EVOLVING AN ENZYME FROM A NON-CATALYTIC SEQUENCE

EVOLVING AN ENZYME FROM A NON-CATALYTIC SEQUENCE

By

RACHEL GYSBERS

B. Sc. McMaster University, Hamilton, Ontario, 2013

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Science

McMaster University

© Copyright by Rachel Gysbers, June 2015

MASTER OF SCIENCE (2015)

McMaster University

(Biochemistry (Astrobiology))

Hamilton, Ontario

TITLE: Evolving an Enzyme from a Non-catalytic Sequence AUTHOR: Rachel Gysbers, B. Sc. (McMaster University) SUPERVISOR: Dr. Yingfu Li NUMBER OF PAGES: 65 (LXV)

Abstract

Life would not exist in the absence of catalysis. The "RNA World" model for the origin of life hinges on the capabilities of ribonucleic acid to encode information and perform catalysis (i.e. self-replication). Previously, functional nucleic acids such as ribozymes and deoxyribozymes (DNAzymes) have been isolated using the process of in *vitro* selection. This method is typically performed by isolating a catalytically active molecule from a large random library, with the assumption being that active molecules are already present in the pool and this method filters them from inactive molecules. However, in vitro selection has never been used to show that a molecule can be evolved from an inactive to an active catalyst. Here we show that the properties of DNA can be exploited to act as a proxy system for the origins of biotic chemistry by isolating a functional catalyst from a previously non-catalytic sequence. This project employs a novel perspective; rather than a random library, a known, non-functional sequence is utilized. Using in vitro selection, this known sequence is gradually evolved into a functional catalyst by solely allowing the existence of sequences that acquire mutations which enhance their function. Deep sequencing analysis of DNA pools along the evolution trajectory has identified individual mutations as the progressive drivers of molecular evolution. Evolving a catalyst from a non-catalyst gives insight into the comprehension of how life originated. This project demonstrates that an enzyme can indeed arise from a sequence of a functional polymer via permissive molecular evolution, a mechanism that may have been exploited by nature for the creation of the enormous repertoire of enzymes in the biological world today.

Acknowledgements

I would like to express my gratitude and appreciation for my supervisor, Dr. Yingfu Li. His enthusiasm for this project, in combination with his guidance and support, has been paramount throughout my graduate studies.

The comprehensive analysis of the deep sequencing data in this project was performed by the Li Lab's research assistant, Jim Gu. His knowledge and assistance were vital to making this project succeed and are greatly appreciated. In addition, two undergraduate students, Lorea Alejaldre and Laura Chan, have worked on related aspects of this project. My thanks; it was a pleasure working with you.

Sincere thanks to my committee members, Dr. Paul Harrison and Dr. Paul Higgs, for their constructive criticism, advice and recommendations throughout. I'm also grateful for other members of the Origins Institute, notably Dr. Ralph Pudritz and Dr. Jon Stone, for imparting their knowledge and guidance.

I'm grateful for all the members of the Li Lab, past and present, who have supported me unconditionally and been like a second family. Special thanks and endless appreciation to my mentor, Kha Tram, who took me under his wing when I was an undergraduate and prevented me from crawling back underneath when times got tough as a graduate student.

Last but not least, thanks to my husband – I couldn't have done this without your unwavering support – and to my family and friends.

Table of Contents

Chapter One: Nucleic Acid Enzymes and the Origin of Life 1
1.1 Catalysts and the Origins of Life 1
1.2 Ribozymes and the RNA World
1.3 Artificial Ribozymes: In Vitro Selection
1.4 Ribozymes: Progress towards an RNA World7
1.5 Deoxyribozymes: An Introduction
1.5.1 DNAzymes: Scope and Use11
1.5.2 RNA-Cleaving Deoxyribozymes
1.6 Conclusion and Thesis Overview
Chapter Two: Developing a Catalyst from a Non-catalyst
2.1 Introduction
2.2 Methods
2.2.1 Materials
2.2.2 In vitro selection
2.2.3 Activity assays of Selected Rounds
2.2.4 Structural Characterization
2.3 Results and Discussion
2.3.1 Development of a Catalyst
2.3.2 Characterizing Sequence Structure
2.4 Summary and Conclusion
Chapter Three: Investigating the Development of a Catalyst
3.1 Sequencing and Fitness Landscapes
3.2 Methods

3.2.1 Deep Sequencing
3.2.2 Activity comparison of top 10 G25 sequences
3.2.3 Activity analysis of MD0-7
3.3 Results and Discussion
3.3.1 Top Sequence Analysis
3.3.2 Developing a Mutation Pathway
3.4 Summary and Conclusion
Chapter Four: Discussion of Additional Experimental Data 46
4.1 Optimization of an RNA-cleaving DNAzyme
4.2 Selection of an RNA-cleaving DNAzyme (II)
4.2.1 Problems and Pitfalls
4.2.2 New Library Design
Chapter Five: Summary and Conclusion 54
Reference List

List of Figures

Chapter One

- 1-1. In Vitro Selection
- 1-2. RNA Cleavage via Transesterification
- 1-3. 8-17 and 10-23 Motifs

Chapter Two

- 2-1. Typical *in vitro* selection vs. selection in this project
- **2-2**. Overall concept schematic
- 2-3. In vitro selection method
- **2-4**. Graph of increasing cleavage activity
- **2-5.** PAGE analysis of selected rounds
- 2-6. Cleavage Improvement: G0 vs. G25
- 2-7. G2501 Secondary Structure Model
- 2-8. Truncation Analysis of G2501
- 2-9. Mutational Analysis of G2501
- 2-10. Trans Structural Verification

Chapter Three

- **3-1**. Composition of G25
- **3-2**. Top ten sequences of G25
- 3-3. Activity analysis of top ten sequences
- 3-4. Time-dependent cleavage of MD0-7
- **3-5**. MD0-7 pathway cleavage vs. time

Chapter Four

- **4-1**. Optimization library design
- **4-2**. Altered primer design (I) cleavage analysis

List of Tables

Chapter One

1-1. Reactions Catalyzed by DNAzymes

Chapter Four

4-1. Library Redesign

List of Abbreviations

ATP	adenosine 5'-triphosphate
BSA	bovine serum albumin
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded deoxyribonucleic acid
FNA	functional nucleic acids
FP	forward primer
GTP	guanosine 5'-triphosphate
LUA	last universal ancestor
LUCA	last universal common ancestor
MRCA	most recent common ancestor
NTP	nucleoside triphosphate
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PNK	polynucleotide kinase
RNA	ribonucleic acid
RP	reverse primer
SB	selection buffer
SELEX	systematic evolution of ligands by exponential enrichment
ssDNA	single-stranded deoxyribonucleic acid

Chapter One: Nucleic Acid Enzymes and the Origin of Life

Life could not exist in the absence of catalysis. The "RNA World" model for the origin of life hinges on the capabilities of ribonucleic acid to encode information and perform catalysis (i.e. self-replication). Functional nucleic acids have previously been isolated using the process of *in vitro* selection, a method typically performed by isolating catalytically active molecules from a large random library. This introduction endeavours to examine and summarize attempts to evolve catalytic nucleic acid systems such as replicases similar to those thought to exist in the RNA world scenario, and aims to set the stage for our project's attempt in understanding life's origins.

1.1 Catalysts and the Origins of Life

The relatively new interdisciplinary field of astrobiology investigates the origin, evolution and distribution of life in our universe. This collaborative enterprise combines expertise from many diverse fields in order to answer the broad questions: "How did we get here?" and "Are we alone?" The first quandary encompasses, on different scales, the emergence of the universe, the elements, and life on Earth. The latter asks if life could have developed elsewhere, and seeks to search for extraterrestrial life.

In any respect, the dilemma of the origin of life is one of the most important problems in biology. If it were to be discovered just how life on Earth came about, it would address both questions above – if we know the conditions for life, we can extrapolate from those conditions to find life on other planets (or moons [1]).

In studying life on Earth, we can work backwards through time. Charles Darwin first proposed that species evolve through the natural selection of heritable traits, which allows individuals to reproduce [2]. He also articulated the idea that all known life descends from one progenitor [2]. Since there are a number of traits shared by all organisms on Earth, Darwin's theory of common descent is logical. The most recent organism from which all life on Earth descends is known as the Last Universal Common Ancestor (LUCA), the Last Universal Ancestor (LUA), or the Most Recent Common Ancestor (MRCA).

Among a variety of properties and characteristics, LUCA had two fundamental characteristics that can be explored with prebiotic chemistry: a genetic code comprised of DNA and the assembly of proteins from amino acids by ribosomes. How could these prebiotic molecules have emerged to make life, and eventually LUCA?

Perhaps the most famous experiment in abiogenesis investigates how molecules essential for life could have arisen from the prebiotic soup. This research, conducted by Stanley Miller and Harold Urey, attempted to reproduce early Earth conditions. Under the assumption that the atmosphere of the early Earth was reducing, molecules of hydrogen, methane, water and ammonia were combined with evaporation and electrical discharge (simulated lightning) and several different amino acids were found to have emerged from the solution [3,4]. This experiment was also performed with an "atmosphere" of nitrogen, methane, water and ammonia, yielding even more variety [5]. Although the environmental conditions assumed in the Miller-Urey experiments may not be valid, as recent studies support a less reducing environment, this research showed that prebiotic molecules as complex as amino acids can be formed abiologically under certain conditions [6]. Other experiments concerning the chemistry of the prebiotic soup focus on the abiotic formation of nucleic acids. For example, the Sutherland group recently showed that under prebiotically plausible conditions, pyrimidine ribonucleotides can be formed; it was previously thought that the formation of ribose and its addition to pyrimidines was near-impossible [7]. The polymerization of nucleotides to form oligomers has been shown to occur with clay catalysts; montmorillonite clay can adsorb RNA molecules, facilitating binding and polymerization [8]. Deamer's group showed that polymerization of nucleotides in an ice microenvironment is another prebiotic possibility [9].

Various fragments of the puzzle of life's origins begin to fit together with the advent of these experiments. The investigation of how life emerged has another dilemma, however. Most life makes use of protein catalysts known as enzymes, molecules that lower the rate-limiting free energy of activation and allow organisms to execute normally slow chemical reactions at enhanced rates. These chemistries crucial for life to exist could not occur without catalysts. But protein catalysts did not magically appear to kick-start life into action. In an origin of life scenario, catalysts must have emerged that were able to facilitate the production of more catalysts – these catalysts could have been completely self-replicating, or co-operative molecules [10]. Enzymes had to start from somewhere, and the RNA World is a likely prospect.

1.2 Ribozymes and the RNA World

The central dogma, first stated by Francis Crick [11], lays out a clear place for nucleic acids in the transfer of information within biological systems. In the central dogma's scheme of "replication, transcription, translation", it puts RNA solidly between the latter two.

However, it was realized in the 1980s that RNA possessed the ability to perform functions other than simply transferring information from DNA to protein. Two researchers were awarded the Nobel Prize in Chemistry in 1989 for their work on the catalytic properties of ribonucleic acid (RNA): Sidney Altman's group discovered that the RNA subunit of ribonuclease P was catalytic [12], while Thomas Cech's group discerned a self-splicing RNA from the organism *Tetrahymena thermophila* [13]. In fact, the most well-known example of ribonucleic acid catalysis exists within all living cells: the linkage of amino acids performed by the RNA within ribosomes, the molecular machines found in all cells [14,15]. Peptide synthesis in the ribosome is a molecular fossil of a time when RNA played an important role.

This concept of RNA molecules forming the basis for life, explored by many [16-18], was first coined "The RNA World" by Walter Gilbert [19], and pivots on the idea that primordial ribonucleic acids could somehow reproduce, copying themselves without biological assistance (i.e. without cellular machinery such as ribosomes) [20,21]. This RNA World scenario encompasses the emergence, replication and evolution of prebiotic polymers. This hypothesis does not provide a clean solution to the multifaceted problem of the origin of life, but whittles it down to a more manageable dilemma.

Leslie Orgel condenses the quandary of the RNA World (under the assumption that it was the first example of organized biochemistry) into four sub-problems, summarized here: 1) The nonenzymatic synthesis of nucleotides; 2) The nonenzymatic polymerization of nucleotides to give random-sequence RNA; 3) The nonenzymatic copying or replication or both, of RNA; and 4) The emergence through natural selection of a set of functional RNA catalysts that together could sustain exponential growth in the prebiotic environment [22].

In the "RNA World" [19], catalysis was performed by RNA, prior to the evolution of proteins [23]. Although there are assorted variants of the RNA World hypothesis, each with advantages and caveats, Robertson & Joyce (2014) sum up all collective RNA World Hypotheses under three premises: 1) Genetic continuity occurred via replication of RNA; 2) A key to replication was Watson-Crick base pairing; 3) catalysts were not genetically-encoded proteins [21].

The catalytic and molecular recognition functions of RNA were first discovered in nature, with the discovery of ribozymes and, later, mRNA with roles in gene regulation. Known as "riboswitches", these mRNA are able to recognize and bind small molecules, resulting in changes in gene expression [24-27]. It wasn't long before artificial methods were developed to isolate functional nucleic acids, each tailored to a specific purpose.

1.3 Artificial Ribozymes: In Vitro Selection

An artificial method to isolate ribozymes *in vitro* had its first success in 1990 [28,29]. This method was termed *in vitro* selection, also known as the Systematic Evolution of Ligands by Exponential Enrichment (SELEX), and involves a large random pool which is subjected to selective pressure and enrichment [30]. Specific RNA sequences capable of catalyzing a wide variety of chemical reactions can be isolated from noncatalytic molecules *via* this technique; the process is depicted in Figure 2 [31,32].

This technique is not exclusive to ribonucleic acid; SELEX has been used to isolate catalytic deoxyribonucleic acid (DNA) molecules as well [33].



Figure 1-1: Schematic of a typical *In Vitro* Selection experiment. (i) A random library is generated and (ii) subjected to selective pressure to carry out a specific reaction. (iii) Catalytic molecules are separated from non-catalytic sequences, (iv) amplified *via* PCR, and (v) the enriched pool is subjected to additional cycles, or rounds, of *in vitro* selection.

The typical SELEX method employs the strategy of "search and evolution", allowing the isolation of a catalytic sequence from a random library [32]. This process begins with a random library of on the order of 10^{15} unique nucleic acid sequences; for a library of length *n*, the number of possible sequences in the library is 4n (four possible bases, A, C, T, G, at each position) [30-32]. To allow for polymerase chain reaction (PCR) amplification, these stretches of variable sequences are flanked by primer-binding sites [34]. Those molecules able to perform catalysis under specific conditions are isolated from the inactive molecules. They are then enriched *via* PCR amplification and

the cycle is repeated. The resulting molecules are known as functional nucleic acids (FNAs).

The use of this combinatorial chemistry technique has been described as "nearlimitless" in its potential for isolating any artificial RNA or DNA catalysts with ideal selection conditions; the "search and evolution" strategy employed by SELEX allows the isolation of functional catalysts in the complete absence of structural knowledge [35]. In addition, the enriched catalyst can be improved through this process with mutations occurring over several cycles, or rounds, of the process [35].

1.4 Ribozymes: Progress towards an RNA World

Since the discovery of the first ribozymes in nature, scientists have been searching for a ribozyme capable of self-replication. A seemingly obvious means to accomplish self-replication is to have a ribozyme perform template-directed synthesis, where nucleotides are successively added to its complementary strand [36]. The method of *in vitro* selection has consequently been applied to engineer systems of RNA mimicking this specific aspect – templated Watson-Crick base pairs – of the proposed RNA World.

Attempts to solve this difficulty are numerous; the search began with RNA sequences able to catalyze ligation, which have been isolated by various groups [31], in the hopes that they could be modified to accomplish processive polymerization essential to a replicase. One such ribozyme capable of the ligation of a nucleotide triphosphate (NTP) substrate to a sequence, analogous to extending a primer, called class I ligase ribozyme, was isolated by the Szostak group [37]. Variants of this ligase went on to be modified and achieved this goal of successive polymerization, described below [38-42].

The variants developed from the class I ligase performed multiple ligations, in effect increasing the polymerase processivity [38-41]. The first ribozyme polymerases derived from class I ligase could only replicate abridged sequences of RNA shorter than the ribozyme itself [38-41] Recently, the Hollinger group isolated the first RNA catalyst capable of polymerizing a sequence greater than its own length [42]. Given the name tC9Y, this sequence is 202 nucleotides in length and is capable of RNA-directed polymerization of up to 206 nucleotides [42]. This specific RNA replicase was selected using mutated ribozymes (extensively modified from the class I ligase) bound to beads and the isolation of the altered ribozyme best able to extend the primer by nucleotide addition [36]. Disappointingly, although it can facilitate the copy of specific sequences, it cannot replicate its own sequence.

An alternative approach to evolving ribozymes for the RNA World scenario also makes use of *in vitro* selection, but eschews the template-directed scheme. Paul and Joyce used the technique to isolate an RNA ligase capable of joining two substrates together [43]. Joyce's group further improved upon this system by making it capable of crossreplication – selecting two ribozymes, each catalyzing the ligation of the other from a total of four substrates [44]. Because the ribozymes compete for the use of substrates, the variants with higher exponential growth rates have a competitive advantage, and they are therefore selected [45]. This system is the only one known to exponentially self-amplify, with the bottleneck being the availability of substrates [21]. However, the system has not undergone Darwinian evolution to attain a novel function; the sequence is too short to support the evolution of variants, and the substrates for lengthier sequences cannot be supplied at the required concentrations, as they would form complex structures and effectively sequester the substrates [21,46,47]. Nonetheless, this is the first system that has undergone directed evolution and is self-sustaining and capable of exponentially amplifying molecules as a result.

Joyce & Sczepanski went a step further with their cross-replicating system. Recognizing that the prebiotic soup may have contained enantiomers in equal concentrations, and that opposing enantiomers in a non-enzymatic heterochiral system actually inhibit template-directed polymerization [48], they attempted to simultaneously evolve two ribozymes of opposing chirality. They used *in vitro* selection to generate two enantiomers, L-RNA and D-RNA; the L-RNA was able to catalyze the ligation of D-RNA substrates onto a D-RNA template and the D-RNA was able to perform the opposite [49]. This was achieved through non-specific tertiary interactions, avoiding Watson-Crick base pairing and instead making use of the recognition of the RNA duplex [49].

Although a cooperative RNA arrangement involving more than one RNA polymerase is plausible under certain conditions [50], it is not the idyllic autocatalytic system of the typical RNA world enthusiast. The most elegant substantiation of the RNA world hypothesis would be the isolation of a ribozyme capable of templated self-replication.

Although some experts believe that enzymatic template replication is the obvious choice for investigating a potential RNA World, others disagree. Leslie Orgel initiated the approach of studying nonenzymatic template-directed replication [51,52], and in

2011, Deck et al. demonstrated that this method can lead to oligonucleotide synthesis in a timeframe of only days [53].

Jack Szostak in particular emphasizes that primordial fatty acid vesicles would be incompatible with ribozymes, as ribozymes require concentrations of divalent metal ions that would destroy the membranous vesicles [54-57]. Szostak's group investigated model protocells which were able to sustain a templated primer extension without a ribozyme; they found that the addition of citrate allowed self-copying while preventing the degradation of the membrane [58].

Both non-enzymatic and enzymatic template replication are reasonable approaches to the quandary of the RNA World Hypothesis. In his set of sub-problems, Orgel sets non-enzymatic copying prior to the development of ribozymes; however, it really is the classic problem, often encountered in origins of life research, of "which came first – the chicken or the egg?" Szostak, who has studied RNA in protocells, is convinced that RNA "couldn't have done it alone" but needed to be compartmentalized [59]. Szostak has an eight-step list to evolve an RNA protocell, three steps of which his group has accomplished and another three purported to be in progress [59].

Although ribozymes and their chemistries lend support to the RNA World, RNA is not the only nucleic acid with functional properties. While never found naturally, DNA has also been selected artificially to catalyze a wide variety of chemical reactions, briefly discussed below.

10

1.5 Deoxyribozymes: An Introduction

As mentioned previously, the method of *in vitro* selection to isolate FNAs is not limited to RNA. The first deoxyribozyme (also known as a DNA enzyme or DNAzyme) was isolated in 1994 by Breaker and Joyce and catalyzed the Pb²⁺-dependent cleavage of an RNA phosphodiester linkage [33]. Subsequently, many DNAzymes have been isolated with a variety of functions, catalyzing a broad array of chemical reactions.

Employing DNA molecules for an *in vitro* selection experiment has several advantages; it is less complicated than using RNA, as RNA replication requires the use of reverse transcriptase for replication. DNA is also inherently more stable and more tolerant to modifications than RNA, as well as less costly and simpler to synthesize [34,60].

Breaker and Joyce were the first to discredit any sort of "DNA World" or the idea of DNA as a precursor molecule to the central dogma [33]. However, the disparities between DNA and RNA are few – the ribose sugar has an additional 2'-hydroxyl compared to deoxyribose, and DNA uses the thymine base instead of uracil. It is therefore not exceptional to use a DNA system as a parallel way to study the origins of life and the RNA World. Our research uses a DNA system to investigate a novel proposition pertaining to the origin of life, as discussed in the next chapter.

1.5.1 DNAzymes: Scope and Use

In the past few decades, many deoxyribozymes have been discovered and/or engineered to be capable of catalyzing a wide variety of biologically relevant reactions [34,61,62]. Many of these reactions involve substrates which also contain nucleic acids [63]. DNAzymes are capable of forming secondary structures, defined as double-helical interactions through Watson-Crick base-pairing, and tertiary interactions, such as triplex interactions [64], guanine quadruplexes [65] and i-motifs [66]. These complex structural arrangements allow deoxyribozymes to catalyze specific reactions; in fact, the *in vitro* selection process itself may be biased towards simpler structures, as it so far has been used with small random libraries (40-100 nt) – deoxyribozymes may be capable of even more complex structures on a larger scale [63].

The table below summarizes the reactions which deoxyribozymes have been shown to be capable of catalyzing. Since the inception of artificially selected catalytic DNA, there has been much progress in the use of these versatile molecules, as facilitators of many biologically relevant chemical reactions, for applications including biosensing and therapeutics [67].

Reaction Catalyzed	Reference
RNA cleavage	[68-74]
RNA ligation	[75-77]
RNA branching	[78,79]
DNA phosphorylation	[80]
DNA capping	[81]
DNA ligation	[82,83]
DNA deglycosylation	[84]
Thymine dimer repair	[85]
Phosphoramidate cleavage	[86]
Porphyrin metalation	[87]

Artificial Deoxyribozymes

Table 1-1: Summary of reactions catalyzed by DNAzymes.

12

1.5.2 RNA-Cleaving Deoxyribozymes

Over the last two decades, deoxyribozymes able to catalyze the particular reaction of RNA transesterification have been characterized extensively. The very first DNAzyme isolated catalyzed a Pb^{2+} -dependent transesterification of RNA with a 10^5 rate enhancement after just five rounds of *in vitro* selection [33]. This transesterification reaction is well-known, as many protein enzymes and ribozymes depend upon this mechanism to cleave RNA [12,88-90]. Traditionally, RNA-cleaving deoxyribozymes have required the presence of divalent metal ions to facilitate the transesterification reaction. These deoxyribozymes share a common reaction mechanism: the nucleophilic attack by the 2'-hydroxyl group of ribose on the neighboring phosphodiester bond results in a 2', 3'-cyclic phosphate and a free 5'-hydroxyl terminus. [91-93].



Figure 1-2: RNA cleavage can be attained via a transesterification reaction.

The Joyce group went on to isolate two well-known DNAzymes, named 8-17 and 10-23, after their cycle and clone number, respectively. 8-17 has been isolated in multiple independent *in vitro* selection experiments and its catalytic core consists of a small 3-base

pair stem and 8 conserved nucleotides [92,94]. The 8-17 DNAzyme motif is also capable of cleaving fourteen out of sixteen possible dinucleotide junctions [95]. 10-23 consists of a 15-nucleotide catalytic core and variable-length substrate-binding arms, and cleaves RNA between a free purine base and a paired pyrimidine [96,97].



Figure 1-3: Structure of the 8-17 motif (left) and the 10-23 motif. Y represents a pyrimidinebased nucleotide and R represents a purine-based nucleotide. The filled triangle indicates the cleavage site of the substrate.

RNA-cleaving DNAzymes have additionally been developed that are able to cleave L-RNA, the enantiomer of D-RNA [98,99], notably useful for employing DNAzymes as biosensors in biological samples, which contain RNA-cleaving enzymes (RNAses). Accordingly, the RNA-cleaving reaction is versatile, useful, and wellcharacterized; an ideal target for the objectives of this project.

1.6 Conclusion and Thesis Overview

The purpose of this project is to make use of the technique of *in vitro* selection, used to isolate many functional nucleic acid enzymes by numerous groups in various past experiments, to evolve a catalytic sequence from a definitively non-catalytic sequence. As a proof of principle, this will show that under specific conditions, a nucleic acid sequence can be evolved into a functional catalyst, mimicking the events that must have occurred in the prebiotic RNA World.

Once a catalyst has been developed using the non-catalytic sequence, deep sequencing will reveal the mutations accrued by the sequence over several rounds of *in vitro* selection. This will allow the assessment of the sequence along various stages of its evolutionary path, showing the route taken to develop catalytic activity.

This experimentation was performed under strict artificial conditions, and isn't presumed to relate to early Earth conditions (unlike the Miller-Urey experiment, for example). However, this experiment shows that under the right conditions, a functional molecule capable of information storage can attain catalytic activity, giving it the properties necessary to survive in an RNA-world-like scenario.

Chapter Two: Developing a Catalyst from a Non-catalyst

Nucleic acid enzymes (NAEs) encompassing both ribozymes and deoxyribozymes have previously been isolated using the process of *in vitro* selection, a method characteristically performed by isolating catalytically active molecules from a large random library. It is assumed that a small fraction of active molecules are present in the pool of random sequences, and this method filters them from inactive molecules.

Does this method mimic evolution? Sequences capable of performing a desired reaction ("high fitness") are selected for, while sequences showing no aptitude for the reaction ("low fitness") are discarded. However, the pool's variation is inherent in its randomness, as a product of the combinatorial chemistry used to create the random library.

The foundation of this project is the proposal that in order to truly mimic the evolutionary process, the starting pool must not be random; instead, it should be a sequence established as *incapable* of the desired reaction. Because DNA polymerase enzymes are not infallible, performing rounds of *in vitro* selection with this sequence would very gradually introduce mutations into the pool, thereby establishing variation and the means by which some sequences achieve higher fitness. This *in vitro* evolution would show that functional molecules are capable, under certain conditions, of attaining catalytic activity.



Figure 2-1: Typical *in vitro* selection (top frame, described in chapter one) vs. the *in vitro* evolution process used in this project (bottom frame). Filled stars represent mutations accruing during the cycles or rounds of *in vitro* selection.

The results of this novel project show that *in vitro* selection (herein termed *in vitro* evolution) can be used to guide the evolution of a molecule from inactive to an active catalyst. With this work, it is revealed that DNA can act as a proxy system for the origins of biotic chemistry by isolating a functional catalyst from a previously non-catalytic sequence.

2.1 Introduction

The RNA World hypothesis, discussed in Chapter One, describes a primordial world where nucleic acids take on roles of both information carriers and replicases. However, the question remains: how could a ribozyme first emerge in the RNA world? An underlying implication of the RNA World Hypothesis is that a ribozyme is capable of emerging from a noncatalytic sequence of RNA under permissive evolutionary conditions. To demonstrate this idea, this project used the method of *in vitro* selection to conduct a test-tube evolution experiment [28,29,100].

The goal of these experiments was to convert a randomly chosen, noncatalytic sequence of single-stranded DNA, a proxy for RNA, into a catalytic DNA with RNase-like activity. Previous ribozymes and DNAzymes have been selected from libraries of completely random sequences [33,49,96,101-103], or randomized versions of a sequence with a dissimilar function [100,104]. In contrast, this study aimed to establish a catalyst from a distinct, distant and non-catalytic corner of sequence space.



Figure 2-2: Overall scheme to evolve a non-catalytic sequence (BTA1) with an attached ribonucleotide-containing substrate (S1) into an RNA-cleaving deoxyribozyme. Sequences of both substrate and starting sequence are shown; initial sequence (green) is flanked by primer-binding sites (grey) to facilitate PCR amplification.

The approach, represented in Figure 2-2, is designed to evolve a noncatalytic DNA strand into a DNAzyme capable of cleaving a single ribonucleotide embedded in a DNA sequence [33]. The starting sequence for the experiment was arbitrarily chosen to be the first 50 nucleotides (green nucleotides) of the *Bos taurus* (cattle) albumin gene (NCBI Reference Sequence: NM_180992.2; GI: 31340937). This candidate sequence was flanked on both 5' and 3' sides by arbitrarily chosen 20-nucleotide sequences (grey nucleotides), intended as primer-binding sites for PCR (which is used both to amplify the cleavage product and introduce a low level of point mutations into the evolving sequence). The entire 90-nucleotide DNA molecule is denoted BTA1. The substrate sequence, S1, contains 28 nucleotides (blue nucleotides, Figure 2-2) with a single ribonucleotide, guanosine ribonucleotide (purple G), located at the 14th position.

Through a modified method of *in vitro* selection, discussed below, S1 is attached to BTA1 and only sequences able to cleave the single guanosine ribonucleotide are permitted to ascend to the next round. Polymerase errors are capable of contributing to mutations in every round, and after several rounds of selection, a catalyst is evolved from a non-catalytic sequence.

2.2 Methods

2.2.1 Materials

All oligonucleotides were synthesized using standard phosphoramidite chemistry (IDT, Coralville, IA) and purified by 10% denaturing polyacrylamide gel electrophoresis (dPAGE). Their concentrations were determined spectroscopically. T4 polynucleotide kinase (PNK), T4 DNA ligase and Tth DNA polymerase were obtained from Thermo Scientific, BioBasic and Biotools, respectively, along with 10× reaction buffers. [γ -³²P]ATP and [α -³²P]dGTP were purchased from Perkin Elmer. All other chemicals were obtained from Sigma-Aldrich. Α S1: 5'-ACTCTTCCTA GCGGAGGTTC GATCAAGA 5' - TACGCAGTCA GTCAGTGTAC ATCTTTTCTA TCAACCCCAA BTA1: AACTTTGGCA CAATGAAGTG GGTGACTTTT GGCTAACTAC CCGAACTTCA FP1: 5'-TACGCAGTCA GTCAGTGTAC RP1: 5'-TGAAGTTCGG GTAGTTAGCC **RP2**: 5'-AAAAAAAAAAAAAAAAAAAAAAA_L-GTAGTTAGCC T1: 5'-TGACTGCGTA TCTTGATCGA В BTA1 + S1 + T1 Ligation (8) dPAGE Ligation



Figure 2-3: (A) Full sequences of all nucleic acid components used in the project. (B) *In vitro* selection methods: (1) Ligation of BTA1 to S1 templated by T1; (2) purification of ligated S1-BTA1; (3) incubation of S1-BTA1 in the presence of metal ions; (4) purification of cleaved products by dPAGE; (5) PCR using FP1 and RP1 as primers; (6) PCR with FP1 and RP2 as primers (note: RP2 contains hexaethyleneglycol spacer and A20 tail at the 5' end (the spacer prevents the poly-A tail from being amplified, making the non-DNAzyme-coding strand 20 nucleotides longer than the coding strand); (7) purification of BTA1 strand by dPAGE; (8) phosphorylating BTA1 and ligating it to S1. This cycle of steps was repeated 25 times in this study.

2.2.2 *In vitro* selection

The steps involved in the in vitro evolution experiment are depicted in Figure 2-3. For the first round of selection, 100 pmol of synthetic BTA1 (5'-TACGC AGTCA GTCAG TGTAC ATCTT TTCTA TCAAC CCCAA AACTT TGGCA CAATG AAGTG GGTGA CTTTT GGCTA ACTAC CCGAA CTTCA-3') was 5'phosphorylated by PNK (10 units) at 37 °C first with 10 µCi [γ -³²P]ATP (20 min) and with 1 mM ATP (20 min). This phosphorylated BTA1 is denoted generation 0 or G0. The phosphorylated DNA was ligated to 110 pmol of S1 (5'-ACTCT TCCTA GCGrGA GGTTC GATCA AGA-3'; rG: guanosine ribonucleotide) in the presence of 110 pmol of LT1 (5'-TGACT GCGTA TCTTG ATCGA-3') and 5 U of T4 DNA ligase (50 µL; 23 °C; 60 min). The ligated S1-BTA1 was purified by 10% dPAGE and used for RNA cleavage reaction (25 µL; 23 °C; 4 h) in 1× selection buffer (100 mM KCl, 300 mM NaCl, 15 mM MnCl₂, 15 mM MgCl₂, 55 mM HEPES, pH 7.5 at 23 °C). Following incubation at room temperature for 4 h, 1 µL of 0.5 M EDTA (pH 8.0) was added. Self-cleavage of S1-BTA1 (118 nt) would generate a 104-nt DNA fragment, which was purified using 10% dPAGE and subjected to DNA amplification by two consecutive polymerase chain reactions (PCR). PCR1 was performed in a 50-µL mixture containing 1× PCR buffer, 0.2 mM each dNTP, and 1 U of Tth DNA polymerase, 2.5 µM each of forward primer FP1 (5'-TACGC AGTCA GTCAG TGTAC-3') and reverse primer RP1 (5'-TGAAG TTCGG GTAGT TAGCC-3'). A small aliquot of the 1st PCR product was further amplified using FP1 and RP2 (5'-AAAAA AAAAA AAAAA AAAAA-iSp18-TGAAG TTCGG GTAGT TAGCC-3'; iSp18 refers to 18-atom hexaethylene glycol spacer) as primers. RP2 contains a poly(A) tail separated by a non-amplifiable linker and thus the 2^{nd} PCR product contains a non-coding strand that is 20-nt longer than the coding strand. This facilitates the separation of coding strand by dPAGE. The second PCR mixture also contained 5 µCi α -GTP for the purpose of radiolabelling. The coding sequence from PCR2 was purified using 10% dPAGE and used as the enriched pool for the second round of selection. The regenerated BTA1 molecules, some of which now contain low levels of mutations acquired during the PCR steps, are ligated again to S1 to initiate the second round of selection and amplification. This procedure was repeated for as many cycles as needed until the BTA1 pool acquired the detectable catalytic activity. A total of 25 cycles of test-tube evolution were conducted.

2.2.3 Activity assays of Selected Rounds

In order to conduct analysis of the activity of specific rounds, G16, 19, 22 and 25, 100 pmol of each radioactively labeled DNA pool-S1 construct was incubated in $1 \times$ selection buffer for 1, 2, 4, 24, 48 and 72 h prior to dPAGE analysis.

To determine the RNA cleavage rate constants of S1, G0 and 25 (Figure 1d), 100 pmol of radioactively labeled S1, G0-S1 and G25-S1 were incubated separately in $1\times$ selection buffer for 14 days. An aliquot was taken out each day and combined with $2\times$ denaturing gel loading dye to quench the reaction. All the samples were analyzed by dPAGE and visualized using Typhoon 9200 (GE Biosciences). The fraction of cleavage was quantified using Image Quant software (Molecular Dynamics). The rate constants

were determined by plotting the natural log of the fraction of uncleaved substrate at various reaction times. The negative slope of the resulting line, generated by a least-squares fit to the data is taken as the rate constant for RNA cleavage.

2.2.4 Structural Characterization

2.2.4.1 Activity comparison of truncated G2501. A series of mutated sequences of G2501 containing sequential 5-nt internal truncation were chemically synthesized and ligated to S1. Each ligated construct was incubated in 1× selection buffer for 24 h. All the samples were analyzed by dPAGE and visualized using Typhoon 9200. The fraction of cleavage was quantified using Image Quant software and normalized by taking G2501 as 100.

2.2.4.2 Activity of DNAzyme constructs in *trans.* Cleavage was tested in *trans* (i.e. where the substrate and sequence are not connected) with various enzyme/substrate constructs (see Figure 2-9). Each radioactive substrate was incubated with each enzyme (S:E = 1:50) for 24 h. All the samples were analyzed by dPAGE and visualized using Typhoon 9200.

2.3 Results and Discussion

2.3.1 Development of a Catalyst

No significant cleavage activity was observed for G0 through G20. PAGE separation yielded an extremely faint band appearing virtually nonexistent at the appropriate size (104 nt). However, weak but more noticeable cleavage activity was observed with all rounds after G21, continuing to intensify for the next 5 cycles, as shown in Figure 2-4.





By G25, a strong cleavage activity was established. Four DNA pools from different rounds of *in vitro* selection, G16, G19, G22 and G25 (each number representing the round number) were subjected to a cleavage assay (Figure 2-5). While both G16 and G19 showed no detectable activity even after 72 hours of incubation, both G22 and G25 exhibited robust cleavage activities in a time-dependent manner.




The cleavage activity was detectable in G25 even after 1 hour of incubation; by 72 hours, a significant fraction of G25 was cleaved. The self-cleaving activities of S1, G0 and G25 were compared using an extended time course (Figure 2-6). While the first-order cleavage rate constant of S1 and G0 was 5.5×10^{-7} /min [96], G25 had a rate constant of 4.3×10^{-4} /min, corresponding to nearly 1,000-fold rate enhancement. These analyses indicate that BTA1 was successfully converted into an RNA-cleaving DNAzyme.



Figure 2-6: A graphical representation of the improvement of cleavage from the initial sequence to the latest round. G0 refers to the initial sequence, prior to the first round of *in vitro* selection. G25 refers to the 25th round of selection. "Substrate only" refers to the substrate incubated alone in selection buffer.

A total of 25 rounds of *in vitro* evolution were conducted, and a DNA pool with significant catalytic activity was established. High-throughput sequencing analysis has identified mutations that have enabled the noncatalytic to catalytic sequence conversion. These findings demonstrate for the first time that an enzyme can arise from a defined sequence of a functional polymer via molecular evolution, a mechanism that may have been exploited by nature to initiate the evolution of enzymes in the RNA world and beyond.

2.3.2 Characterizing Sequence Structure

Previously discovered RNA-cleaving DNAzymes consist of a catalytic motif flanked by two duplex-forming arms that bind to the substrate [33,96]. Supported by this knowledge revealed in prior studies, the top ten sequence classes determined by deep sequencing (see Chapter Three) were used to propose a putative secondary structure of the most abundant sequence in round 25 (Figure 2-7).



Figure 2-7: Proposed secondary structure model of the top class isolated from round 25 of *in vitro* selection.

The structure contains 5 short duplex elements: P1, P2a, P2b, P3 and P4. According to this structural model, four of the seven common mutations, A13G, T50C, T39A and G40A result in more stable P1, P2a and P4 duplex motifs. The catalytic core is made of P1, P2a and P3, along with the single-strand region located between P1 and P3, termed SS1/3. Another well-known RNA-cleaving DNAzyme, 8-17, also contains a stem-loop with three base pairs followed by an unpaired region of 4-5 nt in its catalytic core [96], similar to our proposed structure, which contains four base pairs followed by an unpaired region of seven nucleotides.

Deep sequencing data, explored further in Chapter Three, also reveals that the final structure contains two deletions, both cytosine at positions 17 and 18. Because the catalytic core is made of P1, P2a and P3, along with S1/3, the deletion of C17 and C18 may have been an outcome of properly positioning the catalytic core for either the cleavage site recognition or chemical catalysis or both.

The sequence was examined by truncating consecutive portions of 5-nt, in order to determine which portions of the sequence were crucial to catalytic activity, and which were disposable. This data is shown in Figure 2-8. Generally, deletions occurring within P2b and P4 did not affect the catalytic activity. However, the deletions made at positions within P1, P2a, P3 and SS1/3 dramatically reduced the catalytic performance of the construct.

Certain mutations appear to enhance the structure of the catalyst; the aforementioned A13G, T50C, T39A and G40A add structural integrity to the duplex motifs. According to the structural model, the T6C mutation seems to have occurred as a side effect of the stem-loop structure containing four base pairs. In the original sequence, nucleotides 4 through 7 were all thymidine, allowing them to potentially take the place of one side of the stem-loop duplex, disrupting this structure.

The mutation at the 3' end of the sequence, C50T, evidently facilitates the formation of a duplex with the primer-binding site. This may serve to keep the primer site out of the way, of the main sequence, preventing it from disrupting the rest of the structure.



Figure 2-8: Sequences were tested for activity at 24h as consecutive 5-nt portions were removed from the initial sequence of G2501.

In addition to the truncation experiments, the points of interest where the initial mutations occurred were addressed individually. Figure 2-9 shows the 24h cleavage of G2501 when specific individual mutations are reverted back to the original sequence. This result demonstrates that certain mutations are absolutely essential to activity (e.g.

G13 and the CC deletion) while others have little effect – reverting C6 back to its original thymine actually slightly improves the activity, surprisingly.



Figure 2-9: The activities of the sequences where point mutations have been made in order to revert mutations back to the original. For example, C6T is G2502 with the cytosine at position 6 mutated back to the original sequence's thymine. Where mutations occurred consecutively (AA, CC) they have been tested together.

Based on these observations, a simplified secondary structure of G2501 was derived, which is shown as the first structural model in Figure 2-10. It is a *trans*-acting construct (*trans* specifying that the substrate is not attached covalently) with perfectly matched P1 and P2 and the removal of all catalytically disposable nucleotides. To verify duplex elements within the putative secondary structure of G2501, we designed two additional structural variants. In the first variant, the nucleobases within both P1 and P2 were significantly changed in an attempt to alter the sequence identity but maintain the duplex structure. The second variant was designed on the same principle but targeted P3 for base-pair co-variations in the stem-loop structure. The data provided in Figure 2-10 fully supports the proposed structural model.



Figure 2-10: Simplified secondary structure of G2501 with additional variants, as well as verification of duplex elements within. The first structure represents a *trans*-acting construct with perfectly matched P1 and P2. The refined secondary structure of the DNAzyme is designed based on the structural model in Figure 2-7 and truncation data in Figure 2-8. Specifically, all catalytically disposable nucleotides are truncated; *c*is-acting construct has been made into *trans*-acting format; perfectly base-pairing P1 and P2 are created (with the incorporation of two black nucleotides). The second structure is the altered version of the first structure where significant nucleobase substitutions (dark blue base-pairs) are made within both P1 and P2 in order to change the sequence identity but maintain the duplex structure. The third structure is similar to the first one except that three base-pair co-variations (dark blue base-pairs) are made within P3 in order to change the nucleotide content of P3 but maintain its duplex structure. The PAGE gel shows the testing of the activity of matched and mismatched substrate (S1 and S2) and enzyme strands (E1, E2 and E3).

2.4 Summary and Conclusion

Twenty-five rounds of *in vitro* evolution with a starting sequence incapable of catalysis yielded a catalyst capable of RNA transesterification. The *in vitro* selection method is typically used to isolate catalytically active molecules from a large random library. In contrast, this method began with a sequence established as *incapable* of the desired reaction. The resulting top sequence had a 1000-fold improvement in the transesterification reaction over the original sequence.

The secondary structure of the sequence was determined by modifying aspects of the structure and resolving certain sequence elements as vital or nonessential to the structure. In this way, the active portions of the sequence can be confirmed, and the presumed structure of the evolved DNAzyme can be established as a catalyst.

Chapter Three: Investigating the Development of a Catalyst

Attempts to understand the origins of life using artificial systems and Darwinian evolution are not limited to conventional *in vitro* selection. Recent advances in sequencing allow in-depth analysis of sequence pools and the development of structures for a potential RNA World. A few recent advances in this field are highlighted in the subsection below.

Deep sequencing allowed determination of the precise sequence of the emergent catalyst dominating the last rounds of *in vitro* selection, and study and analysis of each round of our *in vitro* selection experiment allows comparison of different rounds alongside one another. By means of this method, similar sequences that arose from the non-catalyst were compared and a plausible pathway by which the mutations arose *via in vitro* evolution was proposed.

3.1 Sequencing and Fitness Landscapes

The theoretical realm of *in vitro* selection uses a fitness landscape, calculated from the similarity or relatedness of a set of genotypes and their related fitness, as a quantitative measure. The concept of fitness was first used in evolutionary biology by Wright [105] to map the reproductive successes of individuals connected through mutations. Investigation into the fitness landscapes of ribozymes have yielded mixed results [106]. The mutation path between two ribozymes isolated by Schultes & Bartel (2000), one a ligase and the other a self-cleaving RNA, was analyzed; intermediates with two mutations from its closest neighbour retains the activity of its neighbor (e.g. the

ligase with two mutations is still capable of ligation) [106,107]. An intermediate containing 40 mutations relative to both ribozymes (i.e. the variant mutated to be in between both sequences) contains the (reduced) activity of both [106]. This system has no defined "fittest" ribozymes with high fitness peaks.

However, in another analysis, the fitness landscape of the class II ligase ribozyme was characterized by deep sequencing [108]. Also known as next-generation sequencing or high-throughput sequencing, deep sequencing refers to sequencing in which the number of reads of a nucleotide (its depth) is much greater than the number of nucleotides in the sequence. The ligase was mutagenized and the library enriched for highly catalytic ribozymes; sequencing revealed that the best sequence, at the centre of the fitness peak, was the original sequence, surrounded by close relatives [106].

Continuous evolution of a small sample of ligase ribozymes was performed by Arenas & Lehman' and again, the original sequence was at the very centre of the fitness peak, along with closely related sequences with only a few mutations [108,109]. In contrast to the analysis of Schultes & Bartel's ribozymes, this would suggest that the fitness landscape has few peaks (i.e. only a few genotypes have high fitness) [106].

Petrie & Joyce conducted their own experiment where they took two ligase ribozymes and performed continuous evolution with mutagenesis *in vitro* within millions of separate compartments; deep sequencing revealed that in each case, the original sequence was the sequence with the highest fitness, with other closely related sequences surrounding it in a cluster [106]. The fitness landscape is described as one "where high-fitness genotypes are [surprisingly] sparse and well isolated"; not necessarily the rule for

evolution of these systems, but evidence that *in vitro* selection may direct evolution towards one of just a few high fitness options [106].

The fitness landscape explored by these *in vitro* selection experiments highlights that in any system, there could be relatively few ultimate peaks for fitness, funneling the selection experiment towards one or more of the best potential sequences. Alternatively, there could be many fitness peaks in the system, with several ideal sequences the end goal of the directed evolution process. Understanding the fitness landscapes of systems such as these allows the exploration of reactions and molecules that could potentially elucidate the RNA World.

In these experiments, deep sequencing was used to further the understanding of the process of evolution over several rounds of *in vitro* selection. Whether there occur several different fitness peaks remains unclear – only one sequence and its preceding ancestors has been studied extensively. However, this method could be used for future studies to potentially find multiple peaks in fitness across the end results of *in vitro* selection.

3.2 Methods

3.2.1 Deep Sequencing

In the context of this research evolving an RNA-cleaving DNA catalyst from a non-catalytic sequence, deep sequencing technologies sanction a thorough investigation into the populations in each round of *in vitro* selection. This permitted the recognition of

when mutations took place in this selection procedure, as well as their frequencies and accumulation.

The PCR amplicons from Generations 1, 2, 4, 6, 8, 10, 12, 14, 16, 17-25 were sequenced using an Illumina Miseq DNA sequencer at the Farncombe Metagenomics Facility, McMaster University. The raw sequencing data was processed using Illumina's Basespace online NGS platform to sort tagged sequence pools and output sequence data in FASTQ format. Paired-end reads were merged and trimmed of primers using PANDAseq 2.6 analysis [110]. Sequences containing less than perfect complementarity were discarded to minimize sequencing errors in the dataset. FASTA format trimmed sequences were dereplicated and tagged with copy number using USEARCH v7.0.1090_i86linux32 [111]. Trimmed and dereplicated data was then imported into a MySQL 5.6.17 database and annotated with metadata such as mutation name, copy number and selection round. MySQL database queries were used to export sequence data and population level statistics for further analysis in Microsoft Excel 2010. Sequences comprising the limited sequence space used for the determination of mutation pathways were enumerated manually and then used to query the MySQL database for copy number data. Frequencies of each sequence for a given round were calculated in Excel from sequence copy number and total round populations. Sequences with fewer than three copies detected per round were not considered for pathway analysis to reduce spurious signals, weak signals and contamination of early sequencing pools with late round sequences.

3.2.2 Activity comparison of top 10 G25 sequences

100 pmol of each of the top 10 sequences in the G25 pool was chemically synthesized and ligated to S1. Each ligated construct was incubated in $1 \times$ selection buffer for 24 h. All the samples were analyzed by dPAGE and visualized using Typhoon 9200.

3.2.3 Activity analysis of MD0-7

100 pmol of each sequence was ligated to S1 and purified by dPAGE. Each purified construct was incubated in 1× selection buffer for 0.5, 1, 2, 4, 8, 12, 24, 48, 72, 96, 120, 144, 168, 192, 216, 240, 264, 288, and 312 h. All the samples were analyzed by dPAGE and visualized using Typhoon 9200. The fraction of cleavage was quantified using Image Quant software.

3.3 Results and Discussion

3.3.1 Top Sequence Analysis

The end result of *in vitro* evolution, the G25 pool, was subjected to deep sequencing and the top ten sequence classes, denoted G2501-G2510, are listed in Figure 3-2. As shown in Figure 3-1, the top class, G2501, represents 54.83% of the G25 pool; the second most-abundant variant, G2502, comprises 8.17% of the pool and the next 8 sequence classes jointly account for 12.77%.



Figure 3-1: Composition of Round 25 (G25) as determined by deep sequencing.

The top 10 classes each acquired 6-8 mutations (including deletions) within the sequence element of 50 nucleotides (Figure 3-2). Most of the mutations appear to be linked to the selection pressure, rather than random drift, as they occur in all or most of the sequence classes. Interestingly, the following 4 mutations are observed in all top 10 sequences: A13G, Δ C17, Δ C18 and T50C, which suggests that they play either essential or highly important roles in the noncatalytic to catalytic conversion. In addition, T6C, T39A and G40A are observed with most of the top 10 sequences, implying that they act as favourable mutations during the evolution into a catalytic sequence.



Figure 3-2: Top ten sequence classes of G25, normalized to original sequence (green, G0). Divergences from the original sequence (mutations and deletions) are shown in red.

The cleavage activities of the top 10 sequences were compared at 24h and all ten were shown to exhibit similar cleavage activity of between 35-49%, with the top class (G2501) holding a slight edge over the other sequences (Figure 3-3). In contrast, the G25 pool exhibited only 39% cleavage.



Figure 3-3: Activity of top ten sequence classes of G25 after an incubation period of 24h, revealed on dPAGE. G25 is a portion of the pool during the 25^{th} round of selection. The top sequence (2501) is shown on both top and bottom for comparison.

This appears to demonstrate that although the top ten sequences are largely similar with few disparities, these differences have all improved the activity of the catalyst to a great extent.

3.3.2 Developing a Mutation Pathway

The non-catalytic sequence G0 has evolved into a catalyst by acquiring several deletions and mutations. The effectiveness of the catalyst varies somewhat within the top

ten sequences, and this result is presumed to carry through the classes which have less representation in the final pool. However, up until this point the data only shows that G0 evolved into G25. The method of *in vitro* evolution explains the premise of how the sequence evolved, but not the precise roadmap by which this occurred.

The potential evolutionary pathways between G0 and G2501 were thus investigated. For this purpose, deep sequencing analysis of the generation 1, 2, 4, 6, 8, 10, 12, 14, 16, 17-25 pools was conducted. There is a vast number of potential pathways between G0 (renamed MD0) and G2501 (renamed MD7). Thus, the sequence space used to propose a pathway has been redefined as being comprised of all sequences possessing mutations required to mutate MD0 to MD7 with the minimum number of mutational events. An important assumption in this process is that in a given round of selection, only a single mutation occurs, as single mutational events are more likely to occur than multiple simultaneous mutations. For example, there are 6 single-mutation possibilities: T6C, A13G, G17- (or G18- as they are equal), T39A, G40A and T50C.

To determine a mutation pathway, mutants showing a strong signal throughout selection were identified through calculating the proportion of each sequence in the sequence space across all sequenced rounds where a minimum of 3 copies are detected. The 4-mutation sequence A13G/C17-/C18-/T50C (named MD4) was detected in high proportion and persistently relative to other 24 4-mutation sequences. Using MD4 as an anchor, the mutation chain was expanded working in both directions. Figure 3-4 elucidates the changes between sequences along MD0-MD7. Among 16 possible 5-mutation sequences, only 3 fell within the MD4 mutation chain, and of these, A13G/C17-

/C18-/G40A/T50C (MD5) fits best as it becomes persistent shortly after MD4 appears. Given MD5, only two 6-mutation sequences are possible. Mutant T6C/A13G/C17-/C18-/G40A/T50C (MD6) is persistent after the appearance of MD5, therefore we believe the 6^{th} mutation acquired to be T6C. This puts T39A as the final mutation.

Working backwards from MD4, A13G/C17-/C18- (MD3) shows the greatest persistence amongst the 3 possible mutants, suggesting T50C is the 4th mutation. There are two 2-mutation mutants (C17-/C18- and A13G/C17-) that can give rise to MD3. The double deletion C17-/C18- was favourably considered as a deep sequencing analysis of the mutation distribution of the DNA polymerase showed a higher than expected proportion of sequential double deletions than would be expected for independent single deletion events. Therefore it seemed likely that the C17-/C18- (MD2) double deletion occurred very early in its evolution. Based on the above analysis, the following mutation chain was proposed: C17- \rightarrow C18- \rightarrow A13G \rightarrow T50C \rightarrow G40A \rightarrow T6C \rightarrow T39A.

Finally, time-dependent cleavage activity of each mutant sequence – MD0-MD7 – along the proposed evolutionary pathway was analyzed (Figure 3- 5). The activity profile indicates that two mutations along this proposed pathway are the key drivers of the evolution, T50C (4th mutation) and T39A (7th mutation), because each mutation resulted in significant increase in catalytic activity.



Figure 3-4: The time-dependent cleavage activity of MD0-7. MD0 is G0, and MD7 is G2501. MD1-7 contains progressive mutations in the following sequential order: C17-, C18-, A13G, T50C, G40A, T6C and T39A. The time points (0–256 h) are set up to reveal both strong and weak activities expected of these DNA molecules.



Figure 3-5: Percent cleavage vs. reaction time along the MD0-7 pathway.

It is interesting to note that the improvements in activity accompanying certain mutations are larger than in others; this is an exciting prospect. Not only can tiny variations – single mutations and deletions – lead to gradual, almost imperceptible improvements, but evolution sometimes leaps forward, improving a great deal with the addition of just one more position of increased variation.

In terms of a fitness landscape, it appears that much of the final pool of successfully evolved sequences derived from the same pathway. It would be expected

that mutations occurred relatively consistently throughout the multiple rounds, initiating multiple potential pathways – but this variation is not explicitly manifest in the final pool.

It is not inconceivable to propose that several sequences could result from similar *in vitro* evolution experiments, along completely different mutation pathways. However, it appears that in this case, the fitness landscape favoured the sequence that came out on top, with not much room for competitors.

3.4 Summary and Conclusion

By using the comprehensive method of deep sequencing, it is possible to determine the composition of each pool of *in vitro* evolution. This allows tracking the emergence of sequences which engulf the pool as they are enriched over a series of rounds in *in vitro* evolution. Assorted variants, with differences as small as a single nucleotide, can be catalogued by this method.

In addition, deep sequencing allows the comparison of different pools – different rounds of *in vitro* evolution – to be contrasted, and a means by which evolution occurred to be proposed. This capability for organization and classification gives greater credence to this project, showing not only that a catalyst evolved from a non-catalyst, but a probable means by which this occurred.

Chapter Four: Discussion of Additional Experimental Data

The development of a catalyst and its subsequent analysis opens the door to even more questions in the domain of the origin of life. Just how far can this catalyst be pushed to evolve, and how many more mutations will accrue along this pathway? Can this experiment be replicated with a different sequence, a different reaction, or both? These questions were examined subsequent to the evolution of a catalyst from a non-catalyst, and the experiments and results are discussed below.

4.1 Optimization of an RNA-cleaving DNAzyme

Deep sequencing analysis has elucidated the means by which this non-catalytic sequence transformed into a catalyst. However, the best sequence capable of cleaving our substrate still only had a rate constant of 4.3×10^{-4} /min after 25 rounds. This is relatively slow compared with the calculated catalytic rate for 10-23 of ≈ 0.1 /min [96]. In contrast, Breaker and Joyce isolated an RNA-cleaving DNAzyme from a random-sequence pool after merely five rounds of selection, and the derived DNAzyme is capable of achieving 10^{5} -fold rate enhancement for the same reaction [33]. Evidently, the catalyst evolved in this project has a long way to go.

Evolution isn't directed towards a specific goal in mind – it is brought about by disparities in the fitness of different variations. The final catalyst which emerged from these experiments was evolved to perform under specific conditions. Changing the conditions in a way that increases the selective pressure (e.g. giving the sequence only 1 hour to perform the RNA cleavage, as opposed to 4 hours) should, in theory, yield an even more catalytically active sequence.

This theory was tested by an undergraduate student, Laura Chan. A library was designed that took only the parts of G25 that were deemed essential by secondary structure characterization. The catalytic domain of each sequence in the library was partially mutagenized in sequence; each nucleotide in the catalytic domain had a 70% chance of being the original nucleotide, and a 10% chance of being one of the other three nucleotides. Newly designed primers, so as to prevent contamination of the previously evolved library, were also added. *In vitro* selection was carried out as outlined in Chapter Two.



Figure 4-1: New library structure; each N represents a nucleotide that has a 70% chance of remaining original. Primers are different from initial BTA1 primer binding sites.

As of this writing, this library has undergone seven rounds of *in vitro* selection and yields promising results; it awaits further deep sequencing to discern the precise library composition.

This experiment suggests that it is possible to improve the catalyst developed previously; although variation has been introduced artificially, the sequence is statistically still extremely similar to the original sequence, and it will be exciting to observe which changes have led to improvements. These experiments can also further the general understanding of the means of enhancing a nucleic acid enzyme, which would prove useful in applications such as therapeutics and sensing purposes.

4.2 Selection of an RNA-cleaving DNAzyme (II)

This project has taken a definitively non-catalytic sequence and evolved mutations enabling the transformation of the sequence into a catalyst. The sequence that originally had no catalytic activity was randomly chosen, and through selective pressure, evolved into a sequence with a function – the ability to catalyze the cleavage of an RNA substrate.

This experiment has important implications involving the origins of life. Using a DNA system as a model, this project mimics the factors of advantageous mutations and selective pressure of molecular evolution on a molecule containing the ability to transfer genetic information as well as the potential to catalyze chemical and biological reactions. This system acts as an artificial substitute to explore an RNA-World type scenario.

However, the question remains – did this experiment yield the exception to the rule? Is it truly possible, to take any functional sequence (e.g. nucleic acids) and evolve it into a catalytic molecule by means of *in vitro* evolution? This idea was tested by using different sequences.

4.2.1 Problems and Pitfalls

Two recurring issues, outlined and discussed below, plagued the ensuing attempts to evolve a second catalyst from a non-catalytic sequence.

4.2.1.1 Longer Sequences

When selection of BSA1 was initially performed, another selection was executed in parallel with a slightly shorter sequence – the first 30 nucleotides from the mRNA of casein, flanked with primer binding sites on its 5'- and 3'- ends. This sequence was subjected to precisely the same protocol as BSA1. After just three rounds, it was realized that a sequence space of only 30 nucleotides was likely not enough room to evolve a catalyst, based on structural hypotheses of previous RNA-cleaving DNAzymes.

Thus, this parallel selection was repeated with a somewhat longer sequence – the first 70 nt from the mRNA of casein, flanked with primer binding sites on its 5'- and 3'- ends. The rationale behind the different sequence lengths of the two parallel selection experiments was to keep them separate at all times, avoiding potential contamination issues (e.g. a catalyst developing in one pool and cleaving the substate attached to the sequences in the other pool).

During round C21, it was found that somehow during PCR, a non-specific 90 nt sequence was being amplified, resulting in a triple and quadruple banded pattern and making it virtually impossible to isolate the band of the correct size (110 nt). Multiple attempts to repurify sequences and redo PCR amplification yielded the same problem. *In vitro* selection of this sequence was therefore discontinued.

4.2.1.2 Altered Primer Sequences (I)

The mRNA sequence of BSA was deemed reusable as a template for *in vitro* evolution. The two subsequent stretches of 50 nt were flanked them with the same

primers as the BSA-1 sequence, but with a scrambled 5-nt section of the reverse primers to prevent their annealing to BSA1.

Twenty rounds of *in vitro* selection were performed (See protocol in Chapter Two), and the presence of cleavage after four hour selection periods for both sequences looked promising. A time course over a 24-hour period was performed, with the sequences showing much improvement over time (Figure 4-2).



Figure 4-2: Cleavage analysis of round 20 for sequences A and B. The presence of 16% (A) and 8% cleavage (B) at 24 hours hint at the presence of a catalyst.

However, deep sequencing analysis performed on rounds 0 through 18 for both sequences had worrisome results. Unfortunately, contamination of the G2501 sequence was found in the last rounds of both A and B to be sequenced. G2501 made up on the order of 0.01% of the class population of these latest rounds. The presence of the G2501 sequence in the deep sequencing pool suggests that the pools were contaminated with G2501 sequences at one point during *in vitro* selection and this G2501 sequence was somehow being replicated with the other sequences. If also present in the selection pool,

G2501 would have a head start on the A and B sequences, and would always outcompete them.

In vitro selection of this sequence was therefore discontinued.

4.2.1.3 Altered Primer Sequences (II)

The same strategy as above was subsequently used for another attempt at *in vitro* evolution, using completely redesigned primers. The two subsequent stretches of 50 nt BSA mRNA were flanked with 20 nt primer-binding sites. This new selection followed a slightly different method, involving the induction of more mutations *via* PCR before applying selective pressure. A series of multiple PCR reactions were performed to amplify each diluted previous "round" before applying selective pressure. After the equivalent of 18 "partial rounds" of PCR, including 2 full rounds involving the application of selective pressure, *in vitro* selection of this sequence was discontinued. The design of the library was altered so that selection could be performed much more quickly, discussed below.

4.2.2 New Library Design

The successful evolution of a non-catalyst to a catalyst was a relatively timeconsuming process, taking 25 rounds to complete while applying selective pressure at each round. A new library was designed in order to partially bypass this lengthy process. BSA libraries of 50 nt (taken from stretches of BSA) were modified so that each nucleotide had an 85% chance of remaining original, and a 5% chance of being altered to each of the other three nucleotides.

Sequence Description	Sequence
R substrate (Esub)	5'-ACT CTT CCT AGC G rG* A GGT TCG ATC AAG A-3'
<i>In vitro</i> selection	
BSA-2	5' - CAG ACT TAT GTA GGC TCC AG
	ATTTCTCTCTCTCTCTCTCAGCTCTGCTTATTCCAGG
	GGTGTGCT TGC GAA CTG ACA CTT GAG -3'
BSA-3	5' - CAG ACT TAT GTA GGC TCC AG
	TTTCGTCGAGATACACACAAGAGTGAGATTGCTCATC
	GGTTTAAACT TGC GAA CTG ACA CTT GAG -3'
BSA-4	5' - CAG ACT TAT GTA GGC TCC AG
	GATTTGGGAGAAGAACATTTTAAAGGCCTGGTACTGA
	TTGCCTTTCT TGC GAA CTG ACA CTT GAG -3'
E - BSA2 - LT	5'- C ATA AGT CTGTC TTG ATC GA -3'
(Ligation Template)	
E - BSA2 - FP1	5'- CAG ACT TAT GTA GGC TCC AG -3'
(Forward Primer 1 for	
PCR)	
E - BSA2 - RP1	5'- CTC AAG TGT CAG TTC GCA AG -3'
(Reverse Primer 1 for	
PCR)	
E - BSA2 - RP2	5'-AAA AAA AAA AAA AAA AAA AAA AA/iSp18*/ CTC AAG
(Reverse Primer 2 for	TGT CAG TTC GCA AG -3'
PCR)	

Table 4-1: Design of new library. Red portions are 85:5:5:5 where 85% is the original nucleotide.

Initially, this selection went well. However, during round three of selection, a multi-banded pattern appeared on the PCR dPAGE gel intended to separate the coding from the non-coding strand. A look back into agarose gels testing the quality of PCR revealed that the primers were contaminated with a sequence (likely the original library) which was being amplified, even in the control (no template) PCR reaction. The primers were consequently re-purified, and selection began again from round one, free of contamination.

However, multi-banded patterns appeared again in the third round of selection. Both bands represented on the PAGE gel representing the coding ssDNA have been isolated, and *in vitro* selection continued; it remains to be seen whether either, or both, these bands can be amplified in order to continue with the procedure.

If this experiment were to be successful, it would show that a non-catalytic sequence can become catalytic even when minor changes are made to the sequence. This is similar to the initial *in vitro* evolution experiment; the difference is that mutagenesis is artificially induced at a specific ratio, and not randomly introduced by PCR amplification.

Chapter Five: Summary and Conclusion

Many previous studies have reported *in vitro* selection of ribozymes and DNAzymes from random-sequence DNA or RNA pools. This study presents the very first example of turning a non-catalytic DNA sequence into a DNAzyme through *in vitro* evolution. In comparison to the random-library approach, this defined-sequence approach requires a longer evolutionary time, and the derived DNAzyme exhibits a weaker catalytic activity. Specifically, G2501, which only produces 10³-fold rate enhancement for RNA cleavage, was obtained after 25 rounds of selective enrichment. Nevertheless, our study has demonstrated the feasibility to evolve a catalyst from a noncatalytic nucleic acid sequence.

This unique project has investigated evolution at the origins of life with a distinctive method *in vitro*, employing catalytic ability as selective pressure and allowing a non-catalytic sequence to evolve successively over time. The fact that a catalyst evolved from a non-catalytic sequence proffers corroboration of the RNA World Hypothesis, showing that under certain conditions, a sequence which is incapable of catalyzing a certain reaction can evolve to acquire that capacity. This strongly supports the idea suggested by the RNA World: once a functional polymer exists, it can evolve to take on a certain utility.

The catalyst which emerged from the *in vitro* evolution experiment, G2501, has a proposed secondary structure, established by removing partial sections of the sequence and/or truncating the sequence. This sequence is overwhelmingly dominant in the final

selection pools; the evolution pathway analyzed here has apparently outcompeted the other potential contenders. The deep sequencing methods explored in this project can also be exploited for other, similar experiments to distinguish what is occurring in *in vitro* selection.

Where to go from here? Chapter Four has outlined two ongoing projects: 1) the evolution of an improved catalyst, and 2) evolution of a catalyst from a different non-catalytic sequence. Additional research projects along the same lines could potentially be explored. For example, G0 could be evolved with different primers in order to compare pathways; would the pathway discussed here dominate that selection as well? What if the pool of dNTPs during PCR was biased towards purines, or against cytosine? How would this affect the selection?

Another appealing follow-up to this project would be to carry out *in vitro* evolution with a different method of selective pressure, forcing the sequence to perform a different reaction. Table 1-1 lists the variety of reactions known to be performed by artificially selected DNAzymes; evolving a deoxyribozyme to perform a more anabolic reaction, such as ligation or polymerization, would arguably be more relevant to an origins of life situation. However, as a proof of principle for catalytic evolution, RNA-cleaving DNAzymes suit the purpose nicely.

At some point in prebiotic history, RNA must have taken on a catalytic role in order to self-replicate and catalyze other life-supporting chemistries, as described by the RNA World hypothesis. This study shows that when an RNA-like polymer with a defined sequence is given a chance to mutate and reproduce, it can evolve into a purposeful enzyme. This one-of-a-kind artificial evolution experiment was performed in a test tube in a few weeks whereas nature had millions of years to carry out experiments. This experiment suggests that at the outset of life's origins, to evolve powerful enzymes from noncatalytic polymeric molecules, a similar mechanism may have been explored.

Reference List

- 1. Heller R, Williams D, Kipping D, Limbach MA, Turner E, et al. (2014) Formation, Habitability, and Detection of Extrasolar Moons. Astrobiology 14: 798-835.
- Darwin C (1859) On the origins of species by means of natural selection. London: Murray.
- Miller SL (1957) The formation of organic compounds on the primitive earth. Annals of the New York Academy of Sciences 69: 260-275.
- 4. Miller SL (1953) A production of amino acids under possible primitive earth conditions. Science 117: 528-529.
- 5. Miller SL, Orgel LE (1974) The origins of life on the earth: Prentice-Hall Englewood Cliffs, NJ.
- 6. Bada JL (2013) New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. Chemical Society Reviews 42: 2186-2196.
- 7. Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459: 239-242.
- 8. Ferris JP (2005) Mineral catalysis and prebiotic synthesis: montmorillonite-catalyzed formation of RNA. Elements 1: 145-149.
- Monnard P-A, Kanavarioti A, Deamer DW (2003) Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. Journal of the American Chemical Society 125: 13734-13740.
- 10. Higgs PG, Lehman N (2014) The RNA World: molecular cooperation at the origins of life. Nature Reviews Genetics.
- 11. Crick F (1970) Central dogma of molecular biology. Nature 227: 561-563.
- 12. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35: 849-857.
- 13. Zaug AJ, Grabowski PJ, Cech TR (1983) Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage–ligation reaction. Nature 301: 578-583.

- 14. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA (2000) The structural basis of ribosome activity in peptide bond synthesis. Science 289: 920-930.
- Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, et al. (2001) Crystal structure of the ribosome at 5.5 Å resolution. science 292: 883-896.
- Orgel LE (1968) Evolution of the genetic apparatus. Journal of molecular biology 38: 381-393.
- 17. Woese CR (1968) The fundamental nature of the genetic code: prebiotic interactions between polynucleotides and polyamino acids or their derivatives. Proceedings of the National Academy of Sciences of the United States of America 59: 110.
- Crick FH (1968) The origin of the genetic code. Journal of molecular biology 38: 367-379.
- 19. Gilbert W (1986) Origin of life: The RNA world. Nature 319.
- 20. Joyce GF (1989) RNA evolution and the origins of life. Nature 338: 217-224.
- Robertson MP, Joyce GF (2014) Highly efficient self-replicating RNA enzymes. Chemistry & biology 21: 238-245.
- Leslie E O (2004) Prebiotic chemistry and the origin of the RNA world. Critical reviews in biochemistry and molecular biology 39: 99-123.
- Woese CR (2002) On the evolution of cells. Proceedings of the National Academy of Sciences 99: 8742-8747.
- 24. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, et al. (2002) Genetic control by a metabolite binding mRNA. Chemistry & biology 9: 1043-1049.
- 25. Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, et al. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell 111: 747-756.
- 26. Winkler W, Nahvi A, Breaker RR (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419: 952-956.
- Winkler WC, Cohen-Chalamish S, Breaker RR (2002) An mRNA structure that controls gene expression by binding FMN. Proceedings of the National Academy of Sciences 99: 15908-15913.

- Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249: 505-510.
- 29. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. nature: 818-822.
- Joyce GF (2004) Directed evolution of nucleic acid enzymes. Annual review of biochemistry 73: 791-836.
- Joyce GF (2007) Forty years of in vitro evolution. Angewandte Chemie International Edition 46: 6420-6436.
- 32. Chapman KB, Szostak JW (1994) In vitro selection of catalytic RNAs. Current opinion in structural biology 4: 618-622.
- 33. Breaker RR, Joyce GF (1994) A DNA enzyme that cleaves RNA. Chemistry & biology 1: 223-229.
- Baum D, Silverman S (2008) Deoxyribozymes: useful DNA catalysts in vitro and in vivo. Cellular and Molecular Life Sciences 65: 2156-2174.
- 35. Jäschke A, Seelig B (2000) Evolution of DNA and RNA as catalysts for chemical reactions. Current opinion in chemical biology 4: 257-262.
- Lehman N (2013) Origin of life: Cold-hearted RNA heats up life. Nature chemistry 5: 987-989.
- Ekland EH, Szostak JW, Bartel DP (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. Science 269: 364-370.
- Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNAcatalyzed RNA polymerization: accurate and general RNA-templated primer extension. Science 292: 1319-1325.
- McGinness KE, Wright MC, Joyce GF (2002) Continuous in vitro evolution of a ribozyme that catalyzes three successive nucleotidyl addition reactions. Chemistry & biology 9: 585-596.
- 40. Zaher HS, Unrau PJ (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. Rna 13: 1017-1026.

- 41. Wochner A, Attwater J, Coulson A, Holliger P (2011) Ribozyme-catalyzed transcription of an active ribozyme. Science 332: 209-212.
- 42. Attwater J, Wochner A, Holliger P (2013) In-ice evolution of RNA polymerase ribozyme activity. Nature chemistry 5: 1011-1018.
- Paul N, Joyce GF (2002) A self-replicating ligase ribozyme. Proceedings of the National Academy of Sciences 99: 12733-12740.
- 44. Kim D-E, Joyce GF (2004) Cross-catalytic replication of an RNA ligase ribozyme. Chemistry & biology 11: 1505-1512.
- Lincoln TA, Joyce GF (2009) Self-sustained replication of an RNA enzyme. Science 323: 1229-1232.
- 46. Sczepanski JT, Joyce GF (2012) Synthetic evolving systems that implement a userspecified genetic code of arbitrary design. Chemistry & biology 19: 1324-1332.
- 47. Ferretti AC, Joyce GF (2013) Kinetic properties of an RNA enzyme that undergoes self-sustained exponential amplification. Biochemistry 52: 1227-1235.
- 48. Joyce G, Visser G, Van Boeckel C, Van Boom J, Orgel L, et al. (1984) Chiral selection in poly (C)-directed synthesis of oligo (G).
- 49. Sczepanski JT, Joyce GF (2014) A cross-chiral RNA polymerase ribozyme. Nature 515: 440-442.
- 50. Shay JA, Huynh C, Higgs PG (2015) The origin and spread of a cooperative replicase in a prebiotic chemical system. Journal of theoretical biology 364: 249-259.
- 51. Wu T, Orgel LE (1992) Nonenzymic template-directed synthesis on oligodeoxycytidylate sequences in hairpin oligonucleotides. Journal of the American Chemical Society 114: 317-322.
- 52. Hill Jr AR, Orgel LE, Wu T (1993) The limits of template-directed synthesis with nucleoside-5'-phosphoro (2-methyl) imidazolides. Origins of Life and Evolution of the Biosphere 23: 285-290.
- 53. Deck C, Jauker M, Richert C (2011) Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. Nature chemistry 3: 603-608.

- 54. Gebicki J, Hicks M (1973) Ufasomes are stable particles surrounded by unsaturated fatty acid membranes.
- Deamer DW, Dworkin JP (2005) Chemistry and physics of primitive membranes. Prebiotic Chemistry: Springer. pp. 1-27.
- 56. Szostak JW (2012) The eightfold path to non-enzymatic RNA replication. J Syst Chem 3.
- 57. Bowman JC, Lenz TK, Hud NV, Williams LD (2012) Cations in charge: magnesium ions in RNA folding and catalysis. Current opinion in structural biology 22: 262-272.
- Adamala K, Szostak JW (2013) Nonenzymatic template-directed RNA synthesis inside model protocells. Science 342: 1098-1100.
- 59. Service R (2013) The life force. Science (New York, NY) 342: 1032.
- 60. Silverman SK (2009) Deoxyribozymes: Selection Design and Serendipity in the Development of DNA Catalysts[†]. Accounts of chemical research 42: 1521-1531.
- 61. Silverman SK (2008) Catalytic DNA (deoxyribozymes) for synthetic applications current abilities and future prospects. Chemical Communications: 3467-3485.
- 62. Sen D, Geyer CR (1998) DNA enzymes. Current opinion in chemical biology 2: 680-687.
- 63. McManus SA, Li Y (2010) The structural diversity of deoxyribozymes. Molecules 15: 6269-6284.
- 64. Beal PA, Dervan PB (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. Science 251: 1360-1363.
- 65. Zhu L, Li C, Zhu Z, Liu D, Zou Y, et al. (2012) In vitro selection of highly efficient G-quadruplex-based DNAzymes. Analytical chemistry 84: 8383-8390.
- Phan AT, Mergny JL (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson–Crick double helix. Nucleic Acids Research 30: 4618-4625.
- 67. Achenbach J, Chiuman W, Cruz R, Li Y (2004) DNAzymes: from creation in vitro to application in vivo. Current pharmaceutical biotechnology 5: 321-336.
- 68. Breaker RR, Joyce GF (1994) A DNA enzyme that cleaves RNA. Chem Biol 1: 223-229.
- 69. Santoro SW, Joyce GF, Sakthivel K, Gramatikova S, Barbas CF, 3rd (2000) RNA cleavage by a DNA enzyme with extended chemical functionality. J Am Chem Soc 122: 2433-2439.
- 70. Perrin DM, Garestier T, Helene C (2001) Bridging the gap between proteins and nucleic acids: a metal-independent RNAseA mimic with two protein-like functionalities. J Am Chem Soc 123: 1556-1563.
- 71. Sidorov AV, Grasby JA, Williams DM (2004) Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. Nucleic Acids Res 32: 1591-1601.
- 72. Breaker RR, Joyce GF (1995) A DNA enzyme with Mg(2+)-dependent RNA phosphoesterase activity. Chem Biol 2: 655-660.
- 73. Santoro SW, Joyce GF (1997) A general purpose RNA-cleaving DNA enzyme. Proc Natl Acad Sci U S A 94: 4262-4266.
- 74. Geyer CR, Sen D (1997) Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. Chem Biol 4: 579-593.
- 75. Flynn-Charlebois A, Wang Y, Prior TK, Rashid I, Hoadley KA, et al. (2003) Deoxyribozymes with 2'-5' RNA ligase activity. J Am Chem Soc 125: 2444-2454.
- 76. Hoadley KA, Purtha WE, Wolf AC, Flynn-Charlebois A, Silverman SK (2005) Zn2+dependent deoxyribozymes that form natural and unnatural RNA linkages. Biochemistry 44: 9217-9231.
- Purtha WE, Coppins RL, Smalley MK, Silverman SK (2005) General deoxyribozyme-catalyzed synthesis of native 3'-5' RNA linkages. J Am Chem Soc 127: 13124-13125.
- 78. Wang Y, Silverman SK (2003) Deoxyribozymes that synthesize branched and lariat RNA. J Am Chem Soc 125: 6880-6881.

- 79. Coppins RL, Silverman SK (2005) A deoxyribozyme that forms a three-helixjunction complex with its RNA substrates and has general RNA branch-forming activity. J Am Chem Soc 127: 2900-2907.
- 80. Wang W, Billen LP, Li Y (2002) Sequence diversity, metal specificity, and catalytic proficiency of metal-dependent phosphorylating DNA enzymes. Chem Biol 9: 507-517.
- Li Y, Liu Y, Breaker RR (2000) Capping DNA with DNA. Biochemistry 39: 3106-3114.
- Cuenoud B, Szostak JW (1995) A DNA metalloenzyme with DNA ligase activity. Nature 375: 611-614.
- Sreedhara A, Li Y, Breaker RR (2004) Ligating DNA with DNA. J Am Chem Soc 126: 3454-3460.
- Sheppard TL, Ordoukhanian P, Joyce GF (2000) A DNA enzyme with N-glycosylase activity. Proc Natl Acad Sci U S A 97: 7802-7807.
- 85. Chinnapen DJ, Sen D (2004) A deoxyribozyme that harnesses light to repair thymine dimers in DNA. Proc Natl Acad Sci U S A 101: 65-69.
- 86. Burmeister J, von Kiedrowski G, Ellington AD (1997) Cofactor-Assisted Self-Cleavage in DNA Libraries with a 3'–5'-Phosphoramidate Bond. Angewandte Chemie International Edition in English 36: 1321-1324.
- K. Li Y, Sen D (1996) A catalytic DNA for porphyrin metallation. Nat Struct Biol 3: 743-747.
- Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5'terminus of adenovirus 2 late mRNA. Proceedings of the National Academy of Sciences 74: 3171-3175.
- 89. Chow LT, Roberts JM, Lewis JB, Broker TR (1977) A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA: DNA hybrids. Cell 11: 819-836.

- 90. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, et al. (1982) Selfsplicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. cell 31: 147-157.
- 91. Doudna JA, Cech TR (2002) The chemical repertoire of natural ribozymes. Nature 418: 222-228.
- 92. Schlosser K, Li Y (2009) Biologically inspired synthetic enzymes made from DNA. Chem Biol 16: 311-322.
- 93. Silverman SK (2010) DNA as a versatile chemical component for catalysis, encoding, and stereocontrol. Angew Chem Int Ed Engl 49: 7180-7201.
- 94. Liu J, Lu Y (2003) A colorimetric lead biosensor using DNAzyme-directed assembly of gold nanoparticles. J Am Chem Soc 125: 6642-6643.
- 95. Cruz RP, Withers JB, Li Y (2004) Dinucleotide junction cleavage versatility of 8-17 deoxyribozyme. Chem Biol 11: 57-67.
- 96. Santoro SW, Joyce GF (1997) A general purpose RNA-cleaving DNA enzyme. Proceedings of the National Academy of Sciences 94: 4262-4266.
- 97. Santoro SW, Joyce GF (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. Biochemistry 37: 13330-13342.
- 98. Ordoukhanian P, Joyce GF (2002) RNA-cleaving DNA enzymes with altered regio-or enantioselectivity. Journal of the American Chemical Society 124: 12499-12506.
- 99. Tram K, Xia J, Gysbers R, Li Y (2015) An Efficient Catalytic DNA that Cleaves L-RNA.
- 100. Beaudry AA, Joyce GF (1992) Directed evolution of an RNA enzyme. Science 257: 635-641.
- 101. Bartel DP, Szostak JW (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. Science 261: 1411-1418.
- 102. Unrau PJ, Bartel DP (1998) RNA-catalysed nucleotide synthesis. Nature 395: 260-263.
- 103. Rogers J, Joyce GF (1999) A ribozyme that lacks cytidine. Nature 402: 323-325.

- 104. Lorsch JR, Szostak JW (1994) In vitro evolution of new ribozymes with polynucleotide kinase activity. Nature 371: 31-36.
- 105. Wright S (1932) The roles of mutation, inbreeding, crossbreeding, and selection in evolution: na.
- 106. Petrie KL, Joyce GF (2014) Limits of Neutral Drift: Lessons From the In Vitro Evolution of Two Ribozymes. Journal of molecular evolution 79: 75-90.
- 107. Schultes EA, Bartel DP (2000) One sequence, two ribozymes: implications for the emergence of new ribozyme folds. Science 289: 448-452.
- 108. Pitt JN, Ferré-D'Amaré AR (2010) Rapid construction of empirical RNA fitness landscapes. Science 330: 376-379.
- 109. Arenas CD, Lehman N (2010) Quasispecies-like behavior observed in catalytic RNA populations evolving in a test tube. BMC evolutionary biology 10: 80.
- 110. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD (2012) PANDAseq: paired-end assembler for illumina sequences. BMC bioinformatics 13: 31.
- 111. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460-2461.