aromatic stacking and hydrogen bonding, which allows operations on surfaces and at interfaces ^[295]. These properties make nucleic acids a promising material for creating artificial molecular machines that mimic the mechanocatalysis cycles found in protein molecular machines.

This mechanism has been realised in a DNA walker by coupling its mechanical cycle with the catalysed hybridisation reaction of hairpin fuels ^[10]. However, the inherently slow hybridisation and dehybridisation of oligonucleotides make it rather primitive compared to the highly efficient protein motors. The various functions of nucleic acids depend on complex multistep conformational changes in response to external signals. Compared to oligonucleotides, small molecules can trigger more rapid responses. As shown in the

summarised time chart of RNA in Figure 2.1 [296], the small-molecule ligand binding, catalysis, and allosteric processes are all within a timescale of 10^{-7} to 1 s. This makes it possible for a cycle of allosteric binding of small-molecule fuel, catalysis of its chemical reaction, and release of the waste product in seconds.

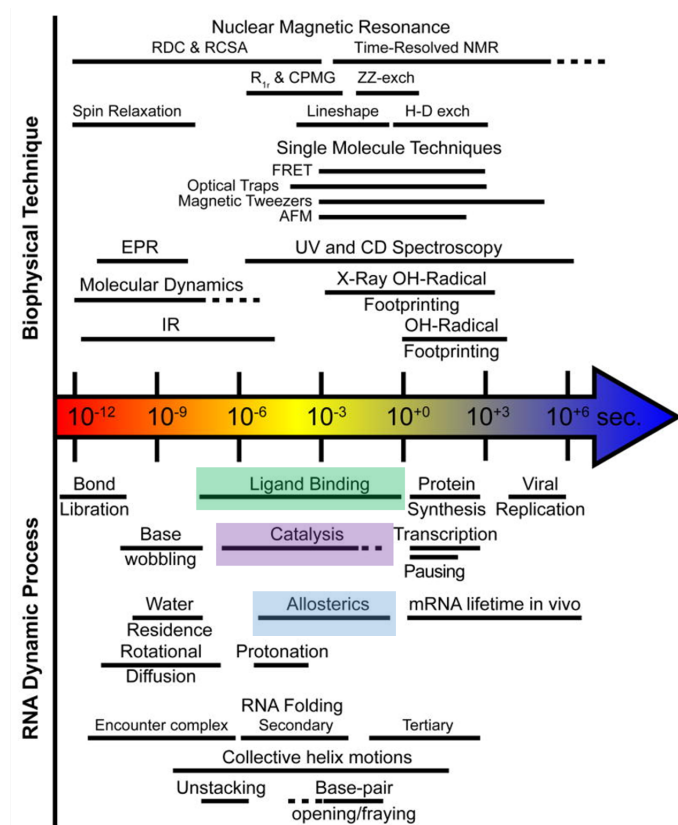


Figure 2.1: Time-chart of dynamic processes in RNA and corresponding biophysical characterisation techniques. Timescales for ligand binding, catalysis and allosterics are highlighted. Image and caption adapted from ref. [296].

2.1 An optimised mechanism

Using the autonomous two-footed walker depicted in Figure 1.9a [10] as a model system, we proposed an optimised mechanism by integrating a mechanocatalyst oligonucleotide as the "feet", where their mechanical states are directly coupled to the chemical catalysis of the small-molecule substrate (Figure 2.2a). We chose nucleoside triphosphates (NTPs) as the substrate, given their prevalence in biological systems and extensive research into functional nucleic acids that interact with them [297–302]. Instead of TMSD, in this system, foot lifting and freeing are driven by allosteric binding of the oligonucleotide to NTP (state 0 to state 1), followed by product (NMP) release after catalysed hydrolysis (state 1 to state

2).

The free energy profiles for the Curtin-Hammett-like step in each coordinated chemomechanical cycle are shown in Figure 2.2b. In state 0, the two feet, A and B, are bound to positions $m - 1$ and m , respectively, through hybridisation. The competition between the two feet for binding to the overlapping hybridisation site on the track favours the right foot (B), revealing the substrate binding site on the left foot (A). This leads to a higher ground-state energy for A in the substrate binding step (from state 0 to state 1), resulting in a lower activation energy for the formation of an oligo-NTP complex (A'). Upon NTP binding, the catalytic site in the oligo is activated and brought closer to the bound NTP, catalysing its hydrolysis (state 1). The product is subsequently released (state 2) because the affinity of the NMP to the oligo is significantly lowered. The newly freed foot L can rebind to the track (state 0), which can either idle by rebinding to $m-1$ (A_1) or take a forward step to bind to $m+1$ (A_3). In this system, NTP hydrolysis is coupled to the mechanical cycle of the walker, and the feet are used as the catalyst. The competition between the feet and the overlapping binding site on the track is responsible for the directionality. Compared to the original system, the foot-lifting step remains the only determinant of directionality. However, the slow branch migration reactions involved in foot-lifting and product release are replaced by more rapid small-molecule binding and product release.

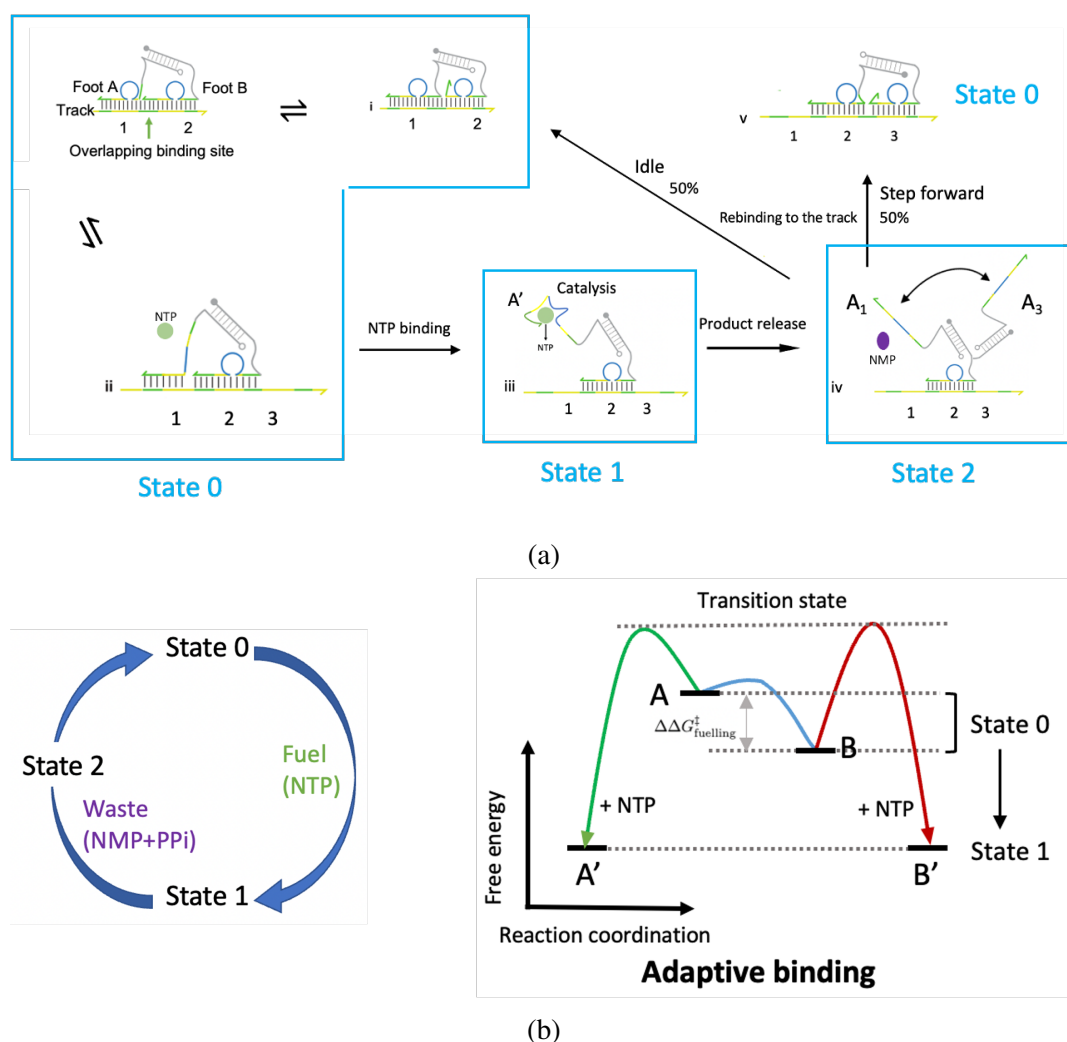


Figure 2.2: (a) Design of the proposed small molecule-driven oligonucleotide walker. For a detailed explanation of its operation, refer to the text. Image adapted from ref. ^[10]. (b) The coordinated chemomechanical cycle and free energy profiles of Curtin-Hammett-like step in each cycle, with green arrows indicating the preferred reaction pathway. For a detailed explanation of its operation, refer to the text.

One may wonder why the free foot would rebind to the track rather than capture another NTP and oscillate between state 1 and state 2 (Figure 2.3a). In reality, the preference can be biased by the engineering and design of the sequence. Wilson and colleagues ^[303] studied the thermodynamic and kinetic behaviour of a common class of aptamer switches designed using an intramolecular strand-displacement strategy. The aptamer switch they investigated was a DNA aptamer for ATP, tethered via a flexible poly-T linker and a short displacement strand (DS) that closed the loop through hybridisation (Figure 2.3b). The authors showed that, by adjusting the length of the DS and the linker between the aptamer and the DS, the thermodynamic and kinetic properties of the aptamer switch can be independently and systematically tuned over a broad range. They showed that an increase in

loop length L_{loop} led to a higher effective concentration, which could increase the on-rate of the displacement strand $k_{\text{on}}^{\text{DS}}$, while the off-rate $k_{\text{off}}^{\text{DS}}$ was negligibly affected. Similarly, foot A (Figure 2.3a) and the track form a hairpin-shaped structure, and the rate of rebinding to the track k_{on}^{T} is also tunable through the design of the loop length and also the length of the single-stranded linker in grey, without impacting the off rate $k_{\text{off}}^{\text{T}}$. Furthermore, the free energy differences between the two feet in state 0 and the ATP-bound state in state 1 can be adjusted by changing the lengths of the overlapping binding site and the total binding length for each foot on the track.

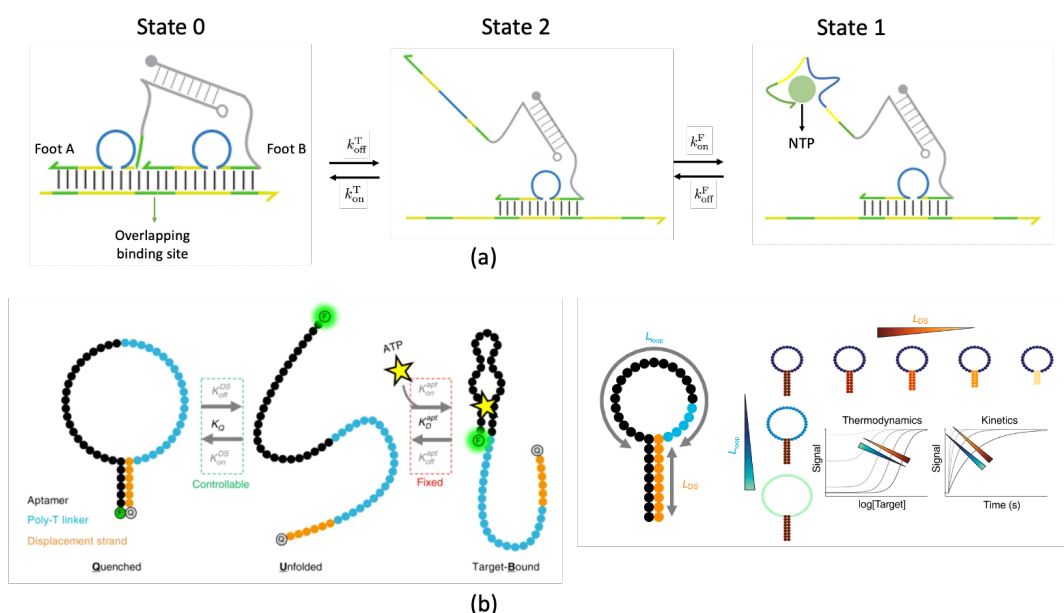


Figure 2.3: (a) Kinetic scheme depicting possible reaction pathways of the free foot. Image adapted from ref. [10]. (b) Design of an intramolecular strand-displacement ATP switch with independent control of its thermodynamic and kinetic properties. Image and caption adapted from ref. [303].

2.2 Two strategies to create the mechanocatalyst

Catalytic nucleic acids for small molecules typically exhibit more structurally complex conformations compared to aptamers, as they must not only bind to the substrate but also catalyse its chemical reaction [140]. While the induced-fit ligand binding model is widespread in aptamers, ribozymes and DNAzymes often have preformed active conformations or only undergo localised structural changes [187, 304–306]. The successful selection of the tyrosine kinase DNAzyme [288], which has discrete aptamer and catalyst domains,

and various nucleozymes [290] developed by the Willner group, demonstrated that explicitly modular catalytic nucleic acids are possible.

Nutiu and Li [307] reported a hybrid DNA–enzyme nanoengine system that operates under a mechanism that also involves allosteric ligand-aptamer interactions and catalysis. Their nanoengine (Figure 2.4c) consists of an ATP aptamer and an antisense DNA strand as the output. They are modified with a FRET pair and are in hybridised duplex form at first. The addition of the input adenosine molecules triggers rapid structural changes of the aptamer strand, producing a free output and an aptamer-input complex. To recycle the aptamer and allow cyclic operations, they used adenosine deaminase, which converts free adenosines in the environment into inosines not bound by the aptamer, thus accelerating the dissociation of the aptamer-adenosine complexes. Each cycle takes about 15 minutes. This system is fundamentally different from the desired mechanism, as the catalysis is performed by an external protein enzyme.

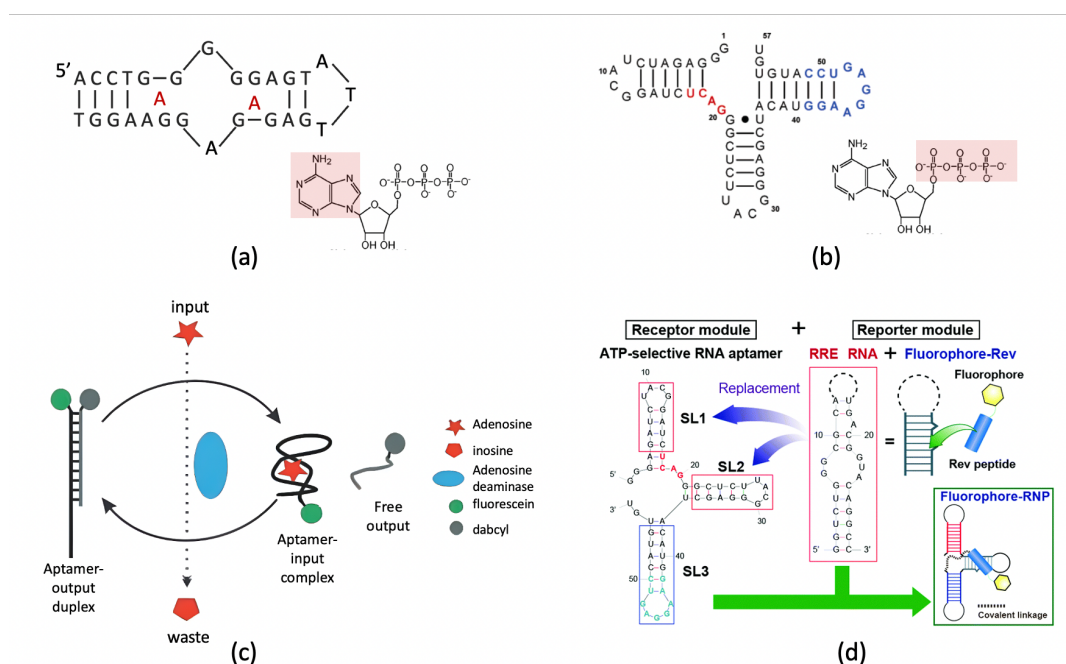


Figure 2.4: (a) The 2D structure of the classic adenosine-binding DNA aptamer with the two bound adenines in red. The recognised moiety of the ATP molecule is highlighted in red. Image adapted from ref. [308]. (b) The 2D structure of the RNA aptamer that recognises the triphosphate moiety of ATP. The coloured nucleotides indicate conserved sites among the selected clones from the final round of selection. The recognised moiety of the ATP molecule is highlighted in red. Image adapted from ref. [270]. (c) A hybrid DNA–enzyme nanoengine. For a detailed explanation of its operation, refer to the text. Image and caption adapted from ref. [307]. (d) The modular construction of a fluorescent sensor that discriminates between ATP and ADP. Image and caption adapted from ref. [309].

The aptamer used in the nanoengine is a 27-nt DNA aptamer (DH25.42) (Figure 2.4a) that recognises ATP and its adenosine derivatives ^[310]. Since its development in the Szostak laboratory in 1995, it has become one of the most studied aptamers and has been extensively used as a structural-switching element in DNA nanotechnology ^[308, 311–313]. One way to avoid using a protein enzyme is to incorporate an adenosine deaminase-mimicking catalytic module. However, to the best of my knowledge, such a ribozyme or DNAzyme has not yet been reported in the literature. Moreover, designing a selection strategy for nucleobase-modifying ribozymes or DNAzymes is quite a challenge.

In 2003, the Szostak lab reported another ATP aptamer (1-1min), this time composed of RNA, which demonstrates strong and specific binding to the triphosphate moiety of ATP ^[270]. Its affinity for ATP was measured to be 64-fold higher than for ADP and 1100-fold higher than that of AMP. Compared with the adenosine aptamer, this one has received much less attention. Nevertheless, ATP-induced allosteric change was demonstrated in a fluorescent ribonucleopeptide sensor reported by Nakano and co-workers (Figure 2.4d) ^[309]. They developed a universal rational modular design of ribonucleopeptide (RNP) sensors by fusing a fluorophore-modified Rev-binding RNA sequence (RRE) to the RNA aptamer of interest. Upon ligand binding, the global structural change in the aptamer module is transduced into changes in fluorescence intensity. By replacing the non-consensus stem-loop regions (SL1 and SL2) in the ATP aptamer with the RRE RNA, they were able to detect differentiable fluorescence intensity changes in response to ATP and ADP.

We devised two strategies for creating the desired mechanocatalyst. The first approach involved enhancing Willner's ATPase-mimicking nucleocatalyst (Figure 1.25b, lower panel) ^[293] by substituting the adenosine-binding aptamer with the triphosphate-binding one. A turnover rate of 6.88 min^{-1} was reported using the adenosine aptamer. However, the ligand-bound and unbound states of this aptamer are independent of the chemical catalysis because its affinities for ATP and ADP are nearly identical. By incorporating the triphosphate-binding aptamer, our aim was not only to improve the turnover rate, but also to establish a coupling between the mechanical state of the aptamer and the chemical catalysis (Figure 2.5a). The second approach involved in vitro selection. We constructed a partially scaffolded library that incorporates the triphosphate-binding

aptamer and selected species that facilitate the conversion of NTP to NMP. We hypothesised that the aptamer would serve as the allosteric binding site, bringing a catalytic site into close proximity with the NTP. Following catalysis, the product would be released due to a significantly reduced affinity (Figure 2.5b).

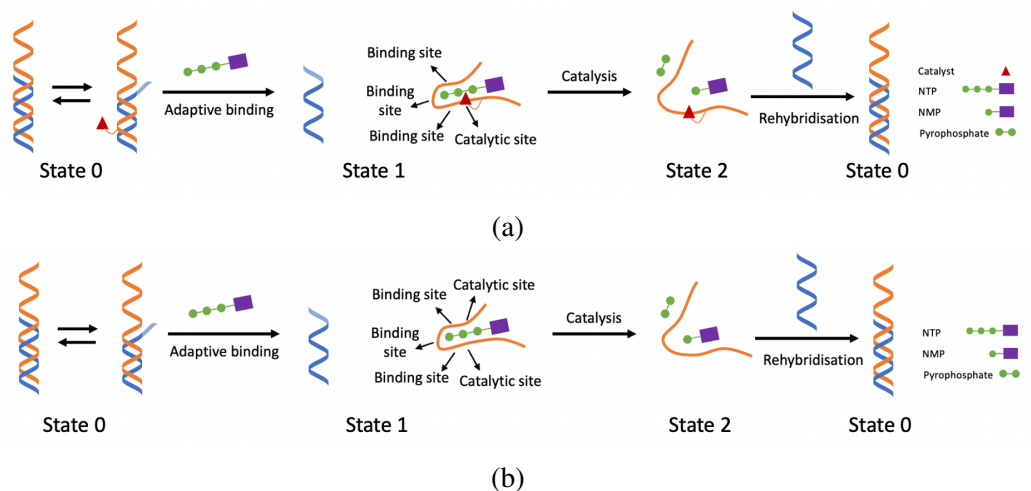


Figure 2.5: Simplified chemomechanical cycles of the two strategies. The competition for a common binding site on the track (not shown) always exposes a single-stranded region on the trailing foot (state 0), which facilitates foot-lifting by the fuel (NTP). Allosteric binding positions either a catalyst or the catalytic site on the foot near the NTP (state 1). Following the conversion of NTP to NMP, the NMP is released (state 2), allowing the free foot to rebind to the track. (a) Operational scheme of a mechanocatalyst comprising an aptamer and a tethered catalyst. (b) Operational scheme of a mechanocatalyst featuring an integrated aptamer and a catalytic oligonucleotide region.

Chapter 3

Tethered catalyst approach

3.1 Introduction

The use of prosthetic molecular catalysts offers an attractive solution because it bypasses the complexities and limitations of in vitro selection. The Willner group constructed a nucleocapsule, consisting of an ATP aptamer linked to a bis-Zn²⁺-pyridyl-salen complex that catalyses ATP hydrolysis with a turnover rate of 6.88 min⁻¹ [293]. Although this is much slower than the rate for natural ATPases like F₁-ATPase (9720 min⁻¹) [314], is considerably higher than the fastest reported synthetic ATPase activity, which was only 0.032 min⁻¹ [315]. In this system (Figure 3.1), the carboxylic-acid-functionalised catalyst is covalently attached to an amino-modified ATP aptamer. Upon binding to ATP, the aptamer captures the substrate and positions it within 18 Å of the catalytic site, allowing the reaction to occur. Neither component is catalytically active on its own.

However, the ligand-bound and unbound states of the aptamer are independent of the chemical catalysis because its affinities for ATP and ADP are nearly identical. We hypothesised that replacing the adenosine-binding aptamer with a triphosphate-binding aptamer may establish a coupling between the mechanical state of the aptamer and the chemical catalysis and could also improve the turnover rate. Since the catalytic complex was not commercially available, it was necessary to synthesise it.

Organic syntheses were performed under the supervision of Dr Rana Abdul Razzak, who also helped greatly in experimental design and NMR data analysis.

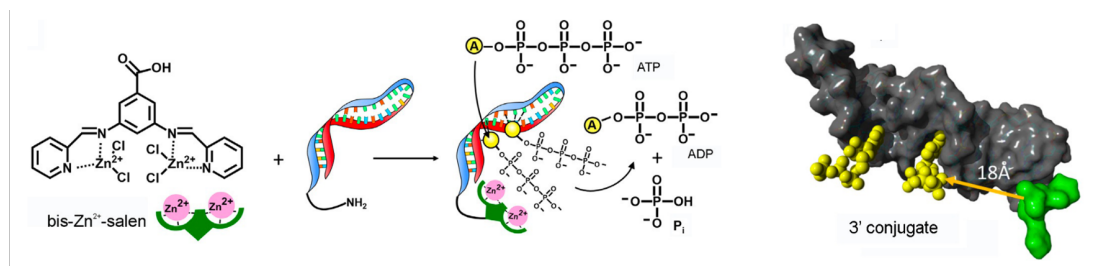


Figure 3.1: Left: Schematic of a nucleozyme with bis-Zn²⁺-pyridyl-salen-functionalised ATP aptamer catalysing ATP hydrolysis to ADP. Right: Molecular dynamics simulation of the ATP/nucleozyme complex, with the catalyst (green) linked to the aptamer's 3'-end (ATP ligand in yellow). Arrows indicate the shortest distances (18 Å) between the catalytic and ATP-binding sites. Images and captions adapted from ref. [293].

The synthesis of the Pyridyl-Salen-Type ligand was successfully reproduced. However, synthesis attempts of the Zn(II) complex yielded inconsistent results. During the initial attempt, we observed a subtle shift in the ¹H NMR resonances, accompanied by line broadening, which is characteristic of Zn-complexes, as reported in the literature. In the second attempt, the shift in the ¹H resonances was even smaller, with reduced line broadening. After leaving the sample in DMSO-*d*₆ for 24 hours and subsequently acquiring a ¹H NMR spectrum, new peaks began to emerge, potentially indicating hydrolysis. This raised uncertainty as to whether the Zn-complex was formed or whether it was formed but remained unstable.

In this chapter, I will present the experimental details and data obtained from these attempts. However, due to COVID-19 restrictions and limited access to laboratory space, along with the reduced availability of the supervising postdoctoral researcher, further investigation into this project was discontinued and I then redirected my efforts towards the *in vitro* selection approach.

3.2 Materials and Methods

All reagents and solvents were obtained from Sigma-Aldrich and used without further purification unless otherwise stated. All reactions were performed under a nitrogen atmosphere using standard Schlenk techniques. Methanol was dried over calcium hydride followed by distillation under nitrogen. Acetonitrile and dichloromethane were purified by passing them through two columns of alumina, the MBRAUN SPS-800 purification

system.

A two-step synthesis procedure was developed (Figure 3.2) based on methods from the literature [316–318]. The first step involved the synthesis of the 3,5-bis((pyridin-2-ylmethylene)amino)benzoic acid ligand. Given the water sensitivity of the imine product, the process was conducted under strictly anhydrous conditions. In the second step, triphenylphosphine (PPh_3) was used to form the precursor complex $[\text{Zn}(\text{PPh}_3)_2\text{Cl}_2]$. This complexation facilitates the solubilisation of ZnCl_2 in the organic solvent and stabilises the transition state during the ligand exchange process, thereby enhancing the efficiency of the desired formation of the zinc-ligand complex.

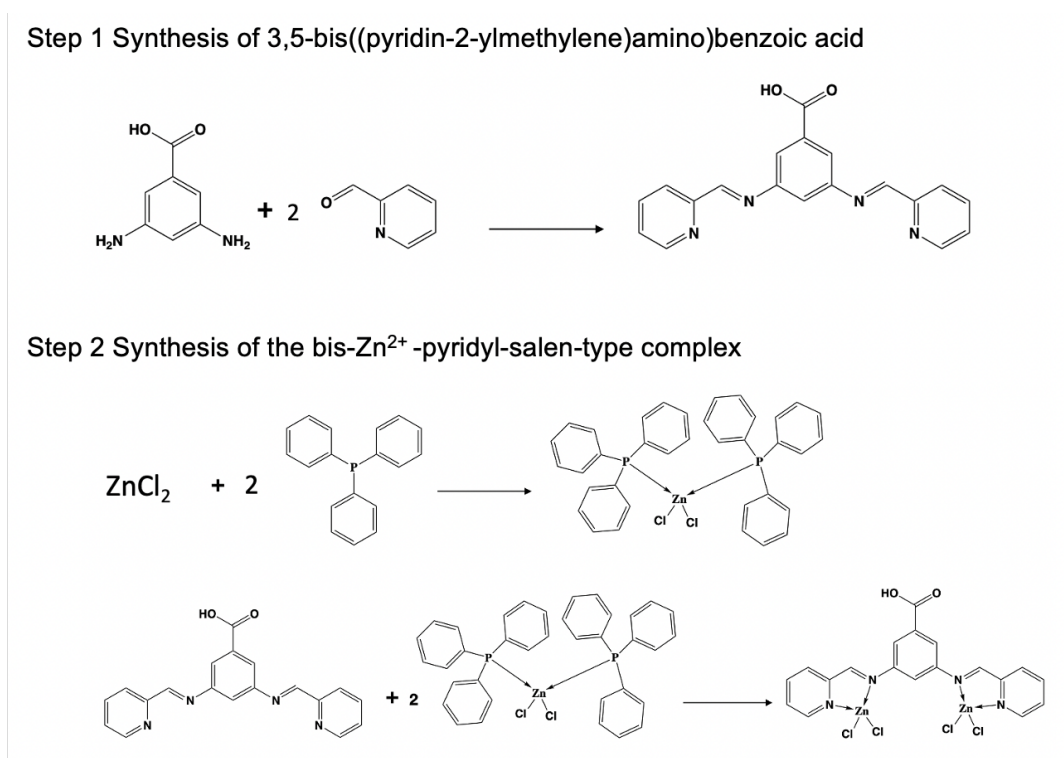


Figure 3.2: Two-step synthesis process of the bis- Zn^{2+} -pyridyl-salen-type complex. Step 1: Synthesis of the ligand 3,5-bis((pyridin-2-ylmethylene)amino)benzoic acid via the condensation of 3,5-diaminobenzoic acid with 2 equivalents of pyridine-2-carbaldehyde under anhydrous conditions. Step 2: Formation of the bis- Zn^{2+} -pyridyl-salen-type complex through the precursor complex $[\text{Zn}(\text{PPh}_3)_2\text{Cl}_2]$.

Synthesis of the ligand 3,5-bis((pyridin-2 ylmethylene)amino)benzoic acid

A three-neck round-bottom flask (100 mL) fitted with a reflux condenser was charged with 0.50 g of freshly ground 3,5-diaminobenzoic acid (98%, powder with small lumps).

Dry methanol (40.00 mL) was added by syringe, followed by the dropwise addition of 0.77 g of 2-pyridinecarboxaldehyde (99%, liquid) at room temperature. Upon complete addition of 2-pyridinecarboxaldehyde, the suspension became a clear amber-colored solution, and an off-white precipitate began to form. The reaction mixture was stirred at reflux temperature (65°C) for four hours, then allowed to cool to room temperature. The precipitate was filtered under nitrogen using a custom filter cannula made from filter paper and a syringe needle, washed with dry MeOH (3 × 20 mL) and dried under vacuum to obtain the product as an off-white powder (0.85 g, 83.3%). The solid product was stored in a glovebox.

Synthesis of the Bis-Zn²⁺-Pyridyl-Salen-Type Complex

A round-bottom flask (50 mL) was charged with anhydrous ZnCl₂ (1 equiv., 1.69 mmol, 0.23 g; ≥97%, anhydrous) and triphenylphosphine (PPh₃; 2.2 equiv., 3.73 mmol, 0.97 g; ≥99.0%, solid). Dry acetonitrile (12.00 mL) was added by a syringe, after which the suspension was heated at 60°C for 20 minutes. After allowing the reaction mixture to cool to room temperature, a suspension of the ligand (0.5 equiv., 0.85 mmol, 0.23 g) in dry DCM (12.00 mL; obtained from the MBraun SPS800 solvent drying system at the Department of Chemistry, University of Oxford) was added in a single portion. The reaction mixture suddenly turned into a bright yellow solution, after which a pale yellow precipitate started to form. The resulting reaction mixture was stirred overnight, then filtered using an in-house made filter cannula, washed with DCM (3 × 20 mL) followed by methanol (3 × 20 mL), and dried under high vacuum to obtain a pale yellow powder (quant. 0.54 g).

3.3 Synthesis outcomes and discussion

Synthesis outcomes and discussion of the ligand

The synthesised product was dissolved in anhydrous DMSO-*d*₆, and ¹H NMR spectra were recorded within 15 minutes of solution preparation (Figure 3.3). The ¹H NMR spectrum (400 MHz, DMSO-*d*₆) displayed the following chemical shifts: δ 13.25 (s,

1H, H₈), 8.76 (ddd, $J = 4.89, 1.74, 0.97$ Hz, 2H, H₄), 8.73 (s, 2H, H₅), 8.20 (ddd, $J = 8.01, 1.17$ Hz, 2H, H₁), 7.99 (ddd, $J = 7.83, 7.76, 1.79$ Hz, 2H), 7.74 (d, $J = 1.97$ Hz, 2H, H₆), 7.61–7.52 (m, 3H). The reaction was repeated three times with slight variations in the experimental conditions, including adjustments to the ratio of 3,5-diaminobenzoic acid to 2-pyridinecarboxaldehyde and the volume of MeOH, with the aim of maximising the yield of the diimine ligand while minimising the formation of the monoimine ligand. Trial 2 produced the highest purity (83.3%) and the reaction conditions are recorded in the Materials and Methods section.

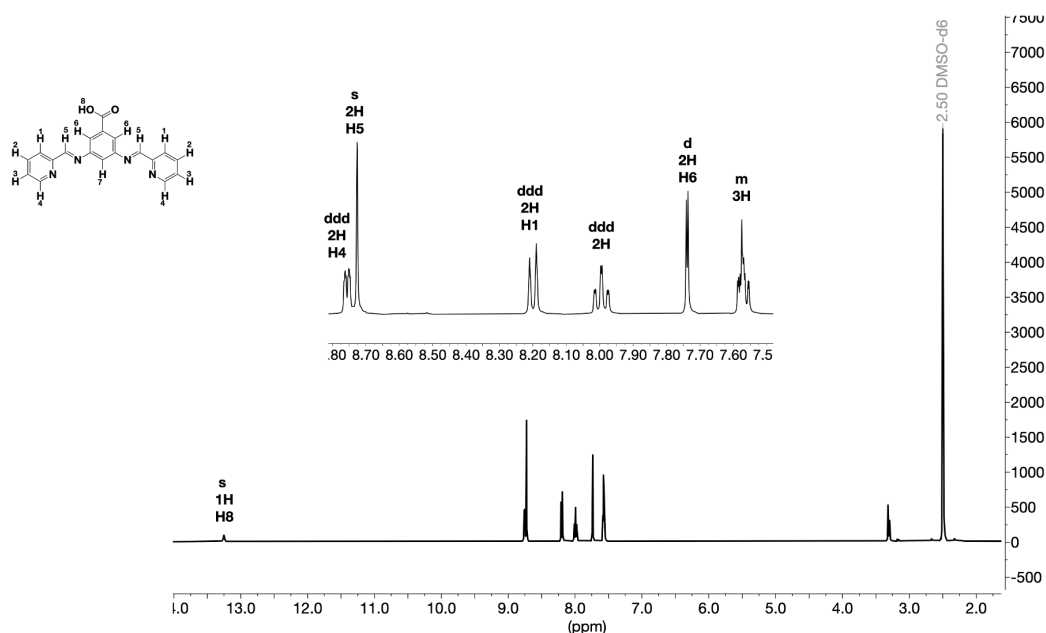


Figure 3.3: ^1H NMR spectra of the ligand 3,5-bis((pyridin-2-ylmethylene)amino)benzoic acid.

Synthesis outcomes and discussion of the complex

The ^1H NMR spectra of the product, after washing with DCM and MeOH, were recorded separately in anhydrous $\text{DMSO-}d_6$. As highlighted in Figure 3.4, the observed NMR shifts for the highlighted protons are subtle. The ^1H NMR spectrum was compared with those of several Zn(II) complexes reported in the literature^[318, 319] to see whether these shifts are typical. The chemical shifts were in agreement with the values in the literature. The spectrum profile changed after washing with MeOH. It is not clear whether washing with MeOH is detrimental to the integrity of the structure of the complex.

The ^1H NMR spectrum exhibits line broadening, a phenomenon that has been pre-

viously observed in Zn(II) complexes (ZnL) reported in the literature [320]. This line broadening in ZnL is suggested to result from aggregation which was shown to be ligand dependent.[321] Salen ligands, which feature a di-imine bridge, constrain the metal centre in a planar geometry. Consequently, ZnL are coordinatively unsaturated and tend to aggregate. In contrast, ZnL complexes with bidentate Schiff base ligands adopt a tetrahedral geometry around the metal centre, making them more likely to remain as discrete units rather than forming aggregates.

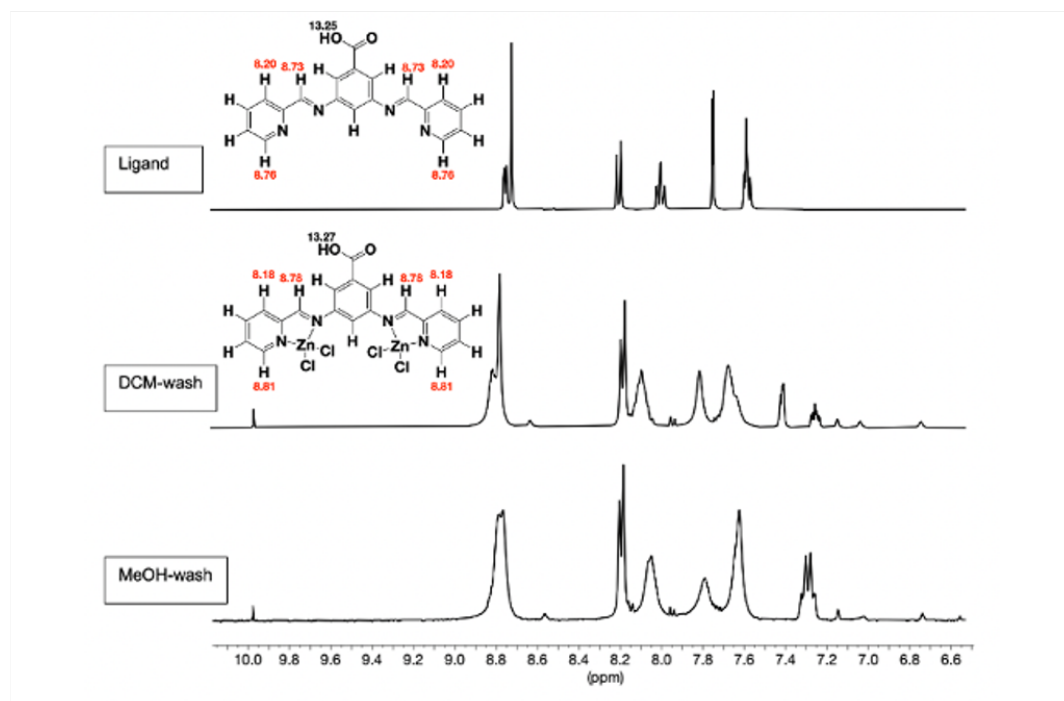


Figure 3.4: Stacked ¹H NMR spectra of the ligand and Zn(II) complex following DCM and MeOH washes.

The reaction was repeated under the same conditions. However, the resulting product exhibited a different ¹H NMR spectral profile (Figure 3.5). The product was then left in anhydrous DMSO-*d*₆ for one day in a desiccator, after which the NMR data revealed significant changes. The emergence of characteristic peaks of the monoimine ligand indicated a tendency towards hydrolysis of the Zn(II) complex. This raises concerns about the stability of the Zn(II) complex in solution because the subsequent attachment reaction to the oligonucleotides is performed in solution. Due to COVID-related restrictions on laboratory access and collaboration, no further work on this synthesis was performed.

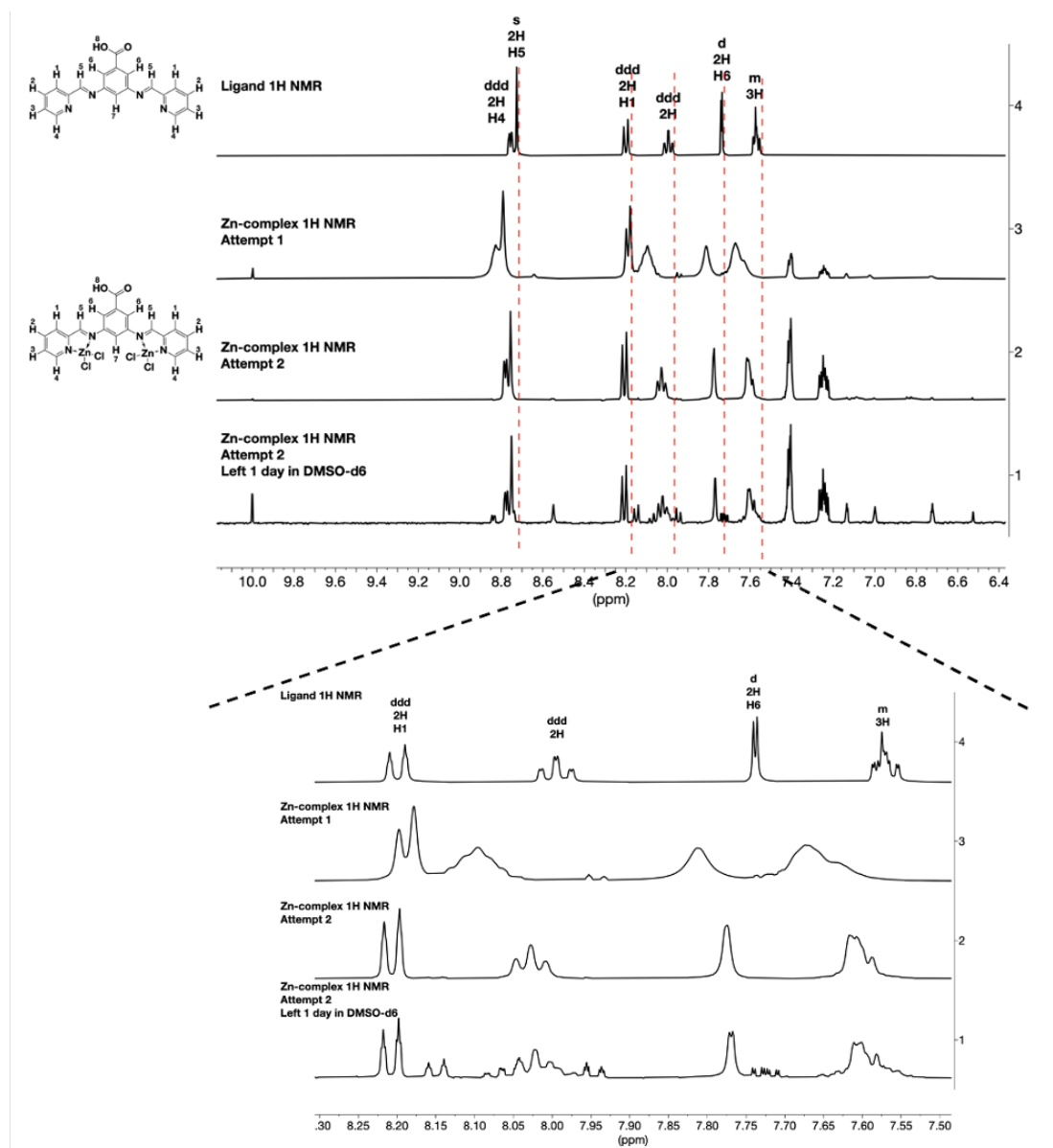


Figure 3.5: ^1H NMR spectra illustrating the variability in profiles observed in reactions involving the Zn(II) complex.

Chapter 4

In vitro selection approach

The in vitro selection process began with a partially scaffolded library based on the triphosphate-binding aptamer, with the objective of selecting a catalyst that facilitates the conversion of GTP to GMP. Our hypothesis was that the aptamer would act as an allosteric binding site, positioning a catalytic site in close proximity to the GTP. Upon catalysis, the product would be released because of its significantly reduced binding affinity. For simplicity in both library construction and the selection process, we designed the selection to target a single-turnover catalyst acting on its own 5' GTP. There was the possibility that the catalyst selected under single-turnover conditions could also perform multiple-turnover catalysis, particularly if the substrate's recognition site overlapped with the reaction site. Even if multiple-turnover activities were not observed initially, truncation and sequence engineering could potentially enable such an activity.

In this chapter, I will present the library design and selection methodology, the methods used for monitoring selection progression, the small-scale Sanger sequencing conducted on the final sub-library, and the analysis of high-throughput sequencing data from the sub-libraries of the last four rounds of selection.

The design of the in vitro selection experiments benefited greatly from the advice of Dr. Phil Holliger, who suggested attaching the GTP substrate to the library and selecting the active variants by ligation. The high-throughput sequencing was performed in his lab during my secondment at the MRC Laboratory of Molecular Biology in Cambridge. Dr. Edoardo Gianni and Samantha Kwok provided the sequencing primers and the protocol for library preparation and sequencing.

4.1 Experimental design

4.1.1 Library design

The library was designed based on the 1-1min aptamer ^[270], which has been shown to exhibit strong and specific interactions with the triphosphate group of ATP. As illustrated in Figure 4.1, the nonconserved UACG nucleotides in one of the aptamer's loops were replaced with a 20-nt random region. Another 20-nt random region was added to the 3' end of the aptamer. The entire construct was flanked by 20-nt primer binding sites at both ends, using established sequences from the literature ^[192, 322].

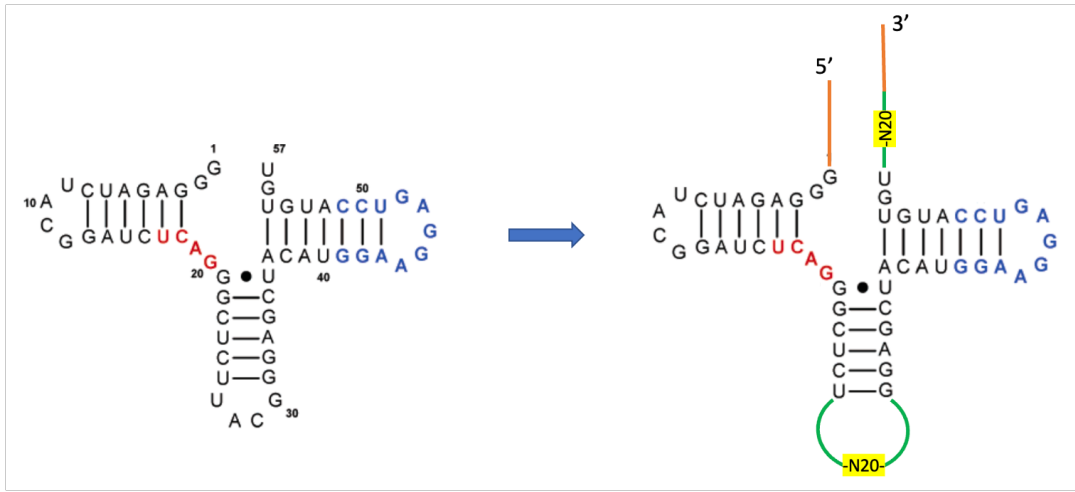


Figure 4.1: Left: Secondary structure of the 1-1min aptamer with conserved nucleotides highlighted in colour. Image and caption adapted from ref. ^[270]. Right: Library construction with the two inserted random regions in green and the primer binding sites in orange.

There have been some investigations and discussions on the influence of the length of the random region on the success of selection ^[323–326]. A library comprising approximately 10^{15} different species is generally considered both practical and sufficient for successful in vitro selection in a test tube ^[327, 328]. Such a library can cover the sequence space of a random region of around 25 nt ($4^{25} \approx 1.12 \times 10^{15}$). When the length of the random region increases to 50 nt, the theoretical sequence space expands by 10^{15} , requiring around 6×10^5 kg of oligonucleotides to cover it ^[326]. Although shorter regions can sample the sequence space more thoroughly, longer regions can allow the formation of more complex structures and increase the likelihood of selecting small active motifs ^[329]. However, the risks of misfolding and masking of active motifs also increase with increased length, while the replication efficiency decreases ^[323]. Short and long random re-

gions present distinct advantages and disadvantages, and successful selections have been reported from libraries with random regions ranging from 12 to over 200 nt [280, 330–332].

The two independent 20-nt random regions in my library offer a good balance between sequence space sampling and structural diversity. Each 20-nt region individually has a sequence space of 1.1×10^{12} , which allows thorough sampling with a typical pool size of 10^{15} . The combined sequence space remains large at 1.2×10^{24} . By using two separate random regions, the likelihood of undesired intramolecular misfolding in longer random regions is reduced, while promoting the formation of a broader range of secondary and tertiary structures. Additionally, shorter random regions are more likely to interact with the conserved regions of the aptamer, encouraging cooperative interactions within a modular ribozyme.

4.1.2 Selection design

The RNA library was generated via in vitro transcription (IVT) from a chemically synthesised DNA library (ordered from IDT) using T7 RNA polymerase. The preference of T7 RNA polymerase for initiating transcription with two guanosine triphosphates (GTP), which results in the formation of a 5' pppGpG structure [333, 334], and the analysis of 1-min binding to other nucleotides, which demonstrated that the aptamer had the highest affinity for GTP with only a 19-fold reduction compared to ATP [270], led us to consider GTP as substrate. This allowed the IVT process to naturally and conveniently link the phenotype to the genotype during selection.

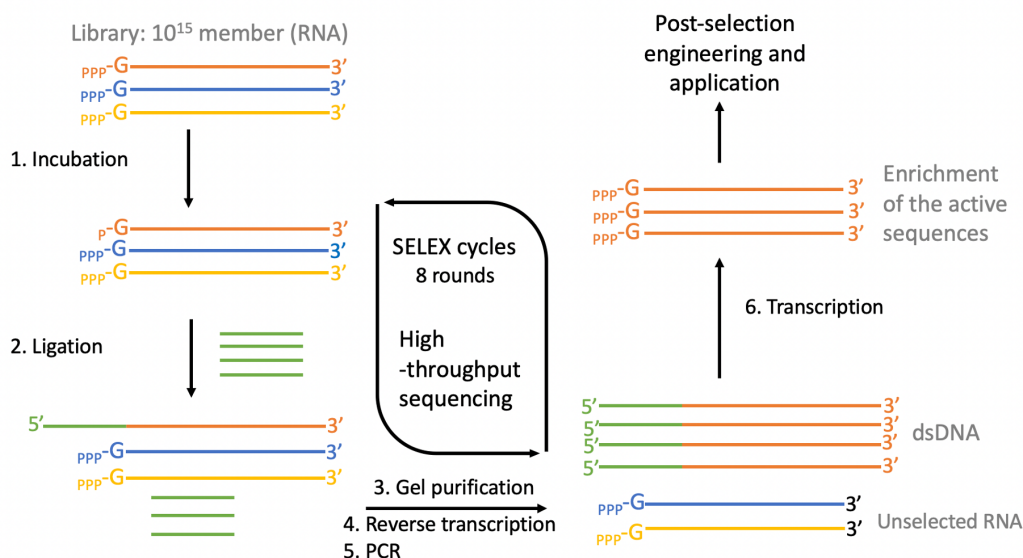


Figure 4.2: Selection workflow. For a detailed explanation of the procedure, refer to the text.

The overview of the selection workflow is illustrated in Figure 4.2. The selection started with an RNA library consisting of approximately 10^{15} different species, each with a GTP at the 5' end. The library was incubated in a customised buffer containing monovalent and divalent metal ions as potential cofactors. After incubation, active species that catalysed the conversion of their 5' GTP to 5' GMP became substrates for T4 RNA ligase 2 [335–338]. Subsequently, these species were isolated via ligation to an acceptor RNA, followed by gel purification. The acceptor RNA provided a new primer binding site for the following reverse transcription and PCR steps, thereby introducing a secondary selection pressure within a single selection round. The amplified species were then transcribed back into RNA for the next round of selection. The incubation duration was progressively reduced after each selection round to intensify the selection pressure. The procedure was repeated eight times, with progression monitored by various methods, including comparison of the bulk activity of sublibraries and melting curve analysis. Following the final round of selection, the PCR products were cloned into plasmids and subjected to Sanger sequencing. The sublibraries from rounds 5, 6, 7, and 8 were analysed using high-throughput sequencing, and the top 10 most enriched sequences were identified.

4.2 Selection procedure

All single-stranded DNA strands were ordered from Integrated DNA Technologies (IDT). The sequences of the oligonucleotides are provided in the Appendix A.1. Buffer and salt stocks were ordered from the Media Glassware Sterilisation Lab at the Department of Biochemistry, University of Oxford, unless otherwise noted. The water used in the experiments was purified using the Millipore Milli-Q laboratory water system.

4.2.1 Library preparation

The initial RNA library was transcribed from a double-stranded DNA (dsDNA) library, which was generated by filling in a single-stranded DNA (ssDNA) library and Primer1 containing a T7 promoter. A total of 2.5 nmol ($\approx 1.5 \times 10^{15}$ species) of the ssDNA library and 5 nmol of Primer1 were combined in a 10 mL PCR reaction, prepared with 1X OneTaq[®] Master Mix (NEB) in 96-well PCR plates (Eppendorf). The reaction was then subjected to the following programme in a thermal cycler (Eppendorf): 94°C for 2 minutes, 50°C for 2 minutes, and 68°C for 15 minutes. After filling in, the double-stranded DNA library was mixed with 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol. The mixture was incubated at -25°C for 2 hours and then centrifuged at 4°C for 1 hour at $13,000 \times g$. The supernatant was removed and the pellets were washed twice with cold 75% ethanol and then air dried for 20 minutes. All precipitated fill-in products were used as templates in a 5 mL in vitro transcription reaction (final concentration: 5,000 U/mL T7 RNA Polymerase (NEB), 1X T7 RNA polymerase buffer (NEB) supplemented with 25 mM MgCl₂, 15 mM DTT, and 6 mM of each NTP). The reaction was incubated at 37°C for 4 hours. Subsequently, TURBO[®] DNase (Invitrogen) was added (2 U per 2 µg of DNA) and incubated at 37°C for 1 hour to remove DNA templates. The transcripts were purified using Monarch[®] RNA Cleanup Columns (500 µg, NEB) following the manufacturer's instructions. RNA concentration was measured and calculated based on A₂₆₀ absorption (Agilent Cary uv-vis spectrophotometer), adjusted to 20 µM, and then stored at -25°C.

4.2.2 In vitro selection

Incubation

For the first round of selection, 15 nmoles of RNA transcripts were incubated in reaction buffer (Buffer1) containing 25 mM HEPES (pH 7.4), 300 mM KCl, 30 mM MgCl₂, and 5 mM MnCl₂. The buffer was freshly prepared using stock solutions of 1 M HEPES (pH 7.4) and 3 M KCl, ordered from the Media Glassware Sterilisation Lab of the Department of Biochemistry, and standard solutions of 1 M MgCl₂ and 1 M MnCl₂ from Sigma. This buffer was designed based on the aptamer buffer ^[270] which consists of 25 mM Tris (pH 7.8), 300 mM NaCl, and 30 mM MgCl₂. Tris in Buffer1 was replaced with HEPES due to its better temperature stability ^[339], and the pH was slightly reduced to a more physiological value (7.4). KCl replaced NaCl because it is particularly effective in stabilising functional structures, such as G-quadruplexes ^[340]. MnCl₂ was included because Mn²⁺, compared to Mg²⁺, can coordinate more ligands within the RNA molecule, including ribonucleotide oxygens and ring nitrogens of the bases ^[341]. In addition, several Mn²⁺-dependent ribozymes and DNAzymes have been reported ^[342, 343]. The incubation temperature was alternated between 25°C (30 minutes) and 37°C (30 minutes) for 6 hours to promote structural changes in each RNA molecule, and then left at 25°C for a total incubation time of 24 hours. The duration of incubation was gradually reduced, and the conditions for each round of selection are listed in the table below (Figure 4.3).

Round	1	2	3	4	5	6	7	8
Input RNA	15 nmol	5 nmol	2 nmol	1 nmol	1 nmol	600 pmol	60 pmol	60 pmol
Time	24 hrs	16 hrs	10 hrs	8 hrs	6 hrs	3 hrs	1.5 hrs	45 mins
Temp.	25°C, 37°C	25°C	25°C	25°C	25°C	25°C	25°C	25°C

Figure 4.3: Incubation conditions for each round of selection.

After incubation, the salts in the reaction were removed and the RNA was concentrated using Monarch[®] RNA Cleanup Columns (500 µg, NEB) following the manufacturer's instructions. RNA concentration was measured and calculated based on A₂₆₀ absorption.

Ligation and gel purification

The species (133 nt) that catalysed the conversion of 5' GTP to 5' GMP were fished out by ligation to another RNA strand, Acceptor1 (36 nt), containing a T7 promoter and a new primer binding site. The mixture of purified RNA, splint DNA Splint1, and Acceptor1 (library : Splint1 : Acceptor1 = 10 μ M : 15 μ M : 22.5 μ M) was heated in annealing buffer (25 mM Tris-HCl pH 7.4, 250 mM NaCl, 1 mM EDTA) at 65°C for 5 minutes, followed by cooling on ice for 10 minutes to form nicked structures. The annealing reaction was subsequently used to set up the ligation reaction as follows: T4 RNA Ligase 2 (NEB) 2.5 U/20 μ L, nicked RNA library 2 μ M, 1X T4 Rnl2 Reaction Buffer (NEB; 50 mM Tris-HCl pH 7.5, 2 mM MgCl₂, 1 mM DTT, 400 μ M ATP), MgCl₂ 8 mM. The ligation reaction was incubated at 25°C for 2 hours and at 16°C for 16 hours. The mixture was precipitated with ethanol, and the pellets were resuspended in a small amount of water.

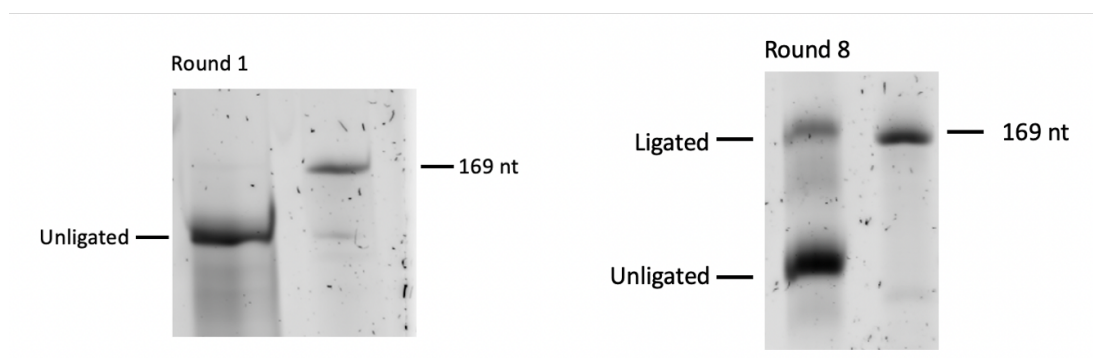


Figure 4.4: Gel images of ligation products from round 1 (left) and round 8 (right) of the selection process. The ligated products (169 nt) migrated to the same position as the 169-nt reference generated by IVT and were excised for further steps.

The resuspended oligonucleotides were run on 8% TBE-urea polyacrylamide gels (details of gel preparation, running conditions, and visualisation are presented in Appendix A.2). The bands corresponding to the ligated products (at the same position as the 169-nt reference RNA) were excised and then eluted in the elution buffer (1 mM EDTA, 0.3 M sodium acetate pH 5.2) through three cycles of freeze-thawing (-25°C for 30 minutes and 37°C for 30 minutes) and then left overnight at room temperature. Subsequently, the eluted RNA in the supernatant was precipitated with ethanol and resuspended in water as described in the above procedure. Figure 4.4 shows the gel images of the ligation reaction after round 1 and round 8 of selection, respectively, demonstrating a significant increase in the bulk activity of the library after 8 rounds of selection. The gel images after ligation

of other rounds are available in the Appendix A.3.

Reverse transcription and PCR

The ligation products were reverse transcribed using M-MuLV reverse transcriptase (NEB). RNA templates (10 μ M) and the primer Primer2 (20 μ M) in water were first denatured at 65°C for 5 minutes and then placed on ice for 2 minutes. The cooled solution was used to set up the reverse transcription reaction as follows: 1X M-MuLV Reverse Transcriptase Reaction Buffer (NEB; 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂, 10 mM DTT), 10 U/ μ L M-MuLV Reverse Transcriptase, 1 mM dNTP, 0.5 μ M RNA template, 1 μ M Primer2. The reaction was incubated at 42°C for 2 hours and then heat inactivated at 65°C for 20 minutes. The reaction mixture was precipitated with ethanol and resuspended in water for the following polymerase chain reaction (PCR).

The PCR reaction after the first round was set as follows: 1X OneTaq[®] Master Mix (NEB), 0.5 μ M Primer2 (from the reverse transcription reaction), 0.5 μ M Primer3, and around 0.25 μ M complementary DNA (cDNA). The reaction was subjected to the following programme in a thermal cycler: 94°C for 2 minutes, optimised cycles (94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 1 minute), followed by 68°C for 10 minutes.

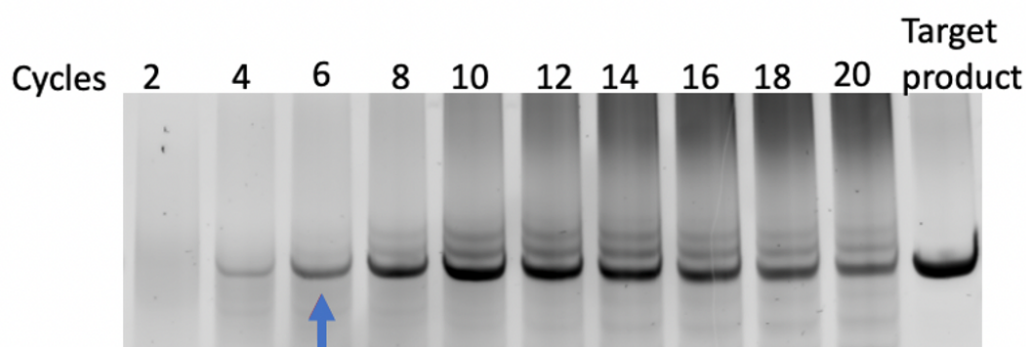


Figure 4.5: Optimisation of PCR cycle number.

To minimise the unavoidable bias of PCR ^[344–346], the number of PCR cycles was optimised by small-scale PCR reactions and kept minimal for each round of selection. In a test PCR reaction, an aliquot was taken after every two cycles of PCR, and the samples were run on a 15% native polyacrylamide gel along with a reference band of the same length as the target product. The sample that first showed a clear and single band determ-

ined the optimal cycle number. For example, Figure 4.5 shows the gel image of the test PCR for Round 4, where the optimal cycle number was determined to be 6. The PCR product was purified using the Monarch[®] PCR & DNA Cleanup Kit (NEB) following the manufacturer's instructions. The concentration of the PCR product was measured and calculated based on A₂₆₀ absorption.

Regeneration of RNA library

As acceptor RNA contains a T7 promoter at the 3' end, the species that were ligated, successfully reverse transcribed, and amplified to dsDNA could be regenerated and amplified by in vitro transcription. The T7 transcription reaction was set up as follows: 400 nM dsDNA, 5,000 U/mL T7 RNA Polymerase (NEB), 1X T7 RNA Polymerase buffer (NEB) supplemented with 25 mM MgCl₂, 15 mM DTT, and 6 mM of each NTP. The reaction was incubated at 37°C for 4 hours, treated with TURBO[®] DNase (Invitrogen), and purified as described in the library preparation section above. RNA concentration was measured and calculated based on A₂₆₀ absorption, adjusted to 20 µM, and then stored at -25°C.

4.3 Evaluation of selection progression

4.3.1 Input-to-Output ratio analysis across selection rounds

The ligation product bands on the gel were quantified using the Fiji software (details of the gel quantification are provided in the Appendix A.2), and the percentage of RNA that was successfully ligated was calculated. The RNA output/input ratios are plotted in Figure 4.6 showing the progression of RNA selection across eight rounds. Despite the gradually increasing stringency of the selection process, the percentage of surviving RNA increased five times after the first round of selection, and the percentage of surviving RNA consistently increased throughout the selection course (although there was a noticeable drop after the fourth round), eventually increasing from less than 1% in the first round to nearly 7% in the eighth round. This progression profile is typical and consistent with those reported in the literature [297, 347].

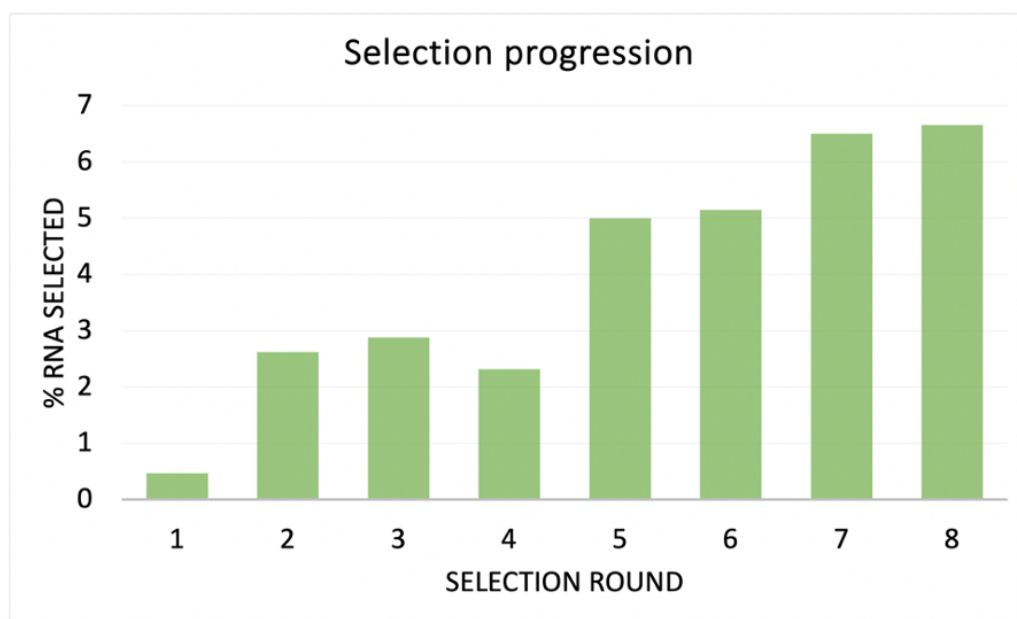


Figure 4.6: Progress of the selection over 8 rounds. Bars represent the percentage of RNA that was ligated after incubation.

4.3.2 Bulk ligation activity of sub-libraries

The bulk ligation activity of each regenerated sublibrary was assessed. In this experiment, T4 RNA Ligase 1 (NEB) was used instead of the T4 RNA Ligase 2 used during the selection process. The primary difference between the two ligases is that T4 RNA Ligase 2 requires a splint oligonucleotide with two regions complementary to the 3' end of the acceptor RNA and the 5' end of the library, respectively, forming a nicked structure. T4 RNA Ligase 2 operates exclusively on perfect nicks (without gaps or bulges), whereas T4 RNA Ligase 1 ligates a 5' monophosphate to any RNA with a 3' OH group. Figure 4.7a illustrates this difference. The use of T4 RNA Ligase 2 during the selection process simplified the gel purification, reverse transcription, and PCR steps, as all ligation products were of the same length. However, there was a risk of losing active species due to poor nick formation. In contrast, T4 RNA Ligase 1 increases the likelihood of an active strand being ligated, as there could be multiple pathways to ligation, leading to the production of several different ligation bands in the gel, as shown in Figure 4.7b.

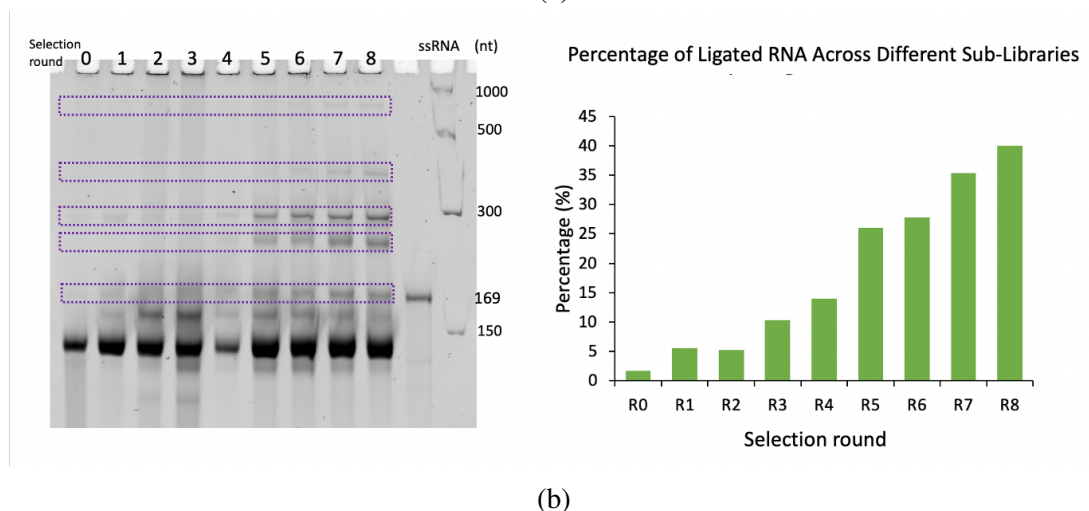
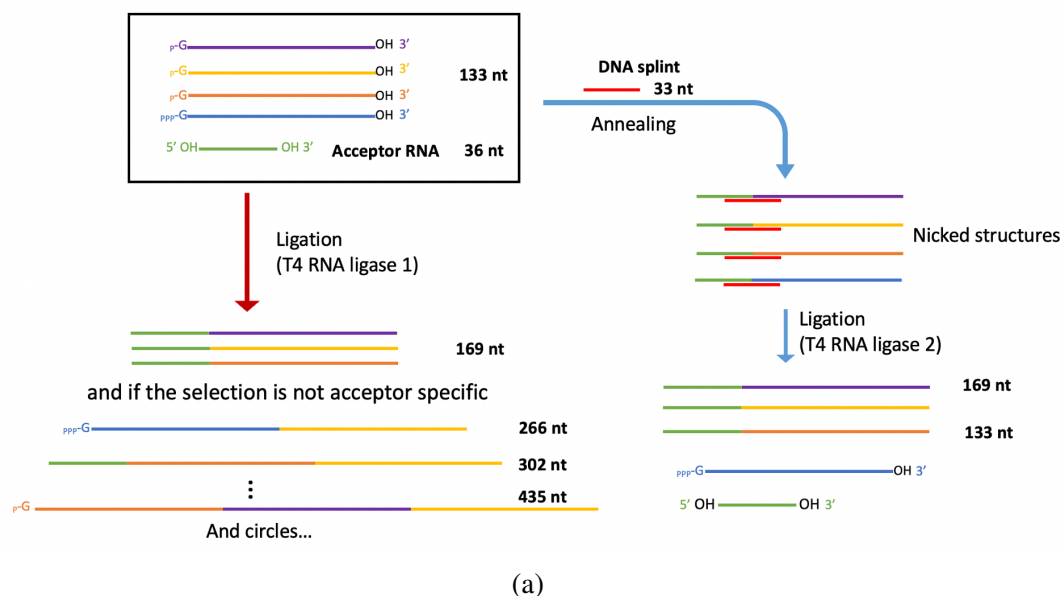


Figure 4.7: (a) Schematic representation of the different mechanisms of T4 RNA Ligase 1 and T4 RNA Ligase 2. T4 RNA Ligase 1 ligates any 3' OH to a 5' GMP, resulting in ligation products of varying lengths. In contrast, T4 RNA Ligase 2 only acts on perfect nicks, producing ligation products of uniform length. (b) Left: Gel image showing the ligation products of sub-libraries after incubation. Bands within the purple squares represent ligation products. Right: Plot illustrating the percentage of RNA ligated in each sub-library.

Incubation reactions were set up for each sub-library (200 μ L, 100 pmol library RNA, 25 mM HEPES pH 7.4, 300 mM KCl, 30 mM MgCl₂, 5 mM MnCl₂) and were incubated at 25°C for 1 hour. After this, the salts were removed and the RNA was concentrated using Monarch[®] RNA Cleanup Columns (10 μ g, NEB) following the manufacturer's instructions. The incubated sublibraries were then tested for ligation activity (10 μ L, 1X Reaction Buffer [50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 1 mM DTT], 10% DMSO, 10 units RNase Inhibitor [Murine; NEB], 5 units T4 RNA Ligase 1 [NEB], 1 mM ATP, 15% PEG 8000, 50 pmol Acceptor1, 5 pmol RNA library). The ligation reactions were incubated at 25°C for 2 hours and the products were visualised on an 8% denaturing urea-polyacrylamide gel.

The relative catalytic activity of each sublibrary is proportional to the ligation efficiencies of the incubated samples. Ignoring the contribution of the additional molecular weight from the adaptor in some ligation products, the apparent percentage of ligated RNA of each sample was calculated according to the equation below:

$$\text{Apparent percentage of ligated RNA (\%)} = \frac{\text{Intensity of ligated products}}{\text{Intensity of ligated products} + \text{Intensity of unligated material}} \times 100$$

As shown in the plot, the percentage of ligated RNA increases consistently over the selection rounds, indicating the enrichment of active species. Despite the use of the same acceptor strand throughout the 8 selection rounds, the appearance of multiple ligation products suggests that the selection was not biased by the acceptor sequence or the ligase.

4.3.3 Melting curve analysis of sublibraries

Melting Curve Analysis (MCA) following a brief reannealing step after PCR can provide valuable information on changes in sequence diversity changes within the library throughout the selection process [348, 349]. RT-PCR reactions were performed in a total volume of 20 μ L in optically clear strip tubes (Thermo Scientific). The reaction mixture included 1X Platinum[™] SuperFi[™] RT-PCR Master Mix (SuperScript[™] IV One-Step RT-PCR System, Invitrogen), 0.5 μ L SuperScript[™] IV RT Mix (SuperScript[™] IV One-Step RT-PCR

System, Invitrogen), 0.5 μM of Primer1, 0.5 μM of Primer2, , 10 pM RNA sub-library (approximately 10^7 RNA strands), and 1 μM EvaGreen® Fluorescent DNA dye (Jena Bioscience). The cycling conditions on the real-time PCR system (Agilent Technologies, Mx3005P) were as follows: reverse transcription at 50°C for 20 minutes, followed by an initial denaturation at 98°C for 2 minutes, then 30 cycles of PCR amplification comprising 10 seconds at 98°C , 15 seconds at 55°C , and 15 seconds at 72°C . After a final elongation step at 72°C for 5 minutes, the reactions were reannealed at 55°C for 15 seconds and then heated to 95°C at a rate of 0.1°C s^{-1} . shown in Figure 4.8. The peaks in the $-R'(T)$ plot (Figure 4.8) correspond to the temperatures at which the most rapid changes in fluorescence occur, indicating the points at which the DNA strands denature most rapidly. Given that all strands in the libraries contain constant regions, two distinct melting peaks can be expected when there is a certain degree of heterogeneity: one peak at a lower temperature, representing imperfectly formed heterogeneous duplexes, and another at a higher temperature, representing more homogeneous duplexes ^[350]. The relative height of these two peaks reflects the proportion of heterogeneous versus homogeneous duplexes in the pool. To facilitate comparison, the melting curves of each sublibrary (red) are plotted (Figure 4.8) alongside the starting library (blue). The starting library exhibited a prominent peak (A) at around 78°C and a smaller peak (B) approximately at 83°C . In a progressively enriched library, the dominant peak is expected to gradually shift to a higher temperature as selection advances. However, in this experiment, the ratio of the heights of the two peaks (F_B/F_A) did not increase consistently over the selection rounds, and the final library still showed a relatively stronger peak (A), indicating a significant proportion of heteroduplexes. This unexpected result may be attributed to overamplification during PCR, leading to the introduction of artifacts when the primers are depleted and the concentration of the PCR product is high. Despite this, previous bulk ligation activity experiments have provided evidence for the progression of the selection process.

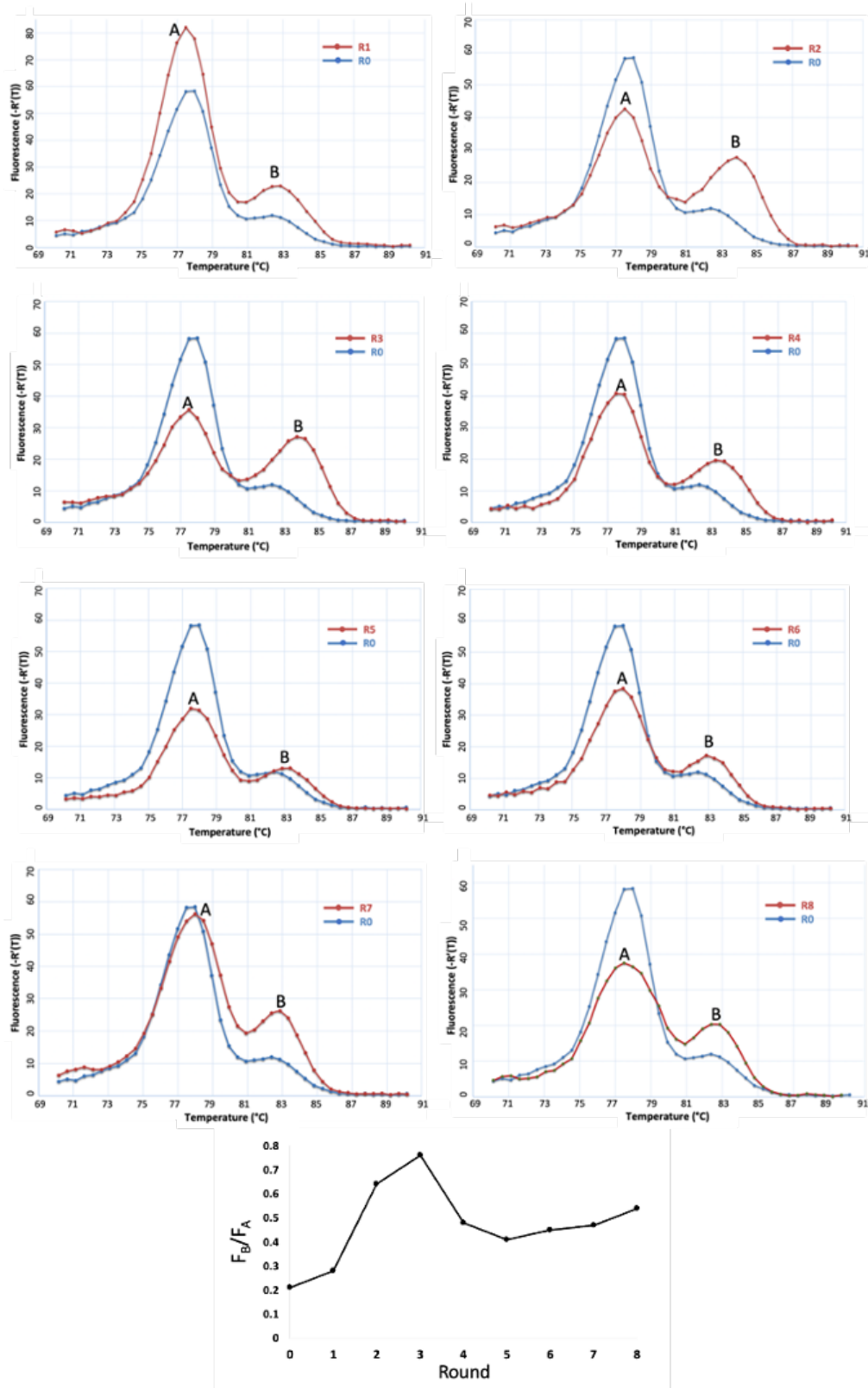


Figure 4.8: Melting curve analysis of sublibraries. Upper panel: Melting curves for the starting library (blue curve) and individual sub-libraries (red curves). $-R'(T)$ represents the negative derivative of fluorescence (R) with respect to temperature (T), which measures the rate of change in fluorescence as the temperature increases. Two distinct peaks indicate the melting of heterogeneous (lower temperature) and homogeneous (higher temperature) duplexes. Lower Panel: A plot of the ratio of the peak heights (F_B/F_A) in the melting curves of sub-libraries across selection rounds, illustrating changes in sequence diversity. The increase or decrease in (F_B/F_A) indicates the increase or decrease in the proportion of homogeneous versus heterogeneous sequences in the library.

4.4 Sanger sequencing

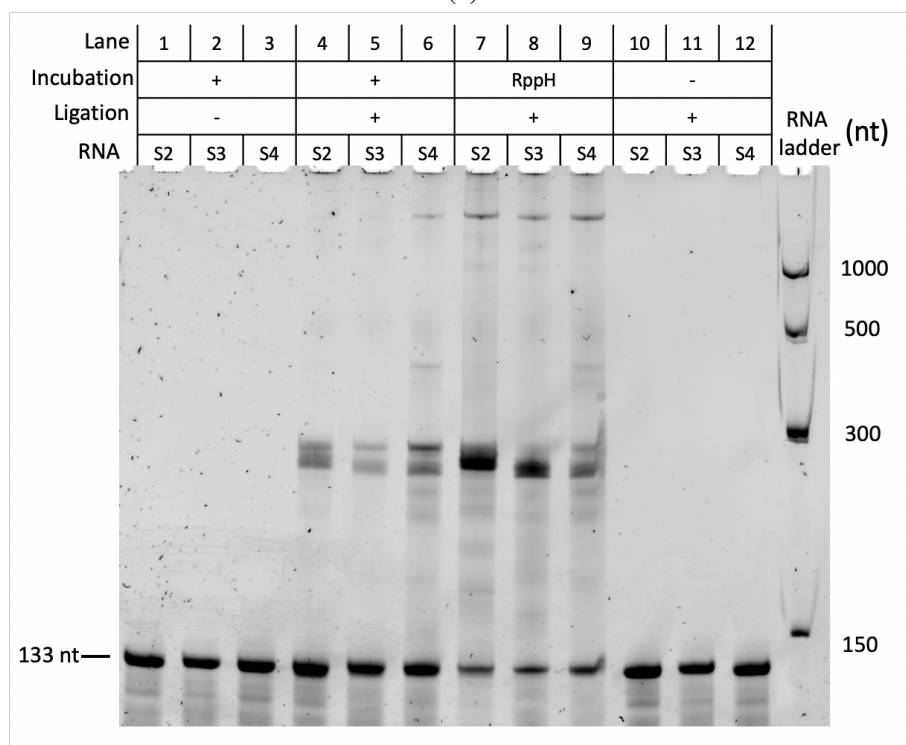
4.4.1 Cloning and sequencing

The sublibrary following the eighth round of selection was reverse transcribed, PCR amplified, and subsequently inserted into a vector using the Zero Blunt™ TOPO™ PCR Cloning Kit for Sanger sequencing (Invitrogen). 2.4 pmol RNA from the eighth sublibrary was reverse transcribed in a 30 µl reaction under conditions described previously. Subsequently, 2 µl of the reverse transcription reaction was used for a 40 µl PCR reaction (1X Q5® High-Fidelity Master Mix (NEB), 0.5 µM Primer2, and 0.5 µM Primer3) to determine the optimal cycle number. Aliquots of 5 µl were taken after every 2 cycles and analysed on a native 15% polyacrylamide gel. The aliquot after 8 cycles of PCR, displaying a strong and sharp band at the expected length, was identified as the optimal cycle number. Following this, an 80 µl PCR reaction was performed using 4 µl of the reverse transcription product. The PCR conditions were as follows: initial denaturation at 98°C for 30 seconds, followed by 8 cycles of 10 seconds at 98°C, 20 seconds at 64°C and 20 seconds at 70°C. The resulting PCR product was purified using Monarch® DNA Cleanup Columns (NEB) according to the manufacturer's instructions. The concentration of the purified PCR product was quantified based on its A₂₆₀ absorption.

50 ng of the purified PCR product was inserted into M13 vectors using the Zero Blunt™ TOPO™ PCR Cloning Kit for Sequencing (Invitrogen) in a 5 µl reaction, and 2 µl of the ligation reaction was transformed into One Shot™ TOP10 Chemically Competent *E. coli* (Invitrogen) following the manufacturer's instructions. The transformed cells in serial dilutions were plated in 6 LB agar petri dishes containing kanamycin and incubated for 16 hours at 37°C. After incubation, 10 colonies were randomly selected and cultured overnight at 37°C in 10 Corning® Falcon® round bottom tubes, each containing 2 ml LB medium supplemented with 50 µg/ml kanamycin. The plasmid DNA was subsequently extracted using the QIAprep Spin Miniprep Kit (QIAGEN) following the provided protocol. The purified plasmid DNA samples were submitted for Sanger sequencing at Source BioScience.

	Domain1	Domain2
S1	5' -AAAGUUA CGGUCC CGCC UGA -	-UCACGAAGAACGCCCC CAUGU -
S2	5' -UGUUUUU UGAGCACCGGUCC U -	-CUUUCUAAGCA UUCAGCUUC -
S3	5' - CAGUGUACUAACUCGAAAGC -	-CUGACUCGAAGAAUAUAACG-
S4	5' - UCAGUACUAACUCGAGAAAC -	-CUACCCAAUU UUCACUAAAGU U -
S5	5' -UCG GUACUAACUCGAAU UAC-	-AGUGCGUGACACUAGAAUGU A -
S6	5' - CAGUACUAACUCGACGUUCC -	-ACCUCUCGAAACAC CAUGUAGU U -

(a)



(b)

Figure 4.9: (a) Six sequences obtained from Sanger sequencing, with common motifs coloured, and insertions highlighted in red. Only the random domains are displayed. (b) Activity analysis of three sequences.

Sequencing data were analysed using the Benchling platform. Among the 10 samples, 6 were determined to be of high quality, despite some containing insertions. The sequences corresponding to their random regions are presented in Figure 4.9a. Despite the extremely small sample size, several motifs (highlighted in colour) were identified, particularly within random domain 1 (Domain1).

4.4.2 Activity analysis

Three representative sequences (S2, S3, S4) were synthesised by PCR and in vitro transcription (IVT) and subsequently their activity was tested. Synthesis and purification were performed according to the previously described protocols. Incubation reactions were set for each sequence in a 200 μ L mixture containing 100 pmol of RNA, 25 mM HEPES (pH 7.4), 300 mM KCl, 30 mM MgCl₂, and 5 mM MnCl₂. The reactions were incubated at 25°C for 1 hour. In parallel, 50 pmol of RNA was treated with RNA 5' pyrophosphohydrolase (RppH, NEB) to remove pyrophosphate from the 5' end of triphosphorylated RNA, producing 5' monophosphate RNA. The RppH reaction was carried out in 1x NEB-uffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM DTT) with 10 units of RppH and incubated at 37°C for 1 hour. Following treatment, RNA samples were purified and concentrated using Monarch[®] RNA Cleanup Columns (10 μ g, NEB) according to the manufacturer's instructions. RNA concentrations were quantified by measuring A₂₆₀ absorption. The purified and treated RNA samples were then tested for ligation activity using T4 RNA Ligase 1 (NEB). Ligation reactions (10 μ L) were prepared in 1x Reaction Buffer (50 mM Tris-HCl, pH 7.8, 10 mM MgCl₂, 1 mM DTT) with 10% DMSO, 10 units of RNase Inhibitor (Murine; NEB), 5 units of T4 RNA Ligase 1 (NEB), 1 mM ATP, 15% PEG 8000, and 5 pmol of treated RNA. The reactions were incubated at 25°C for 2 hours and the products were analysed by electrophoresis on an 8% denaturing urea-polyacrylamide gel.

As shown in the gel image (Figure 4.9b), the three sequences generated less intense but similar ligation patterns (bands above the original 133-nt length) after incubation in selection buffer (lanes 2, 5, and 6) compared to samples treated with RppH (lanes 7, 8, and 9). In contrast, samples without any treatment before ligation (lanes 10, 11, and 12)

did not show ligation activity. These observations suggest that the three sequences function as ribozymes, performing the same catalytic function as the protein enzyme RppH: removing pyrophosphate from 5' GTP and leaving a monophosphate group suitable for ligation. Furthermore, the catalysed reaction occurred in the selection buffer. Lanes 1, 2, and 3 contain samples loaded after incubation in selection buffer and subsequent column purification. The absence of noticeable ligation activity in these samples indicates that the ligation reaction was not self-catalysed by these sequences in the selection buffer.

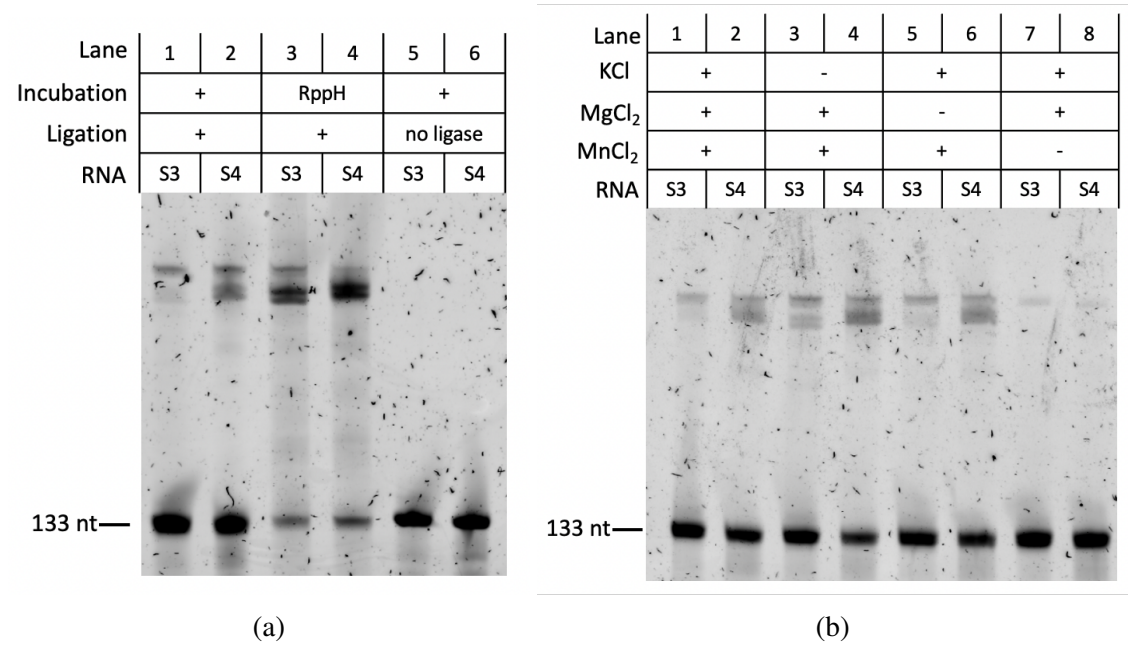


Figure 4.10: (a) Ligation activity analysis of S3 and S4 sequences in various ligation conditions. (b) Ligation activity analysis of S3 and S4 sequences in various buffer conditions.

To further confirm that the selection was not biased towards ligases, another experiment was carried out with three groups of tests using the sequences S3 and S4 (Figure 4.10a). Lane 1, lane 2, lane 3, and lane 4 correspond to samples treated in the same manner as lanes 5, 6, 8, and 9, respectively, in Figure 4.9b. Lanes 5 and 6 were incubated in selection buffer, purified, and subjected to the ligation reaction; however, no ligase was added. The absence of ligated bands in these lanes suggests that the ligation reaction was not self-catalysed by these sequences in the ligation buffer either.

All observations so far indicate that the three sequences catalyse the conversion of 5' GTP to 5' GMP, and this reaction occurs exclusively in the selection buffer. To identify the essential components of the selection buffer for catalysis, an additional experiment was conducted with samples S3 and S4. Each sample was incubated in buffers in which

one component was omitted. The ligation activity of each sample was assessed by gel electrophoresis (Figure 4.10b). Notably, the two samples incubated without MnCl_2 had a greatly decreased ligation activity compared to the other samples.

4.5 High-throughput sequencing

To better understand the selection process, sublibraries from the final four rounds were analysed using high-throughput sequencing (HTS) by MiSeq (Illumina).

4.5.1 Preparation of sequencing templates

1 pmol (approximately 6×10^{11} molecules) of RNA from each sub-library, after 5, 6, 7, and 8 rounds of selection, was reverse transcribed and PCR amplified using the SuperScript™ III One-Step RT-PCR System with Platinum™ Taq DNA Polymerase (Invitrogen, Thermo Scientific) in 50 μL reactions. Each reaction contained 1X reaction mix, 2 μL SuperScript™ III RT/Platinum™ Taq Mix, 0.5 μM forward primer (qianF), and 0.5 μM reverse primer (qianR). The thermocycler was programmed as follows: reverse transcription at 50°C for 45 minutes, initial denaturation at 94°C for 2 minutes, followed by 9 cycles of 94°C for 20 seconds, 52°C for 20 seconds, and 68°C for 1 minute, with a final extension at 68°C for 5 minutes. The PCR products were run on a 4% agarose gel (150 V for 1 hour), and the correct bands were excised and extracted using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions. The concentration of the purified PCR products was measured using a Nanodrop spectrophotometer.

Subsequently, 1 pmol of the purified PCR products was used for a second PCR to add Illumina adapters and barcodes. This was carried out in 50 μL reactions containing 1X Green GoTaq® Flexi buffer, 2 mM MgCl_2 , 0.2 mM of each dNTP, 0.25 μL GoTaq® Hot Start Polymerase, 0.5 μM primer P7, and 0.5 μM primer P5 with barcodes. Thermocycler conditions were as follows: 95°C for 2 minutes, followed by 12 cycles of 95°C for 30 seconds, 52°C for 30 seconds, and 72°C for 1 minute, with a final extension at 72°C for 5 minutes. The products were then purified and quantified as described above.

Each individual library member was diluted to approximately 100 nM, as measured

using a Nanodrop spectrophotometer. To optimise reagent usage, 0.4 pmol of my bar-coded sample was combined with samples from other researchers to form a pooled library totalling 4 pmol at an approximate concentration of 10 nM. The more accurate concentration of the pooled solution was then determined using a Qubit fluorometer (Thermo Scientific). A 1.6 pM sequencing solution was prepared, into which 0.48 μ L of 10 nM PhiX control (Illumina) was added to comprise 15% of the total molar content. Subsequently, 5 μ L of this mixture was combined with 5 μ L of 0.2 N NaOH and incubated at room temperature for 5 minutes to denature the DNA. After incubation, 990 μ L of chilled HT1 hybridisation buffer (Illumina) was added. Finally, 900 μ L of the resulting solution was loaded into the reagent cartridge (MiSeq Reagent Micro Kit v2, Illumina) for sequencing.

4.5.2 Sequencing data analysis

Raw FASTQ reads were first processed using BBDuk and then converted to FASTA format using Reformat.sh from the BBTools suite. The sequences were subsequently demultiplexed based on barcode sequences provided in a separate barcode file. The resulting demultiplexed FASTA files were analysed using a custom Python script, which is included in Appendix A.4 for reference.

Sequences which were identical in each sublibrary were grouped into "clusters". Then these clusters were divided into three categories according to size: those with more than 50 sequences, those containing between 5 and 50 sequences, and those with fewer than 5 sequences. The clusters in each of the size categories were calculated as a percentage of sublibraries as shown in Figure 4.11a. In the earlier rounds (R5 and R6), most of the clusters consisted of less than 5 sequences (blue bars). As selection progressed, a rise in mid-sized clusters (5–50 sequences, orange bars) was observed in R6 and R7, with larger clusters (>50 sequences, green bars) emerging in R7 and R8. The increasing dominance of larger clusters in later rounds suggests an enrichment of specific sequence populations over successive rounds. When the two domains are analysed individually (Figure 4.11b and Figure 4.11c), Domain1 (the random region close to the 5' end) shows a more significant increase in larger clusters (>50 sequences) than Domain2 (the random region close to the 3' end), where most clusters remain small or mid-sized throughout

the selection process. This indicates that Domain1 contributes more significantly to the enrichment of sequences.

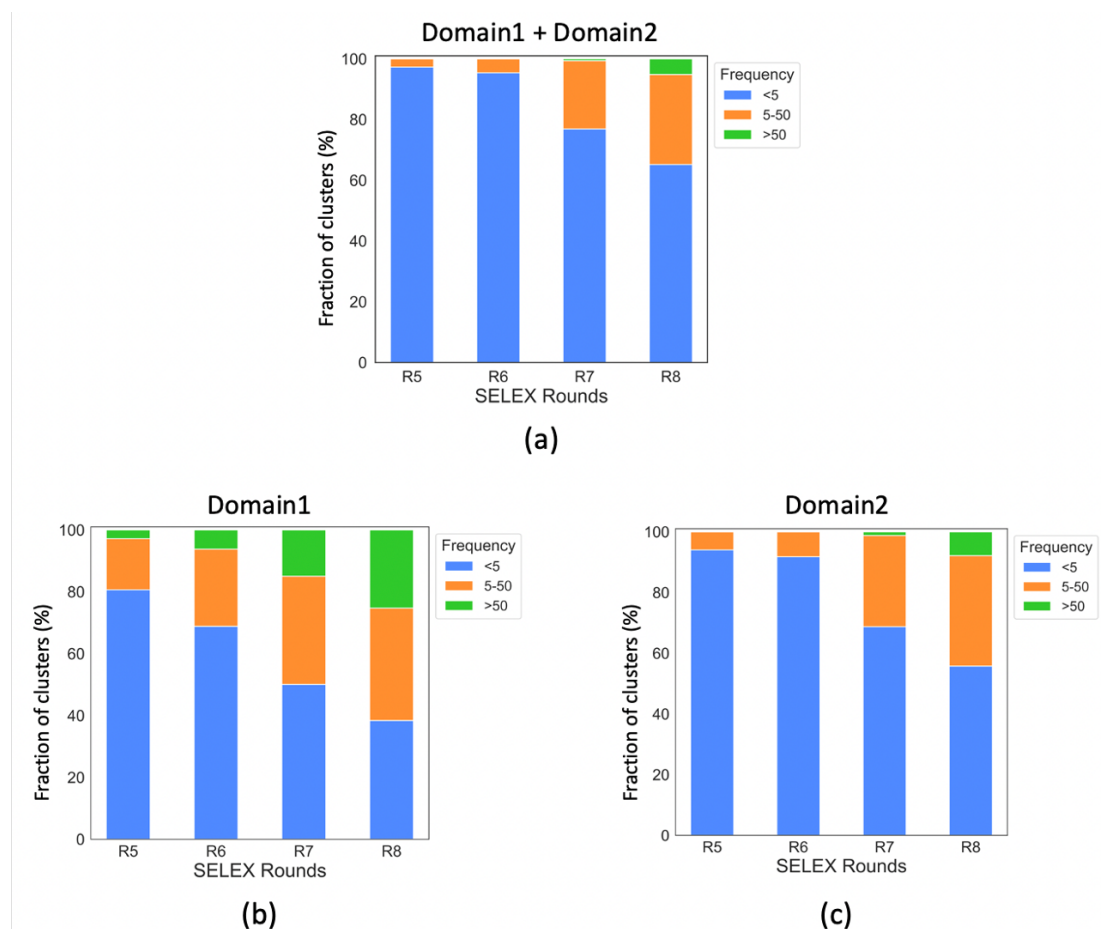


Figure 4.11: Cluster size distribution across selection rounds 5 to 8. (a) Cluster size distribution for both random regions (Domain1 and Domain2 combined). (b) and (c) depict the individual contributions of Domain1 and Domain2, respectively.

Another approach to evaluate the sequence complexity of a library is through k -mer-based analysis [351–353]. A k -mer is a contiguous nucleotide subsequence of length k . In this study, k -mers of three specific lengths—10, 14, and 18 nucleotides—were extracted from the two random regions (Random1 and Random2) within each sequence of the sublibrary after the final round (round 8) of selection. This extraction was performed by sliding a window of length k across the sequences, generating all subsequences of that specific length. The prevalence of each k -mer was determined as the percentage of valid sequencing reads containing that k -mer. The highest prevalence observed for each k -mer length is represented in the bar chart in Figure 4.12a. Both random regions were independently analysed, and their respective k -mer distributions are plotted side by side for

direct comparison. The prevalence of the most enriched k -mers increased as k -mer length decreased in both regions. In particular, the prevalence of the most enriched 10-, 14-, and 18-mers was consistently higher in Random1 than in Random2, suggesting greater sequence convergence and homogeneity in Domain1 than in Domain2.

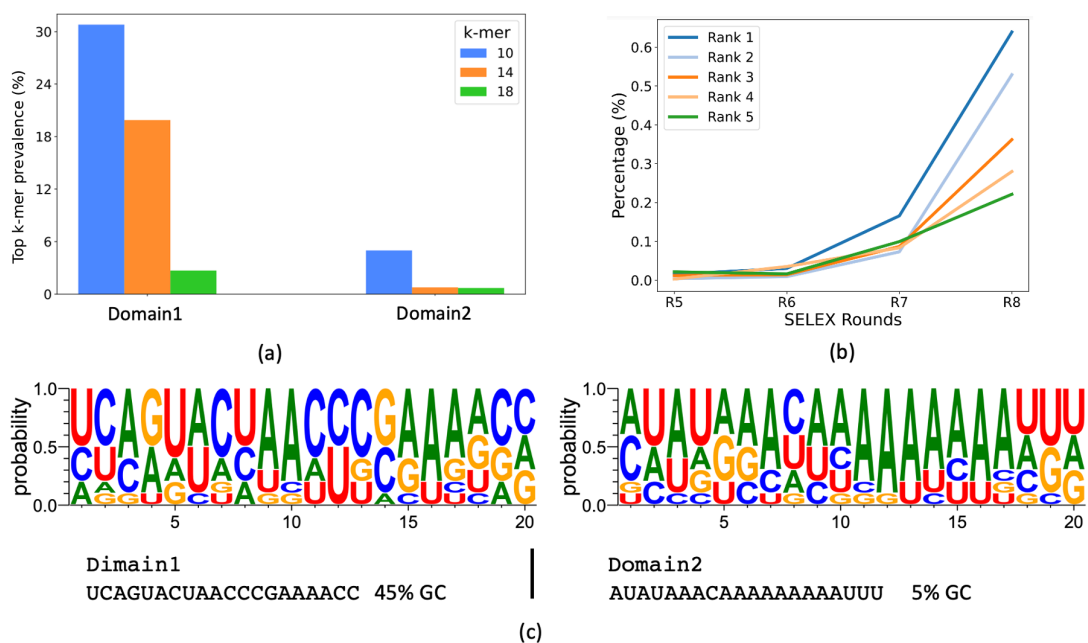


Figure 4.12: (a) k -mer analysis of sequence complexity. (b) Enrichment profile of the top 5 sequences across selection rounds. (c) Sequence Logo of the top 10 Sequences.

The line graph in Figure 4.12b illustrates the enrichment profile of the top five most abundant sequences across selection rounds 5 to 8. In rounds 5 and 6, all five sequences display relatively low abundance. However, a significant increase is observed between rounds 6 and 7, with rapid enrichment continuing into rounds 7 and 8. The most enriched sequence (Rank 1) demonstrates the highest increase, reaching more than 0.6% of the total population by round 8. The remaining four sequences also show substantial enrichment, each exceeding 0.2% by the final round.

The top 10 sequences from the final sublibrary were extracted and used to generate sequence logos (Figure 4.12c for each random region using WebLogo. Each logo consists of stacks of letters, with one stack representing each position in the sequence. The height of each letter within a stack reflects the probability that a specific nucleotide appears at that specific position ^[354]. In Domain1, the sequence logo reveals a relatively balanced nucleotide composition, with a GC content of 45% in the sequence composed of the

most frequent nucleotide at each position. This indicates a certain level of structural stability due to the presence of GC base pairs. In contrast, the sequence logo for Domain2 demonstrates a strong enrichment of A and U nucleotides at almost all positions, leading to a GC content of only 5%. This suggests that selection in Domain2 favoured sequences that are predominantly A/U-rich. This composition likely contributes to weaker secondary structure formation, as the AU base pairs are less stable than the GC base pairs. The clear difference in nucleotide distributions between the two domains suggests different selection pressures or functional requirements.

4.5.3 Activity analysis of the top 10 sequences

The relative catalytic activity of the top 10 sequences from the final sub-library was compared. Since the aptamer scaffold of the initial library has the highest affinity for ATP, sequences beginning with 5' ATP were also synthesised and tested. The 20 sequences were synthesised via PCR and IVT as described above. For transcription of sequences starting with ATP, the T7 class III promoter $\phi 6.5$ (TAATACGACTCACTATA) in the DNA of the transcription template was replaced with the T7 $\phi 2.5$ promoter (TAATACGACTCACTATT) using Primer4 [355]. Incubation reactions were prepared for each sequence in a 200 μ L mixture containing 100 pmol of RNA and either the selection buffer (Buffer 1) or the original aptamer buffer with 5 mM $MnCl_2$ (Buffer 2). The reactions were incubated at 25°C for 1 hour. Incubated samples were purified and concentrated using Monarch® RNA Cleanup Columns (10 μ g, NEB) following the manufacturer's instructions and quantified. The treated RNA samples were then tested for ligation activity using T4 RNA Ligase 1 (NEB) in 10 μ L ligation reactions containing 50 pmol of Acceptor1 and 5 pmol of treated RNA. The reactions were incubated at 25°C for 2 hours, and the products were analysed by gel electrophoresis on an 8% denaturing urea-polyacrylamide gel.

The two gel images in Figure 4.13a show the ligation products of the top 10 sequences with a 5' GTP, incubated in selection buffer (left) and the original aptamer buffer with 5 mM $MnCl_2$ (right), as $MnCl_2$ has been shown to be essential for ribozyme activity as demonstrated in Figure 4.10b. Similarly, the two gel images in Figure 4.13b show the ligation products of the top 10 sequences with a 5' ATP, incubated in selection buffer

(left) and the original aptamer buffer with 5 mM MnCl_2 (right). The bands at the bottom correspond to the unligated RNA (133 nt), while all bands above represent the ligation products. As illustrated in Figure 4.7a, multiple products of different lengths can be generated using T4 RNA ligase 1.

The apparent ligation efficiency of each sample was calculated as previously described. As shown in the bar chart (Figure 4.13c), it is unexpected that the most enriched sequence (H1) did not exhibit the highest ribozyme activity. In addition, the second most enriched sequence (H2) demonstrated background activity similar to that of the random library. H1 displayed similar activity with both 5' GTP and 5' ATP, while the other sequences exhibited significantly higher activity with 5' GTP. Among the two incubation conditions, Buffer2 yielded higher apparent ligation efficiencies for all samples, with the exception of the generally less active samples (the library, H2, and H6). The two buffers differ in pH and monovalent ion composition. Buffer1 contains 25 mM HEPES at pH 7.4 and 300 mM KCl, while Buffer2 contains 25 mM Tris at pH 7.8 and 300 mM NaCl. Based on the ligation results (Figure 4.10b) of a sample incubated in Buffer1 without KCl, it appears that KCl does not significantly affect ligation activity. Therefore, the observed increase in activity is more likely attributed to the higher pH in Buffer2.

The sequences were subsequently re-ranked based on their apparent ligation efficiencies with 5' GTP after incubation in Buffer2. As shown in Figure 4.14a, the tenth most enriched sequence (H10) emerged as the most active among the top ten sequences, while the most enriched sequence (H1) was re-ranked to seventh place. The sequences ranked second (H2) and sixth (H6) were the least active, with H2 showing activity comparable to that of the random library (Figure 4.13c).

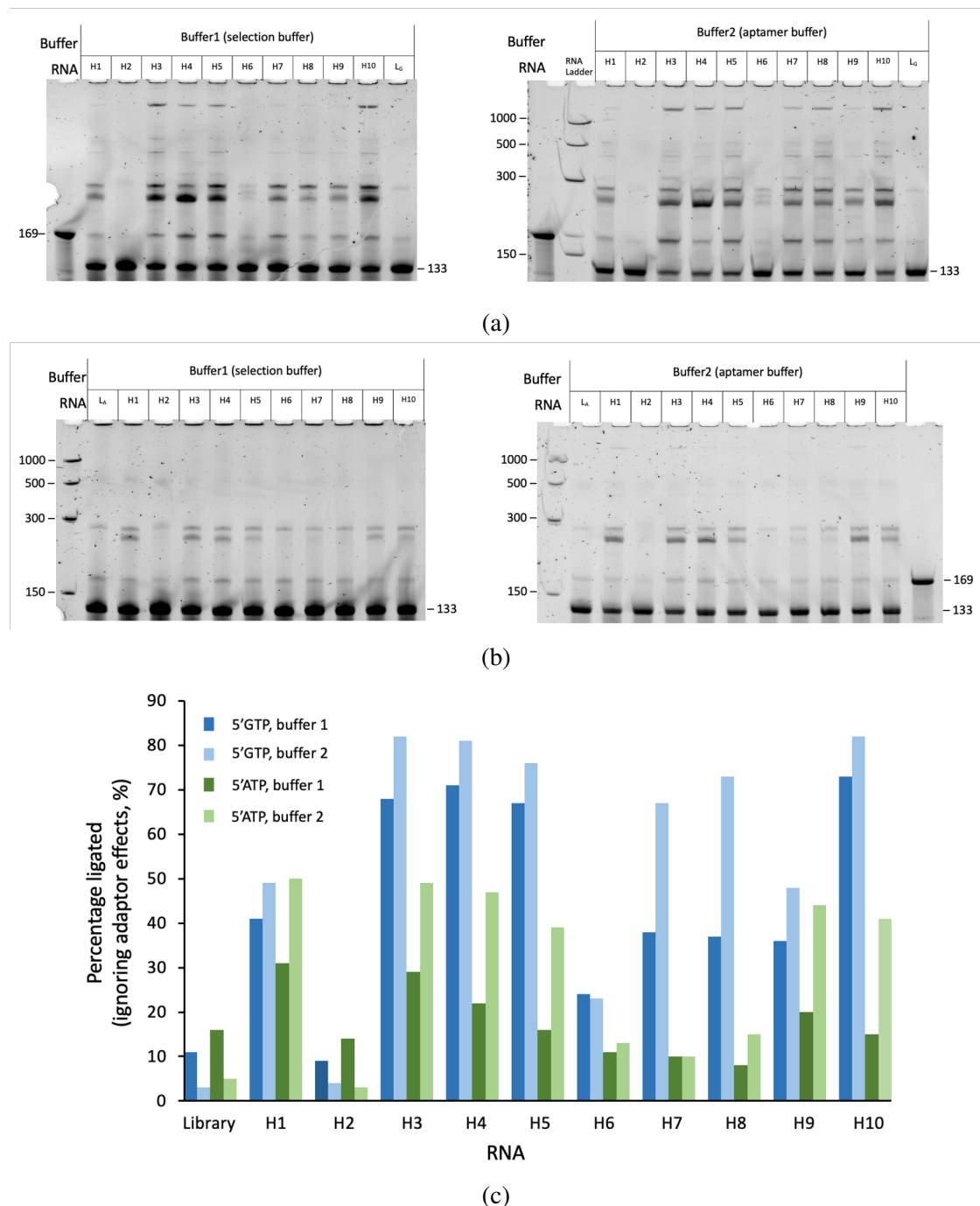


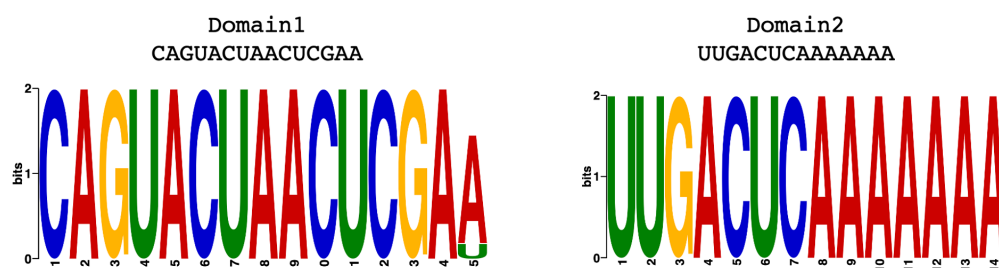
Figure 4.13: (a) Ligation activity analysis of the top 10 sequences with 5' GTP from the final sub-library after incubation in Buffer1 (left) and Buffer2 (right). Buffer1 consists of 25 mM HEPES (pH 7.4), 300 mM KCl, 30 mM MgCl₂, and 5 mM MnCl₂; Buffer2 consists of 25 mM Tris (pH 7.8), 300 mM NaCl, 30 mM MgCl₂, and 5 mM MnCl₂. (b) Ligation activity analysis of the top 10 sequences with 5' ATP from the final sub-library after incubation in Buffer1 (left) and Buffer2 (right). (c) Bar chart of the apparent ligation efficiency of each sequence in different incubation buffers.

The 10 sequences were analysed using the MEME Suite for motif discovery, and two sequence logos (Figure 4.14b) were generated corresponding to motifs in Domain 1 and Domain 2, respectively. The sequence logos visually represent the conservation at each position and the frequency of each nucleotide. The total height of the stack at each position reflects the level of conservation, with higher stacks indicating greater conservation. The height of each individual letter within the stack is proportional to the frequency of that nucleotide at the given position. Both motifs are highly conserved across the sequences, appearing in all except H6 and H2, which rank lowest in activity. This suggests that these motifs are essential contributors to ribozyme activity. The motif in Domain 1 was also found in sequences obtained by Sanger sequencing, while the motif in Domain 2 remains low in GC content (approximately 20%).

Given the high enrichment of H2 and H6 in the library, an additional test was conducted to rule out the possibility of ligation bias contributing to the low levels of ligation products. H2, H6, the starting library, and two highly active sequences (H10 and H4) were tested. Each was treated by incubation in Buffer2 or with the RppH enzyme before being used as substrates for ligation. The ligation products were visualised using gel electrophoresis. As shown in Figure 4.15, the bottom bands represent the unligated material (133 nt), and the above bands correspond to the ligation products. Following RppH treatment, all samples produced strong bands corresponding to ligation products, indicating that there was no significant ligation bias due to sequence or secondary structure. H2 (lane 2), H6 (lane 11), and the starting library (lane 8 and lane 17) displayed undetectable or very low levels of ribozyme activity, while H4 and H10 demonstrated activity for ligation comparable to RppH.

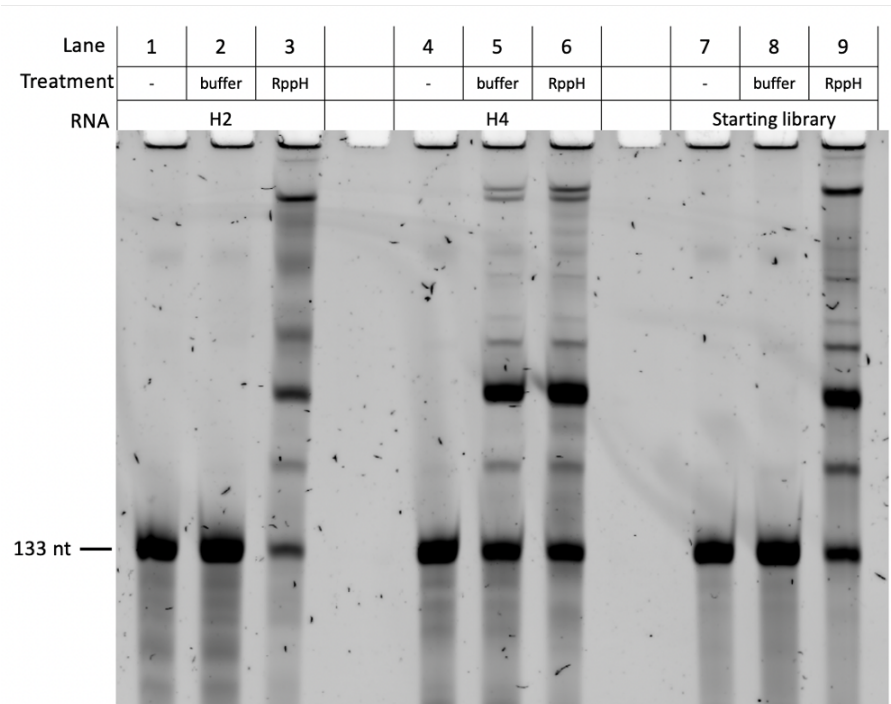
Activity Rank	HTS Rank	Name	Domain1	Domain2
1	10	H10	–U CAGUACUAACUCGAU UUAC–	–CUAGAGAAAAAACUUUUCA–
2	3	H3	–U CAGUACUAACUCGAA AGCG–	–GUAAGCAAAUAAAUUACAGG–
3	4	H4	–U CAGUACUAACUCGAA GACG–	–AAAAAACAAAGUACUAGUA–
4	5	H5	–U CAGUACUAACUCGAA AGCA–	–CA UUGACUC AAAAAAAAAUGU–
5	8	H8	– CAGUACUAACUCGAA AGCAC–	–A UUGACUC AAAAAAAAAUGU–
6	7	H7	–U CAGUACUAACUCGAA AGCA–	–CA UUGACUC AAAAAAAAAUG–
7	1	H1	–CU CAGUACUAACUCGAA AGC–	–ACA UUGACUC AAAAAAAAAUGU–
8	9	H9	–CU CAGUACUAACUCGAA AGC–	–ACA UUGACUC AAAAAAAAAUG–
9	6	H6	–AUCAUUCAAAUCCCGUAUGA–	–UUACGAAUAGGAAAAACUA–
10	2	H2	–AGAAAU GCGGCCGCCACACG –	–CUCUAAAGUUCAUUCAGCUU–

(a)

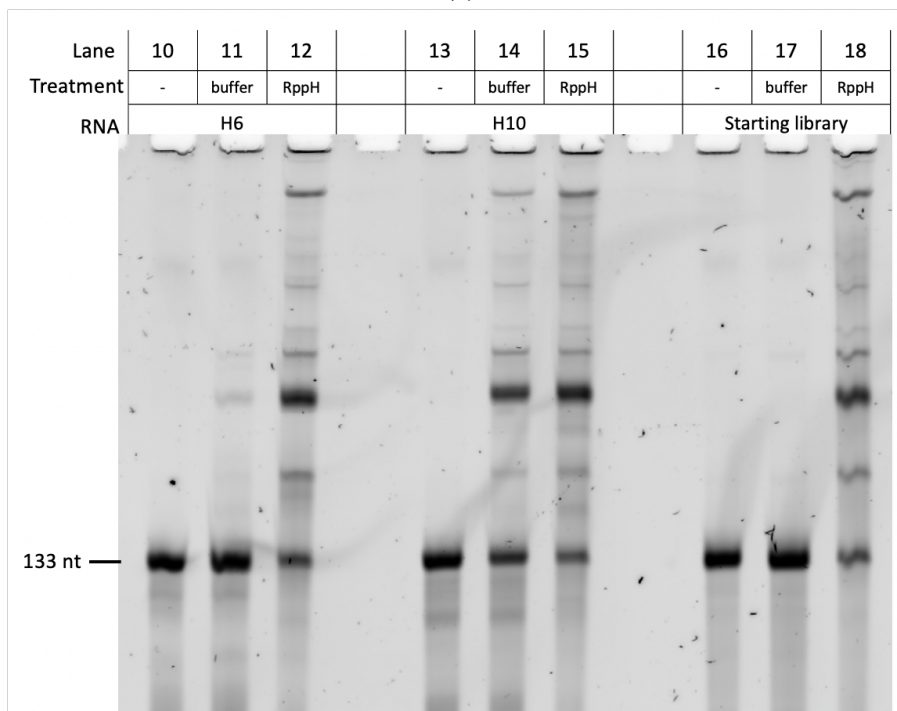


(b)

Figure 4.14: (a) The top 10 most enriched sequences from the final sub-library, with common motifs highlighted in colour. Only the two randomised domains are shown. (b) Sequence logos representing the motifs in Domain1 and Domain2. The height of each letter in a stack is proportional to the sequence conservation (measured in bits) at that position. Bit (also known as shannon) is a unit of information content, and for nucleic acid, the maximum value is 2 [356].

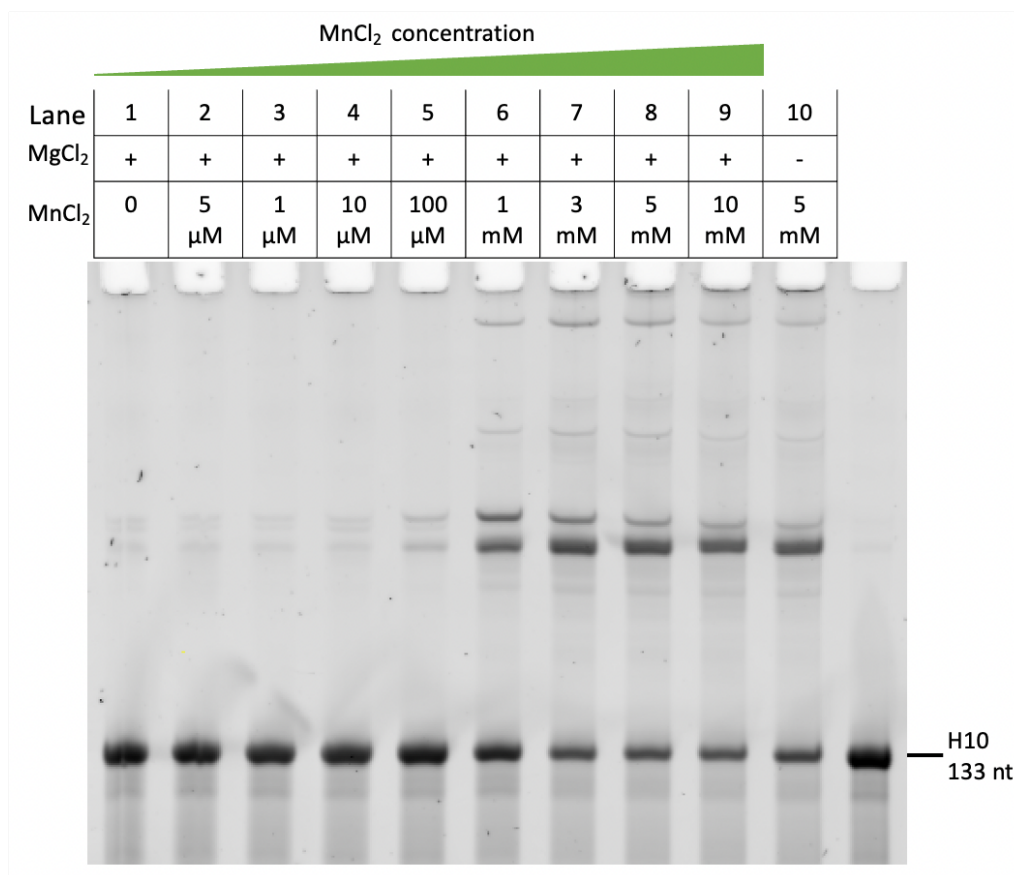


(a)

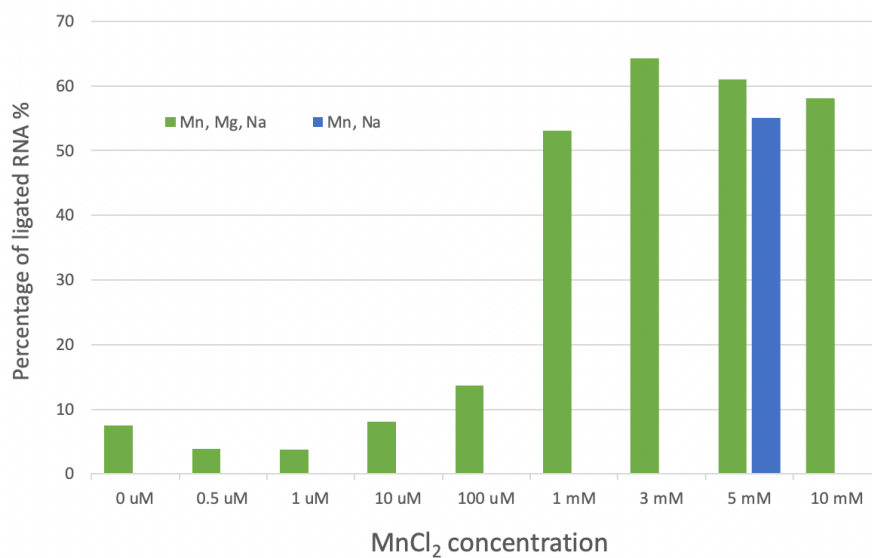


(b)

Figure 4.15: (a) Ligation activity analysis of H2, H4, and the starting library after incubation in Buffer2 or with RppH. (b) Ligation activity analysis of H6, H10, and the starting library after incubation in Buffer2 or with RppH. H2 (lane 2), H6 (lane 11), and the starting library (lanes 8 and 17) exhibited either undetectable or minimal ribozyme activity. In contrast, H4 and H10 demonstrated ligation activities of 0.75 and 0.83, respectively, compared to RppH.



(a)



(b)

Figure 4.16: (a) Ligation activity analysis of H10 in Buffer1 containing different concentrations of MnCl₂. (b) Bar chart of the percentage of ligated RNA of each sample.

Since Mn^{2+} was shown to be important for ribozyme activity, as demonstrated in Figure 4.10b, it was worthwhile to investigate how its concentration influences activity. H10 was incubated for 1 hour at 25°C in a series of Buffer1 solutions containing MnCl_2 concentrations ranging from 0 to 10 mM. The incubated and purified samples were then used as substrates for ligation using T4 RNA ligase 1, following the same reaction conditions as previously described but without the adaptor. All bands above the bottom band at 133 nt (Figure 4.16a) were considered ligation products. The percentage of ligated RNA of each sample was calculated on the basis of band intensity using the same method described above. The results are presented in the bar chart (Figure 4.16b) below. As shown in the chart, the samples exhibited background-level ligation activity when the MnCl_2 concentration was between 0 and 100 μM . A dramatic increase in ligation products was observed as the MnCl_2 concentration increased from 100 μM to 1 mM, with a peak observed between 1 mM and 5 mM, where approximately 65% of the RNA was ligated. At higher concentrations, a slight inhibitory effect was observed. Interestingly, the percentage of ligated RNA was reduced by only 6% in the absence of MgCl_2 , when the sample contained 5 mM MnCl_2 . This implied that MgCl_2 might help stabilise certain secondary or tertiary structures, but was not essential.

Chapter 5

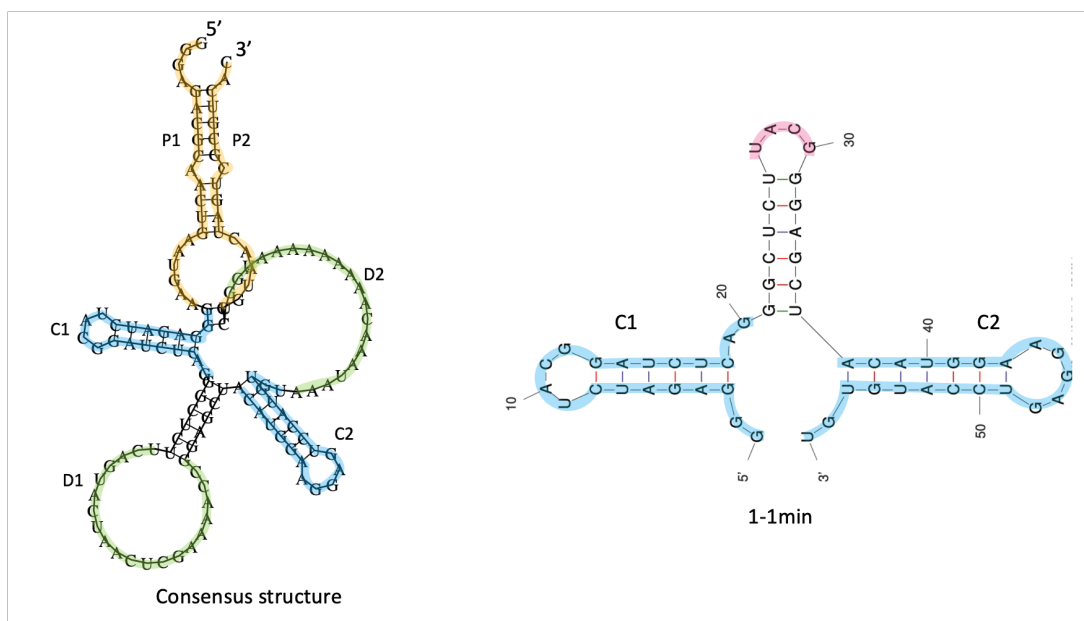
Post-selection engineering and analysis

After eight rounds of selection, the sublibrary exhibited high bulk ligation activity following incubation in the selection buffer. High-throughput sequencing identified the ten most abundant sequences in the final sublibrary, with ligation-based activity analysis revealing H10 as the most catalytically active. In this chapter, a series of truncation and mutation experiments on H10 were conducted to gain insight into its catalytic mechanism and to eliminate redundant regions for easier manipulation. Then kinetic analyses were performed using the shortest ribozyme H10-10. For multiple-turnover kinetic analyses, Dr. James Attwater suggested removing one 5' GTP to create space for free GTP.

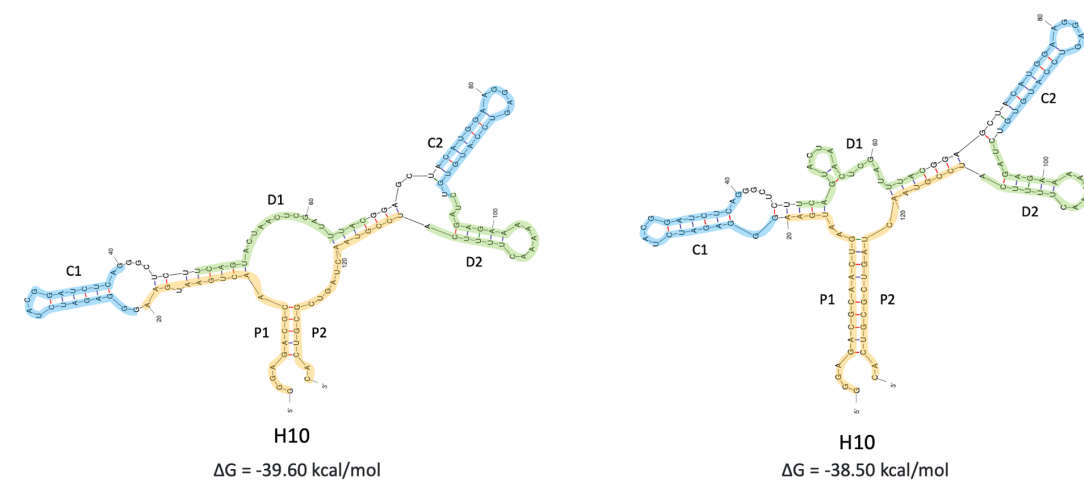
5.1 Sequence engineering

The top 10 sequences were uploaded to the RNAalifold Web server, part of ViennaRNA Web Services, to predict a consensus secondary structure under default conditions (37°C, 1 M NaCl) [357–359]. The resulting structure is shown in Figure 5.1a (left), with the secondary structure of the aptamer scaffold (1-1min) on the right. The two constant regions (C1 and C2) of the aptamer are highlighted in blue, the central stem is shown in black, the primer binding sites (P1 and P2) are highlighted in orange, and the two selected random regions (D1 and D2) are highlighted in green. The loop in pink within the aptamer was replaced by D1 in the selection library. In the consensus secondary structure, the constant regions from the aptamer retained their original secondary structure, while the two random regions lack defined secondary structures. These findings suggest that the selec-

ted ribozymes may be modular, with the aptamer moiety acting as a binding domain that captures the triphosphate, positioning it to interact with the catalytic D1 or D2 regions.



(a)



(b)

Figure 5.1: (a) Predicted consensus secondary structure of the top 10 sequences (left). The secondary structure of the aptamer scaffold (right). (b) Predicted secondary structure of H10. Constant regions from the aptamer are highlighted in blue (C1 and C2) or shown in black, the two random regions (D1 and D2) are highlighted in green, and the primer binding sites (P1 and P2) are highlighted in orange.

The secondary structure of H10 was predicted using mFold under default conditions (37°C, 1 M NaCl) [360]. The structure with the minimum ΔG (-39.6 kcal/mol) is shown in Figure 5.1b (left). The two constant regions, C1 and C2, retain their original stem-loop structures, while the central stem forms partial duplexes with the primer binding sites. The random region D1 hybridises with the primer binding sites at both ends, and D2 forms a

stem-loop structure. In the structure with the second lowest ΔG , D1 forms a small loop and the primer binding sites create a longer stem.

A series of truncation experiments were conducted to identify the minimal functional sequence. These studies revealed which regions of the molecule are essential for activity and which are redundant, providing valuable insights into its mechanism of action. Moreover, shorter sequences are easier to manipulate for subsequent applications, and they may also exhibit enhanced catalytic activity, because they are less prone to misfolding and can adopt their active configuration more rapidly.

Several truncated versions of H10 were synthesised by T7 transcription, and their catalytic activities were tested by incubation in Buffer2 followed by ligation. The HTS results indicated that D2 might not be able to form the stable and complex secondary or tertiary structures required for catalytic function due to its low GC content. Therefore, the D1 region was of primary interest. It was replaced with a 20-nt polyA sequence (H10-6) or truncated (H10-7). The tests revealed that both constructs remained active and exhibited catalytic activity comparable to that of H10 (Figure A.1). When the D1 region of H10-7 was replaced with the longest (16 nt) common motif (CAGUACUAACUCGA), identified using the MEME Suite of the D1 regions of the top 10 sequences, the resulting variant, H10-12, showed significantly reduced activity (Figure A.4). Further truncation of the primer binding sites was performed in H10-2 (D1 and P1 deletion), H10-4 (D1 and P2 deletion) and H10-5 (D1, P1, and P2 deletion); however, none of these variants showed detectable activity (Figure A.1). Primer truncation was also performed on H10 (H10-1 and H10-3), but these constructs did not demonstrate catalytic activity either.

Following the truncation of D2, D1 was also deleted and the loop in pink (Figure 5.1a) was added back in 1-1min-1, leaving the original aptamer and the two primer binding sites. As expected, this construct did not show detectable activity (Figure A.2). The aptamer 1-1min was also tested but showed no catalytic activity (Figure A.2), further demonstrating that D1 is essential for catalysis. In H10-8, the constant regions (shown in black) and part of P2 were truncated (with deleted regions highlighted in purple boxes), and in H10-9, an additional truncation was made (highlighted in the grey box). Both constructs remained highly active (Figure A.2). Then, C1 and C2 were tested (Figure A.4).

It was found that C2 was essential (H10-11), while C1 could be deleted without loss of catalytic activity (H10-10).

By comparing the sequences of the 77-nt H10-10 and the aptamer 1-1min in the black box (Figure 5.2), it can be easily noticed that the four conserved nucleotides in C1 (marked with purple dots) are indeed retained within D1, and the six nucleotides at the 5' end of C1 (highlighted by red dots) appear at the 5' end of P1. These two regions may interact as they do in the aptamer structure, positioning the 5'-triphosphate proximal to the catalytic site in D1. Thus, H10-10 may be a modular ribozyme with a more compact structure.

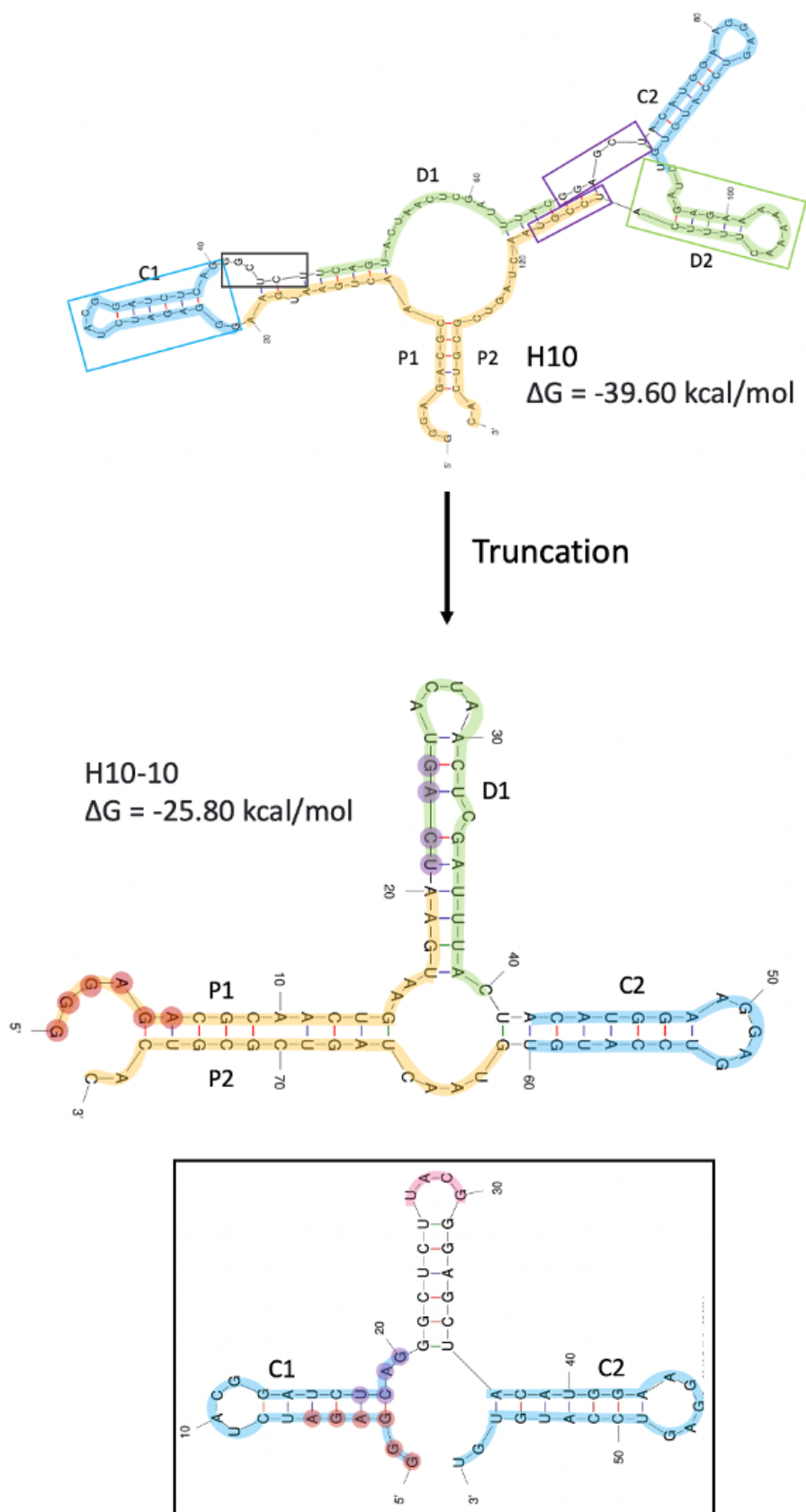


Figure 5.2: Predicted secondary structures of H10 and the truncated sequence H10-10. Truncation sites in H10 are highlighted with square boxes. The predicted secondary structure of the aptamer 1-1min is given inside the black box. Conserved nucleotides in C1 are indicated as coloured dots.

The two primer binding sites, P1 and P2, both appeared to be very crucial for its activity, and are predicted to switch between two main structures: a long stem with an A-C mismatch in the middle or a short stem near the 5' and 3' ends. Therefore, it is worthwhile to investigate further how these structural variations influence catalytic activity. In the case of H10-7, the most stable secondary structure features a short stem, while the alternative long stem structure has a ΔG increased by only 1.1 kcal/mol. Sequences retaining the long stem within the orange square but with different chemical compositions were tested. For example, in H10-7-1, U was swapped with A and G with C; in H10-7-2, U was swapped with G and A with C; and in H10-7-3, U was swapped with C and A with G. If the long-stem structure is functional and only the structure is important, it was expected that changes in the chemical composition would not have a significant impact on activity. However, none of these three variants showed measurable activity (Figure A.3).

For H10-9, the most stable secondary structure contains an extended stem. Changing adenine at position 10 in P1 to guanine (H10-9-1) fixed the mismatch and further decreased ΔG from -36.4 kcal/mol to -42.3 kcal/mol. This single mutation completely abolished the activity of the ribozyme (Figure A.3). While substitution with cytosine had no significant effect (H10-9-2), substitution with uracil led to a faint but detectable activity in the case of H10-9-3, which may be explained by the formation of a weak non-canonical U-C pair (Figure A.3) ^[361]. Replacement of adenine near the 5' end with uracil (H10-9-4) also inactivated the ribozyme (Figure A.3). This may be because the mutation promoted the formation of two base pairs (U-A and G-C) at the 5' end that increased rigidity and prevented the 5' GTP from accessing the catalytic site. H10-9-5 is a variant of the sequence H10-9 in which a single guanine has been deleted at the 5' end. This deletion significantly reduced ribozyme activity, likely due to geometric constraints.

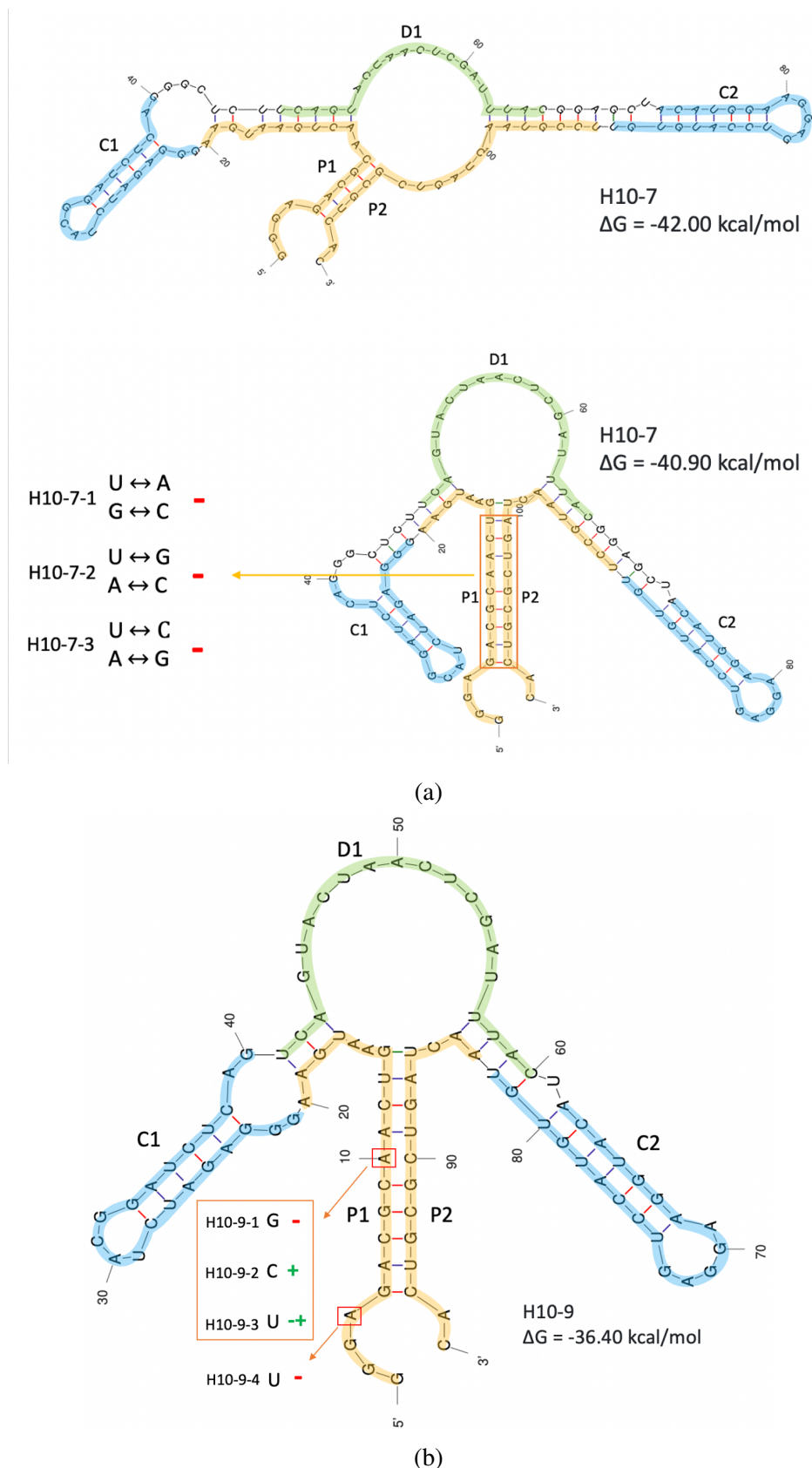


Figure 5.3: (a) Predicted secondary structures of H10-7: the structure with the lowest free energy (top) and the second-lowest free energy (bottom). The box highlights the stem region between P1 and P2, where variations were introduced. The activity levels of these variants (H10-7-1, H10-7-2, and H10-7-3) are indicated, where "-" indicates the loss of activity. (b) The predicted secondary structure of H10-9, highlighting the catalytic activities of mutations at positions A3 and A10. Mutations are labelled: "+" active, "-" inactive, and "-+" with reduced activity.

5.2 3D structure simulation

The 3D structure was predicted for the shortest ribozyme H10-10 (77 nt). The initial structure was generated using RNAComposer^[362], followed by energy minimisation with divalent metal ions (Mg^{2+} and Mn^{2+}) to resolve steric clashes using GROMACS^[363]. The amber99sb-ildn force field and the TIP3P water model were used. The system was equilibrated at 298 K using both NVT and NPT ensembles for 1 ns each, followed by running the molecular dynamics simulation for 50 ns. The generated structure as shown in Figure 5.4 was visualised using PyMOL.

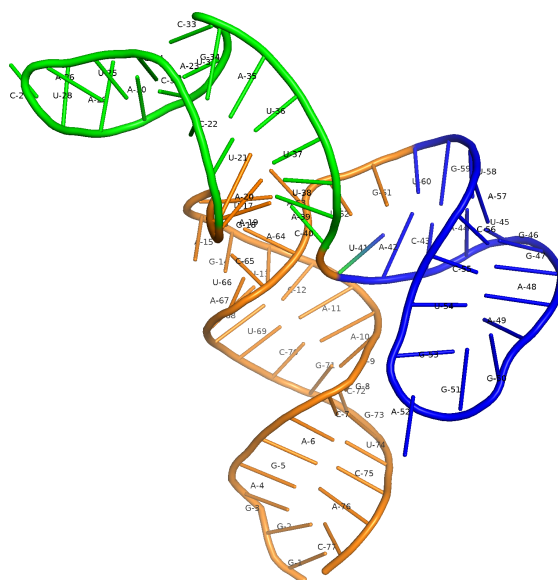


Figure 5.4: Predicted 3D structure of H10-10, with Domain1 in green, C2 in blue, P1 and P2 in orange.

The predicted 3D structure of the H10-10 ribozyme reveals a three-arm architecture, which is consistent with the predicted 2D structure. The molecular dynamics simulation was carried out with a 5' hydroxyl (5' OH) group, as I was unable to successfully incorporate a 5' triphosphate group into the PDB file in a form that could be properly recognised by the simulation software. In the structure with the 5' OH, guanine at position 1 (G1) does not form direct contacts with any residues in Domain 1 or the C2 region, suggesting that significant structural rearrangements may occur in the presence of a 5' triphosphate. The loop in the C2 domain bends towards the termini, with a measured

distance of 34.9 Å between the unpaired adenine (A51) and the 5' hydroxyl group of G1. It is hypothesised that the incorporation of a 5' triphosphate could reduce this distance to approximately 28.8 Å, as suggested by previous structural studies ^[364]. Such a reduction in distance would facilitate the docking of the triphosphate group and may promote closer proximity between Domain 1 and the 5' end, potentially stabilising interactions critical for the catalytic function of the ribozyme.

Further work should prioritise successfully incorporating the 5' triphosphate group into the ribozyme model, as this modification may be essential to accurately model the 5' end, and its interactions within the ribozyme structure could provide information on structural rearrangements and functional mechanisms. The dynamic behaviour of the ribozyme, such as conformational shifts and allosteric changes in the presence of free GTP could also be explored.

5.3 Kinetic analysis

5.3.1 Single-turnover kinetics

Differentiating between the monophosphorylated and triphosphorylated states of H10-10 by direct gel electrophoresis proved to be highly challenging. The test was carried out using a 15% denaturing PAGE gel containing 8 M urea on a 40 × 20 cm plate (Cambridge Electrophoresis), run at constant power (32 W) for 6 hours. Even under these conditions, the two bands remained difficult to distinguish (Figure A.5). This result was not unexpected, as the phosphorylation state of the 5' end has minimal influence on the charge-to-mass ratio of a 77-nt RNA. The difference between a monophosphate (~80 Da) and a triphosphate (~240 Da) is negligible compared to the overall molecular weight (~25,000 Da), introducing only two additional negative charges without inducing a large structural change.

The single-turnover rate was determined indirectly via ligation. A time-course incubation of H10-10 in Buffer3 was performed at 25°C, ranging from 0 to 70 minutes. Reactions (100 µL each) at time points 0, 1, 3, 5, 10, 25, 40, 55, and 70 minutes were quenched by the addition of 10 µL of 500 mM EDTA and subsequently purified using the

Monarch™ RNA Cleanup Kit (NEB). The concentrations of the purified samples were calculated from their A_{260} absorbances. The incubated and purified samples were then subjected to ligation using T4 RNA Ligase 1 (NEB). The RNA substrate was at a concentration of 1.5 μM in a 10 μL reaction volume. No additional adaptor strand was introduced, so the ligation products were formed solely from the reacted RNA strands, which retained a 5' monophosphate after incubation. After 2 hours of incubation at 25°C, 1 μL of each ligation reaction was loaded onto an 8% PAGE gel containing 8 M urea. The gels were run at 300 V for 45–60 minutes using Mini-PROTEAN Tetra Cell (Bio-Rad) and stained with SYBR Gold (Invitrogen, Thermo Fisher Scientific). Three independent trials were performed, and the resulting gel images are presented in Figure 5.5. The bands at the bottom represent the unligated substrate, while the upper bands correspond to various ligation products.

Band intensities were quantified using ImageQuant (Typhoon). Since no adaptor strands were present in the reactions, the intensities reflect the actual amount of substrate RNA at each specific position. Thus, the fraction of the total intensity of the upper bands in each lane represents the actual fraction of ligated RNA. However, the fraction ligated is usually lower than the fraction of molecules available for ligation and is sensitive to reaction conditions [365–368]. A parallel reaction was conducted using RppH, an enzyme that removes pyrophosphate from the 5' triphosphate group. The RppH reaction was incubated at 37°C for 1 hour, double the time recommended in the protocol, to ensure complete removal of pyrophosphate. RppH is a highly effective pyrophosphatase, particularly active on 5' GTP [369]. Consequently, the RppH-treated sample was considered 100% ligatable, and the ligated fraction for each sample was normalised against the RppH-treated control, resulting in a value representing the fraction reacted.

The average fraction reacted for each time point was determined and standard deviations were calculated to account for variability between trials. The data was fit to an exponential association model:

$$FR = F_{\max} \left(1 - e^{-k_{\text{obs}} t} \right) \quad (5.1)$$

where FR is the fraction that reacted at time t , $F_{\max}$ is the maximum fraction that can

react and k_{obs} is the observed rate constant.

Data were processed using nonlinear least squares curve fitting with the `curve_fit` function from the SciPy library in Python. The corresponding Python script is provided in Appendix A.5. The fitting resulted in $F_{\text{max}} = 0.82 \pm 0.03$ and $k_{\text{obs}} = 0.08 \pm 0.01 \text{ min}^{-1}$. As shown in Figure 5.5, most of the reaction occurred within the first 25 minutes and reached a plateau around 30 minutes. The reaction reaches a plateau at F_{max} , suggesting that 82% of the substrate is reacted under these experimental conditions.

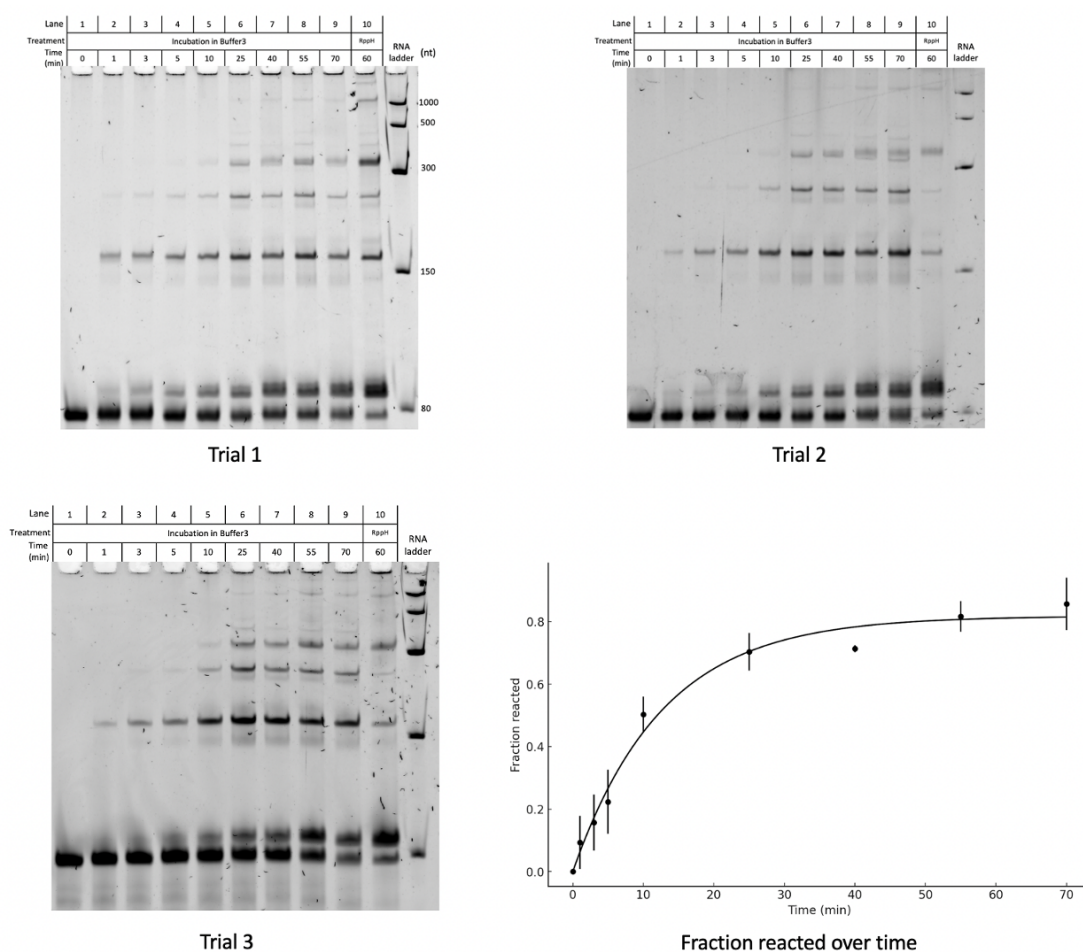


Figure 5.5: Single-turnover kinetics study by ligation. The gel images show the ligation products from three independent trials. The curve represents the data fitted to the equation $FR = F_{\text{max}}(1 - e^{-k_{\text{obs}}t})$, with $F_{\text{max}} = 0.82$ and $k_{\text{obs}} = 0.08 \text{ min}^{-1}$. The points represent the fraction reacted at each time point, and the error bars indicate the standard deviation.

This method is not ideal for determining the rate constant because it does not directly measure product formation and relies on ligation efficiency, which can be influenced by factors such as substrate concentration. When product levels are low, the ligation efficiency may decrease, potentially causing the actual reaction rate to appear slower in the

early stages.

5.3.2 Multiple-turnover kinetics

The ability of the ribozyme to perform multiple turnovers on free GTP was assessed using the PPiLight™ Inorganic Pyrophosphate Assay (Lonza). This assay converts AMP to ATP in the presence of pyrophosphate (PPi), and the generated ATP is then used by luciferase to produce bioluminescence. The intensity of bioluminescence is directly proportional to the amount of PPi in the reaction, within a linear detection range of 0.02 μM to 10 μM . The presence of Mg^{2+} is essential for the luciferase reaction, although excessive concentrations of Mg^{2+} can inhibit the assay. Monovalent salts such as NaCl are generally inhibitory to the luciferase reaction [370]. Since previous experiments have demonstrated that monovalent salts are not essential, they were omitted from this assay.

The impact of the reaction buffer without monovalent metal ions (Buffer4, containing 25 mM Tris-HCl pH 7.8, 30 mM MgCl_2 and 3 mM MnCl_2) and the same buffer with an equimolar concentration of EDTA (33 mM) was assessed. Three sets of four 40 μL samples with increasing concentrations of PPi (0, 0.5, 1.0, and 1.5 μM) were prepared. All samples were tested in triplicate. Group 1 samples were prepared in water, the Group 2 samples in Buffer4 and the Group 3 samples in Buffer4 with 33 mM EDTA. Relative light units (RLU) were recorded and a plot (Figure 5.6a) was generated. All three groups were fitted with linear regression curves. The data show a distinct trend for the three conditions. The samples in water (Group 1) exhibited the highest luminescence in all PPi concentrations, with a strong linear relationship (slope = 8.35×10^5 RLU/ μM PPi). In contrast, the addition of Buffer4 (Group 2) significantly reduced the RLU signals, nearly halving the slope (4.46×10^5 RLU/ μM PPi), indicating a substantial influence of the divalent metal ions Mg^{2+} and Mn^{2+} on the assay signal. The most notable reduction occurred when EDTA was introduced into the buffer system (Group 3). EDTA was tested because it chelates divalent metal ions, effectively halting the ribozyme reaction prior to measuring the PPi levels. The data showed that EDTA almost completely eliminated the detectable signal, with a minimal slope of 0.10×10^5 RLU/ μM PPi. This drastic decrease suggests that EDTA not only inhibited the ribozyme by chelating the required metal ions

but may have also interfered with the assay itself, as the amount of added metal ions should have been sufficient only for chelating the metal ions in the ribozyme reaction.

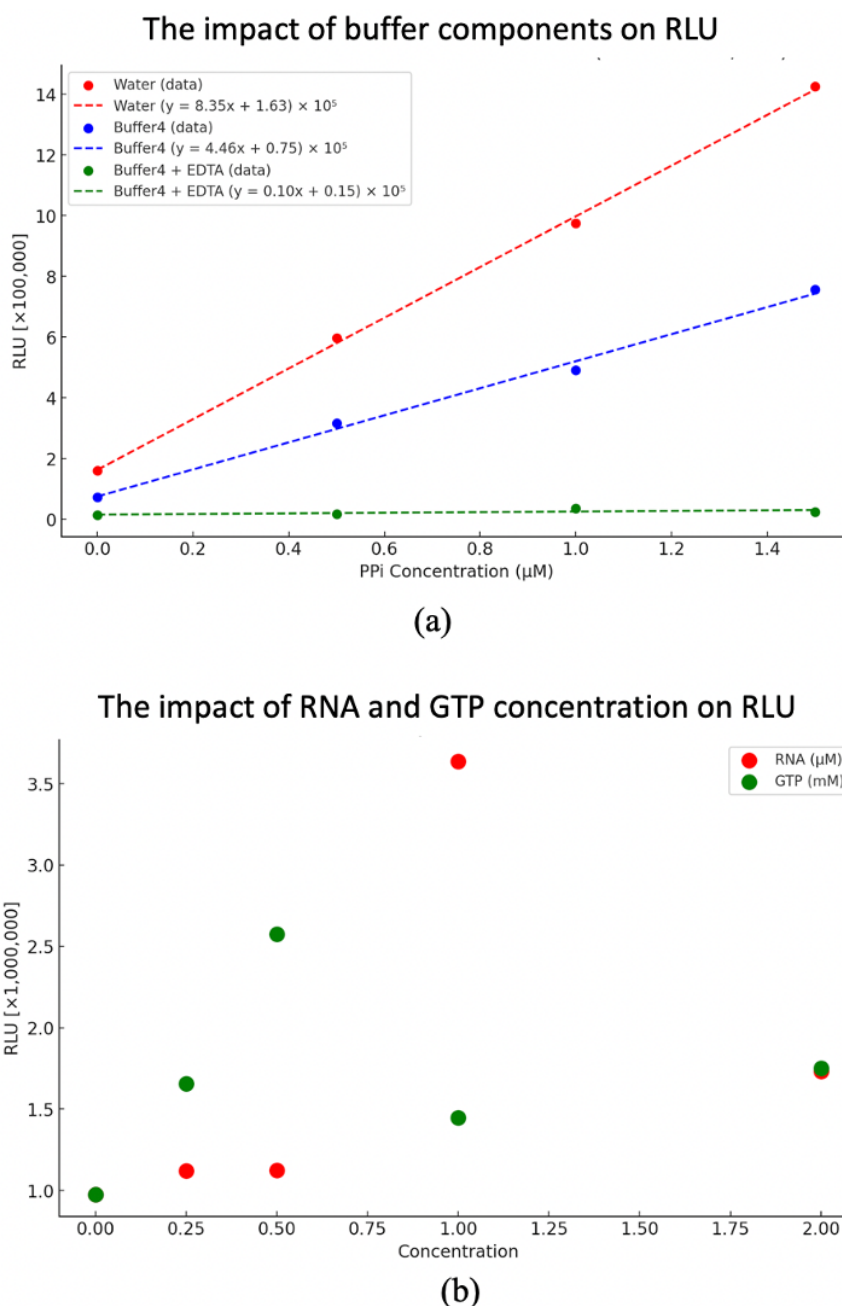


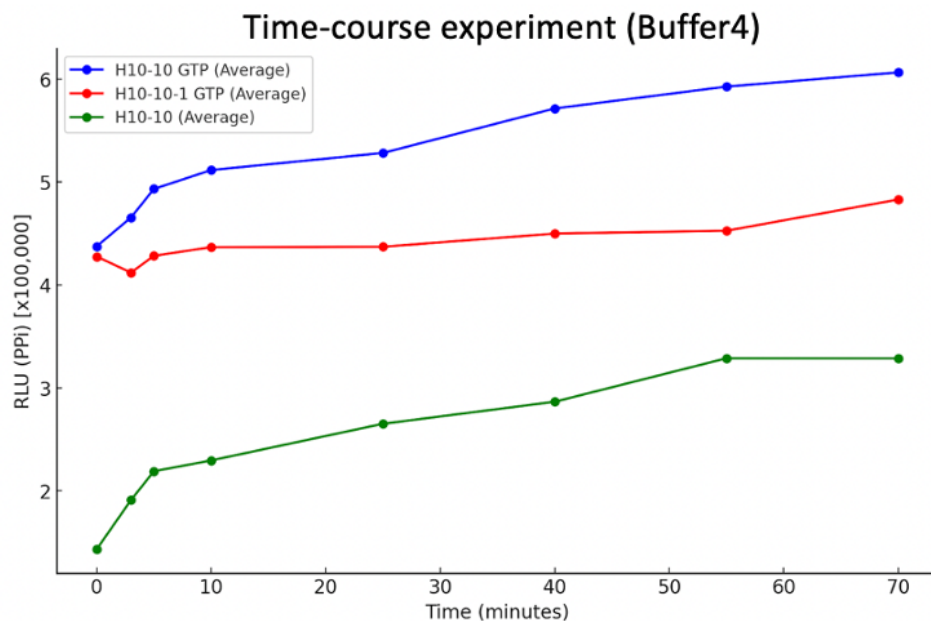
Figure 5.6: (a) The impact of buffer components on relative light units (RLU). (b) The impact of RNA and GTP concentration on RLU (the effects of RNA and GTP concentrations on RLU were measured separately).

Figure 5.6b shows the relationship between RNA and GTP concentrations and RLU produced in the PPiLight™ assay. The red data points represent varying RNA concentrations (0, 0.25, 0.5, 1, 2 μM), while the green data points represent varying GTP concentrations (0, 0.25, 0.5, 1, 2 mM). All samples were prepared in triplicate in water with 1 μM PPI. The results indicate that the addition of RNA and GTP generally increased the

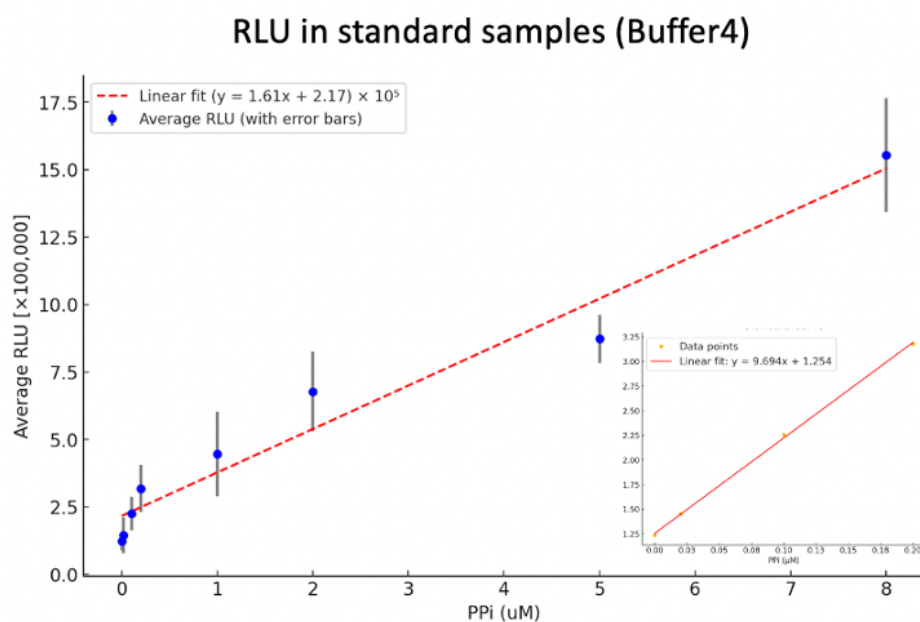
signal but not linearly. In samples with 0.25 μM and 0.5 μM RNA, the signal increased approximately 1.2 times. The sample with 1.0 μM RNA produced the highest RLU value, approximately 3.5×10^6 . However, when the RNA concentration was increased to 2 μM , the RLU decreased to 1.7×10^6 . For GTP, the signal increased linearly within the range of 0 to 0.5 mM, with higher GTP concentrations enhancing the assay response. Interestingly, at 1 mM GTP, the RLU decreased, indicating a potential inhibitory effect. When the concentration of GTP was further increased to 2 mM, the signal recovered slightly but did not reach the level observed at lower concentrations.

Time-course experiments were conducted for H10-10 (77 nt) and H10-10-1 (76 nt, one G deleted from the 5'GGG of H10-10). Detection of generated PPi was performed in the absence (Figure 5.7a) and presence (Figure 5.8a) of EDTA. For the H10-10 ribozyme, both single-turnover conditions (without free GTP) and multiple-turnover conditions (with 1 mM free GTP) were assessed. The variant H10-10-1 was tested only under multiple-turnover conditions, as its activity to catalyse the hydrolysis of its 5' GTP was significantly reduced as suggested by ligation, and it was hypothesised that the removal of the terminal G could allow better recruitment of free GTP.

The reactions were set up by mixing all components except MnCl_2 , leading to final concentrations of 25 mM Tris-HCl (pH 7.8), 30 mM MgCl_2 , and 0.5 μM RNA, with or without 1 mM GTP. The catalytic reaction was initiated by adding 3 mM MnCl_2 at the desired time points resulting in a total reaction volume of 40 μL . After incubation for 3, 5, 10, 25, 40, 55, and 70 minutes, 20 μL of PPiLight™ converting reagent was added to each sample to convert AMP in the reagent into ATP using the PPi generated from the hydrolysis of 5' GTP or free GTP. For the EDTA-treated samples, EDTA was added to a final concentration of 33 mM to stop the reaction before the converting reagent was introduced. After a 30-minute incubation, 20 μL of the detection reagent was added, followed by an additional 30-minute incubation during which luciferase generated light using the ATP produced in the previous step. For the 0-minute time point, Buffer4 and the PPiLight™ converting reagent were added simultaneously. Standard samples containing various concentrations of PPi were prepared in Buffer4 and tested in the same manner (Figure 5.7b and Figure 5.8b).



(a)

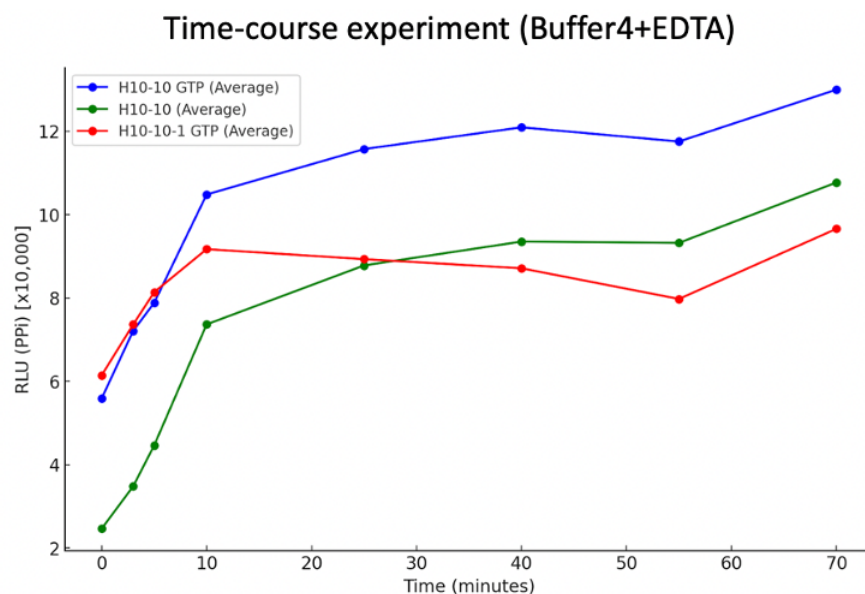


(b)

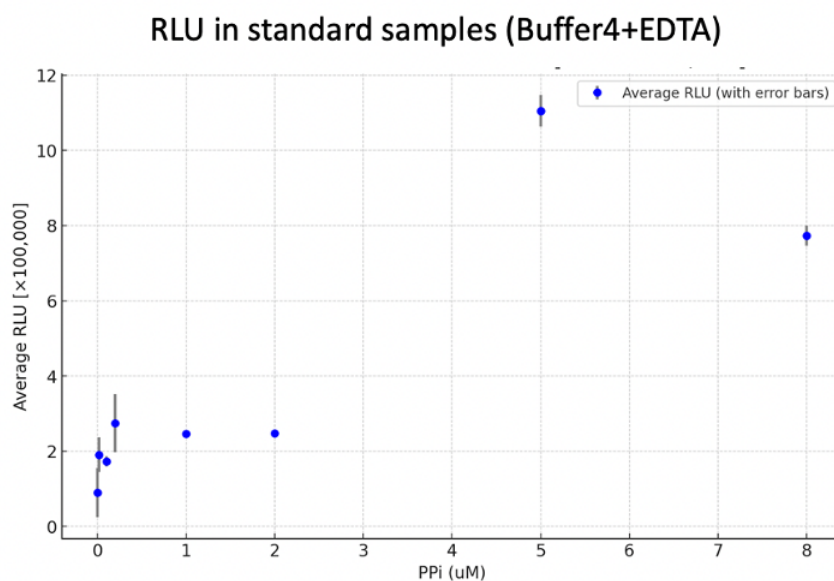
Figure 5.7: (a) RLU generated by the PPI^{Light}™ Inorganic Pyrophosphate Assay from a time-course experiment. The blue curve represents H10-10 under multiple-turnover conditions with 1 mM GTP, the red curve represents H10-10-1 under multiple-turnover conditions with 1 mM GTP, and the green curve represents H10-10 under single-turnover conditions. (b) Standard curve with error bars. All samples were prepared using Buffer4.

In Figure 5.7a, the blue curve represents H10-10 with free GTP, the red curve represents H10-10-1 with free GTP, and the green curve represents H10-10 without free GTP. During the course of 70 minutes, all three samples showed a modest increase in signal. The two samples containing GTP exhibited higher initial signals, consistent with previous findings that the addition of 1 mM GTP increases the signal by approximately 4×10^5 RLU. Interestingly, the standard curve of linearly fitted samples in Buffer4 showed two distinct regions as presented in Figure 5.7b. In the lower concentration range from 0 to 0.2 μM , the signal increased more rapidly. The concentration at the final time point for the green curve was calculated to be 0.21 μM according to the lower concentration region of the standard curve, and 0.696 μM for the higher concentration region. The two curves for H10-10 samples, with (blue) and without (green) free GTP, ran almost parallel, suggesting that H10-10 likely performed only single-turnover catalysis of its own 5' GTP. The H10-10-1 sample with GTP showed an even smaller increase in signal, suggesting that its catalytic activity for 5' GTP hydrolysis is significantly reduced and it is unlikely that it had performed multiple-turnover catalysis on the free GTP.

The results of the measurements with 33 mM EDTA added to the reaction are shown in Figure 5.8a. The blue curve represents H10-10 with free GTP, the red curve represents H10-10-1 with free GTP, and the green curve represents H10-10 without free GTP. The curves for H10-10 in single turnover (green) and multiple turnover (blue) are nearly parallel, showing a trend similar to the signal measured without EDTA. This observation supports the conclusion that H10-10 did not catalyse free GTP under multiple-turnover conditions. The amount of PPi generated showed an initial burst, with the signal plateauing after about 10 minutes of incubation, faster than the 25 minutes observed in the curve generated from ligation products. This evidence supports the earlier conjecture that the observed rate constant (k_{obs}) might be higher than the value measured through ligation, since the ligation efficiency would be considerably reduced in the early stages where the substrate concentration is low. The signal for H10-10-1 samples showed a very slight increase during incubation and is therefore only barely active in catalysis. The signals of the standard samples, compared to those without EDTA, were much lower, and the data points could not fit into a line curve, especially at higher PPi concentrations.



(a)



(b)

Figure 5.8: (a) RLU generated by the PPiLight™ Inorganic Pyrophosphate Assay from a time-course experiment. The blue curve represents H10-10 under multiple-turnover conditions with 1 mM GTP, the red curve represents H10-10-1 under multiple-turnover conditions with 1 mM GTP, and the green curve represents H10-10 under single-turnover conditions. (b) Plotted data points of standard samples. All samples were prepared using Buffer4, with 33 mM EDTA added prior to the PPi conversion and detection assay.

Chapter 6

Conclusions and future directions

6.1 Discussion and conclusions

This thesis explores the application of in vitro selection techniques for the development of nucleic acid-based mechanocatalysts, aimed at improving autonomous nucleic acid-based molecular machines. The introductory chapter discussed the theoretical foundations of artificial molecular machines, structural and dynamic DNA nanotechnology, similarities between protein-based natural molecular machines and functional nucleic acids, and the possible role of nucleic acids as energy-transducing and coordinating elements in molecular machines driven by small-molecule fuels.

Based on the information and discussion in the first chapter, two strategies were proposed. The first strategy involved the adaptation of Willner's ATPase-mimicking nucleozyme by replacing the original aptamer with a triphosphate-specific aptamer to introduce coupling between the mechanical state of the aptamer and the chemical catalysis of the small molecule fuel. Because the catalyst was not commercially available, the chemical synthesis of the tethered catalytic center, a bis-Zn²⁺-pyridyl-salen complex, was performed. However, after several attempts, the synthesis of the catalyst proved challenging due to salen ligand instability, the complexity of its coordination chemistry, and constraints resulting from the covid pandemic. Therefore, this strategy was not pursued further.

The second strategy, which forms the main focus of this thesis, is based on in vitro selection. A selection scheme aimed at isolating a ribozyme capable of catalysing the

hydrolysis of GTP under single-turnover conditions was designed. After eight rounds of selection, the ribozymes identified by high-throughput sequencing (HTS) exhibited catalytic activity comparable to RNA 5' pyrophosphohydrolase (RppH). Ligation-based activity analysis showed that the most active ribozyme ranked 10th (H10) in abundance in the final sub-library, while sequences ranked 2nd (H2) and 6th (H6) exhibited background-level activity. This disparity between abundance and activity is not uncommon in in vitro selection. The relative abundance of certain sequences could have been influenced by factors such as compatibility with the selection mechanism rather than their catalytic efficiency. Consequently, sequences that ranked highly may have been preferentially enriched for traits that were not related to activity, such as ease of ligation or amplification, rather than their true catalytic potential.

Sequence analysis of individual random regions revealed that the random region near the 3' end (D2) exhibited less enrichment compared to the one near the 5' end (D1). The enriched sequences in D2 were predominantly AU rich, which tend to be more structurally flexible, reducing the likelihood of intramolecular interactions across the full-length sequence. This flexibility may also explain the less pronounced peak shifts observed in the melting curve analysis. An increase in AU content within one region could decrease the overall melting temperature despite the increased sequence homogeneity in the other random region. The D2 region was later shown to be non-essential in truncation experiments (H10-7), although the activity for ligation compared to RppH decreased from 0.83 to 0.67. Replacement of the D1 region in H7 with the motif sequence detected among the top 10 sequences in D1 (H10-12) significantly decreased its activity, to 0.23. The further trimming of H10-7 resulted in the identification of a minimal 77 nt ribozyme (H10-10) that exhibited 0.73-fold activity for ligation after a 60-minute incubation in reaction buffer (Buffer2) compared to the sample treated with RppH. H10-10 maintained a predicted secondary structure similar to the original aptamer scaffold (1-1min). The C2 region retained its stem-loop structure. Despite deletion of the C1 region, all conserved nucleotides were preserved. This suggests that H10-10 could function as a modular ribozyme, with distinct but cooperative substrate recognition and catalytic modules. Although the ribozyme was selected under single-turnover conditions, it may be capable of multiple-turnover catalysis

because the binding and reaction sites on the substrate are the same.

A time-course incubation followed by ligation was used in the single-turnover kinetic experiment. The proportion of reacted molecules for each sample was measured by determining the total intensity of the ligation product bands, normalised against that of the RppH-treated control. By fitting the data to a curve, the observed rate constant was calculated to be $k_{\text{obs}} = 0.08 \pm 0.01 \text{ min}^{-1}$. However, this approach is far from ideal because it does not directly measure product formation and is dependent on ligation efficiency, which can be affected by factors such as reaction conditions and substrate concentration. When product levels (monophosphorylated RNA) are low, the ligation efficiency may drop. This would appear to make the rate of reaction proceed more slowly, especially in its early stage. This hypothesis is supported by the results of the later PPiLight™ pyrophosphate detection assay, where the signal plateaued about 15 minutes earlier than was observed in the kinetic analysis by ligation. Under multiple-turnover conditions (with free GTP present in solution), neither H10-10 nor H10-10-1 (with one 5' GTP removed to allow space for incoming GTP) exhibited multiple-turnover activity.

6.2 Future directions

The deletion of a 5' GTP in H10-10 creates space for an additional free GTP molecule, but the H10-10-1 ribozyme produced by T7 IVT still retains a 5' triphosphate. This could result in electrostatic repulsion, which makes it less energetically favourable for free GTP molecules to come into close proximity. Removing 5' triphosphate after T7 IVT may ultimately enable the ribozyme to perform multiple turnover catalysis. This can be achieved using calf intestinal alkaline phosphatase (CIP) [371].

Incubation followed by ligation can give an indication of whether a sequence is catalytically active (converting 5' triphosphate to 5' monophosphate), and offers an intuitive comparison of activity between sequences. However, this is not ideal for quantification, as the efficiency of ligation depends on reaction conditions as well as the sequence and secondary structure of the substrates. Methods including high-performance liquid chromatography (HPLC) [372] and radio thin layer chromatography (radio-TLC) [373] provide more accurate quantification. The PPiLight™ kit is a convenient, fast and non-radioactive

pyrophosphate detection assay. However, test conditions such as buffer composition and improved sample and reagent loading in 96-well plates must be optimised for optimal performance.

The simulated 3D structure of H10-10 with a 5' OH group shows that the 5' end is not in close proximity to any specific regions within the sequence. This would suggest that the presence of a triphosphate group may cause a significant structural rearrangement, perhaps moving it to a position near the binding or catalytic sites. Thus, molecular dynamics simulations with a 5' triphosphate or a free GTP docked at its predicted binding site could provide insight into the functional mechanism of the ribozyme and inform future applications. Additionally, the 3D structure presented in this thesis was derived using a conventionally predicted 2D structure as a starting point. Recent advances in computational tools, such as AlphaFold 3 ^[374], allow the direct prediction of 3D structures from primary nucleotide sequences with potentially improved accuracy.

Although the triphosphate-specific aptamer scaffold used here was designed to predispose ribozymes to multi-turnover activity, such behaviour could not be guaranteed. The selection process was conducted under single-turnover conditions, meaning that the emergence of multiple-turnover behaviour may occur more by chance than by design.

A more effective strategy could be to integrate all the desired functions into the selection pressure. For example, Figure 6.1 illustrates a three-step selection procedure for mechanocatalysts that respond to ATP. A library is prepared with a constant docking region in the middle flanked by two random regions and constant primer binding sites. The first step of selection involves immobilisation of the library onto streptavidin-coated magnetic beads by hybridisation with a biotinylated oligonucleotide ^[272]. Upon the addition of ATP, sequences that undergo significant structural changes, disrupting the hybridisation, are released into the incubation buffer. These sequences are then separated from the rest of the library on the beads using a magnet. The collected sequences are subsequently amplified by RT-PCR and IVT for RNA libraries, or directly by PCR for DNA libraries. For DNA libraries, an ssDNA sublibrary can be prepared using methods such as asymmetric PCR, biotin–streptavidin separation, and lambda exonuclease digestion ^[375–377]. This process can be repeated multiple times to enrich the pool with sequences that ex-

hibit structural changes upon ATP binding. These sequences may include weak binders to the biotinylated oligo, allosteric aptamers for ATP, or catalytic aptamers that had first engaged in allosteric binding. In the second step of selection, the library enriched with allosteric functionality is loaded onto a non-hydrolysable ATP γ S-agarose column. High-affinity aptamers or ribozymes for ATP remain in the column and are subsequently eluted with high concentrations of ATP before being amplified. After these two steps, sequences with both high affinity for ATP and substantial structural switching upon binding should have been enriched. In the last step, the sublibrary is incubated with the ATP-modified agarose column. The flow-through, now containing catalytic aptamers, is collected and the sequences are amplified. Cycling in each step or cycling through all three steps can be done in order to further refine the selection.

The intricate molecular machines made from proteins are the products of approximately 4 billion years of evolution. Countless of these molecular machines, including motor proteins and enzymes, operate within a human body, transforming us into efficient biological factories. These biological machines are essentially mechanocatalysts as they couple their mechanical cycles to the catalysis of ATP hydrolysis. They work repeatedly and efficiently without external regulation. Such autonomy and efficiency are highly desirable in nanotechnology. Autonomy, in particular, avoids the need for complex control systems, which would be impractical in large-scale applications. In addition, autonomy allows molecular machines to function in environments where external stimuli are unavailable or undesirable, such as within living organisms. With their unparalleled programmability, nucleic acids are the ideal material for the engineering of man-made autonomous molecular machines. However, the autonomous nucleic acid-based molecular machines reported so far either rely on protein enzymes or depend on slow strand displacement reactions. We hypothesised that integrating a nucleic acid-based mechanocatalyst capable of acting on small molecules within a motor system could greatly enhance efficiency.

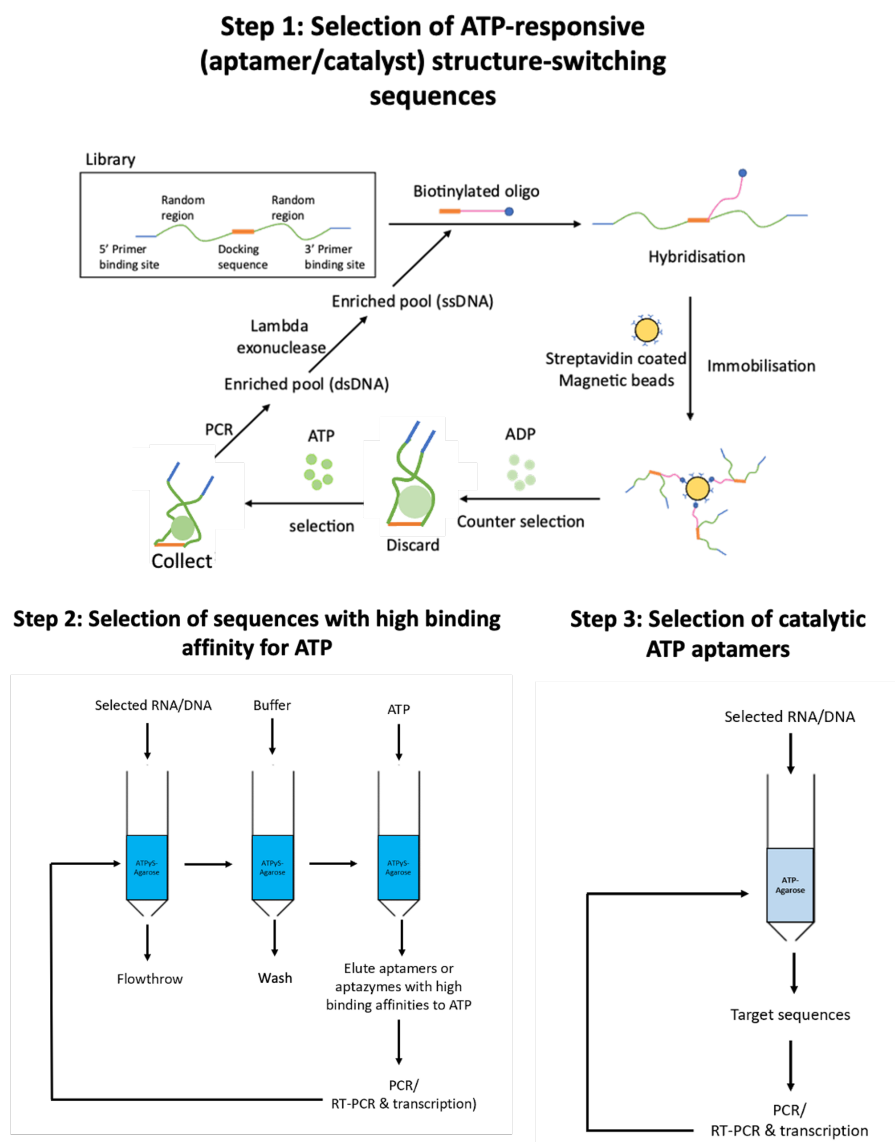


Figure 6.1: Schematic representation of a three-step selection procedure for ATP-responsive allosteric ribozymes or DNAzymes. For a detailed explanation of the procedure, refer to the text.

Although the work presented in this thesis is still far from the creation of the desired nucleic acid-based molecular machine that we proposed, it underlined that nucleic acids can be as functionally versatile and efficient as proteins. This research highlights the potential of nucleic acid-based mechanocatalysts for molecular machines. While the in vitro selection technique is currently the primary route to the discovery of nucleic acids with novel functions, with continuously improved computational tools and predictive modeling, de novo design of functional nucleic acids might be possible in the future.

Appendix A

Appendix

A.1 Lists of oligonucleotides

Name	Sequence 5'- 3'
DNA library	GTGACGCGACTAGTTACGGANNNNNNNNNNNNNNNNNNNNNNNNNNN ACACATGGACTCCTTCCATGTAGCTCCNNNNNNNNNNNNNNNNNN NNNNNNNAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
Primer1	GTATAATACGACTCACTATAGGGAGACGCAACTGAATGAA
Acceptor1	GGACAGCGCGUGUAGAAAUUAAUACGACUCACUAUA
Primer-169	GTATAATACGACTCACTATAGGACAGCGCGUGUAGAAAUU AAUACGACUCACUAUAGGGAGACGCAACTGAATGAA
Primer-153	GTATAATACGACTCACTATAGGACAGCGCGTGTAGAAAT TGGGAGACGCAACTGAATGAA
Splint1	CATTCAGTTGCGTCTCCCTATAGTGAGTCGTAT
Primer2	GTGACGCGACTAGTTACGGA
Primer3	GGACAGCGCGTGTAG
P51_Qian_F	AATGATACGGCGACCACCGAGATCTACACTCTTTTCCCCTACA CGACGCTCTTCCGATCTNNNATCACCGGGGAGACGCAACTGA ATG

Name	Sequence 5' - 3'
P52_Qian_F	AATGATACGGCGACCACCGAGATCTACACTCTTTTCCCCTACA CGACGCTCTTCCGATCTNNN CGATGT GGGGAGACGCAACTGA ATG
P53_Qian_F	AATGATACGGCGACCACCGAGATCTACACTCTTTTCCCCTACA CGACGCTCTTCCGATCTNNN TTAGGC GGGGAGACGCAACTG AATG
P54_Qian_F	AATGATACGGCGACCACCGAGATCTACACTCTTTTCCCCTACA CGACGCTCTTCCGATCTNNN TGACCA GGGGAGACGCAACTGA ATG
P7_Qian_R	CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTNNNGTGACGCGACTAGTTACGG
S1(template)	GTGACGCGACTAGTTACGGAACATGGGGCGTTCTTCGTGAA CACATGGACTCCTTCCATGTAGCTCCTCAGGCGGGACCGTA ACTTTAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
S2(template)	GTGACGCGACTAGTTACGGAGAAGCTGAATGCTTAGAAAGA CACATGGACTCCTTCCATGTTGCTCCAGGACCGGTGCTCAA AAAACAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
S3(template)	GTGACGCGACTAGTTACGGACGTTATATTCTTCGAGTCAGA CACATGGACTCCTTCCATGTAGCTCCGCTTTCGAGTTAGTA CACTGAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
S4(template)	GTGACGCGACTAGTTACGGAACTTTAGTGAAAATTGGGTAG ACACATGGACTCCTCCATGTAGCTCCGTTTCTCGAGTTAGT ACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC

Name	Sequence 5' - 3'
S5(template)	GTGACGCGACTAGTTACGGATACATTCTAGTGTACGCACT ACACATGGACTCCTTCCATGTAGCTCCGTAATTCGAGTTAG TACCGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
S6(template)	TTGACGCGACTAGTTACGGAACTACATGTGTTTCGAGAGGT ACACATGGACTCCTTACATGTAGCTCCGGAACGTCGAGTTA GTACTGAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H1(template)	GTGACGCGACTAGTTACGGAACATTTTTTTGAGTCAATGTA CACATGGACTCCTTCCATGTAGCTCCGCTTTCGAGTTAGTA CTGAGAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H2(template)	GTGACGCGACTAGTTACGGAAAGCTGAATGAACTTTAGAGA CACATGGACTCCTTCCATGTAGCTCCCGTGTGGCGGCCGCA TTTCTAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H3(template)	GTGACGCGACTAGTTACGGACCTGTAATTTATTTGCTTACA CACATGGACTCCTTCCATGTAGCTCCCGCTTTCGAGTTAGT ACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H4(template)	GTGACGCGACTAGTTACGGATACTAGTACTTTGTTTTTTTA CACATGGACTCCTTCCATGTAGCTCCCGTCTTCGAGTTAGT ACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H5(template)	GTGACGCGACTAGTTACGGAACATTTTTTTTGAGTCAATGA CACATGGACTCCTTCCATGTAGCTCCTGCTTTCGAGTTAGT ACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC

Name	Sequence 5' - 3'
H6(template)	GTGACGCGACTAGTTACGGATAGTTTTTTCCTATTCGTAAA CACATGGACTCCTTCCATGTAGCTCCTCATACGGGATTTGA ATGATAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H7(template)	GTGACGCGACTAGTTACGGACATTTTTTTTTTGAGTCAATGA CACATGGACTCCTTCCATGTAGCTCCTGCTTTCGAGTTAGT ACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H8(template)	GTGACGCGACTAGTTACGGAACATTTTTTTTTTGAGTCAATA CACATGGACTCCTTCCATGTAGCTCCGTGCTTTCGAGTTAG TACTGAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H9(template)	GTGACGCGACTAGTTACGGACATTTTTTTTTTGAGTCAATGTA CACATGGACTCCTTCCATGTAGCTCCGCTTTCGAGTTAGTA CTGAGAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H10(template)	GTGACGCGACTAGTTACGGATGAAAAGTTTTTTTCTCTAGA CACATGGACTCCTTCCATGTAGCTCCGTAAATCGAGTTAGTA CTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
Primer4	GTATAATACGACTCACTATTGGGAGACGCAACTGAATGAA
H10-1(template)	TGAAAAGTTTTTTTCTCTAGACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAGCCCTGAGATCCG TAGATCTCCCTTCATTCAGTTGCGTCTCCC
H10-2(template)	ACACATGGACTCCTTCCATGTAGCTCCGTAAATCGAGTTAG TACTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC

Name	Sequence 5' - 3'
H10-6(template)	GTGACGCGACTAGTTACGGATTTTTTTTTTTTTTTTTTTTA CACATGGACTCCTTCCATGTAGCTCCGTAAATCGAGTTAGTA CTGAAGAGCCCTGAGATCCGTAGATCTCCCTTCATTCAGT TGCGTCTCCC
H10-7(template)	GTGACGCGACTAGTTACGGAACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAGCCCTGAGATCCG TAGATCTCCCTTCATTCAGTTGCGTCTCCC
H10-8(template)	GTGACGCGACTAGTTACACATGGACTCCTTCCATGTAGTAA ATCGAGTTAGTACTGAAGAGCCCTGAGATCCGTAGATCTCCC TTCATTCAGTTGCGTCTCCC
H10-9(template)	GTGACGCGACTAGTTACACATGGACTCCTTCCATGTAGTAA ATCGAGTTAGTACTGACTGAGATCCGTAGATCTCCCTTCATT CAGTTGCGTCTCCC
H10-10(template)	GTGACGCGACTAGTTACACATGGACTCCTTCCATGTAGTAA ATCGAGTTAGTACTGATTCATTCAGTTGCGTCTCCC
H10-10-1(template)	GTGACGCGACTAGTTACACATGGACTCCTTCCATGTAGTAA ATCGAGTTAGTACTGATTCATTCAGTTGCGTCTCC
H10-11(template)	GTGACGCGACTAGTTGTAAATCGAGTTAGTACTGACTGAGA TCCGTAGATCTCCCTTCATTCAGTTGCGTCTCCC
H10-12(Fwd)	GTATAATACGACTCACTATAGGGACTGCGCTGATAATGAAG GGAGATCTACGGATCTCAGGGCTCTCAGTACTAACTCG
H10-12(Rev)	GTGACGCGACTAGTTACGGAACACATGGACTCCTTCCATGT AGCTCCTTCGAGTTAGTACTGAGAGCCC
H10-11-1(template)	GTGACGCGACTAGTTGTAAATCGAGTTAGTACTGACTGAGA TCCGTAGATCTCCCTTCATTCAGTTGCGTCTCC
1-1min-1(template)	GTGACGCGACTAGTTACGGAACACATGGACTCCTTCCATGT AGCTCCCGTAAGAGCCCTGAGATCCGTAGATCTCCCTTCATT CAGTTGCGTCTCCC
Primer5	GTATAATACGACTCACTATAGGGAGAUUCUACGGAUCUCAGGG

Name	Sequence 5' - 3'
H10-3(template)	GTGACGCGACTAGTTACGGATGAAAAGTTTTTTCTCTAGA CACATGGACTCCTTCCATGTAGCTCCGTAAATCGAGTTAGTA CTGAAGAGCCCTGAGATCCGTAGATCTCCC
H10-4(template)	GTGACGCGACTAGTTACGGAACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAGCCCTGAGATCCG TAGATCTCCC
H10-5(template)	ACACATGGACTCCTTCCATGTAGCTCCGTAAATCGAGTTAG TACTGAAGAGCCCTGAGATCCGTAGATCTCCC
1-1min(template)	ACACATGGACTCCTTCCATGTAGCTCCCGTAAGAGCCCTGA GATCCGTAGATCTCCC
Primer6	GTATAATACGACTCACTATAGGAGACGCAACTGAATGAA
H10-7-1(Fwd)	GTATAATACGACTCACTATAGGGACTGCGCTGATAATGAAG GGAGATCTACGGATCTCAGGGCTCTTCAGTACTAACTCGATT
H10-7-1(Rev)	GTCTGCGTTGACGTTACGGAACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAGC
H10-7-2(Fwd)	GTATAATACGACTCACTATAGGGATCATACCAGTAATGAAG GGAGATCTACGGATCTCAGGGCTCTTCAGTACTAACTCGATT
H10-7-2(Rev)	GTTTCATATCAGCGTTACGGAACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAGC
H10-7-3(Fwd)	GTATAATACGACTCACTATAGGGAAGTATAGTCGAATGAAG GGAGATCTACGGATCTCAGGGCTCTTCAGTACTAACTCGATT
H10-7-3(Rev)	GTAGTATGGTCAGTTACGGAACACATGGACTCCTTCCATGT AGCTCCGTAAATCGAGTTAGTACTGAAGAG
H10-9-N(template)	GTGACGCGACTAGTTACACATGGACTCCTTCCATGTAGTAA ATCGAGTTAGTACTGAGAGATCCGTAGATCTCCC
H10-9-1(Fwd)	GTATAATACGACTCACTATAGGGAGACGCGACTGAATGAAG GGAGATCTACGGATCTC
H10-9-2(Fwd)	GTATAATACGACTCACTATAGGGAGACGCCACTGAATGAAG GGAGATCTACGGATCTC
H10-9-3(Fwd)	GTATAATACGACTCACTATAGGGAGACGCTACTGAATGAAG GGAGATCTACGGATCTC

Name	Sequence 5' - 3'
H10-9-4(Fwd)	GTATAATACGACTCACTATAGGGTGACGCAACTGAATGAAG GGAGATCTACGGATCTC

Table A.1: Sequences of oligonucleotides ordered from IDT. T7 promoters are underlined. Sequencing barcodes are shown in blue.

Name	Sequence 5' - 3'
RNA library	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCA GGGCUCUNNNNNNNNNNNNNNNNNNNNNNGGAGCUACAUG GAAGGAGUCCAUGUGU NNNNNNNNNNNNNNNNNNNNNNUC CGUAACUAGUCGCGUCAC
169 nt	GGACAGCGCGUGUAGAAAUAUACGACUCACUAUAGGG AGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAGGG CUCUNNNNNNNNNNNNNNNNNNNNNNGGAGCUACAUGGAA GGAGUCCAUGUGUNNNNNNNNNNNNNNNNNNNNNNUCCGU AAUAGUCGCGUCAC
153 nt	GGACAGCGCGUGUAGAAAUUGGGAGACGCAACUGAAUGA AGGGAGAUCUACGGAUCUCAGGGCUCUNNNNNNNNNNNNN NNNNNNNNNGGAGCUACAUGGAAGGAGUCCAUGUGUNNN NNNNNNNNNNNNNNNNNNNNNUCCGUAACUAGUCGCGUCAC
S1	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUAAAGUACGGUCCCGCCUGAGGAGCUACAUGGAA GGAGUCCAUGUGUUCACGAAGAACGCCCCAUGUCCGUAA CUAGUCGCGUCAC
S2	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUGUUUUUUGAGCACCGGUCCUGGAGCAACAUGGA AGGAGUCCAUGUGUCUUUCUAAGCAUUCAGCUUCUCCGUA ACUAGUCGCGUCAC

Name	Sequence 5' - 3'
S3	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCAGUGUACUACUCGAAAGCGGAGCUACAUGGAA GGAGUCCAUGUGUCUGACUCGAAGAAUAUAACGUCCGUAA CUAGUCGCGUCAC
S4	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCUACAGUACUACUCGAGAAACGGAGCUACAUGGAG GAGUCCAUGUGUCUACCCAAUUUUCACUAAAGUUCCGUAA CUAGUCGCGUCAC
S5	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCUCGGUACUACUCGAAUUAACGGAGCUACAUGGAA GGAGUCCAUGUGUAGUGCGUGACACUAGAAUGUAUCCGUA ACUAGUCGCGUCAC
S6	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCACAGUACUACUCGACGUUCCGGAGCUACAUGUAA GGAGUCCAUGUGUACCUCUCGAAACACAUGUAGUUCCGUA ACUAGUCGCGUCAA
H1	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCUCACAGUACUACUCGAAAGCGGAGCUACAUGGAA GGAGUCCAUGUGUACAUUGACUCAAAAAAUGUUCCGUAA CUAGUCGCGUCAC
H2	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCAGAAUGCGGCCGCCACACGGGAGCUACAUGGAA GGAGUCCAUGUGUCUCUAAAGUUCAUUCAGCUUCCGUAA CUAGUCGCGUCAC
H3	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCUACAGUACUACUCGAAAGCGGGAGCUACAUGGAA GGAGUCCAUGUGUGUAAGCAAUAUAUACAGGUCCGUAA CUAGUCGCGUCAC

Name	Sequence 5' - 3'
H4	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAAAGACGGGAGCUACAUGGAA GGAGUCCAUGUGUAAAAAACAAAGUACUAGUAUCCGUAA CUAGUCGCGUCAC
H5	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAAAGCAGGAGCUACAUGGAA GGAGUCCAUGUGUCAUUGACUCAAAAAAAAAAUGUCCGUAA CUAGUCGCGUCAC
H6	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUAUCAUUCAAAUCCCGUAUGAGGAGCUACAUGGAA GGAGUCCAUGUGUUUACGAAUAGGAAAAAACUAUCCGUAA CUAGUCGCGUCAC
H7	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAAAGCAGGAGCUACAUGGAA GGAGUCCAUGUGUCAUUGACUCAAAAAAAAAAUGUCCGUAA CUAGUCGCGUCAC
H8	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCAGUACUACUCGAAAGCACGGGAGCUACAUGGAA GGAGUCCAUGUGUAUUGACUCAAAAAAAAAAUGUCCGUAA CUAGUCGCGUCAC
H9	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUCUCAGUACUACUCGAAAGCAGGAGCUACAUGGAA GGAGUCCAUGUGUACAUUGACUCAAAAAAAAAAUGUCCGUAA CUAGUCGCGUCAC
H10	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGGAA GGAGUCCAUGUGUCUAGAGAAAAAACUUUUCAUCCGUAA CUAGUCGCGUCAC

Name	Sequence 5' - 3'
H10-1	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGGAA GGAGUCCAUGUGUCUAGAGAAAAAACUUUUA
H10-2	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGGAA GGAGUCCAUGUGU
H10-3	GGGAGAUCUACGGAUCUCAGGGCUCUUCAGUACUACUCG AUUUACGGAGCUACAUGGAAGGAGUCCAUGUGUCUAGAGA AAAAAACUUUUAUCCGUAACUAGUCGCGUCAC
H10-4	GGGAGAUCUACGGAUCUCAGGGCUCUUCAGUACUACUCG AUUUACGGAGCUACAUGGAAGGAGUCCAUGUGUUCCGUAA CUAGUCGCGUCAC
H10-5	GGGAGAUCUACGGAUCUCAGGGCUCUUCAGUACUACUCG AUUUACGGAGCUACAUGGAAGGAGUCCAUGUGU
H10-6	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGGAA GGAGUCCAUGUGUAAAAAAAAAAAAAAAAAAUCCGUA ACUAGUCGCGUCAC
H10-7	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGGAA GGAGUCCAUGUGUUCCGUAACUAGUCGCGUCAC
Apt	GGGAGAUCUACGGAUCUCAGGGCUCUUACGGGAGCUAC AUGGAAGGAGUCCAUGUGU
Apt-1	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUC AGGGCUCUUACGGGAGCUACAUGGAAGGAGUCCAUGUG UCCGUAACUAGUCGCGUCAC
H10-7-1	GGGACUGCGCUGAUAAUGAAGGGAGAUCUACGGAUCUC AGGGCUCUUCAGUACUACUCGAUUUACGGAGCUACAUGG AAGGAGUCCAUGUGUUCCGUAACGUCAACGCAGAC

Name	Sequence 5' - 3'
H10-7-2	GGGAUCAUACCAGUAAUGAAGGGAGAUCUACGGAUCUC AGGGCUCUUCAGUACUAAACUCGAUUUACGGAGCUACAUGG AAGGAGUCCAUGUGUUCCGUAACGCUGAUUAUGAAC
H10-7-3	GGGAAGUAUAGUCGAAUGAAGGGAGAUCUACGGAUCUC AGGGCUCUUCAGUACUAAACUCGAUUUACGGAGCUACAUGG AAGGAGUCCAUGUGUUCCGUAACUGACCAUACUAC
H10-8	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG GGCUCUUCAGUACUAAACUCGAUUUACUACAUGGAAGGAGU CCAUGUGUAACUAGUCGCGUCAC
H10-9	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG UCAGUACUAAACUCGAUUUACUACAUGGAAGGAGUCCAUGU GUAACUAGUCGCGUCAC
H10-10	GGGAGACGCAACUGAAUGAAUCAGUACUAAACUCGAUUUAC UACAUGGAAGGAGUCCAUGUGUAACUAGUCGCGUCAC
H10-10-1	GGGAGACGCAACUGAAUGAAUCAGUACUAAACUCGAUUUAC UACAUGGAAGGAGUCCAUGUGUAACUAGUCGCGUCAC
H10-11	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCA GUCAGUACUAAACUCGAUUUACAACUAGUCGCGUCAC
H10-12	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCA GCAGUACUAAACUCGAAACUAGUCGCGUCAC
H10-11-1	GGGAGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCA GUCAGUACUAAACUCGAUUUACAACUAGUCGCGUCAC
H10-9-1	GGGAGACGCGACUGAAUGAAGGGAGAUCUACGGAUCUCAG UCAGUACUAAACUCGAUUUACUACAUGGAAGGAGUCCAUGU GUAACUAGUCGCGUCAC
H10-9-2	GGGAGACGCCACUGAAUGAAGGGAGAUCUACGGAUCUCAG UCAGUACUAAACUCGAUUUACUACAUGGAAGGAGUCCAUGU GUAACUAGUCGCGUCAC
H10-9-3	GGGAGACGCUACUGAAUGAAGGGAGAUCUACGGAUCUCAG UCAGUACUAAACUCGAUUUACUACAUGGAAGGAGUCCAUGU GUAACUAGUCGCGUCAC

Name	Sequence 5' - 3'
H10-9-4	GGGUGACGCAACUGAAUGAAGGGAGAUCUACGGAUCUCAG UCAGUACUAAACUCGAUUUACUACAUGGAAGGAGUCCAUGU GUAACUAGUCGCGUCAC

Table A.2: RNA sequences synthesised by in vitro transcription. In each sequence, Domain1 is shown in green and Domain2 is shown in orange.

A.2 Gel Electrophoresis

Native polyacrylamide gel electrophoresis

For electrophoresis of native polyacrylamide gel, a 10 mL mixture (sufficient for two gels) was prepared by combining 30% 29:1 acrylamide:bis-acrylamide stock solution (Sigma-Aldrich) with Tris-acetate-EDTA (TAE; Thermo Scientific) buffer and water to achieve a final 1× buffer concentration and 15% acrylamide. To this mixture, 100 µL of 10% ammonium persulphate (APS; Sigma-Aldrich) and 10 µL of tetramethylethylenediamine (TEMED; Sigma-Aldrich) were added, and the solution was briefly vortexed. The resulting mixture was poured between two glass plates with a 0.75 mm spacer, a gel comb was inserted, and the gel was left to polymerise for 15 minutes. After polymerisation, the outside of the plates was rinsed with water to remove excess gel, and the comb was carefully removed. The gels were then placed in a Mini-PROTEAN Tetra Cell vertical gel electrophoresis system (Bio-Rad), and the chamber was filled with 1× TAE buffer. Samples (with 1× loading dye; New England Biolabs) were pipetted into each well. The gels were run at 180 V for 45 minutes at room temperature.

Denaturing polyacrylamide gel electrophoresis

For denaturing polyacrylamide gel electrophoresis, a 10 mL mixture (for two gels) was prepared by combining 4.8 g urea (Sigma-Aldrich), 1 mL 10× Tris-Borate-EDTA (TBE; Thermo Scientific), 2.67 mL 30% 19:1 acrylamide:bis-acrylamide stock solution (Sigma-Aldrich), and water to achieve a final buffer concentration of 1×, with 8% acrylamide

and 8 M urea. To this mixture, 100 μ L of 10% APS and 10 μ L of TEMED were added, followed by a brief vortex. The gel casting procedure was the same as that used for the native gel preparation. The gel electrophoresis tank was filled with 1 $\times$ TBE buffer. Before loading, the samples were mixed with an equal volume of 2 $\times$ denaturing buffer (95% formamide, 0.02% SDS, 0.02% bromophenol blue, 0.01% xylene cyanol, and 100 mM EDTA) and denatured at 90°C for 3 minutes. RNA samples were then loaded into each well. The gels were run at 300 V for 45 to 75 minutes depending on the fragment length at room temperature.

Agarose gel electrophoresis

100 mL of 1 $\times$ TBE was added to 3 g of UltraPure™ agarose (Invitrogen, ThermoFisher Scientific) and the mixture was heated in a microwave oven on low power to minimise bubble formation. Once the agarose was fully dissolved, the solution was allowed to cool on the bench to approximately 50°C. Then, 5 μ L of SYBR Safe DNA Gel Stain (Invitrogen, ThermoFisher Scientific) was added and gently mixed using a pipette tip. The mixture was poured into an agarose gel tray and a gel comb was inserted. The gel required approximately 1 hour to fully solidify. Samples (with 1 $\times$ loading dye) were pipetted into each well. The gels were run in 1 $\times$ TBE buffer in a horizontal electrophoresis mini-sub cell (Bio-Rad) at 85 V for 45 minutes.

Gel visualisation

Each gel was soaked in 50 mL of the corresponding running buffer containing 2 μ L of SYBR Gold (Invitrogen, Thermo Fisher Scientific) and gently shaken on a shaker for at least 5 minutes. The gel was then rinsed in water and scanned using an Amersham Typhoon laser scanner. Scanning was performed with a pixel size of 50 μ m, using a PMT voltage determined by pre-scanning. Gel slicing was carried out using a Blue Light Transilluminator (Invitrogen, Thermo Fisher Scientific).

A.3 Supplementary Figures

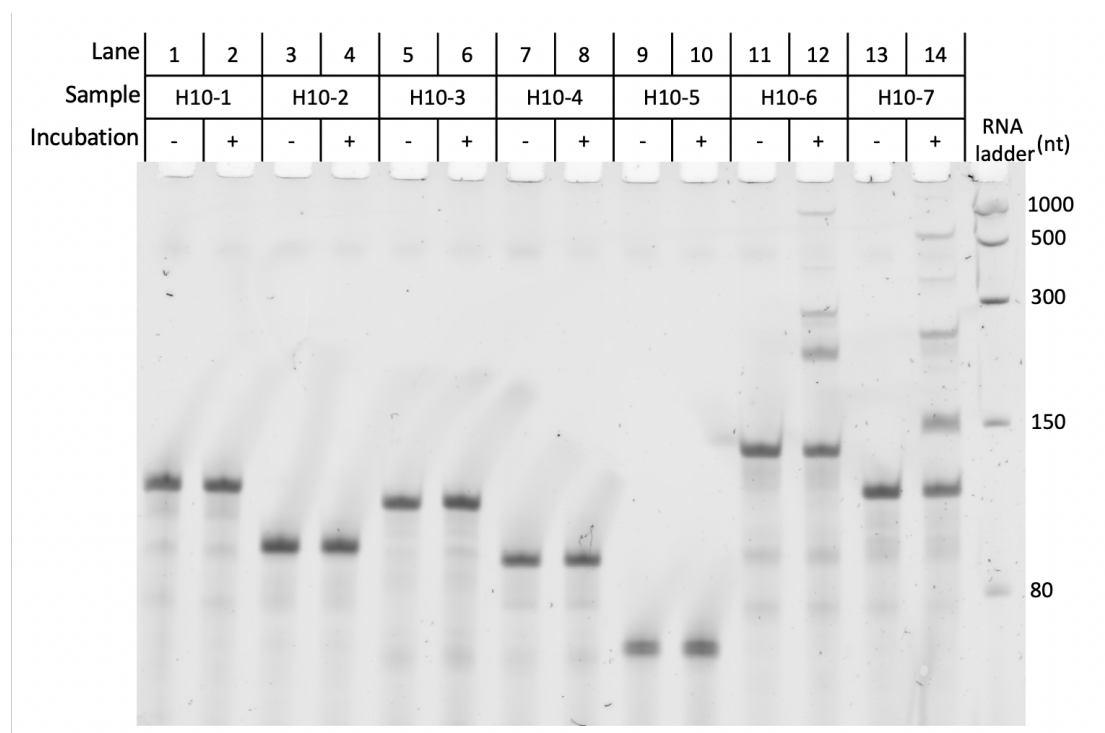


Figure A.1: Gel image showing the ligation products of truncated H10 variants, both incubated and unincubated. H10-1 is H10 with the truncation of P1, while H10-3 has P2 deleted. H10-6 contains a replacement of the D2 region with a 20-nt polyA sequence, and H10-7 has the D2 region truncated. H10-2 has deletions of both D1 and P1, H10-4 involves deletions of D1 and P2, and H10-5 has deletions of D1, P1, and P2. Of these variants, only H10-6 and H10-7 demonstrated significant catalytic activity.

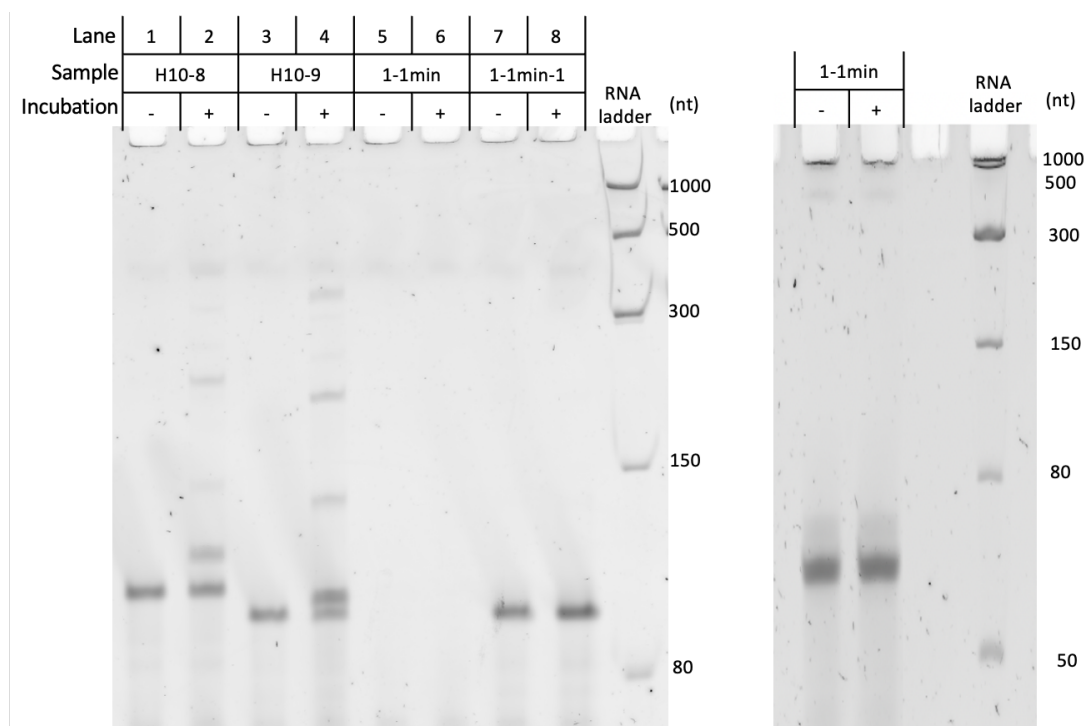


Figure A.2: Gel images showing the ligation products of truncated H10-7 variants (H10-8 and H10-9) and the aptamer with (1-1min-1) and without (1-1min) the two primer binding sites, both incubated and unincubated. In the left gel, the 1-1min aptamer samples ran off due to their short length, so they were re-analysed on the right gel with a shorter run time. The images reveal that neither the aptamer alone nor the aptamer with the primer binding sites exhibited catalytic activity, whereas H10-8 and H10-9 displayed significant and comparable catalytic activity.

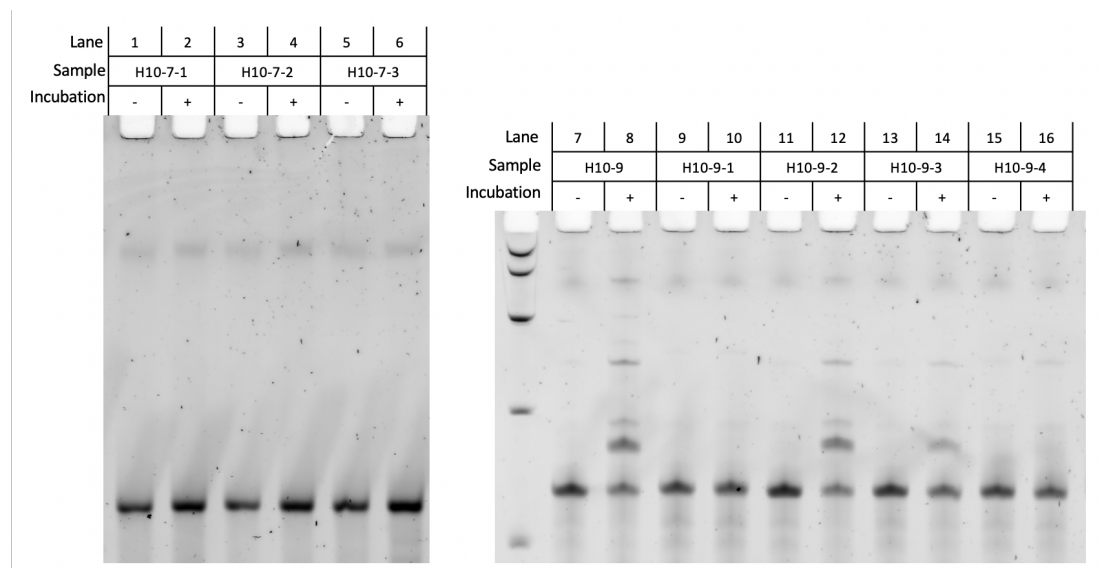


Figure A.3: Left: Gel image showing the ligation products of H10-7 with variations in the primer binding sites, both incubated and unincubated. In H10-7-1, U was substituted with A and G with C; in H10-7-2, U was substituted with G and A with C; and in H10-7-3, U was substituted with C and A with G. None of these three variants exhibited detectable catalytic activity. Right: Gel image showing the ligation products of H10-9 with variations in the primer binding sites, both incubated and unincubated. The ribozyme was deactivated when A10 was replaced with guanine (H10-9-1). It remained active though with reduced activity, when A10 was replaced with cytosine (H10-9-2) or uracil (H10-9-3). When A4 was replaced with uracil (H10-9-4), the ribozyme became inactive.

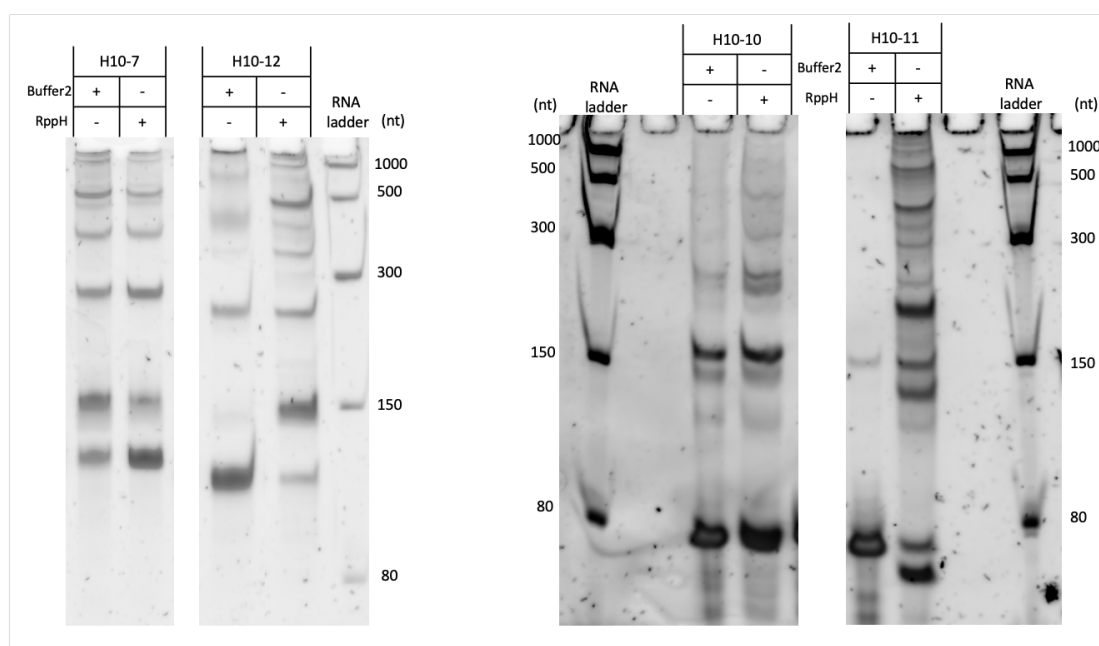


Figure A.4: Left: Gel image comparing the ligation activity of H10-7 and its minimal D1 variant (H10-12) after incubation in Buffer2 or treatment with RppH. H10-7 incubated in Buffer2 displayed 0.67-fold ligation activity compared to the RppH-treated sample, while H10-12 exhibited only 0.23-fold of the ligated product. Middle and Right: Gel images comparing the ligation activity of H10-9 with C1 deleted (H10-10), H10-10 with a 5' G deleted (H10-10-1), and H10-9 with C2 deleted (H10-11). H10-10 incubated in Buffer2 showed 0.73-fold ligation activity compared to the RppH-treated sample, while H10-10-1 and H10-11 displayed reduced ligation activity with 0.036 and 0.038-fold, respectively.

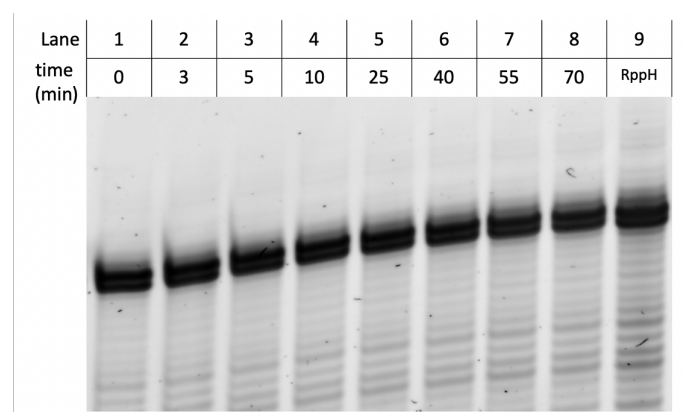


Figure A.5: Gel image of time-course incubation of H10-10 on 15% denaturing PAGE gel after running at constant power (32 W) for 6 hours. Even under these conditions, the two bands remained difficult to distinguish.

A.4 High-throughput sequencing data analysis script

The following Python script was used to analyse demultiplexed HTS data, extract the two random domains, calculate cluster-size distributions and k-mer prevalence, and track enriched sequences across selection rounds.

```
1  #!/usr/bin/env python3
2  """High-throughput sequencing analysis for SELEX rounds 5-8.
3
4  The script starts from one demultiplexed FASTA file per round and generates:
5  1. filtered Domain1 and Domain2 FASTA files;
6  2. cluster-size distributions for Domain1, Domain2, and the D1+D2 pair;
7  3. R8 prevalence of the most common 10-, 14-, and 18-mer in each domain;
8  4. enrichment profiles for the five most abundant R8 variants; and
9  5. the ten most abundant R8 variants printed to the terminal.
10
11  Raw FASTQ quality trimming, FASTQ-to-FASTA conversion, and barcode
12  demultiplexing were performed before this analysis as described in the Methods.
13  """
14
15  from collections import Counter
16  from pathlib import Path
17
18  import matplotlib.pyplot as plt
19  import numpy as np
20
21
22  INPUT_DIR = Path("path/to/demultiplexed_fasta")
23  OUTPUT_DIR = Path("path/to/output")
24  ROUND_NAMES = ["r5", "r6", "r7", "r8"]
25
26  QIAN_F = "GGGAGACGCAACTGAAT"
27  QIAN_R = "CCGTAACTAGTCGCGTCAC"
28  CONSTANT_LEFT = "GGGAGATCTACGGATCTCAGGGCTCT"
29  CONSTANT_MIDDLE = "GAGCTACATGGAAGGAGTCCATGTGT"
30
31  DOMAIN_LENGTH = 20
32  DOMAIN2_OFFSET = 0 # Change to 1 only if one known fixed base precedes Domain2.
33  KMER_LENGTHS = [10, 14, 18]
34  VALID_BASES = set("ACGT")
35
36
37  def read_fasta(fasta_file):
38      """Yield (header, sequence) records from a FASTA file."""
39      header = None
40      sequence_parts = []
41
42      with open(fasta_file, "r", encoding="utf-8") as handle:
43          for line in handle:
44              line = line.strip()
45              if not line:
46                  continue
47              if line.startswith(">"):
48                  if header is not None:
49                      yield header, "".join(sequence_parts).upper()
50                  header = line[1:]
51                  sequence_parts = []
52              else:
53                  sequence_parts.append(line)
54
```

```

55     if header is not None:
56         yield header, "".join(sequence_parts).upper()
57
58
59 def write_fasta(records, fasta_file):
60     """Write (header, sequence) records to a FASTA file."""
61     with open(fasta_file, "w", encoding="utf-8") as handle:
62         for header, sequence in records:
63             handle.write(f">{header}\n{sequence}\n")
64
65
66 def extract_domains(sequence):
67     """Return the two 20-nt random domains, or None if the read is invalid."""
68     qian_f_start = sequence.find(QIAN_F)
69     if qian_f_start == -1:
70         return None
71
72     left_start = sequence.find(CONSTANT_LEFT, qian_f_start + len(QIAN_F))
73     if left_start == -1:
74         return None
75     domain1_start = left_start + len(CONSTANT_LEFT)
76
77     middle_start = sequence.find(CONSTANT_MIDDLE, domain1_start)
78     if middle_start == -1:
79         return None
80     domain1 = sequence[domain1_start:middle_start]
81
82     domain2_start = middle_start + len(CONSTANT_MIDDLE) + DOMAIN2_OFFSET
83     qian_r_start = sequence.find(QIAN_R, domain2_start)
84     if qian_r_start == -1:
85         return None
86     domain2 = sequence[domain2_start:qian_r_start]
87
88     if len(domain1) != DOMAIN_LENGTH or len(domain2) != DOMAIN_LENGTH:
89         return None
90     if not set(domain1).issubset(VALID_BASES):
91         return None
92     if not set(domain2).issubset(VALID_BASES):
93         return None
94
95     return domain1, domain2
96
97
98 def extract_all_domains():
99     """Extract and save valid Domain1 and Domain2 sequences for every round."""
100     OUTPUT_DIR.mkdir(parents=True, exist_ok=True)
101     domains_by_round = {}
102
103     for round_name in ROUND_NAMES:
104         domain1_records = []
105         domain2_records = []
106         both_records = []
107         total_reads = 0
108
109         input_fasta = INPUT_DIR / f"{round_name}.fasta"
110         for header, sequence in read_fasta(input_fasta):
111             total_reads += 1
112             domains = extract_domains(sequence)
113             if domains is None:
114                 continue
115
116             domain1, domain2 = domains
117             domain1_records.append((header, domain1))
118             domain2_records.append((header, domain2))

```

```

119         both_records.append((header, domain1 + domain2))
120
121     write_fasta(domain1_records, OUTPUT_DIR / f"domain1_{round_name}.fasta")
122     write_fasta(domain2_records, OUTPUT_DIR / f"domain2_{round_name}.fasta")
123     write_fasta(both_records, OUTPUT_DIR / f"both_domains_{round_name}.fasta")
124
125     domains_by_round[round_name] = {
126         "domain1": [seq for _, seq in domain1_records],
127         "domain2": [seq for _, seq in domain2_records],
128         "both": [(d1, d2) for (_, d1), (_, d2) in zip(
129             domain1_records, domain2_records
130         )],
131     }
132
133     print(
134         f"{round_name.upper()}: {len(domain1_records)} valid reads "
135         f"of {total_reads} total reads"
136     )
137
138     return domains_by_round
139
140
141 def cluster_size_distribution(sequences):
142     """Return percentages of unique clusters with <5, 5-50, or >50 reads."""
143     cluster_sizes = list(Counter(sequences).values())
144     total_clusters = len(cluster_sizes)
145     if total_clusters == 0:
146         return {"<5": 0.0, "5-50": 0.0, ">50": 0.0}
147
148     return {
149         "<5": sum(size < 5 for size in cluster_sizes) / total_clusters * 100,
150         "5-50": sum(5 <= size <= 50 for size in cluster_sizes) /
151             total_clusters * 100,
152         ">50": sum(size > 50 for size in cluster_sizes) / total_clusters * 100,
153     }
154
155
156 def plot_cluster_size_distributions(domains_by_round):
157     """Plot cluster-size distributions for D1+D2, Domain1, and Domain2."""
158     datasets = {
159         "Domain1 + Domain2": "both",
160         "Domain1": "domain1",
161         "Domain2": "domain2",
162     }
163     categories = ["<5", "5-50", ">50"]
164
165     fig, axes = plt.subplots(1, 3, figsize=(15, 5), sharey=True)
166
167     for ax, (title, domain_key) in zip(axes, datasets.items()):
168         values = {category: [] for category in categories}
169         for round_name in ROUND_NAMES:
170             distribution = cluster_size_distribution(
171                 domains_by_round[round_name][domain_key]
172             )
173             for category in categories:
174                 values[category].append(distribution[category])
175
176         bottom = np.zeros(len(ROUND_NAMES))
177         for category in categories:
178             heights = np.asarray(values[category], dtype=float)
179             ax.bar(ROUND_NAMES, heights, bottom=bottom, label=category)
180             bottom += heights
181
182         ax.set_title(title, fontsize=14)

```

```

183     ax.set_ylim(0, 100)
184     ax.set_xlabel("SELEX rounds", fontsize=12)
185     ax.legend(title="Cluster size", fontsize=10)
186
187     axes[0].set_ylabel("Fraction of unique clusters (%)", fontsize=12)
188     plt.tight_layout()
189     plt.savefig(
190         OUTPUT_DIR / "cluster_size_distribution.pdf",
191         bbox_inches="tight",
192         dpi=300,
193     )
194     plt.close()
195
196
197 def most_prevalent_kmer(sequences, kmer_length):
198     """Return the top k-mer and the percentage of reads containing it."""
199     if not sequences:
200         return None, 0.0
201
202     reads_containing = Counter()
203     for sequence in sequences:
204         kmers_in_read = {
205             sequence[i:i + kmer_length]
206             for i in range(len(sequence) - kmer_length + 1)
207         }
208         reads_containing.update(kmers_in_read)
209
210     if not reads_containing:
211         return None, 0.0
212
213     top_kmer, supporting_reads = reads_containing.most_common(1)[0]
214     prevalence = supporting_reads / len(sequences) * 100
215     return top_kmer, prevalence
216
217
218 def plot_r8_kmer_prevalence(domains_by_round):
219     """Compare top-k-mer prevalence between Domain1 and Domain2 in R8."""
220     r8_domains = {
221         "Domain1": domains_by_round["r8"]["domain1"],
222         "Domain2": domains_by_round["r8"]["domain2"],
223     }
224     fractions = {domain: [] for domain in r8_domains}
225
226     for domain, sequences in r8_domains.items():
227         for kmer_length in KMER_LENGTHS:
228             top_kmer, prevalence = most_prevalent_kmer(sequences, kmer_length)
229             fractions[domain].append(prevalence)
230             print(
231                 f"R8 {domain}, {kmer_length}-mer: {top_kmer}, "
232                 f"{prevalence:.3f}% of valid reads"
233             )
234
235     x = np.arange(2)
236     width = 0.22
237     fig, ax = plt.subplots(figsize=(8, 5))
238
239     for index, kmer_length in enumerate(KMER_LENGTHS):
240         values = [fractions["Domain1"][index], fractions["Domain2"][index]]
241         offset = (index - (len(KMER_LENGTHS) - 1) / 2) * width
242         ax.bar(x + offset, values, width, label=f"{kmer_length}-mer")
243
244     ax.set_ylabel("Reads containing the most prevalent k-mer (%)", fontsize=12)
245     ax.set_xticks(x)
246     ax.set_xticklabels(["Domain1", "Domain2"], fontsize=12)

```

```

247     ax.legend(title="k-mer length", fontsize=10)
248     plt.tight_layout()
249     plt.savefig(
250         OUTPUT_DIR / "r8_top_kmer_prevalence.pdf",
251         bbox_inches="tight",
252         dpi=300,
253     )
254     plt.close()
255
256
257 def variant_percentages(domains_by_round):
258     """Calculate D1+D2 variant percentages in each selection round."""
259     counts_by_round = {
260         round_name: Counter(domains_by_round[round_name]["both"])
261         for round_name in ROUND_NAMES
262     }
263     all_variants = set().union(*(counts.keys() for counts in counts_by_round.values()))
264
265     percentages = {}
266     for variant in all_variants:
267         percentages[variant] = {}
268         for round_name in ROUND_NAMES:
269             total_reads = sum(counts_by_round[round_name].values())
270             count = counts_by_round[round_name].get(variant, 0)
271             percentages[variant][round_name] = (
272                 count / total_reads * 100 if total_reads else 0.0
273             )
274
275     return percentages
276
277
278 def plot_top_five_enrichment(percentages):
279     """Plot R5-R8 abundance of the five most abundant R8 variants."""
280     top_five = sorted(
281         percentages.items(),
282         key=lambda item: item[1]["r8"],
283         reverse=True,
284     )[:5]
285
286     x = np.arange(len(ROUND_NAMES))
287     fig, ax = plt.subplots(figsize=(6, 4))
288
289     for rank, (_, values) in enumerate(top_five, start=1):
290         y = [values[round_name] for round_name in ROUND_NAMES]
291         ax.plot(x, y, label=f"Rank {rank}", linewidth=2)
292
293     ax.set_xticks(x)
294     ax.set_xticklabels(["R5", "R6", "R7", "R8"])
295     ax.set_xlabel("SELEX rounds", fontsize=12)
296     ax.set_ylabel("Abundance in sublibrary (%)", fontsize=12)
297     ax.legend(title="R8 rank", fontsize=10)
298     plt.tight_layout()
299     plt.savefig(
300         OUTPUT_DIR / "top_5_r8_variants_across_rounds.pdf",
301         bbox_inches="tight",
302         dpi=300,
303     )
304     plt.close()
305
306
307 def print_top_ten_r8(percentages):
308     """Print the ten most abundant D1+D2 variants in R8."""
309     top_ten = sorted(
310         percentages.items(),

```

```

311         key=lambda item: item[1]["r8"],
312         reverse=True,
313     )[:10]
314
315     print("Top 10 variants in R8:")
316     for rank, (variant, values) in enumerate(top_ten, start=1):
317         domain1, domain2 = variant
318         print(
319             f"Rank {rank}: {values['r8']:.4f}% | "
320             f"D1={domain1} | D2={domain2}"
321         )
322
323
324 def main():
325     domains_by_round = extract_all_domains()
326     plot_cluster_size_distributions(domains_by_round)
327     plot_r8_kmer_prevalence(domains_by_round)
328
329     percentages = variant_percentages(domains_by_round)
330     plot_top_five_enrichment(percentages)
331     print_top_ten_r8(percentages)
332
333
334 if __name__ == "__main__":
335     main()

```

Listing A.1: High-throughput sequencing data analysis script

A.5 Single-turnover kinetics analysis script

The following script was used to analyse the single-turnover kinetics of the ribozyme. The script fits the experimental data to a kinetic model and generates plots showing the fraction reacted over time along with error bars representing the standard deviation between experimental replicates:

```

1
2 import numpy as np
3 import matplotlib.pyplot as plt
4 from scipy.optimize import curve_fit
5
6 time = np.array([0, 1, 3, 5, 10, 25, 40, 55, 70])
7
8 fraction_reacted_trial1 = np.array([0.00, 0.03, 0.10, 0.25, 0.55, 0.71, 0.72, 0.84,
9                                     0.91])
10 fraction_reacted_trial2 = np.array([0.00, 0.06, 0.11, 0.11, 0.52, 0.76, 0.72, 0.85,
11                                     0.90])
12 fraction_reacted_trial3 = np.array([0.00, 0.19, 0.26, 0.31, 0.44, 0.64, 0.70, 0.76,
13                                     0.76])

```

```

11
12 fraction_reacted_avg = (fraction_reacted_trial1 + fraction_reacted_trial2 +
    fraction_reacted_trial3) / 3
13 fraction_reacted_std = np.std([fraction_reacted_trial1, fraction_reacted_trial2,
    fraction_reacted_trial3], axis=0)
14
15 def fit_func(t, Fmax, k_obs):
16     return Fmax * (1 - np.exp(-k_obs * t))
17
18 popt, pcov = curve_fit(fit_func, time, fraction_reacted_avg, p0=[1, 0.1])
19 t_fit = np.linspace(0, 70, 100)
20 fitted_curve = fit_func(t_fit, *popt)
21
22 plt.errorbar(time, fraction_reacted_avg, yerr=fraction_reacted_std, fmt='o', label=
    'Data with error bars')
23 plt.plot(t_fit, fitted_curve, label=f'Fitted curve: Fmax={popt[0]:.2f}, k_obs={popt
    [1]:.2f}', color='red')
24 plt.xlabel('Time (min)')
25 plt.ylabel('Fraction reacted')
26 plt.title('Single-turnover kinetics of ribozyme')
27 plt.legend()
28 plt.show()
29 Fmax, k_obs = popt
30 print(f"Fmax: {Fmax:.2f}, k_obs: {k_obs:.2f}")

```

Listing A.2: Single-turnover kinetics analysis script

Bibliography

1. Maxwell, J. C., Garber, E., Brush, S. G. & Everitt, C. W. F. (W. F. Maxwell on heat and statistical mechanics: on "avoiding all personal enquiries" of molecules. *Choice Reviews Online* **33**, 33–2107 (1995).
2. Purcell, E. M. Life at Low Reynolds Number. *American Journal of Physics* **45**, 3–11 (1977).
3. Sauvage, J.-P. From Chemical Topology to Molecular Machines (Nobel Lecture). *Angewandte Chemie (International ed. in English)* **56**, 11080–11093 (2017).
4. Feringa, B. L. The Art of Building Small: From Molecular Switches to Motors (Nobel Lecture). *Angewandte Chemie (International ed. in English)* **56**, 11060–11078 (2017).
5. Stoddart, J. F. Mechanically Interlocked Molecules (MIMs)-Molecular Shuttles, Switches, and Machines (Nobel Lecture). *Angewandte Chemie (International ed. in English)* **56**, 11094–11125 (2017).
6. Astumian, R. D. & Derényi, I. Fluctuation driven transport and models of molecular motors and pumps. *European Biophysics Journal* **27**, 474–489 (1998).
7. Kay, E. R., Leigh, D. A. & Zerbetto, F. Synthetic Molecular Motors and Mechanical Machines. *Angewandte Chemie International Edition* **46**, 72–191 (2007).
8. Borsley, S., Leigh, D. A. & Roberts, B. M. W. Molecular Ratchets and Kinetic Asymmetry: Giving Chemistry Direction. *Angewandte Chemie International Edition* **63** (2024).
9. Wilson, M. R. *et al.* An autonomous chemically fuelled small-molecule motor. *Nature* **534**, 235–240 (2016).

10. Green, S. J., Bath, J. & Turberfield, A. J. Coordinated chemomechanical cycles: A mechanism for autonomous molecular motion. *Physical Review Letters* **101**, 20–23 (2008).
11. Vale, R. D. & Milligan, R. A. The way things move: Looking under the hood of molecular motor proteins. *Science* **288**, 88 (2000).
12. Wang, W., Cao, L., Wang, C., Gigant, B. & Knossow, M. Kinesin, 30 years later: Recent insights from structural studies. *Protein Science* **24**, 1047–1056 (2015).
13. Spudich, J. A. & Sivaramakrishnan, S. Myosin VI: an innovative motor that challenged the swinging lever arm hypothesis. *Nature reviews. Molecular cell biology* **11**, 128–37 (2010).
14. Sweeney, H. L. & Houdusse, A. Structural and Functional Insights into the Myosin Motor Mechanism. *Annual Review of Biophysics* **39**, 539–557 (2010).
15. Del Grosso, E., Franco, E., Prins, L. J. & Ricci, F. Dissipative DNA nanotechnology. *Nature chemistry* **14**, 600–613 (2022).
16. Dunn, M. R., Jimenez, R. M. & Chaput, J. C. Analysis of aptamer discovery and technology. *Nature Reviews Chemistry* **1**, 0076 (2017).
17. Ma, L. & Liu, J. Catalytic Nucleic Acids: Biochemistry, Chemical Biology, Biosensors, and Nanotechnology. *iScience* **23**, 100815 (2020).
18. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nature reviews. Molecular cell biology* **5**, 451–63 (2004).
19. Simmel, F. C. in *Natural Computing Series* 17–29 (Springer Science and Business Media Deutschland GmbH, 2023).
20. Feynman, R. P. There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics. *Engineering and Science*, 3–12 (1960).
21. Drexler, K. E. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proceedings of the National Academy of Sciences* **78**, 5275–5278 (1981).

22. Astumian, R. D. Design principles for Brownian molecular machines: how to swim in molasses and walk in a hurricane. *Physical Chemistry Chemical Physics* **9**, 5067 (2007).
23. Ait-Haddou, R. & Herzog, W. Brownian Ratchet Models of Molecular Motors. *Cell Biochemistry and Biophysics* **38**, 191–214 (2003).
24. Astumian, R. D. Thermodynamics and kinetics of a Brownian motor. *Science* **276**, 917–922 (1997).
25. Bustamante, C., Keller, D. & Oster, G. The physics of molecular motors. *Accounts of Chemical Research* **34**, 412–420 (2001).
26. Oster, G. & Wang, H. Rotary protein motors. *Trends in cell biology* **13**, 114–21 (2003).
27. Magnasco, M. O. & Stolovitzky, G. Feynman's ratchet and pawl. *Journal of Statistical Physics* **93**, 615–632 (1998).
28. Lipowsky, R. & Klumpp, S. 'Life is motion': Multiscale motility of molecular motors. *Physica A: Statistical Mechanics and its Applications* **352**, 53–112 (2005).
29. Block, S. M. Kinesin motor mechanics: binding, stepping, tracking, gating, and limping. *Biophysical journal* **92**, 2986–95 (2007).
30. Bhabha, G., Johnson, G. T., Schroeder, C. M. & Vale, R. D. How Dynein Moves Along Microtubules. *Trends in Biochemical Sciences* **41**, 94–105 (2016).
31. Spudich, J. A. The myosin swinging cross-bridge model. *Nature reviews. Molecular cell biology* **2**, 387–92 (2001).
32. Boyer, P. D. The ATP synthase—a splendid molecular machine. *Annual review of biochemistry* **66**, 717–49 (1997).
33. Schliwa, M. & Woehlke, G. Molecular motors. *Nature* **422**, 759–65 (2003).
34. Feynman, R. P., Leighton, R. B. & Sands, M. *Ratchet and pawl. The Feynman lectures on physics* 46.1–46.9 (1963).

35. Barrell, M. J., Campaña, A. G., von Delius, M., Geertsema, E. M. & Leigh, D. A. Light-Driven Transport of a Molecular Walker in Either Direction along a Molecular Track. *Angewandte Chemie International Edition* **50**, 285–290 (2011).
36. Erbas-Cakmak, S. *et al.* Rotary and linear molecular motors driven by pulses of a chemical fuel. *Science* **358**, 340–343 (2017).
37. Thomas, D. *et al.* Pumping between phases with a pulsed-fuel molecular ratchet. *Nature Nanotechnology* **17**, 701–707 (2022).
38. Nguyen, B.-N. T., Grommet, A. B., Tron, A., Georges, M. C. A. & Nitschke, J. R. Heat Engine Drives Transport of an Fe II 4 L 4 Cage and Cargo. *Advanced Materials* **32**, e1907241 (2020).
39. Li, A. *et al.* Precise Control of Radial Catenane Synthesis via Clipping and Pumping. *Journal of the American Chemical Society* **144**, 2085–2089 (2022).
40. Pezzato, C. *et al.* Controlling Dual Molecular Pumps Electrochemically. *Angewandte Chemie International Edition* **57**, 9325–9329 (2018).
41. Kudernac, T. *et al.* Electrically driven directional motion of a four-wheeled molecule on a metal surface. *Nature* **479**, 208–211 (2011).
42. Hwang, W. & Karplus, M. Structural basis for power stroke vs. Brownian ratchet mechanisms of motor proteins. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 19777–19785 (2019).
43. Karplus, M. & Gao, Y. Q. Biomolecular motors: the F1-ATPase paradigm. *Current Opinion in Structural Biology* **14**, 250–259 (2004).
44. Borsley, S., Leigh, D. A. & Roberts, B. M. W. Chemical fuels for molecular machinery. *Nature chemistry* **14**, 728–738 (2022).
45. Amano, S., Borsley, S., Leigh, D. A. & Sun, Z. Chemical engines: driving systems away from equilibrium through catalyst reaction cycles. *Nature nanotechnology* **16**, 1057–1067 (2021).

46. Amano, S. *et al.* Using Catalysis to Drive Chemistry Away from Equilibrium: Relating Kinetic Asymmetry, Power Strokes, and the Curtin–Hammett Principle in Brownian Ratchets. *Journal of the American Chemical Society* **144**, 20153–20164 (2022).
47. Seeman, J. I. Effect of Conformational Change on Reactivity in Organic Chemistry. Evaluations, Applications, and Extensions of Curtin-Hammett/Winstein-Holness Kinetics. *Chemical Reviews* **83**, 83–134 (1983).
48. Seeman, J. I. The curtin-hammett principle and the winstein-holness equation: New definition and recent extensions to classical concepts. *Journal of Chemical Education* **63**, 42 (1986).
49. Ragazzon, G. & Prins, L. J. Energy consumption in chemical fuel-driven self-assembly. *Nature Nanotechnology* **13**, 882–889 (2018).
50. Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin Walks Hand-Over-Hand. *Science* **303**, 676–678 (2004).
51. Astumian, R. D. Irrelevance of the power stroke for the directionality, stopping force, and optimal efficiency of chemically driven molecular machines. *Biophysical Journal* **108**, 291–303 (2015).
52. Wagoner, J. A. & Dill, K. A. Molecular Motors: Power Strokes Outperform Brownian Ratchets. *Journal of Physical Chemistry B* **120**, 6327–6336 (2016).
53. Borsley, S. Membrane Transport, Molecular Machines, and Maxwell’s Demon. *ChemSystemsChem* **6**, e202400004 (2024).
54. Borsley, S., Kreidt, E., Leigh, D. A. & Roberts, B. M. Autonomous fuelled directional rotation about a covalent single bond. *Nature* **604**, 80–85 (2022).
55. Delius, M. V. & Leigh, D. A. Walking molecules. *Chemical Society Reviews* **40**, 3656–3676 (2011).
56. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nature Reviews Materials* **3**, 17068 (2017).
57. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).

58. Neidle, S. & Sanderson, M. *Principles of Nucleic Acid Structure* (Elsevier, 2021).
59. Guckian, K. M. *et al.* Factors contributing to aromatic stacking in water: Evaluation in the context of DNA. *Journal of the American Chemical Society* **122**, 2213–2222 (2000).
60. Hagerman, P. J. Flexibility of DNA. *Annual Review of Biophysics and Biophysical Chemistry* **17**, 265–286 (1988).
61. Abels, J., Moreno-Herrero, F., van der Heijden, T., Dekker, C. & Dekker, N. Single-Molecule Measurements of the Persistence Length of Double-Stranded RNA. *Biophysical Journal* **88**, 2737–2744 (2005).
62. Vella, F. Nucleic Acids in Chemistry and Biology. *Biochemical Education* **19**, 97–98 (1991).
63. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences* **109**, 799–804 (2012).
64. Roth, E., Glick Azaria, A., Girshevitz, O., Bitler, A. & Garini, Y. Measuring the Conformation and Persistence Length of Single-Stranded DNA Using a DNA Origami Structure. *Nano Letters* **18**, 6703–6709 (2018).
65. Hyeon, C., Dima, R. I. & Thirumalai, D. Size, shape, and flexibility of RNA structures. *Journal of Chemical Physics* **125** (2006).
66. DeVoe, H. & Tinoco, I. The stability of helical polynucleotides: Base contributions. *Journal of Molecular Biology* **4**, 500–517 (1962).
67. Crothers, D. M. & Zimm, B. H. Theory of the melting transition of synthetic polynucleotides: Evaluation of the stacking free energy. *Journal of Molecular Biology* **9**, 1–9 (1964).
68. SantaLucia, J. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1460–1465 (1998).
69. Freier, S. M. *et al.* Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Sciences* **83**, 9373–9377 (1986).

70. Allawi, H. T. & Santalucia, J. Thermodynamics and NMR of internal G-T mismatches in DNA. *Biochemistry* **36**, 10581–10594 (1997).
71. Masliah, G., Barraud, P. & Allain, F. H. -. RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cellular and Molecular Life Sciences* **70**, 1875–1895 (2012).
72. Pyle, A. M. & Green, J. B. RNA folding. *Current Opinion in Structural Biology* **5**, 303–310 (1995).
73. Oivanen, M., Kuusela, S. & Lönnberg, H. Kinetics and mechanisms for the cleavage and isomerization of the phosphodiester bonds of RNA by bronsted acids and bases. *Chemical Reviews* **98**, 961–990 (1998).
74. Carlson, R. The changing economics of DNA synthesis. *Nature biotechnology* **27**, 1091–4 (2009).
75. Seeman, N. C. Nucleic acid junctions and lattices. *Journal of Theoretical Biology* **99**, 237–247 (1982).
76. Ke, Y. Designer three-dimensional DNA architectures. *Current opinion in structural biology* **27**, 122–8 (2014).
77. Seeman, N. C. Nanomaterials based on DNA. *Annual review of biochemistry* **79**, 65–87 (2010).
78. Chen, J. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
79. Shih, W. M., Quispe, J. D. & Joyce, G. F. A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **427**, 618–621 (2004).
80. Goodman, R. P. *et al.* Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661–1665 (2005).
81. Aldaye, F. A. & Sleiman, H. F. Modular access to structurally switchable 3D discrete DNA assemblies. *Journal of the American Chemical Society* **129**, 13376–13377 (2007).

82. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–544 (1998).
83. Lin, C., Liu, Y., Rinker, S. & Yan, H. DNA Tile Based Self-Assembly: Building Complex Nanoarchitectures. *ChemPhysChem* **7**, 1641–1647 (2006).
84. Hamada, S. & Murata, S. Substrate-Assisted Assembly of Interconnected Single-Duplex DNA Nanostructures. *Angewandte Chemie International Edition* **48**, 6820–6823 (2009).
85. Tian, C. *et al.* Directed Self-Assembly of DNA Tiles into Complex Nanocages. *Angewandte Chemie International Edition* **53**, 8041–8044 (2014).
86. He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198–201 (2008).
87. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
88. Ashworth, C. DNA nanotechnology: Building big with DNA bricks. *Nature Reviews Materials* **3**, 17092 (2018).
89. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
90. Zhan, P. *et al.* Recent Advances in DNA Origami-Engineered Nanomaterials and Applications. *Chemical reviews* **123**, 3976–4050 (2023).
91. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
92. Jasinski, D., Haque, F., Binzel, D. W. & Guo, P. Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* **11**, 1142–1164 (2017).
93. Geary, C., Grossi, G., McRae, E. K., Rothemund, P. W. & Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nature Chemistry* **13**, 549–558 (2021).
94. Guo, P. The emerging field of RNA nanotechnology. *Nature nanotechnology* **5**, 833–42 (2010).

95. Zhu, Y., Zhu, L., Wang, X. & Jin, H. RNA-based therapeutics: an overview and prospectus. *Cell Death & Disease* **13**, 644 (2022).
96. Kim, Y.-K. RNA therapy: rich history, various applications and unlimited future prospects. *Experimental & Molecular Medicine* **54**, 455–465 (2022).
97. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Current opinion in structural biology* **16**, 279–87 (2006).
98. Schroeder, K. T., McPhee, S. A., Ouellet, J. & Lilley, D. M. A structural database for k-turn motifs in RNA. *RNA* **16**, 1463–1468 (2010).
99. Zhang, C. *et al.* Conformational flexibility facilitates self-assembly of complex DNA nanostructures. *Proceedings of the National Academy of Sciences* **105**, 10665–10669 (2008).
100. Reynaldo, L. P., Vologodskii, A. V., Neri, B. P. & Lyamichev, V. I. The kinetics of oligonucleotide replacements. *Journal of Molecular Biology* **297**, 511–520 (2000).
101. Green, C. & Tibbetts, C. Reassociation rate limited displacement of DNA strands by branch migration. *Nucleic Acids Research* **9**, 1905–1918 (1981).
102. Morrison, L. E. & Stols, L. M. Sensitive Fluorescence-Based Thermodynamic and Kinetic Measurements of DNA Hybridization in Solution. *Biochemistry* **32**, 3095–3104 (1993).
103. Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406**, 605–608 (2000).
104. Simmel, F. C., Yurke, B. & Singh, H. R. Principles and Applications of Nucleic Acid Strand Displacement Reactions. *Chemical Reviews* (2019).
105. Yurke, B. & Mills, A. P. Using DNA to power nanostructures. *Genetic Programming and Evolvable Machines* **4**, 111–122 (2003).
106. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society* **131**, 17303–17314 (2009).

107. Srinivas, N. *et al.* On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucleic Acids Research* **41**, 10641–10658 (2013).
108. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry* **3**, 103–113 (2011).
109. Sherman, W. B. & Seeman, N. C. A precisely controlled DNA biped walking device. *Nano Letters* **4**, 1203–1207 (2004).
110. Gu, H., Chao, J., Xiao, S. J. & Seeman, N. C. A proximity-based programmable DNA nanoscale assembly line. *Nature* **465**, 202–205 (2010).
111. Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J. & Reif, J. H. A Unidirectional DNA Walker That Moves Autonomously along a Track. *Angewandte Chemie* **116**, 5014–5019 (2004).
112. Bath, J., Green, S. J. & Turberfield, A. J. A Free-Running DNA Motor Powered by a Nicking Enzyme. *Angewandte Chemie International Edition* **44**, 4358–4361 (2005).
113. Tian, Y., He, Y., Chen, Y., Yin, P. & Mao, C. A DNAzyme That Walks Processively and Autonomously along a One-Dimensional Track. *Angewandte Chemie* **117**, 4429–4432 (2005).
114. Pan, J., Li, F., Cha, T.-G., Chen, H. & Choi, J. H. Recent progress on DNA based walkers. *Current opinion in biotechnology* **34**, 56–64 (2015).
115. Wickham, S. F. *et al.* Direct observation of stepwise movement of a synthetic molecular transporter. *Nature Nanotechnology* **6**, 166–169 (2011).
116. Cha, T. G. *et al.* A synthetic DNA motor that transports nanoparticles along carbon nanotubes. *Nature Nanotechnology* **9**, 39–43 (2014).
117. Bazrafshan, A. *et al.* Tunable DNA Origami Motors Translocate Ballistically Over μm Distances at nm/s Speeds. *Angewandte Chemie International Edition* **59**, 9514–9521 (2020).
118. Valero, J. & Famulok, M. Regeneration of Burnt Bridges on a DNA Catenane Walker. *Angewandte Chemie International Edition* **59**, 16366–16370 (2020).

119. Wickham, S. F. *et al.* A DNA-based molecular motor that can navigate a network of tracks. *Nature Nanotechnology* **7**, 169–173 (2012).
120. Simmel, F. C. Processive motion of bipedal DNA walkers. *ChemPhysChem* **10**, 2593–2597 (2009).
121. Turberfield, A. J. *et al.* DNA Fuel for Free-Running Nanomachines. *Physical Review Letters* **90**, 4 (2003).
122. Green, S. J., Lubrich, D. & Turberfield, A. J. DNA hairpins: Fuel for autonomous DNA devices. *Biophysical Journal* **91**, 2966–2975 (2006).
123. Bath, J. & Turberfield, A. J. DNA nanomachines. *Nature Nanotechnology* **2**, 275–284 (2007).
124. Riehlman, T. D., Olmsted, Z. T. & Paluh, J. L. in *The Nanobiotechnology Handbook* 73–112 (Garland Science, 2012).
125. Wang, Z. G., Elbaz, J. & Willner, I. DNA machines: Bipedal walker and stepper. *Nano Letters* **11**, 304–309 (2011).
126. You, M., Huang, F., Chen, Z., Wang, R. W. & Tan, W. Building a nanostructure with reversible motions using photonic energy. *ACS Nano* **6**, 7935–7941 (2012).
127. Pumm, A. K. *et al.* A DNA origami rotary ratchet motor. *Nature* **607**, 492–498 (2022).
128. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology* **15**, 342–348 (2005).
129. Nudler, E. & Gottesman, M. E. Transcription termination and anti-termination in *E. coli*. *Genes to Cells* **7**, 755–768 (2002).
130. Hammann, C. & Westhof, E. Searching genomes for ribozymes and riboswitches. *Genome biology* **8**, 210 (2007).
131. Noller, H. F., Hoffarth, V. & Zimniak, L. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416–1419 (1992).

132. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. in *Structural Insights into Gene Expression and Protein Synthesis* 5481, 501–511 (American Association for the Advancement of Science, 2020).
133. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289**, 905–920 (2000).
134. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
135. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
136. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).
137. Mehta, J. *et al.* In vitro selection and characterization of DNA aptamers recognizing chloramphenicol. *Journal of Biotechnology* **155**, 361–369 (2011).
138. Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends in Biochemical Sciences* **41**, 595–609 (2016).
139. Zhou, W., Saran, R. & Liu, J. Metal Sensing by DNA. *Chemical Reviews* **117**, 8272–8325 (2017).
140. Silverman, S. K. in *Functional Nucleic Acids for Analytical Applications* 47–108 (Springer New York, New York, NY, 2009).
141. Micura, R. & Höbartner, C. Fundamental studies of functional nucleic acids: Aptamers, riboswitches, ribozymes and DNAzymes. *Chemical Society Reviews* **49**, 7331–7353 (2020).
142. Pauling, L. Nature of forces between large molecules of biological interest. *Nature* **161**, 707–709 (1948).
143. Batey, R. T., Rambo, R. P. & Doudna, J. A. Tertiary Motifs in RNA Structure and Folding. *Angewandte Chemie International Edition* **38**, 2326–2343 (1999).

144. Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).
145. Varani, G. & McClain, W. H. The G-U wobble base pair diverse biological systems. *EMBO Reports* **1**, 18–23 (2000).
146. Stombaugh, J., Zirbel, C. L., Westhof, E. & Leontis, N. B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Research* **37**, 2294–2312 (2009).
147. Nagaswamy, U. *et al.* NCIR: A database of non-canonical interactions in known RNA structures. *Nucleic Acids Research* **30**, 395–397 (2002).
148. Masquida, B. & Westhof, E. On the wobble GoU and related pairs. *RNA* **6**, 9–15 (2000).
149. Westhof, E. & Auffinger, P. RNA Tertiary Structure. *Encyclopedia of Analytical Chemistry*, 5222–5232 (2000).
150. Ananth, P., Goldsmith, G. & Yathindra, N. An innate twist between Crick’s wobble and Watson-Crick base pairs. *RNA* **19**, 1038–1053 (2013).
151. Hermann, T. & Westhof, E. Non-Watson-Crick base pairs in RNA-protein recognition. *Chemistry & Biology* **6**, R335–R343 (1999).
152. Taylor, K. & Sobczak, K. Intrinsic Regulatory Role of RNA Structural Arrangement in Alternative Splicing Control. *International Journal of Molecular Sciences* **21**, 5161 (2020).
153. Butcher, S. E. & Pyle, A. M. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Accounts of Chemical Research* **44**, 1302–1311 (2011).
154. Fiore, J. L. & Nesbitt, D. J. An RNA folding motif: GNRA tetraloop-receptor interactions. *Quarterly Reviews of Biophysics* **46**, 223–264 (2013).
155. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: A new RNA secondary structure motif. *EMBO Journal* **20**, 4214–4221 (2001).

156. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: The A-minor motif. *Proceedings of the National Academy of Sciences* **98**, 4899–4903 (2001).
157. Doherty, E. A., Batey, R. T., Masquida, B. & Doudna, J. A. A universal mode of helix packing in RNA. *Nature Structural Biology* **8**, 339–343 (2001).
158. Pleij, C. W. RNA pseudoknots. *Current Opinion in Structural Biology* **4**, 337–344 (1994).
159. Spiegel, J., Adhikari, S. & Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry* **2**, 123–136 (2020).
160. Misra, V. K. & Draper, D. E. On the role of magnesium ions in RNA stability. *Biopolymers* **48**, 113–135 (1999).
161. Tinoco, I. & Bustamante, C. How RNA folds. *Journal of Molecular Biology* **293**, 271–281 (1999).
162. Draper, D. E. A guide to ions and RNA structure. *RNA* **10**, 335–43 (2004).
163. Manning, G. S. The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Quarterly Reviews of Biophysics* **11**, 179–246 (1978).
164. Sun, L. Z., Zhang, D. & Chen, S. J. Theory and Modeling of RNA Structure and Interactions with Metal Ions and Small Molecules. *Annual Review of Biophysics* **46**, 227–246 (2017).
165. Woodson, S. A. Metal ions and RNA folding: a highly charged topic with a dynamic future. *Current Opinion in Chemical Biology* **9**, 104–109 (2005).
166. Draper, D. E. RNA folding: Thermodynamic and molecular descriptions of the roles of ions. *Biophysical Journal* **95**, 5489–5495 (2008).
167. Du, X. *et al.* Insights into Protein-Ligand Interactions: Mechanisms, Models, and Methods. *International journal of molecular sciences* **17**, 144 (2016).
168. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, 2985–2993 (1894).

169. Sacanna, S., Irvine, W. T., Chaikin, P. M. & Pine, D. J. Lock and key colloids. *Nature* **464**, 575–578 (2010).
170. Odriozola, G. & Lozada-Cassou, M. Statistical mechanics approach to lock-key supramolecular chemistry interactions. *Physical Review Letters* **110**, 105701 (2013).
171. Sacanna, S. *et al.* Shaping colloids for self-assembly. *Nature Communications* **4**, 1–6 (2013).
172. Jin, Z. & Wu, J. Hybrid MC-DFT method for studying multidimensional entropic forces. *Journal of Physical Chemistry B* **115**, 1450–1460 (2011).
173. Chang, H. Y. *et al.* Assembly of Lock-and-Key Colloids Mediated by Polymeric Depletant. *Langmuir* **31**, 13085–13093 (2015).
174. Munzar, J. D., Ng, A., Corrado, M. & Juncker, D. Complementary oligonucleotides regulate induced fit ligand binding in duplexed aptamers. *Chemical Science* **8**, 2251–2256 (2017).
175. Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences* **44**, 98–104 (1958).
176. Bosshard, H. R. Molecular recognition by induced fit: How fit is the concept? *News in Physiological Sciences* **16**, 171–173 (2001).
177. Hariharan, P. & Guan, L. Insights into the inhibitory mechanisms of the regulatory protein IIA(Glc) on melibiose permease activity. *The Journal of biological chemistry* **289**, 33012–9 (2014).
178. Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* **12**, 88–118 (1965).
179. Zhou, H. X. From induced fit to conformational selection: A continuum of binding mechanism controlled by the timescale of conformational transitions. *Biophysical Journal* **98**, L15–L17 (2010).
180. Greives, N. & Zhou, H.-X. Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proceedings of the National Academy of Sciences* **111**, 10197–10202 (2014).

181. Bucher, D., Grant, B. J. & McCammon, J. A. Induced fit or conformational selection? the role of the semi-closed state in the maltose binding protein. *Biochemistry* **50**, 10530–10539 (2011).
182. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. *Science* **287**, 820–5 (2000).
183. McKeague, M. & Derosa, M. C. Challenges and opportunities for small molecule aptamer development. *Journal of nucleic acids* **2012**, 748913 (2012).
184. Munzar, J. D., Ng, A. & Juncker, D. Duplexed aptamers: history, design, theory, and application to biosensing. *Chemical Society Reviews* **48**, 1390–1419 (2019).
185. Lackey, H. H., Peterson, E. M., Harris, J. M. & Heemstra, J. M. Probing the Mechanism of Structure-Switching Aptamer Assembly by Super-Resolution Localization of Individual DNA Molecules. *Analytical Chemistry* **92**, 6909–6917 (2020).
186. You, M., Litke, J. L. & Jaffrey, S. R. Imaging metabolite dynamics in living cells using a Spinach-based riboswitch. *Proceedings of the National Academy of Sciences* **112**, 2756–65 (2015).
187. Zhang, J., Lau, M. W. & Ferré-D'Amaré, A. R. Ribozymes and riboswitches: Modulation of RNA function by small molecules. *Biochemistry* **49**, 9123–9131 (2010).
188. Li, S. & Breaker, R. R. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Research* **41**, 3022–3031 (2013).
189. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
190. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**, 1167–1171 (2006).
191. Kaiser, C., Schneider, J., Groher, F., Suess, B. & Wachtveitl, J. What defines a synthetic riboswitch?-Conformational dynamics of ciprofloxacin aptamers with similar binding affinities but varying regulatory potentials. *Nucleic Acids Research* **49**, 3661–3671 (2021).

192. Groher, F. *et al.* Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. *Nucleic Acids Research* **46**, 2121–2132 (2018).
193. Munzar, J. D., Ng, A. & Juncker, D. Comprehensive profiling of the ligand binding landscapes of duplexed aptamer families reveals widespread induced fit. *Nature Communications* **9**, 1–15 (2018).
194. Lilley, D. M. J. The origins of RNA catalysis in ribozymes. *Trends in biochemical sciences* **28**, 495–501 (2003).
195. Li, Y. & Breaker, R. R. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2 γ -hydroxyl group. *Journal of the American Chemical Society* **121**, 5364–5372 (1999).
196. Breaker, R. R. *et al.* A common speed limit for RNA-cleaving ribozymes and deoxyribozymes. *RNA* **9**, 949–957 (2003).
197. Lilley, D. M. J. Classification of the nucleolytic ribozymes based upon catalytic mechanism. *F1000Research* **8**, 1462 (2019).
198. Wilcox, J. L., Ahluwalia, A. K. & Bevilacqua, P. C. Charged nucleobases and their potential for RNA catalysis. *Accounts of Chemical Research* **44**, 1270–1279 (2011).
199. Wilson, T. J., McLeod, A. C. & Lilley, D. M. A guanine nucleobase important for catalysis by the VS ribozyme. *EMBO Journal* **26**, 2489–2500 (2007).
200. Nakano, S. I., Chadalavada, D. M. & Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493–1497 (2000).
201. Liu, Y., Wilson, T. J. & Lilley, D. M. The structure of a nucleolytic ribozyme that employs a catalytic metal ion. *Nature Chemical Biology* **13**, 508–513 (2017).
202. Wilson, T. J. *et al.* Nucleobase-mediated general acid-base catalysis in the Varkud satellite ribozyme. *Proceedings of the National Academy of Sciences* **107**, 11751–11756 (2010).

203. Kath-Schorr, S. *et al.* General acid-base catalysis mediated by nucleobases in the hairpin ribozyme. *Journal of the American Chemical Society* **134**, 16717–16724 (2012).
204. Wilson, T. J., Liu, Y., Domnick, C., Kath-Schorr, S. & Lilley, D. M. The Novel Chemical Mechanism of the Twister Ribozyme. *Journal of the American Chemical Society* **138**, 6151–6162 (2016).
205. Gaines, C. S. & York, D. M. Model for the Functional Active State of the TS Ribozyme from Molecular Simulation. *Angewandte Chemie International Edition* **56**, 13392–13395 (2017).
206. Zheng, L. *et al.* Structure-based insights into self-cleavage by a four-way junctional twister-sister ribozyme. *Nature Communications* **8**, 1–12 (2017).
207. Gong, B. *et al.* Direct measurement of a pKa near neutrality for the catalytic cytosine in the genomic HDV ribozyme using raman crystallography. *Journal of the American Chemical Society* **129**, 13335–13342 (2007).
208. Wilson, T. J. *et al.* Comparison of the Structures and Mechanisms of the Pistol and Hammerhead Ribozymes. *Journal of the American Chemical Society* **141**, 7865–7875 (2019).
209. Cech, T. R. Structure and mechanism of the large catalytic RNAs: group 1 and group 2 introns and ribonuclease P. *The RNA World* **24**, 239–269 (1993).
210. Klemm, B. P. *et al.* The Diversity of Ribonuclease P: Protein and RNA Catalysts with Analogous Biological Functions. *Biomolecules* **6**, 27 (2016).
211. Reiter, N. J. *et al.* Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**, 784–791 (2010).
212. Stahley, M. R. & Strobel, S. A. Structural biology: Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* **309**, 1587–1590 (2005).
213. Gordon, P. M., Fong, R. & Piccirilli, J. A. A Second Divalent Metal Ion in the Group II Intron Reaction Center. *Chemistry and Biology* **14**, 607–612 (2007).

214. Shan, S.-o., Yoshida, A., Sun, S., Piccirilli, J. A. & Herschlag, D. Three metal ions at the active site of the Tetrahymena group I ribozyme. *Proceedings of the National Academy of Sciences* **96**, 12299–12304 (1999).
215. Lilley, D. M. J. Mechanisms of RNA catalysis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2910–2917 (2011).
216. Sievers, A., Beringer, M., Rodnina, M. V. & Wolfenden, R. The ribosome as an entropy trap. *Proceedings of the National Academy of Sciences* **101**, 7897–7901 (2004).
217. Leung, E. K. Y., Suslov, N., Tuttle, N., Sengupta, R. & Piccirilli, J. A. The mechanism of peptidyl transfer catalysis by the ribosome. *Annual Review of Biochemistry* **80**, 527–555 (2011).
218. Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Molecular cell* **20**, 437–48 (2005).
219. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. *Annual review of biochemistry* **68**, 611–47 (1999).
220. Conn, M. M., Prudent, J. R. & Schultz, P. G. Porphyrin metalation catalyzed by a small RNA molecule. *Journal of the American Chemical Society* **118**, 7012–7013 (1996).
221. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54–57 (1997).
222. Sengle, G., Eisenführ, A., Arora, P. S., Nowick, J. S. & Famulok, M. Novel RNA catalysts for the Michael reaction. *Chemistry and Biology* **8**, 459–473 (2001).
223. Fusz, S., Eisenführ, A., Srivatsan, S. G., Heckel, A. & Famulok, M. A ribozyme for the aldol reaction. *Chemistry and Biology* **12**, 941–950 (2005).
224. Chen, X., Li, N. & Ellington, A. D. Ribozyme Catalysis of Metabolism in the RNA World. *Chemistry & Biodiversity* **4**, 633–655 (2007).
225. Coey, C. T. & Drohat, A. C. Kinetic Methods for Studying DNA Glycosylases Functioning in Base Excision Repair. *Methods in enzymology* **592**, 357–376 (2017).

226. Emilsson, G. M., Nakamura, S., Roth, A. & Breaker, R. R. Ribozyme speed limits. *RNA* **9**, 907–18 (2003).
227. Canny, M. D. *et al.* Fast cleavage kinetics of a natural hammerhead ribozyme. *Journal of the American Chemical Society* **126**, 10848–10849 (2004).
228. Roychowdhury-Saha, M. & Burke, D. H. Extraordinary rates of transition metal ion-mediated ribozyme catalysis. *RNA* **12**, 1846–1852 (2006).
229. Santoro, S. W. & Joyce, G. F. Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry* **37**, 13330–13342 (1998).
230. Herschlag, D. & Cech, T. R. Catalysis of RNA cleavage by the *Tetrahymena thermophila* ribozyme. 1. Kinetic description of the reaction of an RNA substrate that forms a mismatch at the active site. *Biochemistry* **29**, 10172–10180 (1990).
231. Agresti, J. J., Kelly, B. T., Jäschke, A. & Griffiths, A. D. Selection of ribozymes that catalyse multiple-turnover Diels–Alder cycloadditions by using in vitro compartmentalization. *Proceedings of the National Academy of Sciences* **102**, 16170–16175 (2005).
232. Park, C. Visual Interpretation of the Meaning of K_{cat} / K_M in Enzyme Kinetics. *Journal of Chemical Education* **99**, 2556–2562 (2022).
233. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
234. Ouldrige, T. E., Sulc, P., Romano, F., Doye, J. P. K. & Louis, A. A. DNA hybridization kinetics: zippering, internal displacement and sequence dependence. *Nucleic acids research* **41**, 8886–95 (2013).
235. Nelson, J. W. & Tinoco, I. Comparison of the Kinetics of Ribooligonucleotide, Deoxyribooligonucleotide, and Hybrid Oligonucleotide Double-Strand Formation by Temperature-Jump Kinetics. *Biochemistry* **21**, 5289–5295 (1982).
236. Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences* **94**, 4262–4266 (1997).

237. Zaug, A. J., Grosshans, C. A. & Cech, T. R. Sequence-specific Endoribonuclease Activity of The Tetrahymena Ribozyme: Enhanced Cleavage of Certain Oligonucleotide Substrates That Form Mismatched Ribozyme—Substrate Complexes. *Biochemistry* **27**, 8924–8931 (1988).
238. Beebe, J. A. & Fierke, C. A. A Kinetic Mechanism for Cleavage of Precursor tRNA^{Asp} Catalyzed by the RNA Component of *Bacillus subtilis* Ribonuclease P. *Biochemistry* **33**, 10294–10304 (1994).
239. Butcher, S. E., Heckman, J. E. & Burke, J. M. Reconstitution of hairpin ribozyme activity following separation of functional domains. *Journal of Biological Chemistry* **270**, 29648–29651 (1995).
240. Bertrand, E., Pictet, R. & Grange, T. Can hammerhead ribozymes be efficient tools to inactivate gene function? *Nucleic Acids Research* **22**, 293–300 (1994).
241. Miller, O. J. *et al.* Directed evolution by in vitro compartmentalization. *Nature Methods* **3**, 561–570 (2006).
242. Silverman, S. K. in *Wiley Encyclopedia of Chemical Biology* 1–17 (Wiley, 2008).
243. Mao, X., Liu, M., Li, Q., Fan, C. & Zuo, X. DNA-Based Molecular Machines. *JACS Au* **2**, 2381–2399 (2022).
244. Krissanaprasit, A., Key, C. M., Pontula, S. & LaBean, T. H. Self-Assembling Nucleic Acid Nanostructures Functionalized with Aptamers. *Chemical reviews* **121**, 13797–13868 (2021).
245. Jepsen, M. D. *et al.* Development of a genetically encodable FRET system using fluorescent RNA aptamers. *Nature Communications* **9**, 1–10 (2018).
246. Liu, Z. *et al.* Self-assembly of responsive multilayered DNA nanocages. *Journal of the American Chemical Society* **137**, 1730–1733 (2015).
247. Muscat, R. A., Bath, J. & Turberfield, A. J. Small molecule signals that direct the route of a molecular cargo. *Small* **8**, 3593–3597 (2012).
248. Funck, T., Liedl, T. & Bae, W. Dual Aptamer-Functionalized 3D Plasmonic Metamolecule for Thrombin Sensing. *Applied Sciences* **9**, 3006 (2019).

249. Chang, A. L., McKeague, M., Liang, J. C. & Smolke, C. D. Kinetic and equilibrium binding characterization of aptamers to small molecules using a label-free, sensitive, and scalable platform. *Analytical Chemistry* **86**, 3273–3278 (2014).
250. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell* **18**, 49–60 (2005).
251. Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nature Communications* **8**, 1–8 (2017).
252. Spöring, M., Finke, M. & Hartig, J. S. Aptamers in RNA-based switches of gene expression. *Current opinion in biotechnology* **63**, 34–40 (2020).
253. Li, D., Shlyahovsky, B., Elbaz, J. & Willner, I. Amplified analysis of low-molecular-weight substrates or proteins by the self-assembly of DNAzyme-aptamer conjugates. *Journal of the American Chemical Society* **129**, 5804–5805 (2007).
254. Lu, Y. & Liu, J. Functional DNA nanotechnology: emerging applications of DNAzymes and aptamers. *Current Opinion in Biotechnology* **17**, 580–588 (2006).
255. Willner, I., Shlyahovsky, B., Zayats, M. & Willner, B. DNAzymes for sensing, nanobiotechnology and logic gate applications. *Chemical Society Reviews* **37**, 1153–1165 (2008).
256. Zhao, X. H. *et al.* Versatile DNAzyme-based amplified biosensing platforms for nucleic acid, protein, and enzyme activity detection. *Analytical Chemistry* **85**, 3614–3620 (2013).
257. Liu, M. *et al.* A DNAzyme Feedback Amplification Strategy for Biosensing. *Angewandte Chemie International Edition* **56**, 6142–6146 (2017).
258. McConnell, E. M. *et al.* Biosensing with DNAzymes. *Chemical Society Reviews* **50**, 8954–8994 (2021).
259. Sando, S., Sasaki, T., Kanatani, K. & Aoyama, Y. Amplified Nucleic Acid Sensing Using Programmed Self-Cleaving DNAzyme. *Journal of the American Chemical Society* **125**, 15720–15721 (2003).

260. Elbaz, J. *et al.* DNA computing circuits using libraries of DNzyme subunits. *Nature nanotechnology* **5**, 417–22 (2010).
261. Orbach, R., Willner, B. & Willner, I. Catalytic nucleic acids (DNzymes) as functional units for logic gates and computing circuits: From basic principles to practical applications. *Chemical Communications* **51**, 4144–4160 (2015).
262. Yang, S. *et al.* DNA as a universal chemical substrate for computing and data storage. *Nature Reviews Chemistry* **8**, 179–194 (2024).
263. Zheng, X. *et al.* Allosteric DNzyme-based DNA logic circuit: Operations and dynamic analysis. *Nucleic Acids Research* **47**, 1097–1109 (2019).
264. Li, J. *et al.* Zippered G-quadruplex/hemin DNzyme: exceptional catalyst for universal bioanalytical applications. *Nucleic Acids Research* **49**, 13031–13044 (2021).
265. Pressman, A., Blanco, C. & Chen, I. A. The RNA World as a Model System to Study the Origin of Life. *Current Biology* **25**, R953–R963 (2015).
266. Orgel, L. E. Evolution of the genetic apparatus. *Journal of Molecular Biology* **38**, 381–393 (1968).
267. Lorsch, J. R. & Szostak, J. W. Chance and Necessity in the Selection of Nucleic Acid Catalysts. *Accounts of Chemical Research* **29**, 103–110 (1996).
268. Szostak, J. W. & Ellington, A. D. in *COLD SPRING HARBOR MONOGRAPH SERIES* 511–511 (1993).
269. Komarova, N. & Kuznetsov, A. Inside the Black Box: What Makes SELEX Better? *Molecules* **24**, 3598 (2019).
270. Sazani, P. L., Larralde, R. & Szostak, J. W. A small aptamer with strong and specific recognition of the triphosphate of ATP. *Journal of the American Chemical Society* **126**, 8370–8371 (2004).
271. Nutiu, R. & Li, Y. In Vitro Selection of Structure-Switching Signaling Aptamers. *Angewandte Chemie International Edition* **44**, 1061–1065 (2005).

272. Stoltenburg, R., Nikolaus, N. & Strehlitz, B. Capture-SELEX: Selection of DNA Aptamers for Aminoglycoside Antibiotics. *Journal of Analytical Methods in Chemistry* **2012**, 1–14 (2012).
273. Nikolaus, N. & Strehlitz, B. DNA-aptamers binding aminoglycoside antibiotics. *Sensors* **14**, 3737–3755 (2014).
274. Famulok, M. & Mayer, G. Aptamer modules as sensors and detectors. *Accounts of Chemical Research* **44**, 1349–1358 (2011).
275. Shaban, S. M. & Kim, D.-H. Recent Advances in Aptamer Sensors. *Sensors* **21**, 979 (2021).
276. Feagin, T. A., Maganzini, N. & Soh, H. T. Strategies for Creating Structure-Switching Aptamers. *ACS Sensors* **3**, 1611–1615 (2018).
277. Yoshikawa, A. M. *et al.* A massively parallel screening platform for converting aptamers into molecular switches. *Nature Communications* **14**, 1–14 (2023).
278. Fedotova, T. A. & Kolpashchikov, D. M. Liquid-to-gel transition for visual and tactile detection of biological analytes. *Chemical Communications* **53**, 12622–12625 (2017).
279. Nutiu, R. & Li, Y. Structure-Switching Signaling Aptamers. *Journal of the American Chemical Society* **125**, 4771–4778 (2003).
280. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
281. Griffiths, A. D. & Tawfik, D. S. Man-made enzymes — from design to in vitro compartmentalisation. *Current Opinion in Biotechnology* **11**, 338–353 (2000).
282. Chandra, M. & Silverman, S. K. DNA and RNA can be equally efficient catalysts for carbon-carbon bond formation. *Journal of the American Chemical Society* **130**, 2936–2937 (2008).
283. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16**, 652–656 (1998).

284. Höbartner, C. & Silverman, S. K. Engineering a Selective Small-Molecule Substrate Binding Site into a Deoxyribozyme. *Angewandte Chemie International Edition* **46**, 7420–7424 (2007).
285. Zelin, E., Wang, Y. & Silverman, S. K. Adenosine is inherently favored as the branch-site RNA nucleotide in a structural context that resembles natural RNA splicing. *Biochemistry* **45**, 2767–2771 (2006).
286. Flynn-Charlebois, A. *et al.* Deoxyribozymes with 2′–5′ RNA Ligase Activity. *Journal of the American Chemical Society* **125**, 2444–2454 (2003).
287. Khosla, C. & Harbury, P. B. Modular enzymes. *Nature* **409**, 247–252 (2001).
288. Dokukin, V. & Silverman, S. K. A modular tyrosine kinase deoxyribozyme with discrete aptamer and catalyst domains. *Chem. Commun* **50**, 9317 (2014).
289. Sachdeva, A., Chandra, M., Chandrasekar, J. & Silverman, S. K. Covalent Tagging of Phosphorylated Peptides by Phosphate-Specific Deoxyribozymes. *ChemBioChem* **13**, 654–657 (2012).
290. Golub, E., Albada, H. B., Liao, W.-C., Biniuri, Y. & Willner, I. Nucleoapzymes: Hemin/G-Quadruplex DNAzyme–Aptamer Binding Site Conjugates with Superior Enzyme-like Catalytic Functions. *Journal of the American Chemical Society* **138**, 164–172 (2015).
291. Wulf, V. & Willner, I. Nucleoapzymes: Catalyst-aptamer conjugates as enzyme-mimicking structures. *Emerging Topics in Life Sciences* **3**, 493–499 (2019).
292. 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? *Critical reviews in biochemistry and molecular biology* **46**, 478–92 (2011).
293. Biniuri, Y. *et al.* A Bis-Zn ²⁺ -Pyridyl-Salen-Type Complex Conjugated to the ATP Aptamer: An ATPase-Mimicking Nucleoapzyme. *ChemBioChem* **21**, 53–58 (2020).
294. Vázquez-González, M., Zhou, Z., Biniuri, Y., Willner, B. & Willner, I. Mimicking Functions of Native Enzymes or Photosynthetic Reaction Centers by Nucleoapzymes and Photonucleoapzymes. *Biochemistry* **60**, 956–965 (2021).

295. Stangherlin, S. & Liu, J. Nanomaterials enabled and enhanced DNA-based biosensors. *Journal of materials chemistry. B* **11**, 6994–7003 (2023).
296. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Current Opinion in Structural Biology* **18**, 321–329 (2008).
297. Davis, J. H. & Szostak, J. W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences* **99**, 11616–11621 (2002).
298. Li, Y. & Liu, J. Aptamer-based strategies for recognizing adenine, adenosine, ATP and related compounds. *Analyst* **145**, 6753–6768 (2020).
299. Wachowius, F., Porebski, B. T., Johnson, C. M. & Holliger, P. Emergence of ATP- and GTP-Binding Aptamers from Single RNA Sequences by Error-Prone Replication and Selection**. *ChemSystemsChem* **5**, e202300006 (2023).
300. Saran, D., Held, D. M. & Burke, D. H. Multiple-turnover thio-ATP hydrolase and phospho-enzyme intermediate formation activities catalyzed by an RNA enzyme. *Nucleic Acids Research* **34**, 3201–3208 (2006).
301. Cojocaru, R. & Unrau, P. J. in *Ribozymes: Volume 1 and Volume 2* 331–358 (John Wiley & Sons, Ltd, 2021).
302. Akoopie, A., Arriola, J. T., Magde, D. & Müller, U. F. A GTP-synthesizing ribozyme selected by metabolic coupling to an RNA polymerase ribozyme. *Science Advances* **7** (2021).
303. Wilson, B. D., Hariri, A. A., Thompson, I. A., Eisenstein, M. & Soh, H. T. Independent control of the thermodynamic and kinetic properties of aptamer switches. *Nature Communications* **10**, 1–9 (2019).
304. Serganov, A. *et al.* Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nature Structural and Molecular Biology* **12**, 218–224 (2005).
305. Ferré-D'Amaré, A. R. The glmS ribozyme: Use of a small molecule coenzyme by a gene-regulatory RNA. *Quarterly Reviews of Biophysics* **43**, 423–447 (2010).

306. Woodson, S. A. Structure and assembly of group I introns. *Current Opinion in Structural Biology* **15**, 324–330 (2005).
307. Nutiu, R. & Li, Y. A DNA–Protein Nanoengine for “On-Demand” Release and Precise Delivery of Molecules. *Angewandte Chemie International Edition* **44**, 5464–5467 (2005).
308. Zhang, Z., Oni, O. & Liu, J. New insights into a classic aptamer: Binding sites, cooperativity and more sensitive adenosine detection. *Nucleic Acids Research* **45**, 7593–7601 (2017).
309. Nakano, S., Shimizu, M., Dinh, H. & Morii, T. Highly selective dual sensing of ATP and ADP using fluorescent ribonucleopeptide sensors. *Chemical Communications* **55**, 1611–1614 (2019).
310. Huizenga, D. E. & Szostak, J. W. A DNA Aptamer That Binds Adenosine and ATP. *Biochemistry* **34**, 656–665 (1995).
311. Deng, J. & Walther, A. ATP-Responsive and ATP-Fueled Self-Assembling Systems and Materials. *Advanced Materials* **32**, 2002629 (2020).
312. Li, Q. *et al.* ATP-Triggered, Allosteric Self-Assembly of DNA Nanostructures. *Journal of the American Chemical Society* **142**, 665–668 (2020).
313. Slavkovic, S. *et al.* Thermodynamic analysis of cooperative ligand binding by the ATP-binding DNA aptamer indicates a population-shift binding mechanism. *Scientific Reports* **10** (2020).
314. Ariga, T., Masaike, T., Noji, H. & Yoshida, M. Stepping Rotation of F1-ATPase with One, Two, or Three Altered Catalytic Sites That Bind ATP Only Slowly. *Journal of Biological Chemistry* **277**, 24870–24874 (2002).
315. Bazzicalupi, C. *et al.* A zinc(II)-based receptor for ATP binding and hydrolysis. *Chemical Communications*, 2630–2632 (2005).
316. Buffin, B. P., Fonger, E. B. & Kundu, A. Palladium(II) and platinum(II) complexes containing hydrophilic pyridinylimine-based ligands. *Inorganica Chimica Acta* **355**, 340–346 (2003).

317. Tzimopoulos, D. *et al.* Synthesis and study of triorganostannyl esters of 3-,4- and 3,5-pyridinylimino substituted aminobenzoic acids: Crystal structures of dimorphs of aqua-trimethyltin 3-pyridinyliminobenzoate. *Journal of Molecular Structure* **965**, 56–64 (2010).
318. Tzimopoulos, D. *et al.* Synthesis and structural characterization of heteroleptic Cu(I) complexes with α -diimines derived from pyridinyliminobenzoic acids. *Journal of Coordination Chemistry* **65**, 393–402 (2012).
319. Saha Roy, A. *et al.* Unsymmetrical diimine chelation to M(II) (M = Zn, Cd, Pd): Atropisomerism, pi-pi stacking and photoluminescence. *Dalton Transactions* **40**, 7375–7384 (2011).
320. Ardizzoia, G. A., Colombo, G., Therrien, B. & Brenna, S. Tuning the Fluorescence Emission and HOMO-LUMO Band Gap in Homoleptic Zinc(II) Complexes with N,O-Bidentate (Imidazo[1,5-a]pyrid-3-yl)phenols. *European Journal of Inorganic Chemistry* **2019**, 1825–1831 (2019).
321. Consiglio, G., Oliveri, I. P., Failla, S. & Di Bella, S. On the Aggregation and Sensing Properties of Zinc(II) Schiff-Base Complexes of Salen-Type Ligands. *Molecules* **24**, 2514 (2019).
322. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA mimics of green fluorescent protein. *Science* **333**, 642–646 (2011).
323. Coleman, T. M. & Huang, F. Optimal random libraries for the isolation of catalytic RNA. *RNA Biology* **2**, 129–136 (2005).
324. Knight, R. & Yarus, M. Finding specific RNA motifs: Function in a zeptomole world? *RNA* **9**, 218–230 (2003).
325. Sabeti, P. C., Unrau, P. J. & Bartel, D. P. Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. *Chemistry and Biology* **4**, 767–774 (1997).
326. Pobanz, K. & Lupták, A. Improving the odds: Influence of starting pools on in vitro selection outcomes. *Methods* **106**, 14–20 (2016).

327. Vant-Hull, B., Payano-Baez, A., Davis, R. H. & Gold, L. The mathematics of SELEX against complex targets. *Journal of Molecular Biology* **278**, 579–597 (1998).
328. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX—A (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomolecular Engineering* **24**, 381–403 (2007).
329. Joyce, G. F. Directed evolution of nucleic acid enzymes. *Annual review of biochemistry* **73**, 791–836 (2004).
330. Lozupone, C., Changayil, S., Majerfeld, I. & Yarus, M. Selection of the simplest RNA that binds isoleucine. *RNA* **9**, 1315–1322 (2003).
331. Vorobyeva, M. A., Davydova, A. S., Vorobjev, P. E., Pyshnyi, D. V. & Venyaminova, A. G. Key aspects of nucleic acid library design for in vitro selection. *International Journal of Molecular Sciences* **19**, 470 (2018).
332. Hall, B. *et al.* Design, synthesis, and amplification of DNA pools for in vitro selection. *Current Protocols in Molecular Biology* **39**, 1–27 (2009).
333. Studier, F. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* **189**, 113–130 (1986).
334. Jia, Y. & Patel, S. S. Kinetic mechanism of GTP binding and RNA synthesis during transcription initiation by bacteriophage T7 RNA polymerase. *Journal of Biological Chemistry* **272**, 30147–30153 (1997).
335. Uhlenbeck, O. C. T4 RNA ligase. *Trends in Biochemical Sciences* **8**, 94–96 (1983).
336. Yin, S., Ho, C. K. & Shuman, S. Structure-function analysis of T4 RNA ligase 2. *Journal of Biological Chemistry* **278**, 17601–17608 (2003).
337. El Omari, K. *et al.* Molecular architecture and ligand recognition determinants for T4 RNA ligase. *Journal of Biological Chemistry* **281**, 1573–1579 (2006).
338. England, T. E. & Uhlenbeck, O. C. 3'-terminal labelling of RNA with T4 RNA ligase. *Nature* **275**, 560–1 (1978).

339. Stoll, V. S. & Blanchard, J. S. Buffers: Principles and Practice. *Methods in Enzymology* **182**, 24–38 (1990).
340. Marchand, A. & Gabelica, V. Folding and misfolding pathways of G-quadruplex DNA. *Nucleic Acids Research* **44**, 10999–11012 (2016).
341. Pyle, A. M. Ribozymes: A Distinct Class of Metalloenzymes. *Science* **261**, 709–714 (1993).
342. Dange, V., Arra, R. B. V. A. N. & Hecht, S. M. A Mn²⁺-Dependent Ribozyme. *Science*, 585–589 (1985).
343. Walsh, S. M., Sachdeva, A. & Silverman, S. K. DNA Catalysts with Tyrosine Kinase Activity. *Journal of the American Chemical Society* **135**, 14928–14931 (2013).
344. Czerny, T. High primer concentration improves PCR amplification from random pools. *Nucleic Acids Research* **24**, 985–986 (1996).
345. Tolle, F., Wilke, J., Wengel, J. & Mayer, G. By-product formation in repetitive PCR amplification of DNA libraries during SELEX. *PLoS ONE* **9**, 1–12 (2014).
346. Kohlberger, M. & Gadermaier, G. SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry* **69**, 1771–1792 (2022).
347. Vaish, N. K., Larralde, R., Fraley, A. W., Szostak, J. W. & McLaughlin, L. W. A Novel, Modification-Dependent ATP-Binding Aptamer Selected from an RNA Library Incorporating a Cationic Functionality. *Biochemistry* **42**, 8842–8851 (2003).
348. Schütze, T. *et al.* A calibrated diversity assay for nucleic acid libraries using DiStRO- a Diversity Standard of Random Oligonucleotides. *Nucleic Acids Research* **38**, e23–e23 (2009).
349. Vanbrabant, J., Leirs, K., Vanschoenbeek, K., Lammertyn, J. & Michiels, L. Remelting curve analysis as a tool for enrichment monitoring in the SELEX process. *Analyst* **139**, 589–595 (2014).
350. Luo, Z., He, L., Wang, J., Fang, X. & Zhang, L. Developing a combined strategy for monitoring the progress of aptamer selection. *Analyst* **142**, 3136–3139 (2017).

351. Pei, S., Slinger, B. L. & Meyer, M. M. Recognizing RNA structural motifs in HT-SELEX data for ribosomal protein S15. *BMC Bioinformatics* **18**, 1–14 (2017).
352. Beier, R., Boschke, E. & Labudde, D. New strategies for evaluation and analysis of SELEX experiments. *BioMed Research International* **2014**, 849743 (2014).
353. Hoinka, J., Berezhnoy, A., Sauna, Z. E., Gilboa, E. & Przytycka, T. M. AptaCluster - A Method to Cluster HT-SELEX Aptamer Pools and Lessons from its Application. *Research in Computational Molecular Biology* **8394**, 115–128 (2014).
354. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Research* **14**, 1188–1190 (2004).
355. Coleman, T. M., Wang, G. & Huang, F. Superior 5' homogeneity of RNA from ATP-initiated transcription under the T7 phi 2.5 promoter. *Nucleic acids research* **32**, e14–e14 (2004).
356. Schneider, T. D. & Stephens, R. M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Research* **18**, 6097–6100 (1990).
357. Hofacker, I. L., Fekete, M. & Stadler, P. F. Secondary structure prediction for aligned RNA sequences. *Journal of Molecular Biology* **319**, 1059–1066 (2002).
358. Bernhart, S. H., Hofacker, I. L., Will, S., Gruber, A. R. & Stadler, P. F. RNAalifold: Improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* **9**, 1–13 (2008).
359. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic acids research* **36**, W70–W74 (2008).
360. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406–3415 (2003).
361. Tanaka, Y. *et al.* Solution structure of an RNA duplex including a C-U base pair. *Biochemistry* **39**, 7074–7080 (2000).
362. Popenda, M. *et al.* Automated 3D structure composition for large RNAs. *Nucleic Acids Research* **40**, e112–e112 (2012).

363. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).
364. Martick, M., Lee, T. S., York, D. M. & Scott, W. G. Solvent Structure and Hammerhead Ribozyme Catalysis. *Chemistry and Biology* **15**, 332–342 (2008).
365. Song, Y., Liu, K. J. & Wang, T. H. Elimination of ligation dependent artifacts in T4 RNA ligase to achieve high efficiency and low bias microRNA capture. *PLoS ONE* **9**, e94619 (2014).
366. Ho, C. K., Wang, L. K., Lima, C. D. & Shuman, S. Structure and Mechanism of RNA Ligase. *Structure* **12**, 327–339 (2004).
367. Munafó, D. B. & Robb, G. B. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA* **16**, 2537–2552 (2010).
368. Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y. & Robb, G. B. Structural bias in T4 RNA ligase-mediated 3-prime adapter ligation. *Nucleic Acids Research* **40**, e54–e54 (2012).
369. Almeida, M. V., de Jesus Domingues, A. M., Lukas, H., Mendez-Lago, M. & Ketting, R. F. RppH can faithfully replace TAP to allow cloning of 5-prime triphosphate carrying small RNAs. *MethodsX* **6**, 265–272 (2019).
370. Rodionova, N. S. & Petushkov, V. N. Effect of different salts and detergents on luciferin-luciferase luminescence of the enchytraeid *Fridericia heliota*. *Journal of Photochemistry and Photobiology B: Biology* **83**, 123–128 (2006).
371. Richardson, C. D., Ray, G. J., Bray, N. L. & Corn, J. E. Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nature Communications* **7**, 1–7 (2016).
372. Hannan, J. P. *et al.* HPLC method to resolve, identify and quantify guanine nucleotides bound to recombinant ras GTPase. *Analytical Biochemistry* **631**, 114338 (2021).
373. Deana, A., Celesnik, H. & Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* **451**, 355–8 (2008).

- 374. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
- 375. Avci-Adali, M., Paul, A., Wilhelm, N., Ziemer, G. & Wendel, H. P. Upgrading SELEX technology by using lambda exonuclease digestion for single-stranded DNA generation. *Molecules* **15**, 1–11 (2010).
- 376. He, C. Z. *et al.* Single-primer-limited amplification: A method to generate random single-stranded DNA sub-library for aptamer selection. *Analytical Biochemistry* **440**, 63–70 (2013).
- 377. Nehdi, A. *et al.* Novel Strategies to Optimize the Amplification of Single-Stranded DNA. *Frontiers in Bioengineering and Biotechnology* **8**, 1–13 (2020).