UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Charting Evolutionary Paths and Alternate Outcomes of the Origin of Life

Permalink

https://escholarship.org/uc/item/19z2k7pk

Author

Kenchel, Joshua Aaron

Publication Date

2022

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Santa Barbara

Charting Evolutionary Paths and Alternate Outcomes of the Origin of Life

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biochemistry and Molecular Biology

by

Joshua Aaron Kenchel

Committee in charge:

Professor Irene A. Chen, University of California Los Angeles, Chair

Professor Cheryl J. Briggs

Professor Enoch Yeung

June 2022

The dissertation of Joshua Aaron Kenchel is approved.

_____

Cheryl J. Briggs

_____

Enoch Yeung

_____

Irene A. Chen, Committee Chair

June 2022

Charting Evolutionary Paths and Alternate Outcomes of the Origin of Life

Copyright © 2022

by

Joshua Aaron Kenchel

DEDICATION

For Kendra Chan, who deserved this more than I ever will.

ACKNOWLEDGEMENTS

"None of us got where we are solely by pulling ourselves up by our bootstraps. We got here because somebody – a parent, a teacher, an Ivy League crony or a few nuns – bent down and helped us pick up our boots."

– Thurgood Marshall

Too many people have helped me with my boots to possibly list them all or enumerate their contributions. A subset of them, in no particular order, are Viviane Schultz, Jonas Schultz, Garnet Kenchel, Lisa Kenchel, Roger Kenchel, Jilli Kenchel, Jordan Janowiak, Leanne Friedrich, Laura Schultz, Joel Rothman, Nico Hartmann, Matt Rosales, Ryan Yamamoto, Kiel Jindra, Andy Rosales, Kevin Rosales, Patti Rosales, Tom Rosales, Julie Harding, Todd Roerig, Lin Chao, Camilla Rang, Audrey Proenca, Susan Golden, Ryan Simkovsky, Irene Chen, Cherie Briggs, Matthieu Louis, Enoch Yeung, Chris Richardson, Terano Mari, Matsuno Makoto, Kumon Ayako, Itō Mari, Zayn Croft, Jenna Malcom, Jeff Schultz, Jaymie Udan, Lee Darnell, Joanna Young, Mika Adamson, Steph Sherman, Broc Sewell, Georgia Pollard, Frank Joyce, Sofía Arce Flores, Kendra Chan, Karen Sinclair, Barbara Orelo, Jeremy Schrader, Annika Nabors, Shohei Burns, Aaron Goodman, Gordon Lau, Chad Oliver, Todd Kontje, Evan Janzen, Yei-Chen Lai, Sam Verbanic, Claire Tran, Celia Blanco, Jen Mobberly, Beto Vázquez-Salazar, Nadine Wells, Krishna Brunton, Ranajay Saha, Kelly Shi, Alanna Stull, Huan Peng, Yuning Shen, Ray Borg, Steven Yang, Vickie Cheng, Mitch Owens, Steven Horton, Ameer Bajwa, David Fitz, Kevin Park, Victor Contreras, Anya Diamond, Don Ripatti, Nate Wieneke, Sam Sánchez, Victoria Walton, Marvin Andrade, James Hsu, Guilhem Dardier, Ethan Ewe, Lisa Fetter, Lulu Velázquez, Hunter McCrea, Sam Wilson, Cesar Orellana, Matt Smith, Lawrence Wei, Hiromi Mori, Tomoko Yamauchi, Joe Ricasa, Franklin Cheng, Dylan Peterson, Joanna Hung, Liang Stotler, Kika Shah, Fujimoto Atsushi, Fujimoto Miyuki, Yamashita Mari, Nishioka Naoki, Ishī Kazuyoshi, Horibe Terumi, Chris Riedl, Akiyo Riedl, Philip Wong, David Tadres, Tatum Katz, Grace DiRenzo, Nicole Chan, Allison Gabbert, Payton DeMarzo, Hannah Gustafson, Aimal Khankhel, Anna Nguyen, Geoff Meyerhoff, Jon Vu, Duncan Proctor, Guillermo Najarro, Zach Aralis, Nerea Muniozguren, Austin Miller, Soham Chowdhury, Caroline Ackley, Katharine Dickson, Adam Christenson, Maeda Kazuyo, Takemichi Mika, Sakamoto Yū, Miguel Leitón, María Leitón Bello, Diego Armando Bello Delgado, *et al.* No person is an island, and science is a collaborative endeavor, so in this dissertation I consistently use the royal “we” to describe the subjects who performed the work. I consider the people here to also be included in this royal “we”.

VITA OF JOSHUA AARON KENCHEL
June 2022

EDUCATION

Bachelor of Science in Biochemistry and Environmental Systems, University of California, San Diego, June 2014 (cum laude)
Master of Science in Biology, University of California, San Diego, June 2015
Doctor of Philosophy in Biochemistry and Molecular Biology, University of California, Santa Barbara, June 2022 (expected)

PROFESSIONAL EMPLOYMENT

2013-15: Teaching Assistant, Division of Biological Sciences, University of California, San Diego
2015-17: Assistant Language Teacher, Miyoshi City Board of Education, Tokushima Prefecture, Japan
2021: Teaching Assistant, Department of Chemical and Biomolecular Engineering, University of California, Los Angeles
2021-22: Teaching Assistant, Department of Molecular, Cellular, and Developmental Biology, University of California, Santa Barbara

PUBLICATIONS

Janzen, E., Shen, Y., Liu, Z., Blanco, C., Kenchel, J., and Chen, I. A. (2022). Error minimization and specificity could emerge in a genetic code as by-products of prebiotic evolution. *Nature Communications*, 2022; *bioRxiv* 2021.05.14.444235
Ewe, C. K., Sommerman, E. M., Kenchel, J., Flowers, S. E., Maduro, M. F., and Rothman, J. H. (2022). Recursive feedforward regulatory logic underlies robustness of the specification-to-differentiation transition and fidelity of the terminal cell fate during C. elegans development. *Development*, 2022; *bioRxiv* 2021.08.24.457588
Janzen, E., Blanco, C., Peng, H., Kenchel, J., and Chen, I. A. (2020). Promiscuous Ribozymes and Their Proposed Role in Prebiotic Evolution. *Chemical Reviews*, 120, 11, 4879-4897
Kenchel, J. (2015). A model for exponential elongation in *Escherichia coli* and evidence of inert polar cellular regions. *UC San Diego*. ProQuest ID: Kenchel_ucsd_0033M_14905. Merritt ID: ark:/20775/bb0274822c.
Kenchel, J. (2014). Microbial Islands: Soil Microbe Diversity in Canopy Soil Patches. *Saltman Quarterly*, 11, 37-42

FIELDS OF STUDY

Major Field: Astrobiology with Professor Irene A. Chen
Studies in Disease Ecology and Microbiology with Professor Cheryl J. Briggs
Studies in Neuroscience with Professor Matthieu Louis
Studies in Developmental Biology with Professor Joel Rothman

ABSTRACT

Charting Evolutionary Paths and Alternate Outcomes of the Origin of Life

by

Joshua Aaron Kenchel

How did life originate on Earth? How might it have happened differently? The answers to these fundamental questions will inform efforts to develop useful synthetic organisms and empower the search for extraterrestrial life. The “RNA World” hypothesis posits that ribonucleic acid (RNA) once played the roles of both information carrier and catalyst in primitive living systems. In the transition from the RNA World to DNA-based heredity and catalysis by left-handed (L-)proteins, a likely essential step would have been the emergence of substrate-specific self-aminoacylating RNA enzymes (ribozymes) functioning in an early translation system. How these early self-replicating systems evolved toward modern cells, and what other outcomes might have been possible, are open questions. Evolution is conceptualized as a biased random walk through genetic space, in which each genotype is associated with a fitness (the “fitness landscape”). Mapping the fitness landscape for a given genetic element enables prediction of evolutionary paths. A recently developed method maps fitness landscapes for self-aminoacylating ribozymes by selection from a random library using prebiotically plausible activated amino acid analogs. In this work, we

map the landscapes for self-aminoacylating ribozymes under variable environmental conditions and assess the effect of the environment on the topography and connectivity of fitness landscapes. We also evaluate the stereoselectivity of self-aminoacylating ribozymes to determine whether life's transition to homochiral L-proteins was deterministic, or whether a right-handed (D-)protein world was possible. We find that a dynamic environment improves the connectivity of fitness landscapes, increasing the number of evolutionary paths available and enhancing evolution's ability to optimize function. We also find that self-aminoacylating ribozymes can be either L- or D-selective, suggesting alternate possible outcomes for the chirality of life. These results represent incremental but meaningful contributions to human understanding of the emergence and evolution of primitive life.

TABLE OF CONTENTS

LIST OF FIGURES

Introduction

How life can emerge from nonliving matter is one of the most important questions in biology. The nature of this phenomenon, called abiogenesis, has implications that span the natural sciences and border on the philosophical. Specifically, knowledge of how the living systems found on Earth came to be, and what other outcomes are possible, is crucial to the fields of astrobiology and synthetic biology.

A widely accepted theory of the origin of life on Earth involves the "RNA World" hypothesis, which posits that ribonucleic acid (RNA) once functioned as both genetic material and catalyst in primitive self-replicating systems[1–4]. This has been supported by a growing body of circumstantial evidence, including the ongoing discovery of the versatility of RNA enzymes (ribozymes)[5–8], and the centrality of RNA to the gene expression system in all organisms[9,10]. Although some other genetic system might have preceded RNA[1,11], the RNA World provides a useful model for studying how early self-replicating systems can emerge and evolve[12]. This knowledge stands to aid efforts in synthetic biology, including the development of minimal cells[13] with applications in e.g. manufacturing, medicine, and environmental remediation.

At some point, the RNA World would have given way to modern cells dominated by protein catalysts. Although RNA enjoys the advantage of template replication by base-pairing, the selective advantage of protein enzymes is clear: with 20 chemically diverse amino acid monomers, compared to the 4 bases available in RNA, proteins can fit a much larger catalytic repertoire into a polymer of the same length. A necessary step in this transition would have been the advent of self-aminoacylating ribozymes: RNAs that

self-modify by 2'/3' esterification with an activated amino acid[14]. If different RNA sequences were specific to different amino acids, these ribozymes could have formed the basis for a genetic code in a primitive translation system. In modern cells, this role is played by transfer RNAs (tRNAs), and anticodon specificity is ensured by aminoacyl-tRNA synthetases[15].

Aminoacyl-tRNA synthesis is driven by an energetically activated aminoacyl-adenosine monophosphate (AMP) intermediate[16]. A number of analogous, prebiotically plausible activated amino acids have been proposed to fill this role in the RNA World[17,18], including 5(4H)-oxazolones. These were found to react with mononucleotides to form 2'/3' esters[19]. The late RNA World may have seen the evolution of self-modifying, substrate-specific catalysts of this reaction.

How could a system as complex as the genetic code evolve? Multiple substrate-specific self-aminoacylating ribozymes might have arisen independently, or one common ancestor might have evolved into substrate-specific variants. The latter case seems more likely, dependent on whether there existed viable evolutionary paths from one function to another. Evolution can be conceptualized as a biased random walk through a genotype space in which each genotype is associated with a fitness. This framework is known as a "fitness landscape"[20]. Starting with a population at a given point in the landscape, mutants explore the adjacent space, and mutants with higher fitness are selected; the population therefore tends to move "uphill" to "peaks" of fitness. Evolutionary paths from one point to another are only possible if they do not pass through any non-functional (zero-fitness) intermediates. The structure of the fitness landscape therefore determines the trajectories available for evolution to produce new or improved functions. For example, the "neutral theory" of evolution posits that the vast majority of mutations have no effect on fitness[21,22]; in

other words, the landscape is mostly very flat. Other theoretical fitness landscapes suggest more "ruggedness", which would place constraints on evolutionary paths[23,24].

Though fitness landscapes have historically been a theoretical framework, there has been recent interest in mapping empirically measured fitness landscapes, fueled by the increasing power and decreasing cost of genetic sequencing[6,25–29]. Knowledge of the fitness landscape of a given genetic element can not only elucidate how it evolved, but also enable prediction of future evolution[30,31]. As empirical fitness landscapes garner more focus, previously unanswerable questions about the nature of fitness landscapes demand attention. One such question concerns the effect of the environment on fitness landscapes. It stands to reason that the landscape should change under different environmental conditions, with formerly nonfunctional regions of the genotype space becoming functional. Thus, a dynamic environment might create viable paths between otherwise isolated fitness peaks, in essence forming a hyperdimensional bridge between them. However, this has yet to be tested empirically for a molecular fitness landscape.

Efforts to map fitness landscapes generally involve generating a large number of mutants at a defined locus, exposing them to selection pressure over multiple generations, and tracking the winners and losers by high throughput sequencing; more successful genotypes correspond to points of higher fitness on the landscape[31]. In the context of RNA evolution, a technique called Systematic Evolution of Ligands by Exponential Enrichment (SELEX), also known as *in vitro* selection, is used[32–34]. This technique couples selection from a randomized RNA library with reverse transcription-polymerase chain reaction (RT-PCR) to amplify selected sequences. Studying RNA fitness landscapes is attractive because of the astronomical size of genotype space: since nucleic acids consist of 4 different bases, a nucleic

acid polymer of length n has an n-dimensional genotype space of size 4^n. For a protein-coding gene, which usually consists of ~10^2-10^3 nucleotides, only a very small region of the space can be explored. However, because even very small RNAs (~15-50 nucleotides) can form functional structures[28,35–40], mapping a comprehensive fitness landscape has become feasible.

Recent work has focused on constructing fitness landscapes for self-aminoacylating ribozymes. The study that formed the conceptual foundation for the work in this dissertation used biotinylated 5(4H)-oxazolones of a tyrosine analog to select for self-aminoacylating ribozymes from an RNA library with a 21-nucleotide random region[41]. By starting with enough RNA to cover the entire genotype space as defined, the authors were able to map a comprehensive fitness landscape for the function. The results sketched a picture of a rugged landscape with three remote peaks of fitness, each corresponding to a functional motif, separated by large valleys of baseline fitness. These valleys would seem to prevent optimization by stepwise mutation and selection, producing a "frustrated" evolutionary network.

These findings raised some immediate questions. First was whether the functional sequence motifs that emerged were specific to tyrosine; this was later found to be the case for two of the three motifs[42]. Second was whether there might be other motifs specific to other amino acids. Such substrate-specific motifs might form the basis for a primitive genetic code, which could be useful in a translation system in a synthetic minimal cell. Furthermore, the presence of accessible paths between substrate-specific ribozymes would suggest evolvability from a common ancestor. A third question concerned the structure of the fitness landscape. The selection was performed under the constant environmental conditions of the

laboratory, at room temperature. This is unlikely to be reflective of prebiotic conditions, which probably experienced fluctuations[43]. How would the landscape change if the system were subjected to dynamic conditions? Would the emergence of new fitness peaks "bridge the gap" between previously isolated peaks, increasing the connectivity of the landscape?

The ribozymes discovered through this work also presented the opportunity to address a major question in the field of astrobiology, that of the chirality of life. All life on Earth is homochiral, consisting almost exclusively of right-handed (D-) sugars and left-handed (L-)amino acids[44,45]. This contrasts with the observation that prebiotic chemistry is chirally symmetric; in the astrobiology field, chiral asymmetry is considered to be one of the hallmarks of life[46]. There exist multiple selection pressures that would drive an initial chiral imbalance to completion; for example, secondary structures such as the α-helix can only be formed by a homochiral chain of amino acids[47]. However, a major open question is why life on Earth went in the direction of L- rather than D-amino acids. It is possible that there was an initial stochastic imbalance that could have just as likely gone the other way. On the other hand, there might have been deterministic factors. Might early RNA catalysts, such as self-aminoacylating ribozymes, have preferred one stereoisomer over the other, creating a chiral imbalance?

The work presented in this dissertation uses techniques developed for the selection of self-aminoacylating ribozymes to answer some of these key open questions about fitness landscapes and the origin and early evolution of life. The dissertation is divided into two chapters. In the first chapter, we select for self-aminoacylating ribozymes specific to valine under both constant and dynamic environmental conditions, resolving fitness landscapes for both. We find that, as hypothesized, a fitness peak that emerges under dynamic conditions

provides enhanced connectivity between tyrosine-specific and likely valine-specific peaks that are otherwise isolated. This constitutes a novel experimental demonstration of environmental conditions affecting the evolvability of a fitness landscape at the molecular level. In the second chapter, we test some previously identified self-aminoacylating ribozymes for stereoselectivity and discover that every ribozyme tested is stereoselective. Furthermore, we find that enantiomeric preference clusters by sequence motif, with two motifs entirely L-selective and one motif entirely D-selective. To our knowledge, this is the first documentation of D-selective self-aminoacylating ribozymes. Taken together, the research presented here represents an incremental expansion of human understanding of the possible evolutionary paths and alternate outcomes of the origin of life.

CHAPTER 1

Evolutionary Paths

1.1. Author Contributions

This chapter will be prepared as a manuscript tentatively titled

"Dynamic environmental conditions enhance adaptability between substrate-specific self-aminoacylating ribozymes"

by

Josh Kenchel[1,2], Alberto Vázquez-Salazar[2], Weiwei Li[3], Evan Janzen[1,4], Krishna Brunton[5,6], Lisa Fetter[1], and Irene A. Chen[2]

The text in this chapter was written by Josh Kenchel. The experiments were designed with contributions from all authors and performed by Josh Kenchel, Alberto Vázquez-Salazar, Weiwei Li, Krishna Brunton, and Lisa Fetter.

1. Program in Biomolecular Science and Engineering, University of California, Santa Barbara
2. Department of Chemical and Biomolecular Engineering, University of California, Los Angeles
3. Bren School of Environmental Management, University of California, Santa Barbara
4. New England Biolabs, Ipswich, Massachusetts, USA
5. Department of Chemistry and Biochemistry, University of California, Santa Barbara
6. Department of Bioengineering, University of California, San Diego

1.2. Abstract

Fitness landscapes are a theoretical concept underpinning evolutionary biology. Populations are visualized as evolving along paths through the landscape by stepwise mutation and selection toward "peaks" of fitness. Elucidating the basic principles governing fitness landscapes will enhance directed and predictive evolution, and aid evaluation of possibilities for life on exoplanets. The RNA World hypothesis is a useful model for studying fitness landscapes using *in vitro* selection. Previously, we resolved fitness landscapes for a likely important role in the RNA World, self-aminoacylating ribozymes, using a tyrosine oxazolone substrate. In this chapter, we expand on previous *in vitro* selection methods to evaluate how the landscape changes under different environmental conditions. Starting with the same RNA library, we selected for self-aminoacylating ribozymes using a valine oxazolone substrate at constant temperature or under dynamic temperature. Comparing these landscapes, we found that otherwise-disconnected tyrosine- or valine-specific peaks are connected by sequences that only become fit under dynamic temperature. This result suggests that dynamic environmental conditions reveal additional evolutionary paths on fitness landscapes, improving the ability of evolution to optimize function.

1.3. Introduction

Evolutionary space can be idealized as a "fitness landscape"[20], in which each genotype is associated with a fitness and populations evolve along the landscape toward "peaks" of fitness. Environmental adaptation being a primary determinant of fitness, the topography of the landscape is in principle dependent on the environment. Thorough understanding of the structure and topography of fitness landscapes and the forces governing them is foundational knowledge. Among other applications, it would aid the design and execution of directed evolution projects in synthetic biology, and enable predictive evolution[30] of e.g. viral pathogens or endangered species. In astrobiology, fitness landscapes of protocatalysts in prebiotic conditions are important to understanding the plausibility of different hypotheses for the origin of life on exoplanets or on the early Earth.

Due to the astronomical size of genotype space, studies of fitness landscapes have long been limited to examinations of a few alleles[25,26,48,49], or otherwise confined to the domain of theoretical evolution[20,50,51]. Recent improvements in high throughput sequencing (HTS) capability have enabled fine-grained, comprehensive elucidation of small slices of genotype space, also called sequence space, at the molecular level[27,28,41,42]. While still many orders of magnitude away from the power needed to resolve genome- or gene-level fitness landscapes, current sequencing methods can cover the entire sequence space of e.g. some small noncoding ribonucleic acids (RNAs)[52–55]. Because even short (~15-50 nucleotides) RNAs can function in catalysis ("ribozymes")[35–37] or specific binding ("aptamers")[28,38–40], comprehensive fitness landscapes for RNAs performing certain functions can be resolved.

Recent work on molecular fitness landscapes has focused on the RNA World hypothesis[1–4], which posits that early life on Earth originated in or passed through a stage in which RNA was both genetic material and catalyst. The transition from the RNA World to cells with DNA and protein, which require a translation system, would likely have relied on self-aminoacylating ribozymes that functioned as self-charging tRNAs[14]. The first comprehensive fitness landscape for a ribozyme was resolved for this function using *in vitro* selection on a random RNA library[41] with a biotinylated tyrosine 5(4*H*)-oxazolone (BYO), a prebiotically plausible high-energy amino acid analog[19]. In this method, catalytic sequences in the RNA pool are amplified over several rounds of selection, the pools from all rounds are sequenced, and fitness is evaluated by abundance or enrichment in later rounds. The fitness landscape of the 21-nucleotide sequence space (about 4×10^{12} possible sequences) consisted of three sequence motifs (peaks) separated by large "valleys" of baseline fitness[41]. In this "frustrated evolutionary network", the global optimum was separated from the broadest peak. Therefore, given a randomly drawn catalyst, optimization along an evolutionary path by stepwise mutation and selection was implied to be highly unlikely.

In the work presented in this chapter, we expand upon the previous work to experimentally test hypotheses about the effect of environmental conditions on fitness landscapes. First, we were motivated to repeat the *in vitro* selection on the same library with a different amino acid analog of the same substrate, biotinyl-Val-5(4*H*)-oxazolone (BVO). We hypothesized that valine-specific variant peaks in different regions of the fitness landscape might be connected to tyrosine-specific peaks, creating paths that would enable the evolution of substrate-specific variants from a single starting point.

A major limitation of the previous work was the use of constant environmental conditions in a laboratory setting, which is likely not reflective of a prebiotic environment[43]. Using temperature as an example of an environmental variable, some RNA sequences may only fold into catalytic structures at lower temperatures, while others may only become appreciably catalytic at higher temperatures. Furthermore, the eutectic phase of an aqueous solution might promote molecular crowding that could enhance catalysis[56]. We therefore hypothesized that a dynamic environment, simulated by a temperature ramp, would enable identification of new peaks of fitness by causing more sequence families to become amplified. We further hypothesized that fitness peaks isolated under a constant environment, as in the BYO landscape, would be better connected under a variable environment, with viable paths passing through peaks only available at different temperatures (Figure 1.1c).

We sought to test these hypotheses in a single *in vitro* selection experiment, modifying the previous method by changing the substrate to BVO and exposing the RNA pool to a temperature ramp during aminoacylation (Figure 1.1a-b). We performed duplicate selections of this test condition, and of a control condition at room temperature. We were therefore able to separately assess changes in the fitness landscape caused by the dynamic environment or by the change in substrate, by comparing the test condition to the control, or by comparing to the original BYO landscape.

1.4. Methods

General synthesis procedures

Synthesis of the biotinyl-Val-5(4*H*)-oxazolone substrate (BVO) followed previously published methods[19,41,42,57] that are described here in brief. The method is also mentioned in Chapter 2 and nearly identical to the schematic in Figure 2.1. We obtained all reagents and solvents from Millipore Sigma or Thermo Fisher and used them without purification, unless otherwise noted. We prepared samples for proton nuclear magnetic resonance (^{1}H NMR) by dissolving dry analyte in deuterated dimethyl sulfoxide (DMSO), recorded spectra using a Varian Unity Inova AS600 (600 MHz), and analyzed them with MestReNova software[58], comparing observed shifts to those predicted using ChemDraw software version 20.1.

Preparation of biotinyl valine (BV)

To biotinylate L-valine methyl ester (VME), we dissolved dry VME to 0.1 M concentration in pyridine with 0.2 M N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide (EDC) and 0.01 M dimethylaminopyridine (DMAP) and incubated under nitrogen with stirring at room temperature overnight. We then evaporated the solvent under reduced pressure and dissolved the residue in dichloromethane (DCM). We washed this solution three times with equal volumes of saturated sodium bisulfate and three times with equal volumes of saturated sodium bicarbonate, and dried it with solid sodium sulfate. We evaporated the

DCM under vacuum and analyzed the resulting biotinyl valine methyl ester (BVME) residue by ^{1}H NMR (Figure 1.2). Dry BVME was stored at -20 °C indefinitely.

To demethylate the carboxyl group of BVME, we dissolved dry BVME in 2:1 isopropanol:water to ~0.05 M concentration with three molar equivalents of NaOH and stirred at room temperature overnight. We then evaporated the solvent under reduced pressure, dissolved the residue in water, and added 1 M HCl dropwise until biotinyl valine (BV) crystallized and precipitated. We separated the precipitate by vacuum filter, washed it three times with water, and dried it in a desiccator with dry phosphorus pentoxide for at least three days. We analyzed the dry BV by ^{1}H NMR (Figure 1.3) and stored it at -80 °C indefinitely.

*Preparation of biotinyl-5(4*H*)-Val oxazolone (BVO)*

We dissolved dry BV to ~0.1 M concentration in DCM with EDC in excess and stirred at 4 °C overnight. We then washed the solution with equal volumes of water (twice), saturated sodium bicarbonate (once), and saturated sodium chloride (once), and dried it with solid sodium sulfate. We evaporated the solvent under vacuum, analyzed the dry BVO by ^{1}H NMR (Figure 1.4), and stored it at -80 °C for up to two years. Stock solution for aminoacylation reactions was prepared by dissolving dry BVO in anhydrous acetonitrile, which was stored at -20 °C for up to six months.

BVO degradation rate determination

We prepared BVO stock solution in anhydrous acetonitrile and mixed it to 15 mM concentration in 1X selection buffer (100 mM HEPES (pH 7), 100 mM NaCl, 100 mM KCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$). The mixture was incubated at 0, 22, or 50 °C. At different time points, we determined [BVO] and [BV] by ultrahigh performance liquid chromatography-mass spectrometry (UHPLC-MS) on an Agilent InfinityLab 1290 Infinity II Series system. Briefly, 2 μL of sample was injected into a mobile phase of 20%:80% A:B, where A was 20 mM ammonium formate in water adjusted to pH 3 with formic acid and B was 90% acetonitrile and 10% A. The mobile phase flowed through the stationary phase (100 mm Agilent HILIC-Z column) at 0.5 mL/min. The MS peaks were integrated and concentrations determined by comparison to a standard curve generated using the same method. We measured experimental triplicates at each time point and fit [BVO] over time at each temperature to an exponential decay model by least-squares regression. The log-adjusted half-lives versus temperature fit an inverse linear relationship, again determined by least-squares regression.

Ribozyme melting curves by circular dichroism (CD) spectroscopy

For each ribozyme, we prepared a solution of 1X selection buffer with 4 μM RNA in a quartz cuvette. Ellipticity from 175 nm to 900 nm was recorded at 5 °C intervals from 5 °C to 95 °C using a JASCO J-1500 spectropolarimeter. We fit the ellipticity at 264 nm to the

Boltzmann sigmoidal equation $\theta = \theta_{min} + (\theta_{max} - \theta_{min})/(1 + e^{(Tm - T)/s})$, where θ is the ellipticity, T is temperature, and T_m is the melting temperature.

In vitro *selection of self-aminoacylating ribozymes*

Wherever possible, the selection scheme (Figure 1.1) adhered to previously published methods[41,42] for determining molecular fitness landscapes of self-aminoacylating ribozymes. The method is briefly summarized here and deviations are noted.

A DNA library of sequence 5′-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGAATTC-N21-TTCACTGCAGACTTGACGAAGCTG-3′, where N21 denotes the 21-nucleotide random region and nucleotides upstream of the transcription start site are underlined, was chemically synthesized (IDT, PAGE purification) and used as the starting (R0) pool.

We performed 15 rounds of selection. For the first round, we transcribed 200 pmol DNA (~30X sequence space coverage) using the New England Biolabs HiScribe T7 polymerase system and purified full-length RNA by urea polyacrylamide gel electrophoresis (PAGE) and ethanol precipitation. 100 pmol RNA (~15X coverage) was mixed with 500 μM BVO in 200 μL of 1X selection buffer (100 mM HEPES (pH 7), 100 mM NaCl, 100 mM KCl, 5 mM $MgCl_2$, 5 mM $CaCl_2$). We stopped the reaction and transferred the RNA into Tris buffer (pH 7.4) by desalting with Bio-Rad P30 size exclusion spin columns. Aminoacylated RNA was selected by pulldown with Promega Streptavidin MagneSphere magnetic beads. Beads were prepared by washing with bead buffer (phosphate-buffered saline, pH 7, with

0.01% Triton-X 1000), 20 mM NaOH, and again with bead buffer. According to manufacturer specifications, enough bead solution was mixed with each reaction mixture to provide at least one streptavidin binding site per RNA molecule. This mixture was incubated at room temperature with rotation for 15 min, then the beads were washed three times with bead buffer, resuspended in ultrapure H_2O, and heated at 70 °C to disrupt the biotin-streptavidin interaction. RNA in the supernatant was reverse-transcribed using SuperScript IV reverse transcriptase (Thermo Fisher), and the cDNA amplified by PCR with Phusion high-fidelity DNA polymerase (Thermo Fisher).

Duplicates of each of two selection conditions were run in parallel. In the variable temperature condition (selections V1 and V2), the RNA/BVO reaction mixture was subjected to a quasi-linear temperature ramp from 60 °C to -20 °C over 8 h. First, the temperature was changed linearly from 60 °C to 0 °C over 6 h in a Bio-Rad C1000 thermal cycler. The tubes containing the reaction mixture were then immediately transferred to a 0 °C insulated container, which was then stored at -20 °C. The reaction mixture reached -20 °C after an additional 2 h. In the constant temperature (control) condition (selections C1 and C2), the reaction mixture was incubated at room temperature for 8 h. In both conditions, three 1.5-μL aliquots of BVO were added to the mixture at consistent time points in order to keep [BVO] approximately constant over the course of the reaction.

Subsequent rounds of selection were performed identically to the first round except for the following modifications. 50 pmol RNA (7.5X coverage) was added to 50 μL of 1X selection buffer. To gradually increase selection stringency, [BVO] was reduced to 250, 50, and 5 μM following rounds 2, 5, and 9 respectively. The reaction time was reduced to 1.5 h

following round 13, with the temperature ramp in the variable temperature condition adjusted accordingly.

High-throughput sequencing of selection pools

DNA pools from all 15 rounds of selection for both duplicates of both conditions, plus the R0 pool (61 total pools) were adapted and barcoded using the Illumina Nextera XT primer set, then multiplexed. Sequencing by synthesis (SBS) and demultiplexing were performed using the Illumina NovaSeq 6000 system (Technology Center for Genomics and Bioinformatics, UCLA).

Bioinformatic analysis of fitness landscapes

We used EasyDIVER[59] to filter for quality and length, isolate the randomized region of the ribozymes, and sum the counts of each unique sequence. We then used ClusterBOSS[42] to identify homologous sequence families, which reflect peaks on the fitness landscape. Briefly, ClusterBOSS treats the highest-abundance unclustered sequence as the center of a peak, clusters all sequences within a provided edit distance into that peak, then repeats until no sequences remain. Evolutionary paths between peaks and other sequences of interest were identified by Peak-Pather[41], a modified A* algorithm implemented in Python. Briefly, this algorithm identifies best paths between two sequences by minimizing a cost function

determined by the sum of a growing path's length and the remaining edit distance to the target sequence. Sequences present in the round 15 (R15) pools with at least 2 counts were included in the set of possible node sequences.

Aminoacylation assays of pools and individual ribozymes

Depending on the experiment, we adjusted RNA and substrate concentrations and reaction volumes as needed. Generally, we added 500 ng of RNA (4 μM) to a 50 μL reaction mixture of 500 μM biotinylated amino acid oxazolone in 1X selection buffer. We incubated this mixture for 1.5 h at room temperature or under the temperature ramp used in rounds 14-15 of the selection, then stopped the reaction by desalting with Bio-Rad P30 size exclusion spin columns. The RNA/aminoacyl-RNA mixtures eluted into Tris buffer, pH 7.4. The fraction of aminoacylated RNA was assessed by electrophoretic mobility shift assay (EMSA) with streptavidin, or by reverse transcription-quantitative polymerase chain reaction (RT-qPCR).

Streptavidin electrophoretic mobility shift assays (EMSA)

We mixed post-reaction RNA with streptavidin (1 mg/mL stock solution, New England Biolabs) at a 1:4 molar ratio of RNA:streptavidin and incubated at room temperature for 5 min. We then mixed these samples with native gel loading dye and separated them by

8% native polyacrylamide gel at 300 V for 15 min. We stained the gels with SYBR Gold (Thermo Fisher) and visualized them using an Amersham Typhoon (General Electric). Aminoacylation was observed as high molecular weight-shifted bands due to binding of streptavidin to the biotinylated amino acid. When appropriate, we quantified the fraction of aminoacylated RNA by comparison of band density to the bands of non-aminoacylated RNA.

Catalytic activity measurement by reverse transcription and quantitative polymerase chain reaction (RT-qPCR)

We mixed RNA with BVO at a range of concentrations as previously described. We separated the aminoacylated RNA using streptavidin-bound magnetic beads and reverse-transcribed it to cDNA, following the procedure used in the selection. We quantified the cDNA by qPCR using the Bio-Rad C1000 system and SYBR Green intercalating dye (Thermo Fisher). We calculated the fraction of aminoacylated RNA as the quantity measured by qPCR divided by the quantity of RNA added to the reaction mixture. We measured experimental triplicates for each concentration of BVO and fit these data to the first-order rate equation $F = A(1 - e^{k[BVO]t})$, where F is the fraction aminoacylated, A is the maximum fraction aminoacylated, k is the effective rate constant, and t is the reaction time[41]. Errors in these measurements were assessed by bootstrapped 95% confidence intervals, and statistical differences in catalytic activity between two pools or sequences were also determined by bootstrapping. The product of k and A has been previously found to be the best metric of catalytic enhancement[41,42,60], so we generally used kA to assess catalytic activity.

1.5. Results and Discussion

Temperature ramp protocol design

We sought to expose the entire RNA library to a linear temperature ramp, such that all sequences would be exposed to each temperature for the same amount of time. In theory, any sequence with catalytic aminoacylation activity active at any temperature in the range would react and be selected, so the resulting fitness landscape would be the union of all fitness landscapes from all temperatures in the range. For simplicity, we decided to start at the high temperature and ramp down one time to the low temperature, with the reasoning that RNA secondary and tertiary structures would start mostly unfolded and different sequences would slowly fold as they crossed their melting thresholds. For the endpoints, we sought the highest reasonably achievable temperature at which RNA would not spontaneously degrade[61,62], and an achievable low temperature below the eutectic point of the reaction mixture. In our experience, RNA degradation in aqueous solution with divalent salts becomes appreciable on the timescale of minutes above 70 °C, so we chose 60 °C as the high endpoint. We measured the freezing point of the solution at about -18 °C, so we chose -20 °C as the low endpoint. Given that previous selections had used 90 min reactions at room temperature, and taking human and equipment availability into account, we chose 8 h as the reaction time, for an easily memorable ramp rate of -10 °C/h.

Commercially available dry baths and thermal cyclers cool samples to 0 or -10 °C at minimum, and circulating wet baths or liquid nitrogen-based temperature control equipment solutions were cost-prohibitive. Balancing cost and efficacy, we decided that exponential

cooling from 0 °C to -20 °C would approximate a linear ramp and allow the reaction mixture to slowly cross its eutectic point. Therefore, after linear decrease from 60 °C to 0 °C in a thermal cycler, we transferred the samples to an insulated container equilibrated at 0 °C and stored the container at -20 °C. We measured the coefficient of heat transfer of the container to be such that the half-life of the temperature of the samples by Newton's law of cooling[63] was 29 min (Figure 1.5), so they descended below -19 °C in 2 h.

We tested the stability of the RNA over the planned temperature ramp, and verified that aminoacylation activity would be possible, by subjecting a model ribozyme, the known generalist S-2.1-a[41,42], to the above protocol in 1X selection buffer with BVO. We observed aminonacylated RNA, but no evidence of spontaneous RNA degradation, via an electrophoretic mobility shift assay (EMSA) with streptavidin (Figure 1.9).

Maintaining [BVO] within a consistent range

A major concern in designing the temperature ramp (variable, or V selection condition) was keeping [BVO] relatively constant over the course of the reaction, such that the average [BVO] over a given ~10 °C interval would be about the same as over any other similar interval. This ensured that temperature was the only environmental variable experienced by the RNA pool, and the resulting fitness landscape represented the union of the landscapes from every temperature. Additionally, the protocol was designed such that the average [BVO] in the constant temperature (control, or C selection condition) reaction was the same as the average [BVO] in the V reaction.

BVO is depleted in aqueous solution by a side reaction in which the oxazolone hydrolyzes[19], producing the precursor biotinyl valine (BV). By tracking the fraction of BVO remaining in 1X selection buffer over time by ultrahigh performance liquid chromatography-mass spectrometry (UHPLC-MS), we determined the half-life of BVO at 0, 22, and 50 °C to be 20, 109, and 471 min respectively (Figure 1.6), with respective bootstrapped 95% confidence intervals (19, 21), (107, 110), and (435, 520). These data followed an expected log-linear relationship between reaction rates and temperature (Figure 1.7). The half-life of BVO at a given temperature can therefore be closely estimated by the equation $t_{1/2} = ae^{bT}$ in which $t_{1/2}$ is the half-life, T is temperature, a is 460±22 min, and b is -0.063±0.001 $°C^{-1}$. Using this range of possible values for the relationship between temperature and BVO degradation, we conducted bootstrap simulations of the likely range of [BVO] over the course of the 8 h temperature ramp (Figure 1.8). We selected time points for the addition of supplemental BVO to keep [BVO] within a factor of 4 of the starting concentration. Under this protocol, the average [BVO] experienced by the RNA pool over any 10 °C interval was no less than 36% and no higher than 61% of the starting concentration. Bootstrapped 99% confidence intervals of the average [BVO] over the course of the entire reaction were (49%, 56%) and (40%, 51%) for C and V respectively.

Following this procedure, the C reaction mixture was given supplementary 1.5-µL aliquots of BVO equal to 10, 10, and 50% of the starting concentration at 0.5, 1.5, and 2.5 h respectively (Figure 1.8). The V reaction mixture was given BVO supplements equal to 75% of the starting concentration, at the same time points. Due to the use of anhydrous acetonitrile to prepare the BVO stock solution, this protocol resulted in a final acetonitrile concentration of 12% in the reaction mixture. To check whether that level of acetonitrile in solution might

inhibit aminoacylation activity, we tested a previously identified model ribozyme with biotinyl tyrosine oxazolone (BYO) in 1X selection buffer with 12% acetonitrile and assessed aminoacylation activity by electrophoretic mobility shift assay (EMSA). We observed no reduction in aminoacylation (Figure 1.10).

Evidence for temperature-dependent aminoacylation activity

We reasoned that essential secondary structures in some or most of the ribozymes in the library might melt at temperatures nearing 60 °C. We recorded melting curves for four previously characterized self-aminoacylating ribozymes by circular dichroism (CD) and found the melting temperatures of all four ribozymes to be between 60 and 70 °C (Figure 1.11). We then reasoned that although the secondary structure might be stable below 60 °C, there may be some tertiary structures that melt at lower temperatures. We tested one model ribozyme, S-1A.1-a, with BYO at 22, 40, and 50 °C and assayed aminoacylation activity by EMSA to assess whether aminoacylation was disrupted at higher temperatures. We observed a modest reduction of aminoacylation activity at 40 °C and complete ablation at 50 °C (Figure 1.12), suggesting disruption of some essential interaction, likely a tertiary structure, just above 40 °C. Whatever the cause, this finding lent additional confidence to our hypothesis that aminoacylation activity was temperature-dependent within the range of the proposed temperature ramp, and the specific temperature dependence was likely determined by sequence.

In vitro *selection progress*

Each round, we tracked the approximate fraction of aminoacylated RNA based on back-calculated cDNA recovery (Figure 1.13). We noticed that the variable temperature (V) condition consistently produced two to four times as much aminoacylated RNA as the control (C) condition. The aminoacylated fraction of the pool appeared to increase rapidly each round, so we increased the selection stringency by decreasing the substrate concentration. We also tracked the activity of the pool across rounds by electrophoretic mobility shift assay (EMSA) with streptavidin, which indicated that RNA in the V condition was aminoacylated at about three times the rate of RNA in the C condition (Figure 1.14). Based on previous selections for self-aminoacylating ribozymes starting with the same library[41,42], we expected to detect aminoacylation over the background rate in the round 0 (R0) pool by EMSA within 6 rounds. However, we detected no aminoacylation over the R0 pool by R7 (Figure 1.14). It also became clear that the back-calculated cDNA recovery was correlating poorly with activity observed by EMSA. Sequencing data from R7 indicated that the pools were still very diverse, with some sequences countable over background in the C condition, including at least one previously identified promiscuous ribozyme, S-2.1-a. V-selected R7 pools were statistically indistinguishable from R0, consistent with the hypothesis that selection stringency was lower in the V condition due to more sequences becoming aminoacylated. We continued the selection until we observed enhanced aminoacylation over background in the C condition by EMSA, which was after R15 (Figure 1.15).

We measured the activity of the R15 pools under C and V conditions by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) and compared them to the baseline R0 activity (Figure 1.16). Consistent with our observation of higher aminoacylation rate in the V condition measured by back-calculated cDNA or by EMSA, the adjusted background catalytic rate kA under the V condition was about two times higher than under the C condition. We measured significant increases in catalytic activity in both C-selected and V-selected pools under both reaction conditions. Furthermore, the catalytic enhancement of the C-selected R15 pool was significantly higher under the C condition than under V, and the reverse was true for the V-selected R15 pool. This suggested that each pool was specialized to its selection condition and likely consisted of a unique population of sequence families.

Sequencing data from all rounds revealed that no sequence family comprised more than 1% of the pool before round 13 in the C condition, and before round 14 in the V condition (Figure 1.17), again consistent with the observed higher diversity in the V-selected pools due to lower selection stringency. In the C-selected pools, the most abundant sequence was S-2.1-a, a ribozyme originally identified by selection with BYO[41] but later found to be a generalist[42]. S-2.1-a comprised more than 5% of the R15 pool, with the top three sequences comprising nearly 15%. By comparison, no unique sequence comprised more than 0.02% of the V-selected R15 pools. The round-over-round enrichment of top families in each condition also aligned with these observations, averaging 400% for the top two C-selected families and 50% for the top three V-selected families.

Fitness landscape peaks and paths

Based on the experimental and sequencing data, we were satisfied that the constant temperature (C)-selected R15 pools were not overselected. The variable temperature (V)-selected pools, while perhaps under-selected due to the comparatively low selection stringency, had enough identifiable sequence families amplified above background to explore the fitness landscape-related questions that we had initially set out to answer. The C-selected landscape consisted of 11 families, 10 of which were previously unidentified; the remaining family was the generalist Motif 2 identified by selection with BYO and containing the aforementioned S-2.1-a[41]. The V-selected landscape consisted of 25 families, none of which were dominated by previously characterized sequences. Each set of duplicates had greater than three-quarters of sequence families in common, while only one-quarter of sequence families were selected by both temperature conditions. Both the increased diversity in the V condition and the appearance of C-selected families in the V-selected pools aligned with our expectation that room temperature-optimized ribozymes would be selected in the V condition in addition to ribozymes active at other temperatures.

Of critical importance to our initial hypotheses was the presence or absence of evolutionary paths between fitness peaks under each condition. Due to the immense complexity of the highly multidimensional sequence space and the associated computational intensity, we focused our search on paths between the center sequences of peaks (that is, the most abundant sequence of each family). To evaluate paths between substrate-specific peaks, we used fitness landscape data from the selection performed on the same library with BYO, the tyrosine analog of BVO. To our surprise, one of the V-specific peaks supported both of

our initial hypotheses. Two of the top C-selected sequences, referred to here as Val-sC1 and Val-sC2, differ by 12 edit distance and are separated by a gap of 7 edit distance along the best possible path if only C-selected sequences are available (Figure 1.17). When allowed to travel through V-selected sequences, however, the best path travels through a V-specific peak with center sequence named Val-sV2, and has a maximum gap of 3 edit distance. Val-sC1 and Val-sC2 are separated from a BYO-selected ribozyme named S-3.1-a, which is substrate-specific for aromatic amino acids[42], by 12 and 13 edit distance, respectively. The best paths between them must cross gaps of 5 or 8 edit distance, respectively. When allowed to travel through V-selected sequences, the best path passes through the same Val-sV2 peak, again reducing the maximum gap to 3 edit distance.

Consistent with previously published methods[41], only sequences with at least 2 counts in R15 were included as possible nodes for paths. About 5×10^5 C-selected sequences and 1.7×10^6 V-selected sequences met this criterion. We considered that any addition of the same number of possible nodes would also create better paths between peaks, so as a control we searched for paths between C-selected sequences and 1.7×10^6 randomly generated 21-nucleotide sequences. With the exception of the path between Val-sC2 and S-3.1-a, for which the minimum gap decreased from 8 to 7 edit distance, no better paths were found (Figure 1.18). We also considered that the observed patterns might be dependent on the chosen counts threshold, which is ultimately arbitrary. When considering only sequences with at least 5 counts as possible nodes, the best paths between Val-sC1, Val-sC2, and S-3.1-a still pass through the family of Val-sV2, with maximum steps of 4-5 edit distance; without V-selected sequences, the maximum steps increase to 6-8 edit distance (Figure 1.19).

Validation of newly identified candidate ribozymes

We tested a subset of peak sequences for aminoacylation activity in order to check our assumption that the selected sequence families were real catalysts. We observed detectable aminoacylation in the constant temperature (C)-selected sequences Val-sC1 and Val-sC2 by EMSA, with the fraction aminoacylated greater in C reaction conditions (Figure 1.20). However, the V-selected sequences Val-sV1 and Val-sV2 had no detectable aminoacylation above background by EMSA under either reaction condition. We are awaiting results of qPCR measurement of catalytic enhancement over background for these ribozymes; the EMSA technique is not sensitive enough to detect small but significant increases in catalytic rate. As with the V-selected R15 pools, it is possible that Val-sV1 and Val-sV2 are catalysts with enhanced rates detectable by qPCR but not by EMSA.

Another remaining question is whether the newly identified sequence families are substrate-specific, either exclusively for valine or for the class of small nonpolar amino acids. In the case of Val-sC1 and Val-sC2, this question is especially interesting because of the evolutionary paths between those sequences and the aromatic-specific S-3.1-a[41,42]. Given that these sequence families were not amplified in previous selections using the tyrosine and phenylalanine substrate analogs, it is unlikely that they self-aminoacylate with those substrates at the same rate as with valine. This would then be the first study to our knowledge to identify a plausible evolutionary path between substrate-specific ribozyme variants by using *in vitro* selection to resolve a comprehensive fitness landscape.

Implications for early molecular evolution

This study was motivated by questions about the topography of molecular fitness landscapes, inspired by previous studies by our group using the same RNA library and experimental technique for resolving fitness landscapes by *in vitro* selection[41,42]. Those found that there were perhaps few solutions in the sequence space as defined, and any evolutionary paths between those solutions were scarce and improbable, producing a frustrated evolutionary network. The evolution and optimization of essential functions in the very early stages of life were therefore implied to be very difficult, rare events. However, those previous studies utilized constant selection conditions, which do not reflect environmental changes in early Earth conditions over evolutionary time. We hypothesized that varying the selection conditions, by selecting with a novel substrate and by changing the reaction temperature, would select for new solutions in the same sequence space which might provide better paths between otherwise-isolated peaks of fitness.

Though we are still awaiting experimental confirmation of catalysis by sequences selected under variable temperature and the substrate specificity of the ribozymes newly identified in this work, the landscapes resolved by high-throughput sequencing of the selection pools appear to support our hypothesis. There are several lines of evidence indicating that selection under variable temperature conditions increased the aminoacylated fraction of the RNA pool, reducing the selection stringency and increasing the number of sequence families enriched. The variable temperature-selected families provide paths between peaks otherwise unavailable at constant temperature alone. Furthermore, they also provide paths between likely substrate-specific sequence motifs. Based on these findings, it

appears that environmental variability can significantly improve the ability of stepwise mutation and selection to find and optimize solutions to essential functions.

Of course, the limitations of our experimental system, dictated by current technological limitations, necessitate a number of caveats about the strength and significance of the conclusions that can be drawn from this study. Among these limitations are the size of the sequence space, the necessity of constant flanking regions on the RNA, and the coverage by high-throughput sequencing. We readily acknowledge that the scope of this experiment is limited to one highly constrained slice of the nearly infinite domain of possible genetic space; e.g., there could be unapparent solutions in 20- or 22-nucleotide space that also provide paths between peaks isolated in 21-nucleotide space. Furthermore, we only explore evolution by stepwise mutation and selection, and disregard the possible effects of gene duplication or recombination. With those considerations in mind, we expect that it is possible to draw appropriately-scaled conclusions about the topography of fitness landscapes from our findings.

The observation that dynamic environmental conditions increase the available space in the fitness landscape echoes an ecological theory known as the intermediate disturbance hypothesis[64], which posits that biodiversity is highest when moderate levels of environmental disturbance prevent communities from reaching equilibrium. It might be that in this chapter we have resolved fitness landscapes at low and intermediate levels of disturbance. It would be interesting to observe the landscape at a high level of disturbance, for example by selecting at low temperature for one round, then at high temperature the next, and so on. How many sequence families would survive in both conditions, and how connected would the landscape be?

Reflections on selection strategy

As noted, based on previous selections with the same library, we expected to observe enhanced aminoacylation activity over background between rounds 5 and 10. However, those previous selections used a reaction time of 90 min and 50 μM substrate concentration. Stemming from a desire to avoid losing catalytic sequences to genetic drift, the selection scheme in this work started with a >5X greater reaction time and 10X greater substrate concentration. We increased the selection stringency as the selection progressed, but in hindsight, the need for more rounds of selection than expected was likely due to keeping the selection stringency lower than necessary for longer than necessary. *In vitro* selection design often involves much trial and error[65,66]. Excessively high selection stringency may eliminate functional sequences or their relatives by genetic drift (the "bottleneck" effect[67]), while excessively low stringency will fail to amplify functional sequences on a reasonable time scale. Most *in vitro* selection experiments are discovery-oriented, meaning that their goals are achieved if even one sufficiently functional sequence is identified. Keeping in mind the goal of resolving a comprehensive fitness landscape, we tried to avoid losing catalytic sequences, and erred on the side of lower stringency.

The major technological limitation of our method of resolving fitness landscapes by *in vitro* selection is still sequencing throughput. Currently, the most powerful method outputs on the order of 10^{10} reads[53], requiring that sequences from our sequence space of size 4×10^{12} be amplified by at about four orders of magnitude in order to be detectable. Though we expect that there are ~10^6 catalytic solutions in this sequence space, it is possible that some lower-activity ribozymes fail to be included in this "snapshot" of the landscape. If the ceiling

of sequencing throughput could be raised by about three orders of magnitude, this would cease to be a concern. A related issue is that even in later rounds, between 10% and 50% of sequences appearing with 1 count may not be true catalysts. Introducing some method, e.g. a targeted digestion-based method, to remove non-aminoacylated sequences as part of the selection scheme would likely free up much sequencing depth to detect true catalysts.

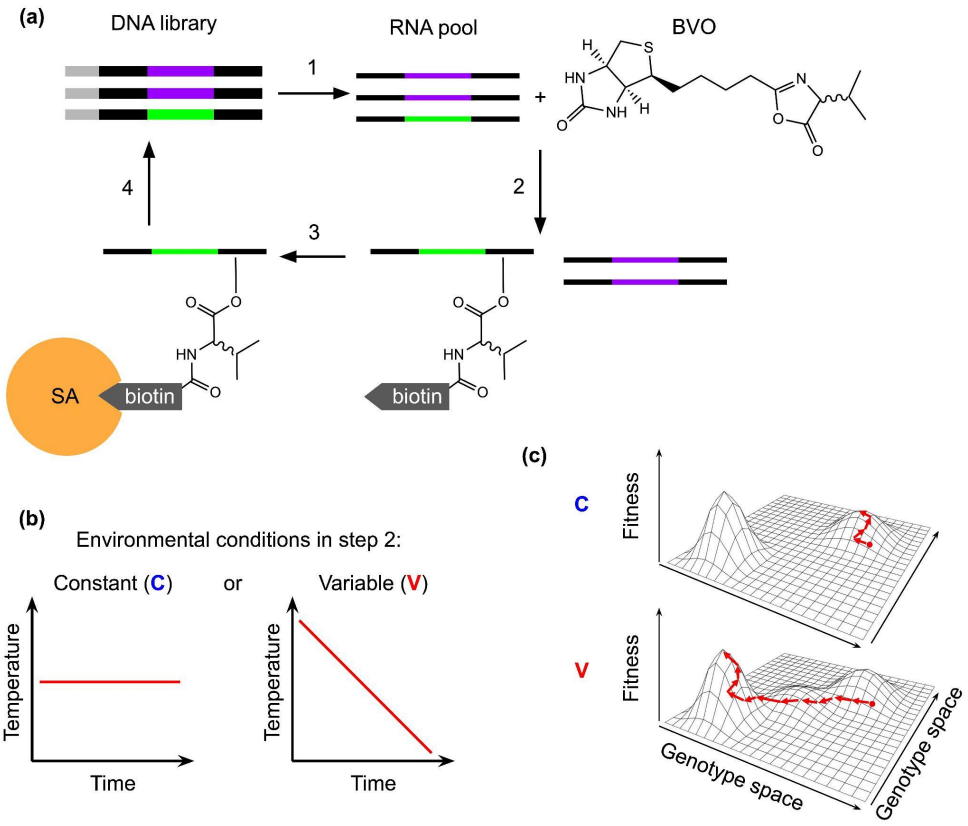

Figure 1.1. Method of mapping fitness landscapes by *in vitro* selection and hypothesized effect of environment on the landscape. a) *In vitro* selection method. A random DNA library is transcribed (1) to RNA, which is incubated with biotinyl-Val-5(4H)-oxazolone (BVO) and catalytic sequences (green) self-aminoacylate (2); streptavidin-coated magnetic beads separate the biotin-tagged aminoacyl-RNAs (3) from non-catalytic sequences (purple), allowing them to be selectively reverse-transcribed (4) and PCR-amplified. b) The RNA pool was exposed to either constant temperature (as a control) or a linear temperature ramp during the aminoacylation step. c) Because all sequences in the pool have equal opportunity to react with BVO at any temperature in the ramp, we expected to find more peaks of fitness in the landscape at variable temperature, connecting otherwise-isolated peaks.

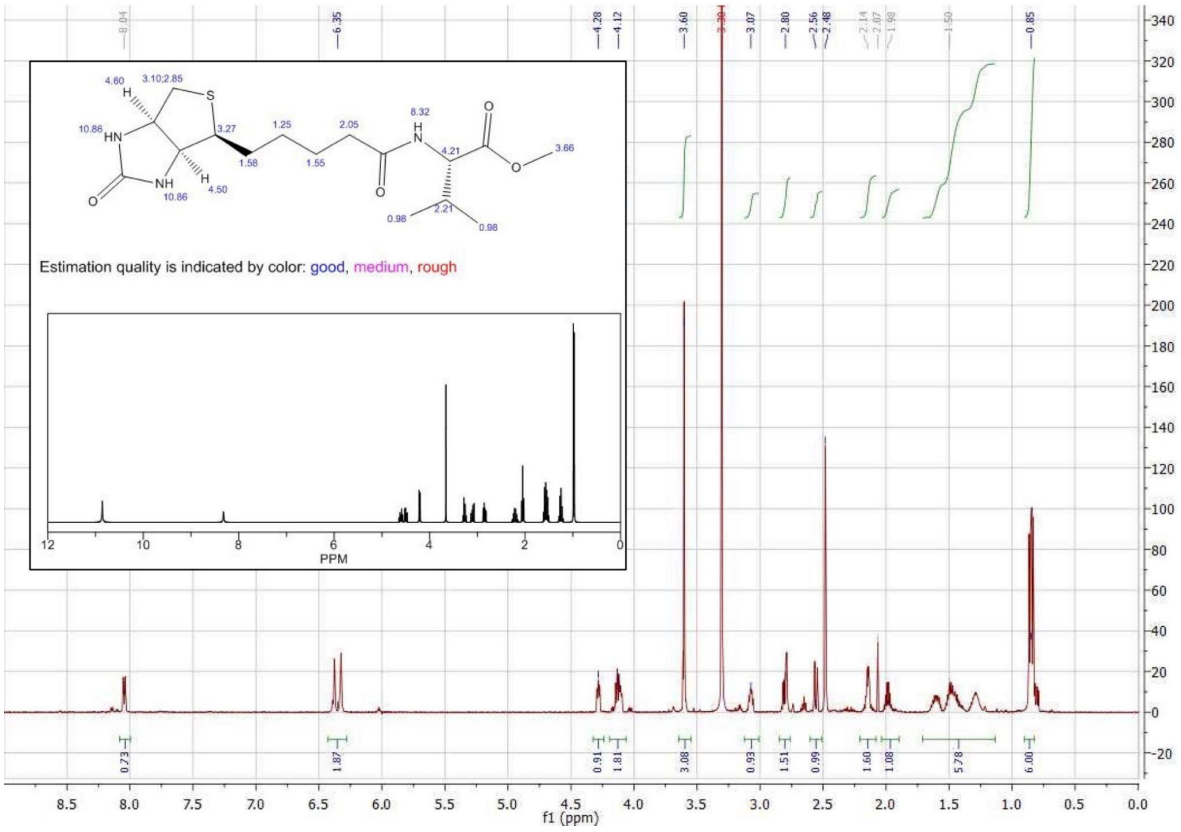

Figure 1.2. Proton nuclear magnetic resonance (^{1}H NMR) spectrum of biotinyl valine methyl ester (BVME) in deuterated dimethyl sulfoxide (DMSO). Peak centers of chemical shifts are labeled at the top. Peak integrations are labeled at the bottom. Inset: predicted shifts generated using ChemDraw software.

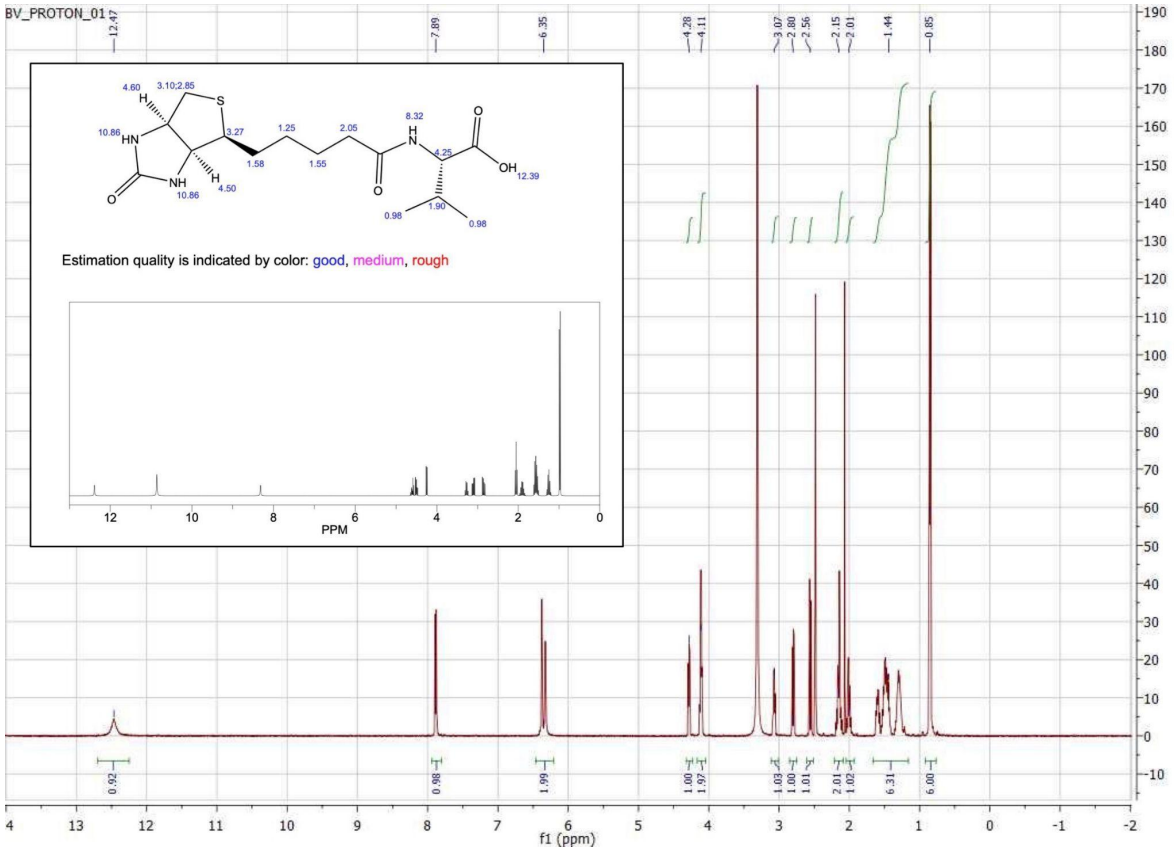

Figure 1.3. ^{1}H NMR spectrum of biotinyl valine (BV) in deuterated DMSO. Peak centers of chemical shifts are labeled at the top. Peak integrations are labeled at the bottom. Inset: predicted shifts generated by ChemDraw.

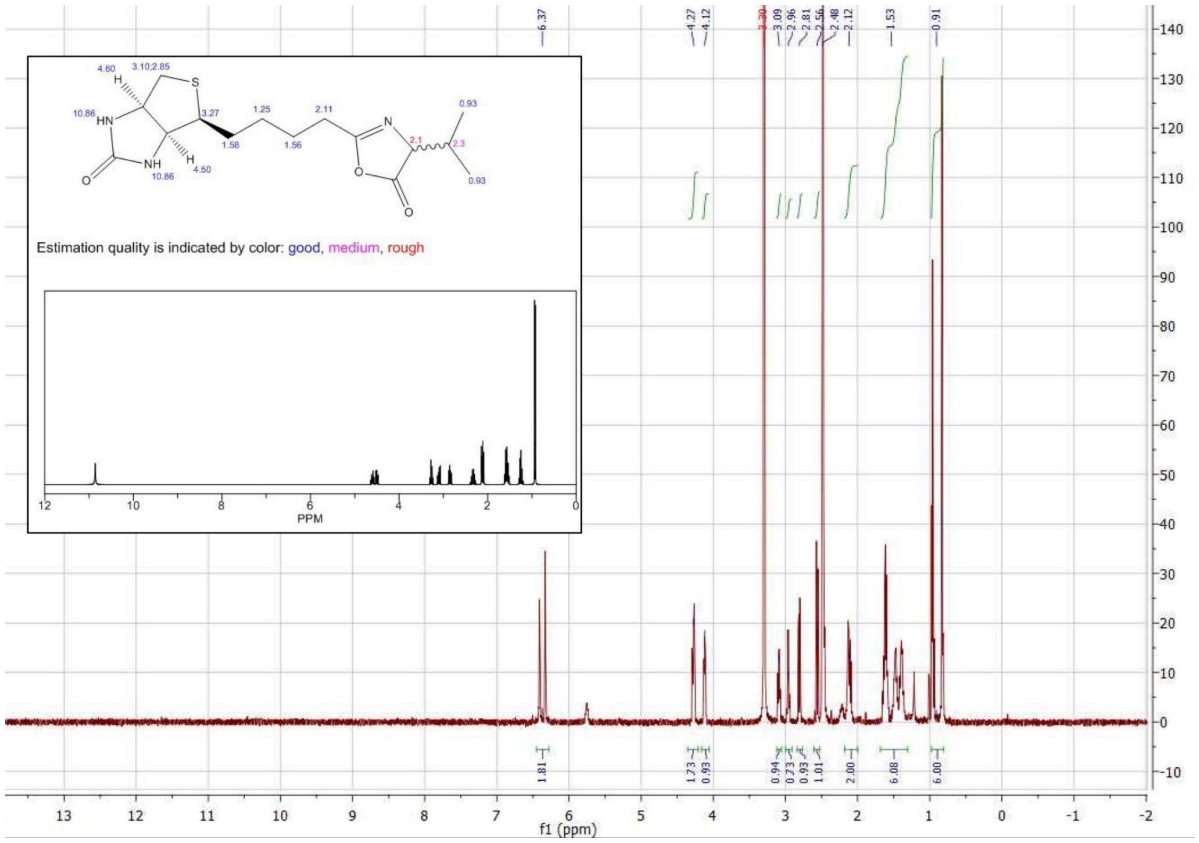

Figure 1.4. ^{1}H NMR spectrum of biotinyl-Val-5(4*H*)-oxazolone (BVO) in deuterated DMSO. Peak centers of chemical shifts are labeled at the top. Peak integrations are labeled at the bottom. Inset: predicted shifts generated by ChemDraw.

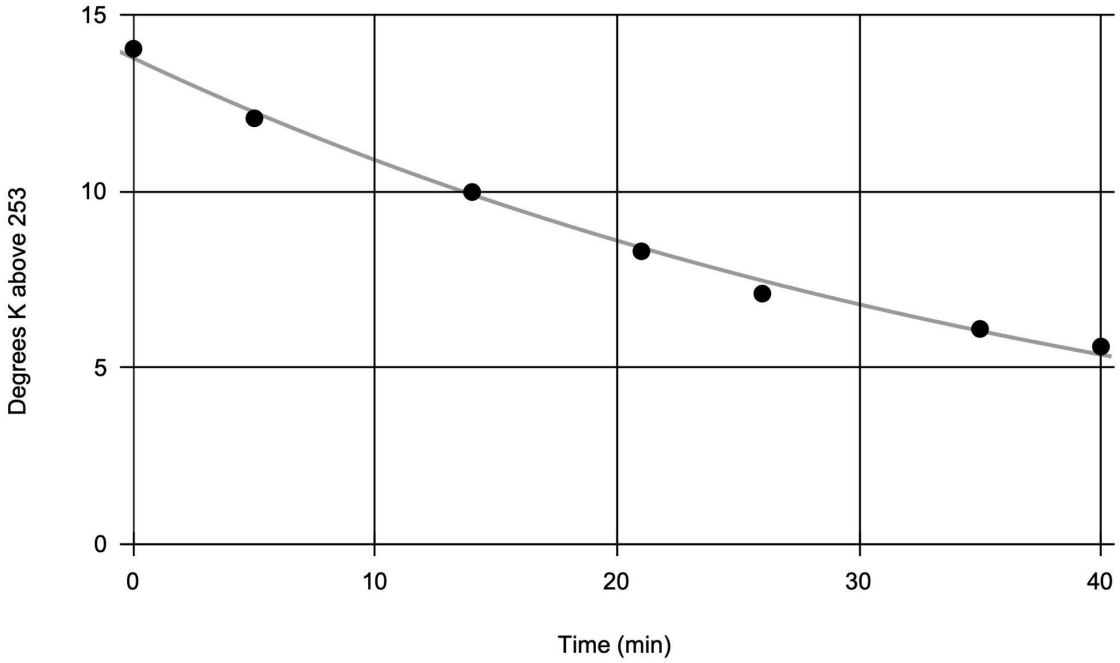

Figure 1.5. Calculation of thermal conductivity constant for the insulated container used in the variable temperature selection condition. We measured the temperature in the container using a digital temperature probe after placing the container at -20 °C. Points display the recorded temperature at time points thereafter. The gray curve displays the best fit by least squares regression ($R^2 = 0.995$) to the model $T = T_0e^{rt}$, where T is temperature, T_0 is initial temperature, t is time, and r is the rate of exponential cooling, which here is -0.0235 and corresponds to a half-life of 29 min.

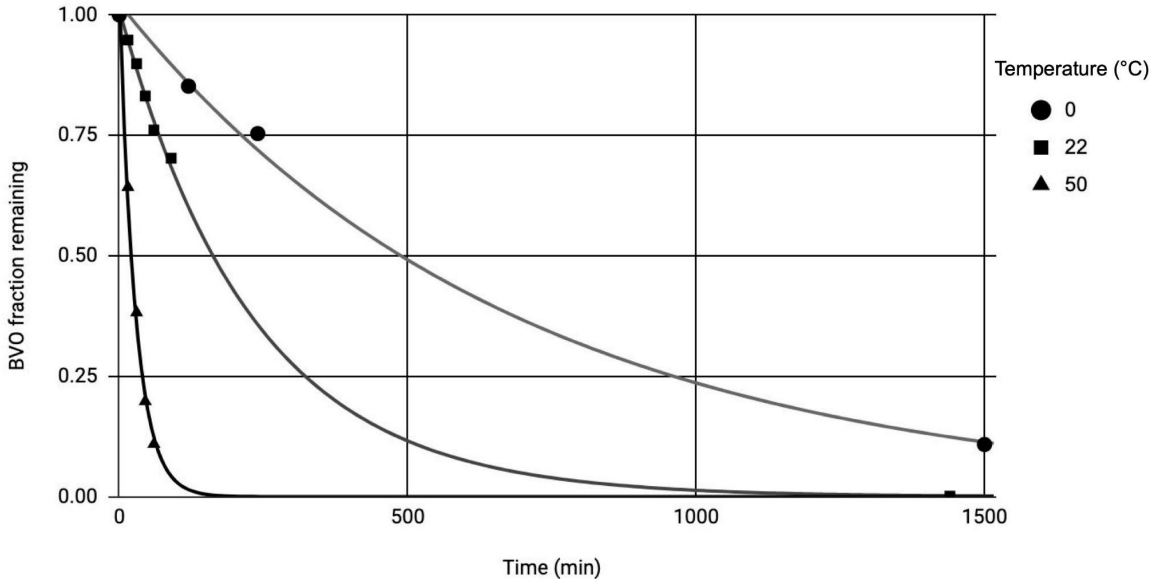

Figure 1.6. Temperature-dependent rate of substrate hydrolysis. We mixed 15 mM biotinyl valine oxazolone (BVO) in 1X selection buffer, incubated at each temperature without agitation, and measured the fraction of BVO remaining at different time points by UHPLC-MS. The average of triplicates at each time point is shown. Curves indicate the best-fit to an exponential decay model by least-squares regression ($R^2 > 0.99$ for all).

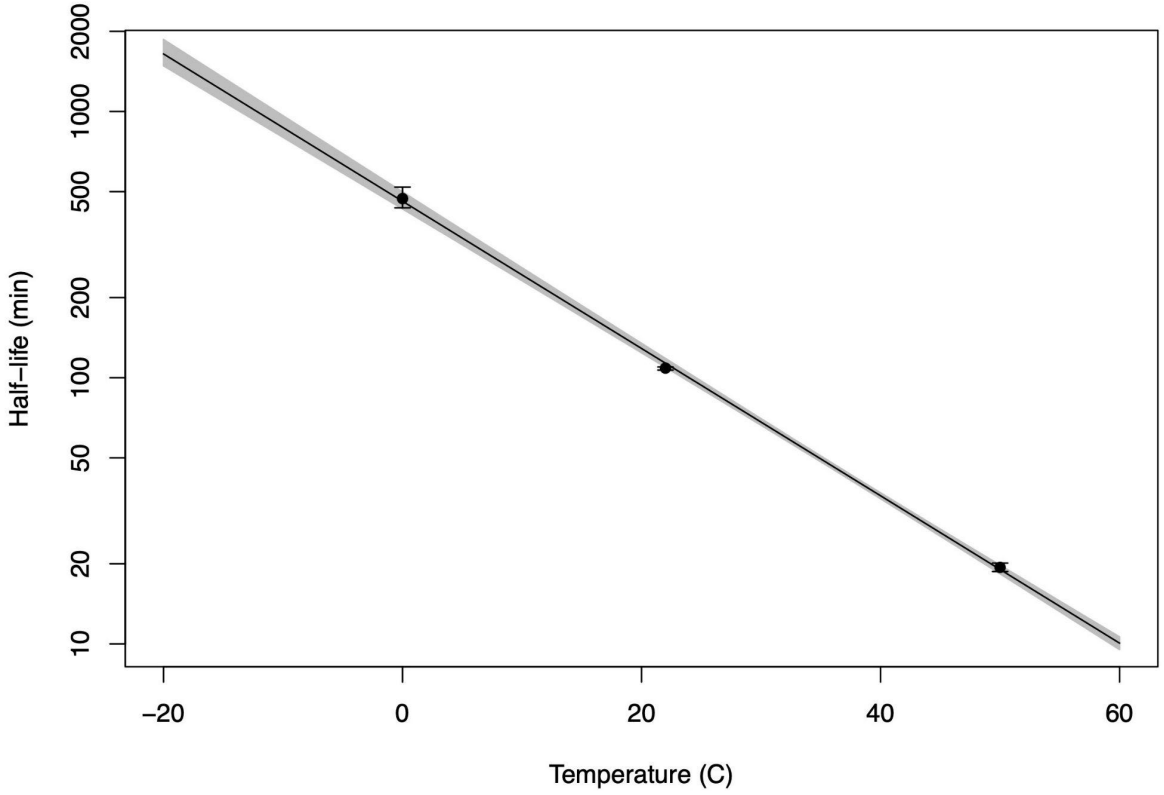

Figure 1.7. Log-linear relationship between substrate hydrolysis rate and temperature used to simulate [BVO] over the course of the variable temperature reaction. We used replicate measurements of BVO hydrolysis by UHPLC-MS (Fig. 1.5) to calculate median (points) and 95% confidence intervals (error bars) of half-life at each temperature by bootstrapping. These data conform to an expected log-linear relationship between temperature and first-order reaction rates. The solid line indicates the best-fit by least squares regression ($R^2 > 0.99$) and the ribbon indicates a bootstrapped 95% confidence interval of the relationship.

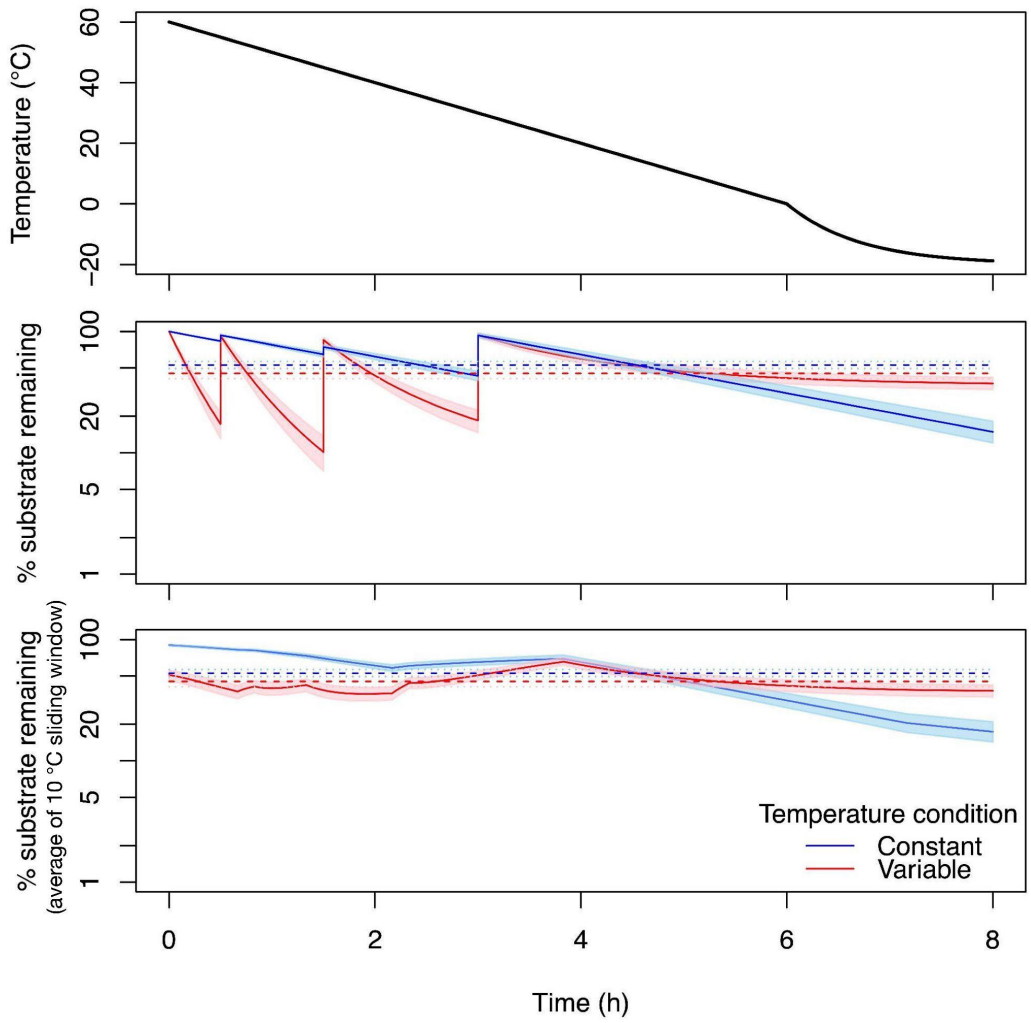

Figure 1.8. Biotinyl valine oxazolone (BVO) concentration over the course of the 8 h reaction in the protocol used for the selection in the constant (C) and variable (V) temperature conditions, simulated using the determined log-linear relationship between half-life and temperature (Figure 1.7). Top: pseudo-linear temperature ramp experienced by the RNA pool in the V condition. Middle: [BVO] over the 8 h reaction. Ribbons indicate bootstrapped 99% confidence intervals. Dotted lines indicate average concentration for the entire 8 h reaction and 99% CI. Bottom: same as middle, but showing average [BVO] experienced by the pool during a 10 °C (1 h) sliding window.

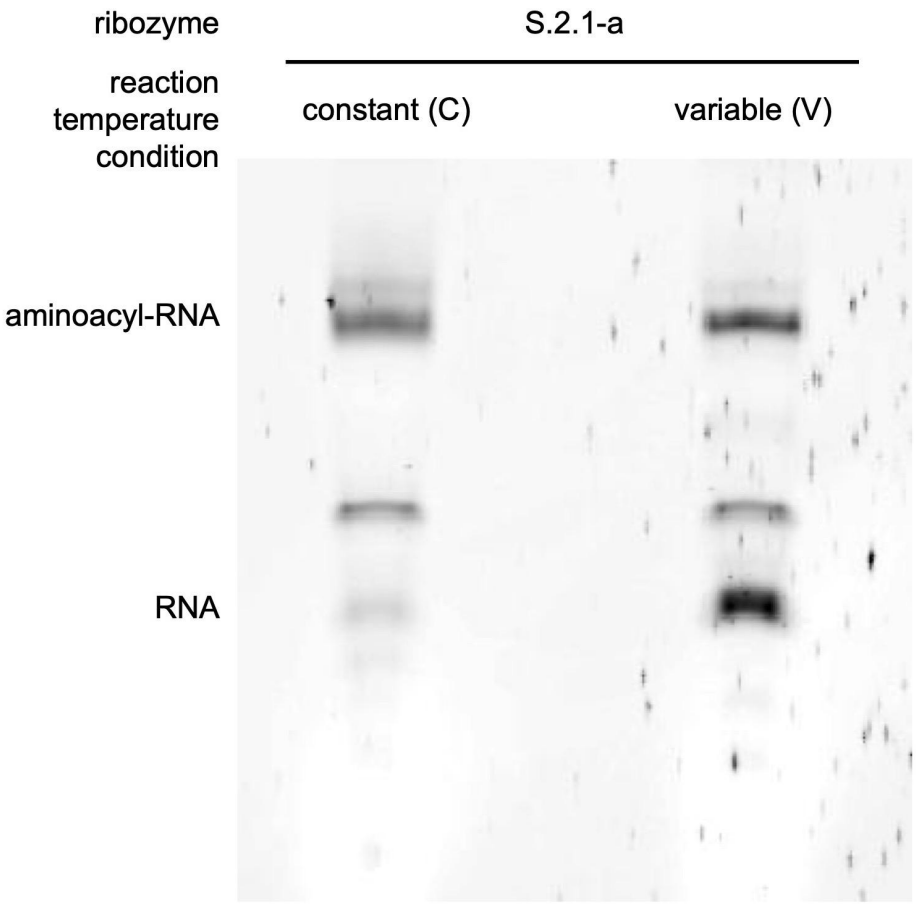

Figure 1.9. Streptavidin electrophoretic mobility shift assay for self-aminoacylation activity under the reaction conditions used in the selection. S-2.1-a is a previously documented[41] promiscuous[42] self-aminoacylating ribozyme known to react with BVO at room temperature. We mixed S-2.1-a with 500 μM BVO at constant (C) room temperature, or variable (V) temperature, a quasi-linear ramp from 60 °C to -20 °C, over 8 h. After desalting, we mixed the RNA with streptavidin and size-separated it by native PAGE. Aminoacylated RNA appears as a higher molecular weight band. Degraded RNA would appear as a lower molecular weight smear.

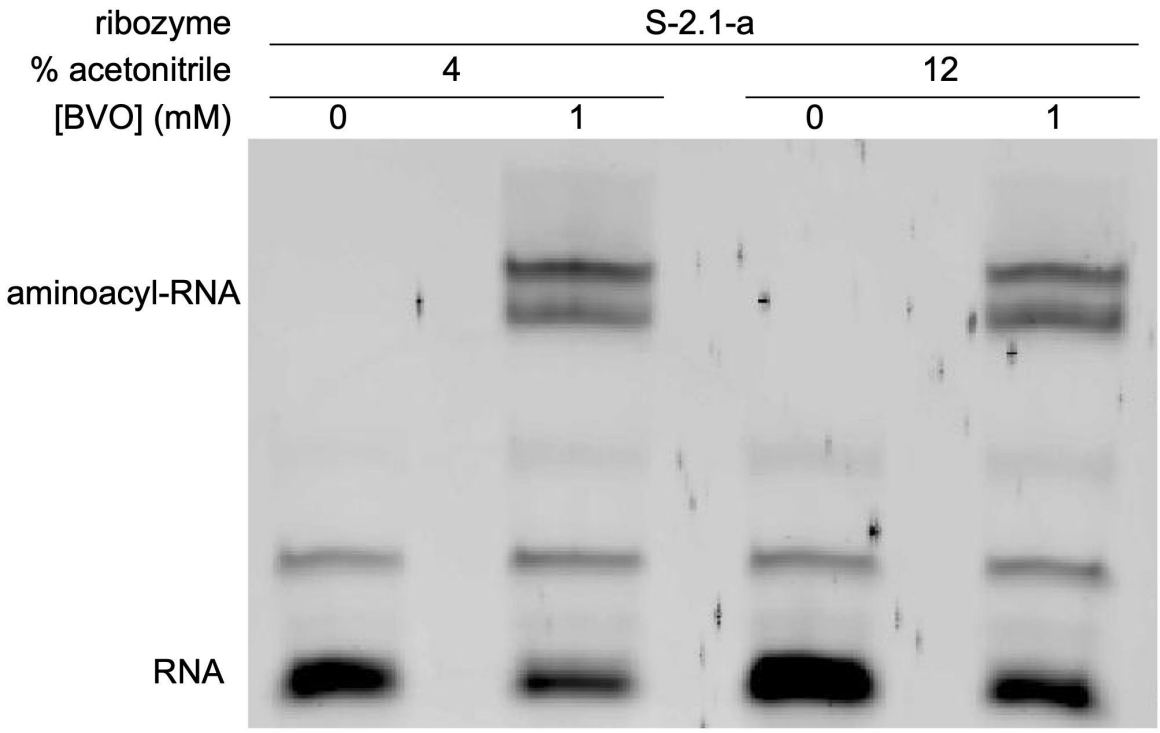

Figure 1.10. Streptavidin electrophoretic mobility shift assay of self-aminoacylation activity in the presence of the maximum acetonitrile concentration experienced by the RNA pool in the selection protocol. We mixed the previously identified generalist ribozyme S-2.1-a with BVO and stopped the reaction by desalting after 90 min incubation at room temperature, then mixed the RNA with streptavidin and size-separated it by native PAGE.

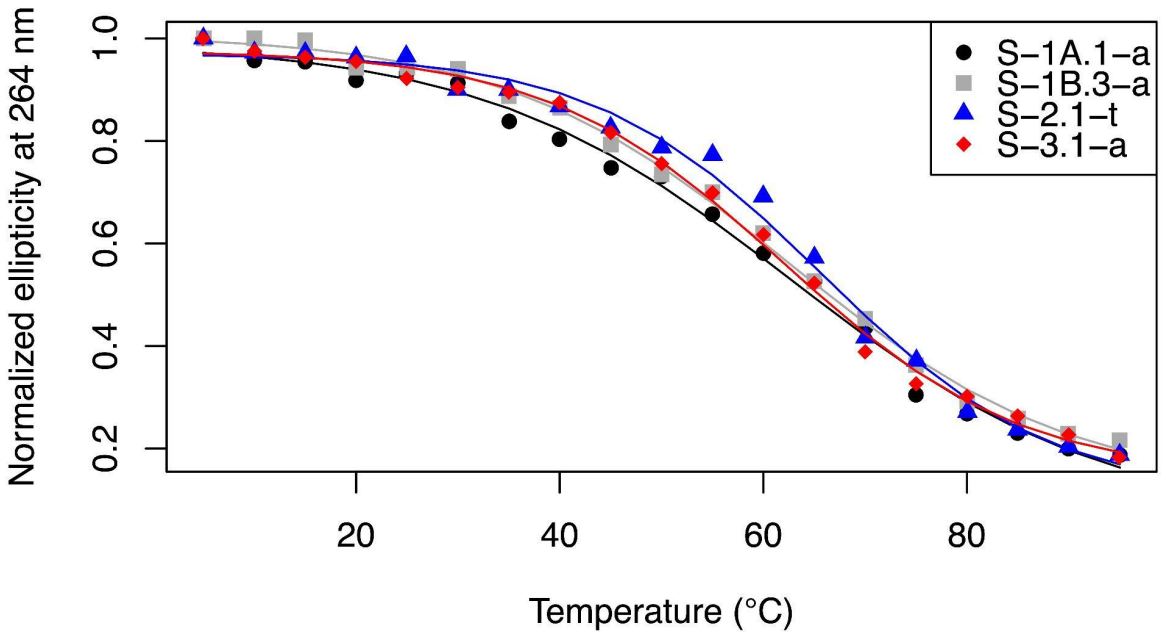

Figure 1.11. Melting curves by circular dichroism (CD) for some previously identified[41] model ribozymes. We mixed 4 μM RNA in 1X selection buffer in a quartz cuvette and measured the elliptical absorbance at 264 nm in 5 °C intervals with a JASCO J-1500 spectropolarimeter, then fit to the Boltzmann sigmoidal equation. T_m for all ribozymes measured fell in between 60 °C and 70 °C.

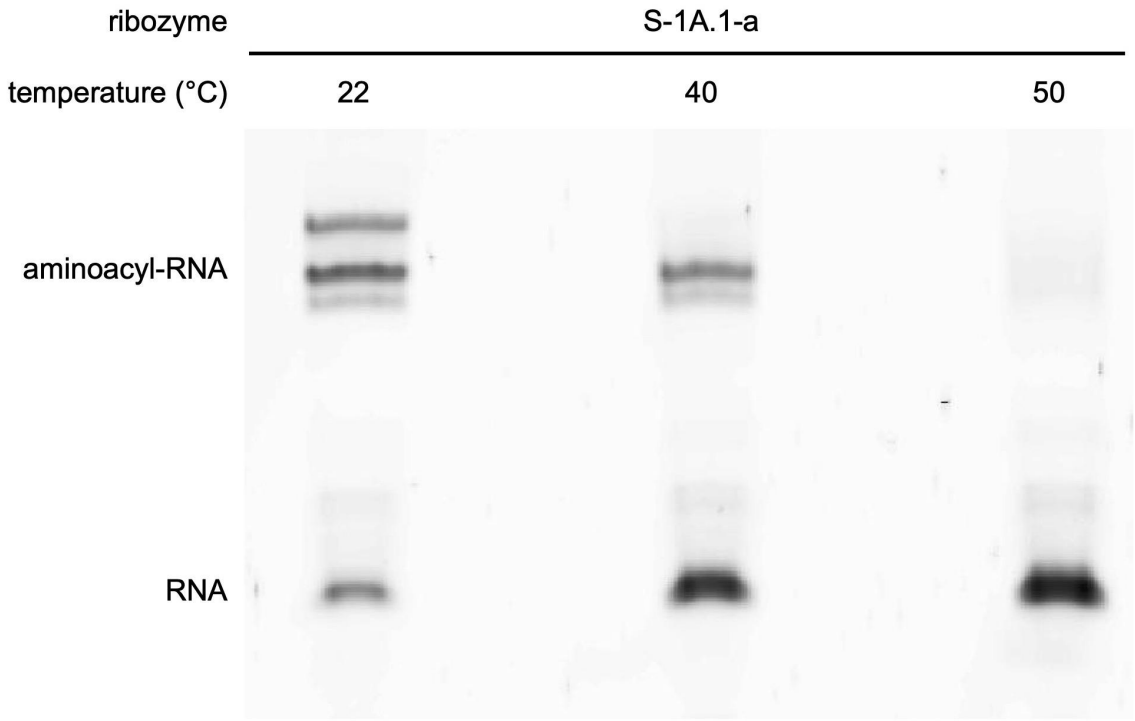

Figure 1.12. Temperature-dependent self-aminoacylation activity detected by streptavidin electrophoretic mobility shift assay. We mixed a previously characterized model ribozyme, S-1A.1-a, with 1 mM biotinyl tyrosine oxazolone (BYO), incubated for 90 min at different temperatures, then desalted the RNA, mixed it with streptavidin and size-separated by native PAGE. Streptavidin-bound aminoacyl-RNA appears as a higher molecular weight band.

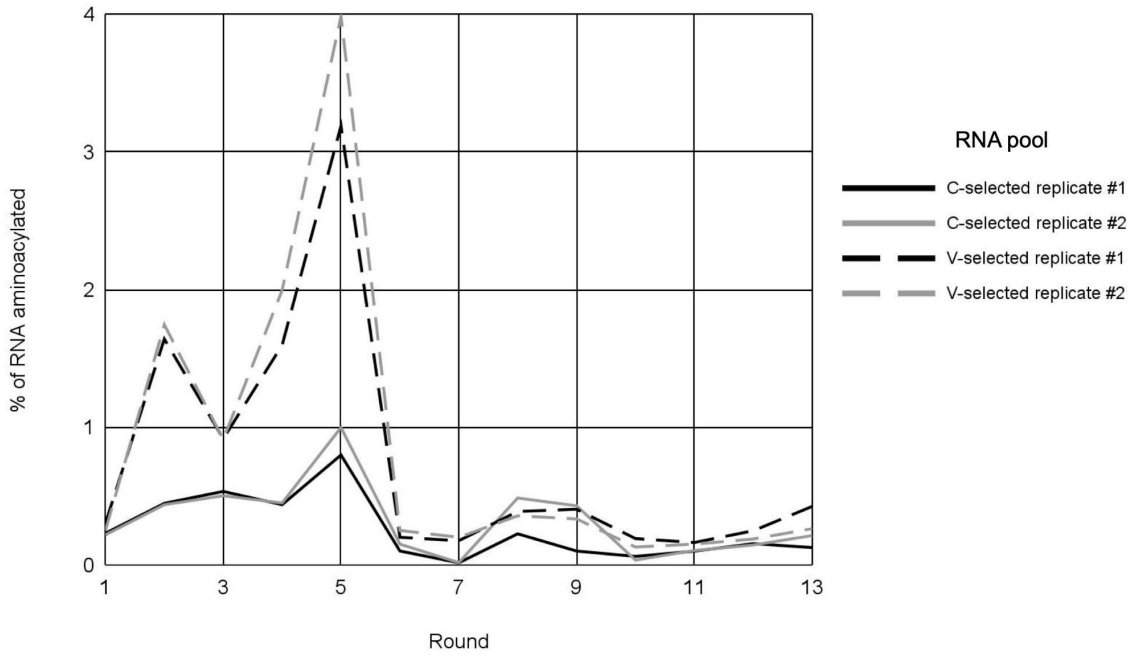

Figure 1.13. Fraction of each RNA pool aminoacylated each round based on back-calculated cDNA recovery following amplification by PCR. We measured the amount of DNA recovered after PCR by absorbance at 260 nm and back-calculated the initial quantity of cDNA based on the number of rounds of PCR performed. We increased the selection stringency following rounds 2, 5, 9, and 13. Though the relative fraction aminoacylated compared between C-selected and V-selected pools was consistent with electrophoretic mobility shift assays (EMSA), the round-over-round fraction aminoacylated correlated poorly with EMSA; thus, we stopped recording these data after round 13.

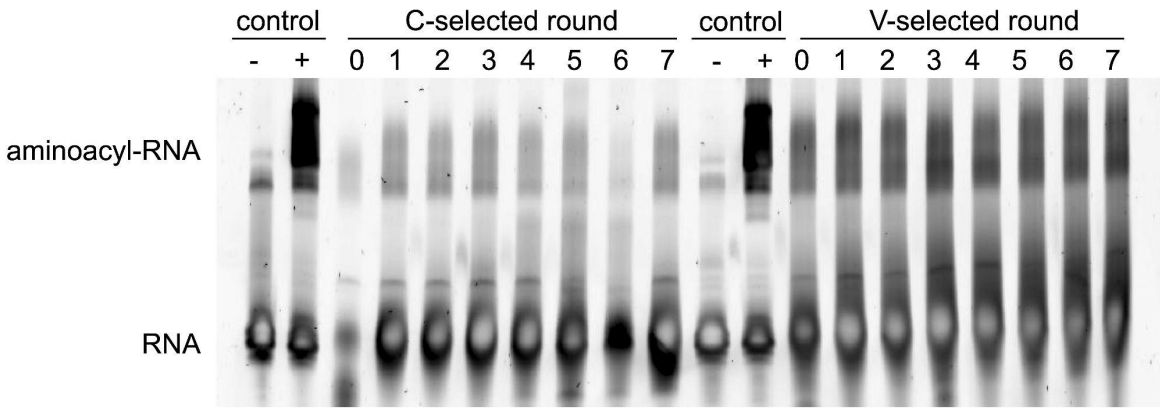

Figure 1.14. Streptavidin electrophoretic mobility shift assay of selection progress by round 7. We mixed RNA pools from replicate #1 selected under the constant temperature (C) condition or the variable temperature (V) condition with 500 μM BVO and subjected them to their respective temperature conditions. All lanes in this assay were overloaded, but no increase in aminoacylation over the round 0 background is detectable. Due to overloading, hyperstaining of lanes causes bands to appear as smears, and the middle of some bands appear white due to lack of penetration by SYBR Gold intercalating fluorescent dye. The control lanes are S-2.1-a mixed with 0 μM (-) or 500 μM (+) BVO.

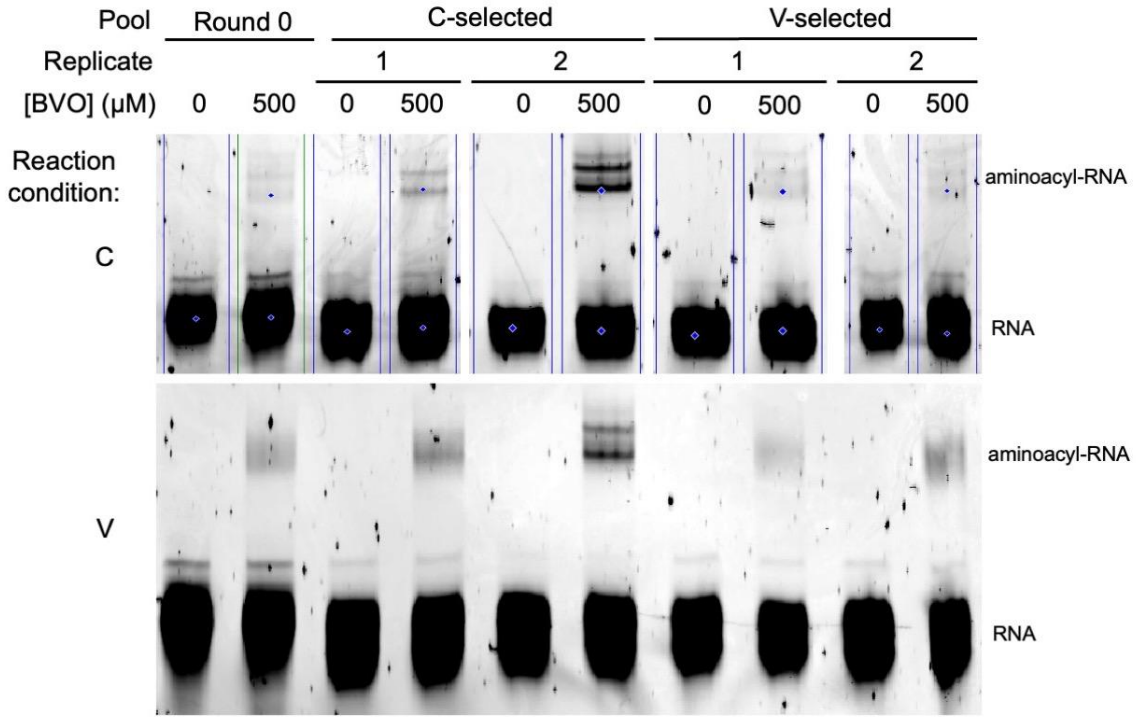

Figure 1.15. Streptavidin electrophoretic mobility shift assay of round 15 pools. We mixed RNA with BVO or acetonitrile and incubated at room temperature (control or "C" condition) or under a quasi-linear temperature ramp (variable or "V" condition) from 60 °C to -20 C for 90 min, then stopped the reaction by desalting, mixed with streptavidin, and size-separated on native PAGE. Blue lines and points on the C condition images are artifacts from imaging software denoting lanes and bands.

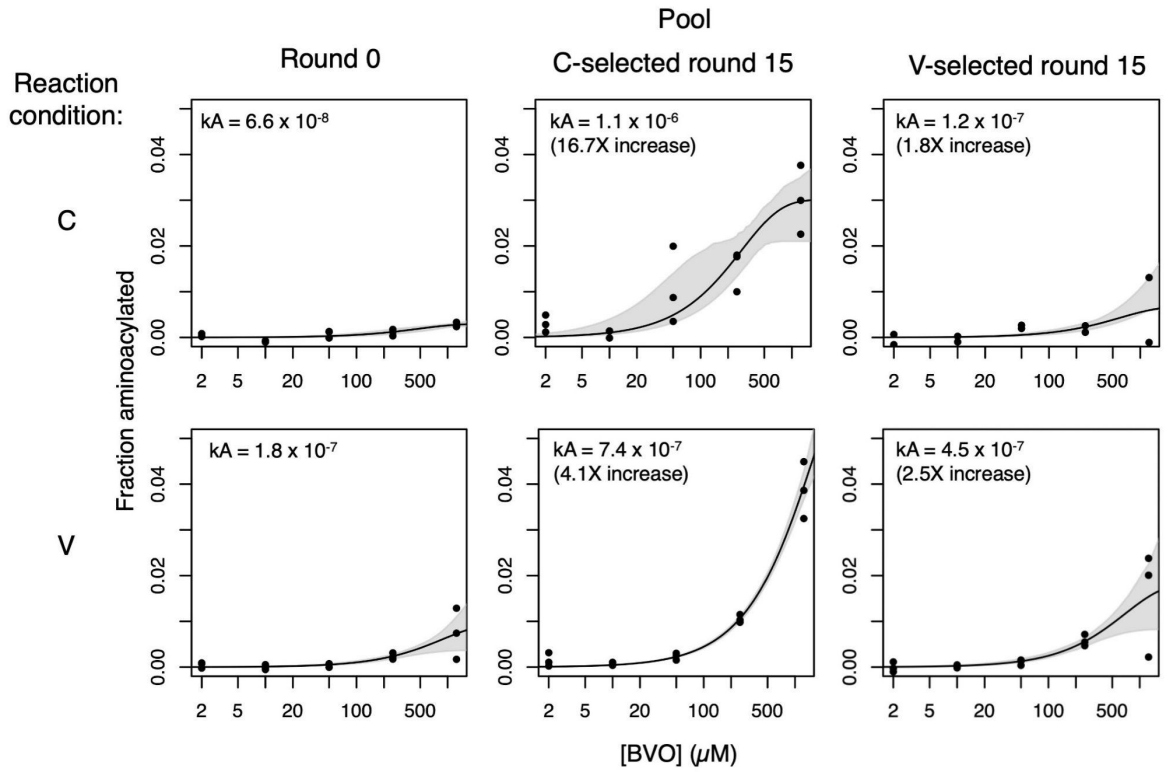

Figure 1.16. Catalytic activity of round 15 pools measured by reverse transcription and quantitative polymerase chain reaction (RT-qPCR). We mixed RNA with BVO at different concentrations for 90 min at room temperature (control or "C" condition) or under a quasi-linear ramp from 60 °C to -20 °C (variable or "V" condition), stopped the reaction by desalting, then separated the aminoacylated RNA with streptavidin-coated magnetic beads and quantified it by RT-qPCR. We measured triplicates at each concentration and fit the fraction aminoacylated (F) versus [BVO] to the first-order rate law $F=A(1-e^{k[BVO]t})$ to generate estimates of k and A. Black curves display the best fits and gray ribbons indicate bootstrapped 95% confidence intervals. The product kA is the most consistent metric of catalytic enhancement[41,42,60]; fold-increases over round 0 background are shown. All pairwise comparisons are significantly different (bootstrap, $p < 0.01$).

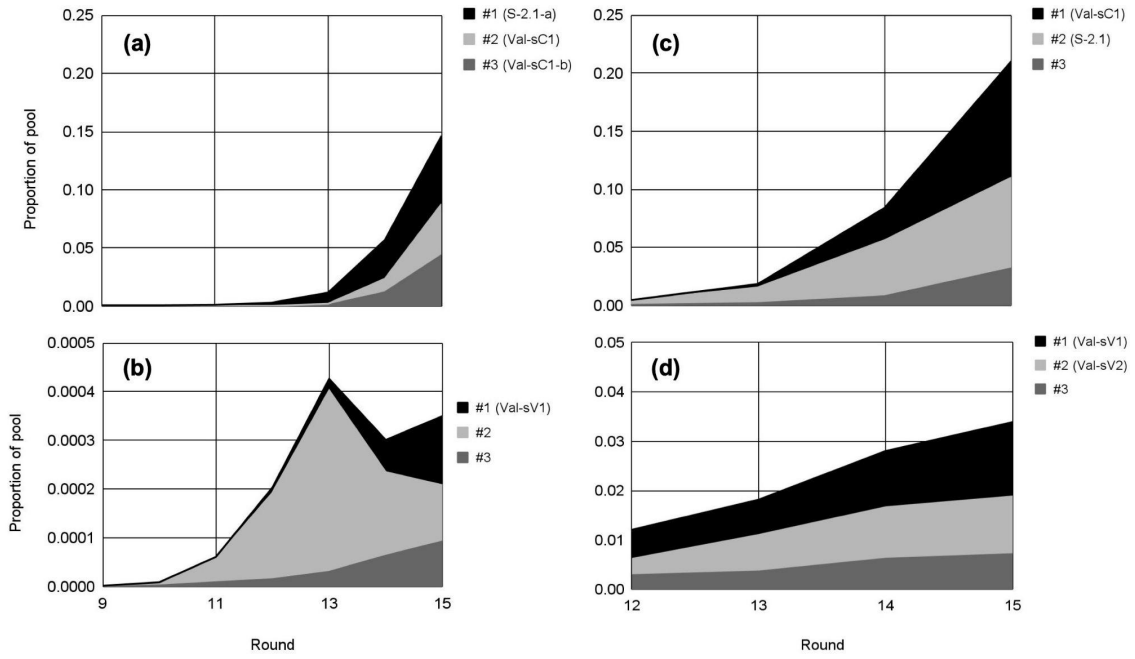

Figure 1.17. Enrichment of top sequences and sequence families in later rounds of the *in vitro* selection. Shown are the top three a) constant temperature (C)-selected sequences; b) variable temperature (V)-selected sequences; c) C-selected families; d) V-selected families by abundance after round 15. Sequences or families named elsewhere in the text are noted where appropriate.

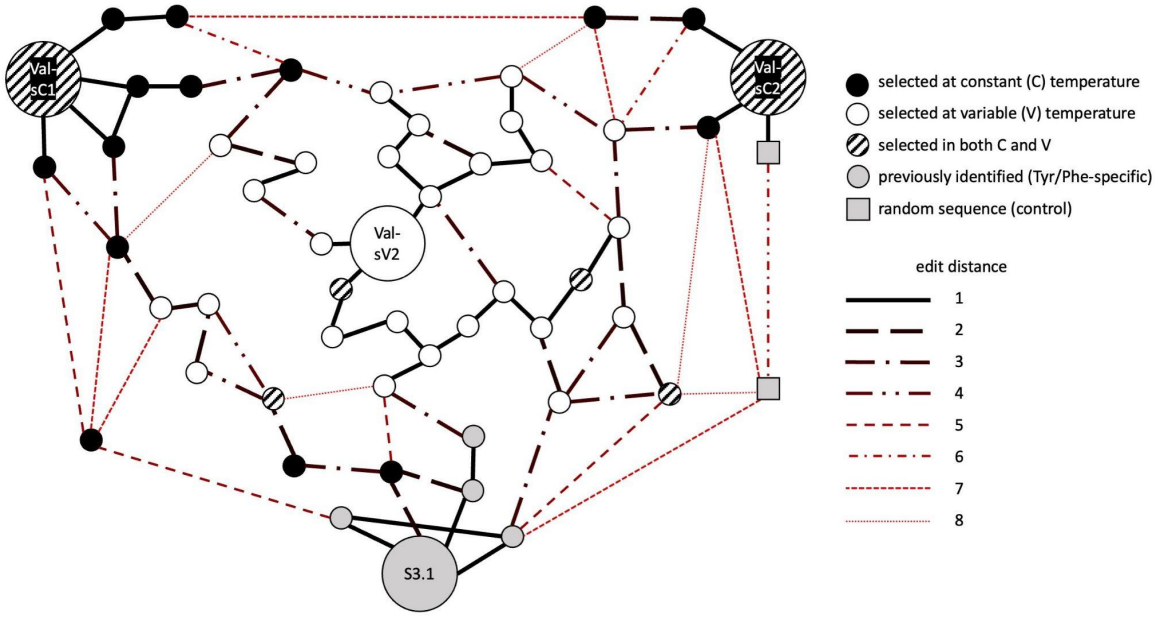

Figure 1.18. Best paths, using a modified A* algorithm, between the aromatic-specific ribozyme S-3.1-a[41,42] and valine-selected sequences Val-sC1 and Val-sC2, using only C-selected sequences, C- and V-selected sequences, or C-selected and random sequences as the set of possible nodes. All sequences with at least 2 counts after round 15 were considered. Without V-selected sequences, the best paths must make multiple steps of 5-8 edit distance. When V-selected sequences are included, the best paths travel through a V-specific peak, Val-sV2, reducing the maximum step size to 3-4 edit distance.

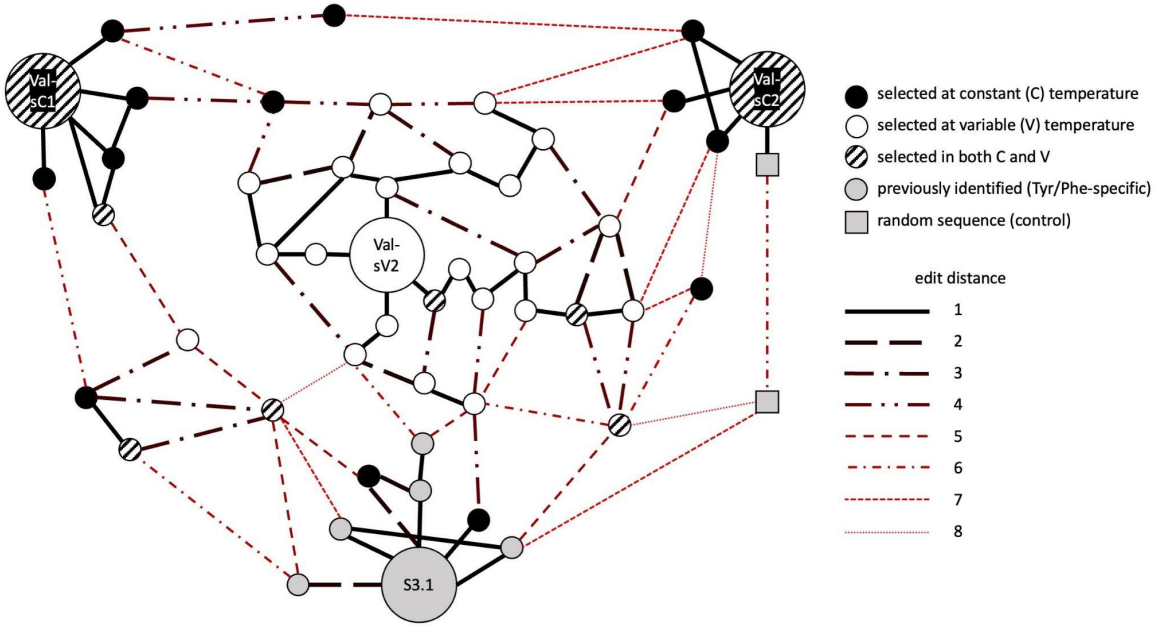

Figure 1.19. Same as Figure 1.18, but considering only sequences with at least 5 counts after round 15 as the set of possible nodes. Without V-selected sequences, the best paths require multiple steps of 6-8 edit distance. With V-selected sequences included, the best paths travel through the V-specific peak Val-sV2 and have maximum steps of 4-5 edit distance.

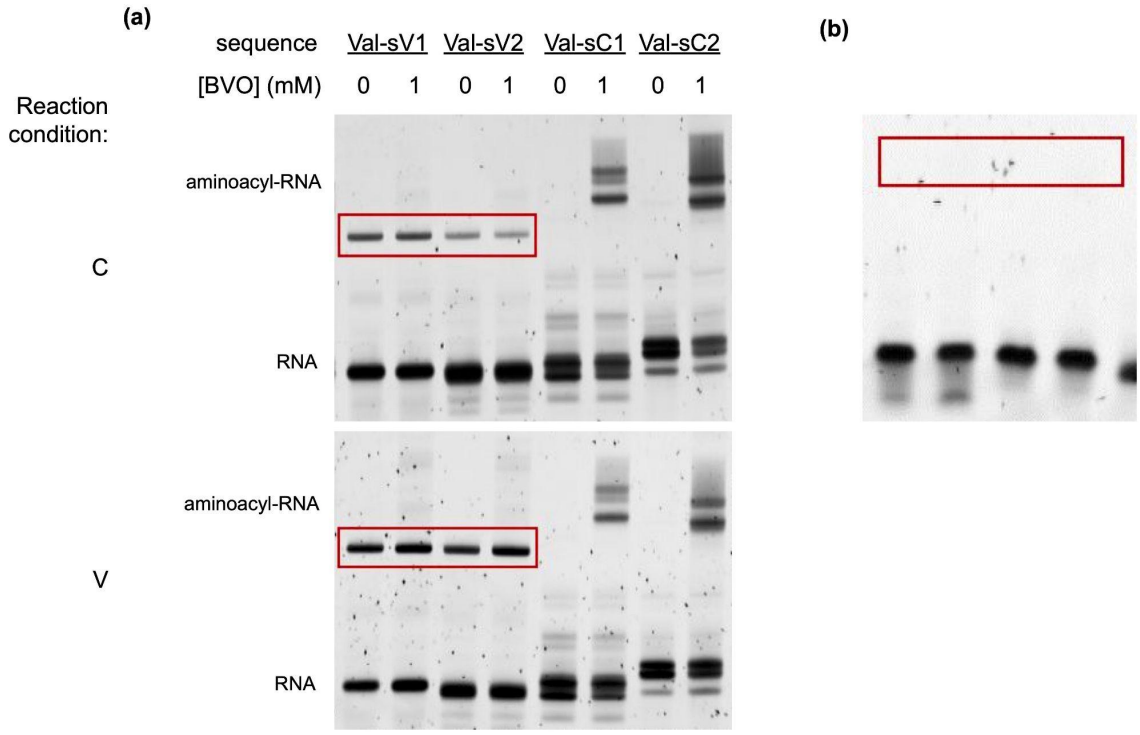

Figure 1.20. Streptavidin electrophoretic mobility shift assays of candidate ribozymes identified by *in vitro* selection in this chapter. 90-min reactions and native PAGE were conducted as before. a) Self-aminoacylation activity assay of two sequences selected in constant temperature (C) reaction conditions (Val-sC1 and Val-sC2) and two sequences selected in variable temperature (V) reaction conditions (Val-sV1 and Val-sV2). Aminoacylation activity is indicated by a higher molecular weight streptavidin-bound aminoacyl-RNA band after mixing with BVO that does not appear without BVO. The box highlights bands that appear streptavidin-bound, but disappear after heating the samples at 60 °C for 5 min (b), suggesting that they are some higher-order or alternate RNA structure.

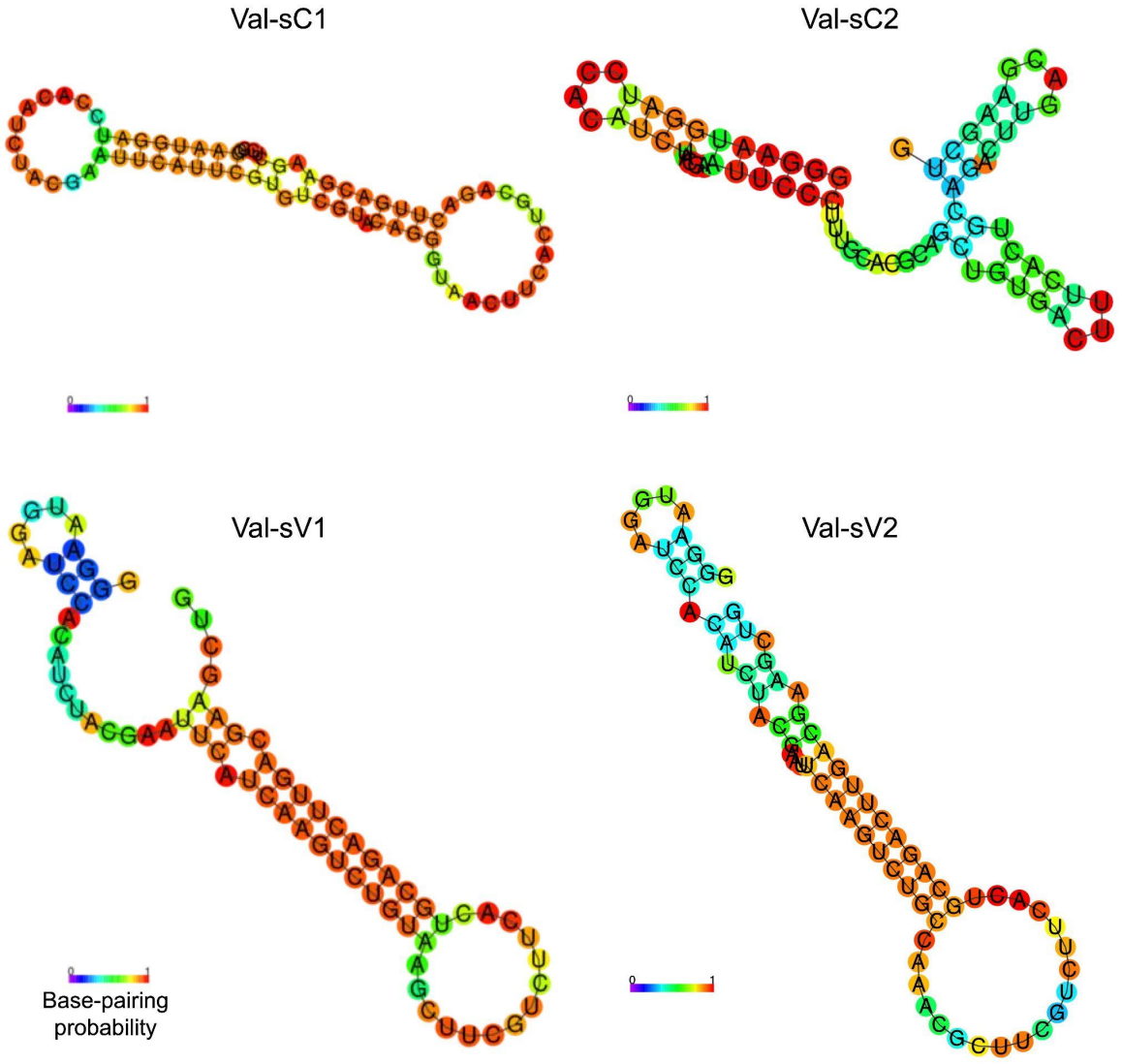

Figure 1.21. Predicted secondary structures for the individual sequences identified and tested in this chapter, generated by RNAfold[68].

Table 1.1. Individual sequences identified and tested in this chapter.

Name	Selected in	Variable region coding DNA sequence
Val-sC1	C, V	ATTCGTGTCGTACAGGGTAAC
Val-sC2	C, V	CCTTTGCACGCAGCTGTGACT
Val-sV1	V	ATCAAGTCTGTAAGCTTCGTC
Val-sV2	V	AAGTCTGCCAAACGCTTCGTC

1.7. Acknowledgements

We thank Ranajay Saha and Celia Blanco for technical assistance with CD spectroscopy and *in vitro* selection and Yei-Chen Lai, Ziwei Liu, and Robert Pascal for technical discussions. NMR was performed in the MRL Shared Experimental Facilities, which are supported by the MRSEC Program of the NSF under Award No. DMR 1720256; a member of the NSF-funded Materials Research Facilities Network (www.mrfn.org). Funding from NASA (NNX16AJ32G and 80NSSC21K0595) and the Simons Foundation Collaboration on the Origin of Life, and the UCSB Mellichamp Graduate Fellowship in Systems Biology and Bioengineering are acknowledged.

CHAPTER 2

Alternate Outcomes

2.1. Author contributions

This chapter was adapted from the manuscript

"Prebiotic chiral transfer from self-aminoacylating ribozymes may favor either handedness"

by

Josh Kenchel[1,2], Nadine Wells[3], Krishna Brunton[3], Evan Janzen[1,3], Weiwei Li[4], Eric T. Parker[5], Jason Dworkin[5], and Irene A. Chen[1,2,3]

The text was primarily written by Josh Kenchel and Irene A. Chen, with contributions from Weiwei Li. The experiments were designed with contributions from all authors and performed by Josh Kenchel, Nadine Wells, Krishna Brunton, Evan Janzen, Weiwei Li, and Eric T. Parker.

1. Program in Biomolecular Science and Engineering, University of California, Santa Barbara
2. Department of Chemical and Biomolecular Engineering, University of California, Los Angeles
3. Department of Chemistry and Biochemistry, University of California, Santa Barbara
4. Bren School of Environmental Management, University of California, Santa Barbara
5. Goddard Space Flight Center, National Aeronautics and Space Administration

2.2. Abstract

Modern life is essentially homochiral, consisting of D-nucleic acids and L-proteins. Since coded proteins are believed to have developed from a prebiotic RNA World, the homochirality of L-proteins observed in all known life presumably resulted from chiral transfer from a homochiral D-RNA World. This transfer would have been mediated by aminoacyl-RNAs defining the genetic code. Previous work on aminoacyl transfer using tRNA mimics has suggested that aminoacylation using D-RNA may be inherently biased toward reactivity with L-amino acids, implying a deterministic path from a D-RNA World to L-proteins. Using a model system of self-aminoacylating D-ribozymes and epimerizable activated amino acid analogs, we tested the chiral selectivity of 15 ribozymes derived from an exhaustive search of sequence space. All of the ribozymes exhibited detectable selectivity, and nearly half (6/15) reacted preferentially to produce the D-enantiomer of the product. Furthermore, chiral preference was conserved within sequence families. These results are consistent with the transfer of chiral information from RNA to proteins, but do not support an intrinsic bias of D-RNA for L-amino acids. Different aminoacylation structures result in different directions of chiral selectivity, such that L-proteins need not emerge from a D-RNA World.

2.3. Introduction

All known life is homochiral, consisting almost exclusively of right-handed (D-) sugars and left-handed (L-) amino acids[69]. Indeed, homochirality is considered to be a hallmark of life itself, and enantiomeric ratios of organic monomers are used as a biosignature informing the search for life elsewhere[46,70–72]. The selective advantages of homochirality, namely structural specificities that confer chemical efficiencies, could drive natural selection toward a system with nearly complete chiral asymmetry[73]. Therefore, factors determining the direction of the imbalance, and the exact mechanisms responsible for creating, amplifying, and perpetuating the imbalance, are important subjects of study[74].

The emergence of homochirality from a racemic abiotic milieu can be considered in three stages: symmetry breaking, in which the imbalance is introduced; amplification, in which the initial imbalance increases until homochirality is reached; and chiral transfer, in which one homochiral molecular type (e.g., RNA) passes the imbalance to another type (e.g., peptides), breaking the symmetry of the latter[75]. While recent research has substantially advanced the understanding of symmetry breaking and chiral amplification[45,76,77], chiral transfer is still not understood well.

The RNA World hypothesis posits that at some point in its early development, life went through a stage in which RNA played both roles of information carrier and catalyst in primitive living systems[2,3,78]. Since modern life consists of D-sugars, the RNA World is assumed to have shared this handedness. Interestingly, enantioenrichment of glyceraldehyde (a prebiotic sugar precursor) can be promoted by amino acids and dipeptides, indicating the potential for chiral transfer between small molecules even before the RNA World[79,80] became

established. The emergence of protein translation in the RNA World represents a major transition, during which the chiral nature of RNA could have been transferred to coded proteins. A critical bridge during this transition would have been the "adaptor" molecules that matched individual codons with specific amino acids. While tRNAs, charged by aminoacyl-tRNA synthetases, perform the role of adaptor molecules today, self-aminoacylating ribozymes[14,41,81,82] may have been the RNA World precursors of this system. Given the modern observation that nucleic acids have D-sugars and proteins have L-amino acids, it has been hypothesized that D-RNA might have inherent affinity or reaction with L-amino acids or their precursors, and therefore a D-RNA World would lead deterministically to an L-protein world. Indeed, earlier studies suggested that D-RNA would favor L-amino acids, based on the observation that an RNA minihelix mimicking the acceptor stem of tRNA reacts more quickly with an aminoacyl phosphate oligonucleotide when the aminoacyl chirality is L- rather than D- [83,84]. Furthermore, aminoacyl transfer between strands of a model tRNA acceptor stem also showed stereoselectivity preferring the L-enantiomer in two different transfer mechanisms. [57,85].

These studies on structural tRNA mimics suggest an intrinsic interaction of D-RNA with L-amino acids. On the other hand, some variants of an aminoacyl transferase ribozyme (the flexizyme) are capable of promiscuous activity, including catalyzing reaction between tRNA mimics and D-substrates[86], indicating the potential for chiral transfer from D-ribozymes to D-peptides. However, because the flexizyme was originally selected for reaction with L-substrates[87], these observations cannot address whether D-ribozymes would have an intrinsic preference for L- versus D-substrates. Therefore, the question of whether a D-RNA World would lead to an L-peptide world, due to an intrinsic preference of D-RNA,

remains open. In particular, in an alternative RNA World, utilizing different catalytic structures for aminoacylation, would D-RNA still favor reaction of L-amino acids?

To probe this question, we analyzed several self-aminoacylating ribozymes originally discovered through *in vitro* selection for reaction with an activated amino acid, biotinyl-Tyr(Me)-5(4*H*)-oxazolone (BYO)[19]. The oxazolone substrates epimerize[57,83,84] at the α-carbon due to tautomerization, so both L- and D-chirality at the α-carbon are equally represented in the reaction mixture. The starting library of this selection was a pool of RNA containing a random sequence flanked by regions of constant sequence[41], where the random region had nearly complete coverage of sequence space. The ribozymes catalyze the formation of an ester bond with an internal 2'-OH to form aminoacyl-RNA[85], fixing the chirality of the α-carbon upon reaction. Three distinct sequence motifs, having distinct aminoacylation sites, emerged from the selection, including several subfamilies and many individual ribozymes. All ribozyme motifs were found to react at internal 2'-OH sites (i.e., not at a strand terminus), unlike the modern tRNA acceptor stem and previously studied systems. The ribozymes also tolerate some changes to the substrate side chain, particularly for aromatic side chains[42]. The unusual reactive sites on the RNA, as well as the fact that the ribozymes were discovered through an exhaustive search of sequence space, indicate that these ribozymes are orthogonal to natural and tRNA-inspired systems. This orthogonality offers an opportunity to test whether the propensity for D-RNA to react to give L-aminoacyl products is in fact generalizable. In this work, we analyzed the ratio of D- to L-aminoacyl products in the reactions between a representative set of these ribozymes and the substrate biotinyl-Phe-5(4*H*)-oxazolone (BFO), to test the existence of an intrinsic, deterministic stereoselectivity governing chiral transfer from D-RNA to L-amino acids.

2.4. Methods

Ribozyme sequences and substrate

For each of the 15 D-ribozyme sequences investigated in this paper, chemical synthesis (Integrated DNA Technologies, PAGE-purified) was used to obtain single-stranded template DNA with the following sequence: 5′-GATAATACGACTCACTATAGGGAATGGATCCACATCTACGAATTC-N_x-TTCACTG CAGACTTGACGAAGCTG-3′, where N_x denotes the variable region (Table S1) and nucleotides upstream of the transcription start site are underlined. The ssDNA template was converted to dsDNA by Phusion polymerase (New England Biolabs) and a reverse primer of sequence 5'-CAGCTTCGTCAAGTCTGCAGTGAA-3', in a 50 μL mixture containing 1.5 μg template, 10 μM primer, 1 mM dNTPs, and 2 U enzyme in 1X HF buffer supplied by the manufacturer, with 68 °C annealing temperature and 72 °C synthesis temperature for 30 min.

RNA was transcribed from the dsDNA reaction mixture using T7 RNA Polymerase (New England Biolabs), in a 30 μL mixture containing 1 μg template, 6.7 mM each NTP, and 2 U enzyme in 1X buffer supplied by the manufacturer, at 4 °C overnight. The full-length product was purified by 7 M urea, 8% 29:1 mono:bis-polyacrylamide gel electrophoresis (PAGE) with the "crush-and-soak" method of gel extraction[88,89], followed by ethanol precipitation[90]. Purified RNA was dried in a vacuum centrifuge for 1.5 h and redissolved in TE buffer (10 mM Tris pH 8, 1 mM EDTA) and stored at -20 °C for up to one year. The purity of the full-length product was verified by PAGE with the same conditions.

L-RNA of ribozyme S-1A.1-a (Table 2.1) was obtained by chemical synthesis with HPLC purification (Bio-Synthesis, Inc.). The lyophilized pellet was resuspended in 10 mM Tris buffer (pH 8) and stored at -20 °C for up to one month.

Synthesis of substrate biotinyl-Phe-oxazolone (BFO) was performed according to a previously described procedure[41,42] (Figure 2.1). Reagents and solvents were obtained from Millipore Sigma and used without further purification. The identity and purity of the product were verified by ^{1}H NMR (Figure 2.2). Vacuum-dried solid BFO was dissolved in anhydrous acetonitrile with sonication, to a concentration of 25 mM, and stored at -20 °C for up to three months.

Aminoacylation reaction

To 6 mL of an aqueous buffer solution of 100 mM HEPES (pH 7) containing 100 mM NaCl, 100 mM KCl, 5 mM $MgCl_2$, and 5 mM $CaCl_2$, BFO stock solution was added to the desired concentration and mixed thoroughly. Then, RNA was added to a resultant concentration of 10 ng/μL (~400 nM). Because the reaction with the lowest [BFO] contained 50 μM BFO, all reactions contained >100-fold molar excess of substrate over RNA. The reaction was incubated on a rotator at room temperature overnight (Figure 2.3). Since the apparent catalytic rates of the ribozymes are ≥0.01 min^{-1} at the substrate concentrations tested, reactions reached saturation (>99% of maximum aminoacylation) within 10 hours.

Hydrolysis of reacted RNA to isolate biotinyl-phenylalanine

The reaction solution after incubation was concentrated to ~50 μL (Amicon Ultra-4 Centrifugal Filter Units, Millipore Sigma), after which the RNA was exchanged into 60 mM $Na_2B_4O_7$ (pH 9.5) by size exclusion chromatography (Bio-Gel P-30, Bio Rad), also removing excess substrate. The sample was then incubated overnight at 45 °C, during which time the mildly alkaline buffer selectively hydrolyzed the aminoacyl ester bond to liberate biotinyl-Phe from the RNA[91] (Figure 2.3). 1-μL aliquots of the RNA were taken before and after hydrolysis and analyzed by electrophoretic mobility shift assay (EMSA) (Figure 2.4). EMSAs were performed by mixing samples with 2 μM streptavidin (aqueous stock solution, New England Biolabs), incubating at room temperature for 5 min, adding 6X native loading dye (New England Biolabs), and loading without heating into native 8% 29:1 mono:bis-PAGE cast and run with the Bio-Rad minigel system. Gels were stained with SYBR Gold (Thermo Fisher Scientific) and imaged with an Amersham Typhoon scanner (General Electric). Aminoacylated RNA migrates more slowly than unmodified RNA due to streptavidin binding, resulting in a lower-mobility band. After successful hydrolysis of biotinyl-Phe was verified by the disappearance of the high-MW band, the RNA was removed from the hydrolysis reaction by size-based filtration (Amicon Ultra Centrifugal Filter Units, 10 kDa, Millipore Sigma), to isolate biotinyl-Phe.

Quantification of L- and D-phenylalanine by chiral chromatography

The concentrations of L- and D-phenylalanine were determined by two methods. In the first method, biotinyl-Phe samples were desiccated by vacuum centrifuge overnight, dissolved in 100 µL 6 M HCl, and incubated at 100 °C for 3 h to remove biotin[92] (Figure 2.3). Samples were then neutralized with equimolar NaOH prior to column separation (Figure 2.5). Samples were diluted 1:1 with acetonitrile and analyzed by ultrahigh performance liquid chromatography and mass spectrometry with an Agilent InfinityLab 1290 Infinity II Series system. L- and D-phenylalanine were separated on a 150 mm Agilent InfinityLab Poroshell 120 Chiral column at a flow rate of 0.5 mL/min, with 2 µL injection. The isocratic mobile phase was 70:30 aqueous:organic solvent, in which the aqueous component was 20 mM ammonium formate (pH 3) and the organic component was LC-MS grade methanol. The total run time for each sample was 10 min. Needle wash with isopropyl alcohol was done between injections. The transition (from precursor ion to fragment ion) for both L/D-Phe was 166 m/z to 120 m/z, with retention time of L-Phe at 5 min and D-Phe at 5.8 min. Six-point calibration standards ranging from 1 ng/mL to 100 ng/mL were used for absolute quantitation.

A second, more sensitive method was used to validate the accuracy of the first method (Figure 2.7). Biotinyl-Phe samples were desiccated and subjected to acid vapor hydrolysis at 150 °C for 3 h, using 6 M double-distilled hydrochloric acid[92,93], to remove biotin. Samples were then desalted by cation exchange chromatography[92,94] and prepared for analysis by pre-column derivatization with *o*-pthaldialdehyde/*N*-acetyl-L-cysteine as previously described[92,93]. Racemic tryptophan was used as an internal standard during

execution of ion exchange chromatography. Sample analysis was performed using ultrahigh performance liquid chromatography with fluorescence detection and ultrahigh resolution Orbitrap mass spectrometry. Chromatography of target analytes was achieved using a 150 mm Waters ACQUITY CSH Phenyl Hexyl column, a 100 mm Waters ACQUITY CSH C18 column, and a 150 mm Waters ACQUITY CSH Phenyl Hexyl column in series. Analytes were eluted using an isocratic flow (130 µL/min) of 60:40 aqueous:organic, where the aqueous eluent was 46 mM ammonium formate (pH 9.00) with 7% LC-MS methanol, and the organic eluent was LC-MS grade methanol. Fluorescence was detected with 340 nm excitation and 450 nm emission wavelength (Figure 2.6). Mass spectrometry was performed at a resolution setting of 60,000 (at full-width-half-maximum for m/z 400). Additionally, accurate mass analysis was facilitated using a polysiloxane compound (m/*z* 371.10124, $[(C_2H_6SiO)_5 + H]^+$ found in ambient air as an internal lock mass, which resulted in a typical mass accuracy of <1 ppm.

Determination of enantioselectivity index

We measured the chiral output (L-Phe vs. D-Phe) for multiple experimental replicates of reaction with each ribozyme, at multiple substrate concentrations, with overnight incubations, as described above. The fraction of total RNA aminoacylated by each stereoisomer, in each reaction, was calculated as the product of the fraction of L or D-Phe [$c_L/(c_L+c_D)$ and $c_D/(c_L+c_D)$, where c_X is the concentration of isomer X measured by LC-MS], multiplied by the fraction of total RNA aminoacylated, as measured by EMSA. To calculate

the ratio of the rate constants (k_D/k_L), a kinetic model was developed (Supplementary Text). The total aminoacylation reaction was assumed to follow pseudo-first-order kinetics with [BFO] >> [RNA], and [BFO] was assumed to be constant[41,60]. Since the substrate epimerizes[57,85,95] and was confirmed to contain a 50:50 mixture of D and L isomers (Figure 2.7, negative control), and [BFO] >> [RNA], [D] was assumed to be equal to [L]. In this model, the integrated rate equation is $F_X([BFO]) = \frac{Ak_X}{k_L+k_D}(1 - e^{-[BFO](k_L+k_D)t})$, where F_X is the fraction of total RNA aminoacylated with the *X*-Phe isomer (*X* = D or L), *A* is the maximum fraction of total RNA aminoacylated, and *t* is the incubation time (in these experiments, *t* = 16-20 h). To determine the ratio k_D/k_L, first the total amount of aminoacylated RNA (F_D+F_L) was fit to the form $C(1 - e^{-[BFO](k_L+k_D)t})$to estimate $k_D + k_L$. Then, this value of $k_D + k_L$ was fixed in the integrated rate equation, and data for F_X over [BFO] was fit to estimate Ak_x for each isomer, allowing calculation of k_D/k_L for each ribozyme. Note that k_D/k_L is insensitive to the choice of reaction time *t* (Supplementary Text). The log of this ratio [$\log_2(k_D/k_L)$], here termed the enantioselectivity index (i_{enant}), was used to characterize the selectivity of each ribozyme. Errors were calculated as 95% confidence intervals, determined by a bootstrapping procedure sampling replicate measurements with replacement and calculating k_D/k_L (*n* = 1,000).

2.5. Results

Ribozyme reactions with biotinyl-Phe-oxazolone substrate

Prior work from an exhaustive search of sequence space (random region of 21 nucleotides) had identified three ribozyme sequence motifs.[41] These ribozymes catalyze self-aminoacylation with biotinylated 5(4*H*)-oxazolones with aromatic side chains[42]. Ten ribozymes representing a diverse set of the most active sequences from the three major motifs had been previously validated for activity (Table 2.1). These included the most abundant representative from each major family or subfamily (eight sequences designated by the '-a' suffix), as well as two other sequences with significant activity (one from Motif 1 and one from Motif 2). In addition, we chose five more sequences from the most active family of Motif 2. To choose these, we first used electrophoretic mobility shift assays (EMSA) with streptavidin to assay the activity of 12 candidate sequences from Motif 2, based on results from a previous high-throughput assay (kinetic sequencing, or *k*-seq[41,60]) which had indicated high activity[42]. Of those 12 candidates, we chose the five ribozymes with the greatest self-aminoacylation by biotinyl-Phe-5(4*H*)-oxazolone (BFO), determined by EMSA, to include in the present analysis (Figure 2.8; Table 2.1). Thus, a total of 15 sequences (six from Motif 1, eight from Motif 2, and one from Motif 3) were tested in the following experiments.

Although these ribozymes had been originally identified by reaction with biotinyl-Tyr(Me)-5(4*H*)-oxazolone (BYO), BFO is a preferable substrate for chiral study, given established protocols for downstream chiral analysis of the products. In particular, methods for separating and detecting enantiomers of phenylalanine have been well

described[92,96,97]. The fifteen ribozymes were confirmed to have catalytic activity with BFO by EMSA (Figure 2.4), consistent with prior results[41,42].

As a 5(4*H*)-oxazolone, BFO is subject to epimerization and the rate of degradation by hydrolysis was expected to be similar for both stereoisomers[57,85]. Therefore, the BFO substrate was expected to contain a 1:1 mixture of both stereoisomers. To verify this, samples of BFO alone were analyzed by adding BFO stock solution (500 nM resultant concentration) to 60 mM $Na_2B_4O_7$ (pH 9.5) for hydrolysis, and incubating at room temperature on a rotator overnight. The resulting biotinyl-Phe was then hydrolyzed under strongly acidic conditions to yield Phe (L and D). The ratio of L-Phe to D-Phe was then verified to be indistinguishable from 1:1 using chiral chromatography (Figure 2.7).

Ribozymes display a range of stereoselectivities

We measured the stereoselectivity of the 15 ribozymes by conducting overnight aminoacylation reactions with different concentrations of substrate BFO. The amount of L- and D-isomers incorporated during aminoacylation was determined after each reaction, by mild basic hydrolysis of aminoacylated RNA to release biotinyl-Phe, followed by acid hydrolysis to release Phe (L and D). LCMS with a chiral column was used to quantify L-Phe and D-Phe (Figures 2.5 and 2.6). A more sensitive protocol involving derivatization was also used to validate the results (Figure 2.7). The amount of RNA aminoacylated by each stereoisomer over the concentration series was used to calculate the ratio of catalytic rates (k_D/k_L) (Figure 2.9), according to a pseudo-first-order rate law with excess substrate, given a

1:1 ratio of BFO isomers (Methods). The enantioselectivity index (i_e) of a ribozyme sequence was defined as $\log_2(k_D/k_L)$, such that a negative value of i_e indicates bias favoring the L isomer and a positive value of i_e indicates a bias favoring the D isomer (Figures 2.10 and 2.11). Of the ribozymes tested, nearly half (6/15) showed clear preference (>2-fold bias) for the D isomer of BFO. Of the nine L-selective ribozymes, five showed a strong preference (>4-fold bias) while four displayed relatively weak selectivity (<2-fold bias).

Stereoselectivity is consistent within sequence families

The 15 ribozymes represented three sequence motifs[41], with two motifs (1 and 2) represented by multiple distinct ribozymes. Within each motif, the direction of selectivity was consistent, with all members of Motif 1 preferring D-Phe and all members of Motif 2 preferring L-Phe (Figures 2.10 and 2.12). The difference in enantioselectivity between two ribozymes was generally greater when comparing ribozymes from different motifs (up to $\Delta i_e \approx 8$) rather than ribozymes from the same motif (up to $\Delta i_e \approx 4$) (Figure 2.13).

Stereoisomer preference is attributable to the aminoacyl α-carbon

Because BFO contains a biotin group, which itself has chiral centers (Figures 2.1-3 and 2.5), we considered the possibility that the measured ribozyme preferences might be dependent on stereocenters other than the aminoacyl α-carbon. In theory, this possibility

could be tested by reversing the stereocenters in the biotin moiety of BFO and assaying ribozyme selectivity, or, equivalently, by assaying the selectivity in a reaction of BFO with an L-ribozyme. The D-isomer of ribozyme S-1A.1-a prefers D-Phe with $i_e = 2.1$ (95% CI: 1.4-2.7), and yields 80 ± 3% aminoacylation at [BFO] = 1 mM ($i_e = 2.7 \pm 0.8$ [2σ] based on triplicates at 1 mM done in parallel with the L-RNA). If the ribozyme's activity is independent of the stereocenters of biotin, then the L-enantiomer of S-1A.1-a is expected to prefer L-Phe, with $i_e \cong -2.1$. The reaction of the L-ribozyme showed a preference for L-Phe with $i_e = -3.3 \pm 0.6$ and 79±3% aminoacylation based on triplicate measurements at [BFO] = 1 mM (Figure 2.14). Neither i_e nor % aminoacylation was significantly different (t-test, p = 0.11 and 0.29 respectively). These results are consistent with the primary determinant of selectivity being the chirality at the α-carbon rather than biotin.

2.6. Discussion

Previous work on this subject supports the idea that aminoacyl transfer reactions inspired by the transfer RNA (tRNA) structures found in modern cells show chiral selectivity in agreement with that found in nature, with right-handed (D-)RNA favoring reaction with left-handed (L-)amino acids. However, until now it has been unclear whether this selectivity would be generalizable to alternative structures from the RNA World. The ribozymes studied here catalyze aminoacylation at internal sites using an epimerizable substrate, 5(4*H*)-oxazolones, and thus represent a system orthogonal to tRNA-inspired structures. Several of the D-ribozymes analyzed in this work, specifically those from ribozyme Motif 1, were selective for D-phenylalanine over L-phenylalanine. This finding contradicts the hypothesis that D-RNA has an intrinsic general preference for L-amino acids. Instead, chiral preference appears to depend on the specifics of the system.

The homochiral L-protein world observed on Earth today is believed to have arisen from a homochiral D-RNA World, given the important advantages of homochirality, such as greater efficiency of replication[98]. The results from our model system suggest that either outcome (D-protein or L-protein) would have been possible, depending on the structures used for aminoacylation in the D-RNA world. Some of the observed selectivities are not large, with the greatest selectivity less than 16-fold. However, small imbalances could have been amplified by natural selection and/or physical resolution mechanisms[74]. Chiral preference was preserved within the ribozyme motifs studied, so evolution of aminoacylation ribozymes would not affect the direction of imbalance as long as the general motif was conserved. When replaying the Gouldian “tape of life” from a homochiral D-RNA World, a

homochiral L-protein world would not be guaranteed. Given the likely importance of chance in ribozyme discovery[28,41,60], the ultimate emergence of an L- vs. D-protein world may have been as random as a coin flip.

2.7. Figures

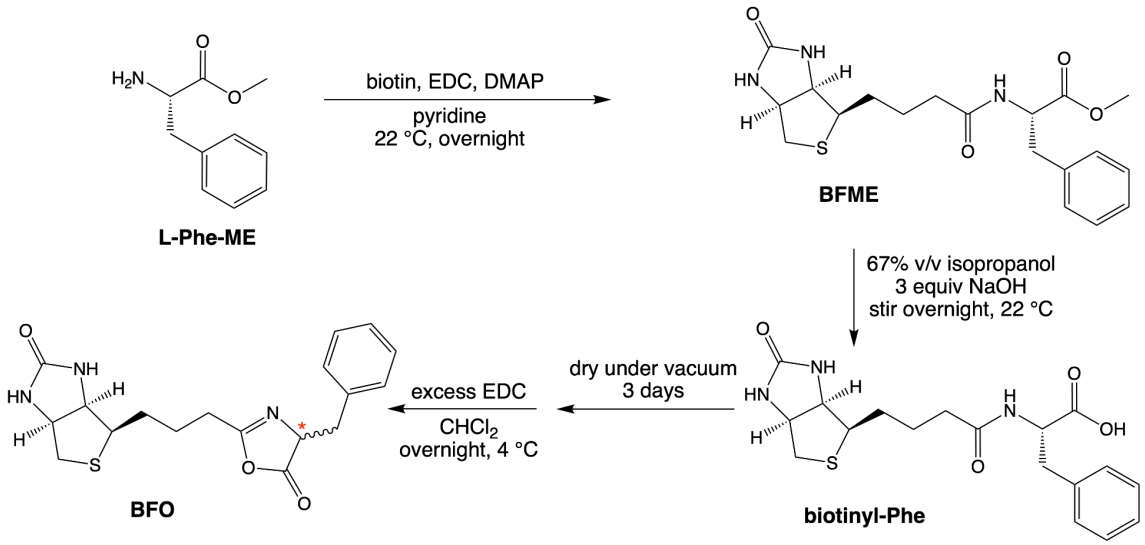

Figure 2.1. Synthesis of biotinyl-Phe-5(4*H*)-oxazolone (BFO)[42]. The chiral center corresponding to the α-carbon is indicated with a red asterisk.

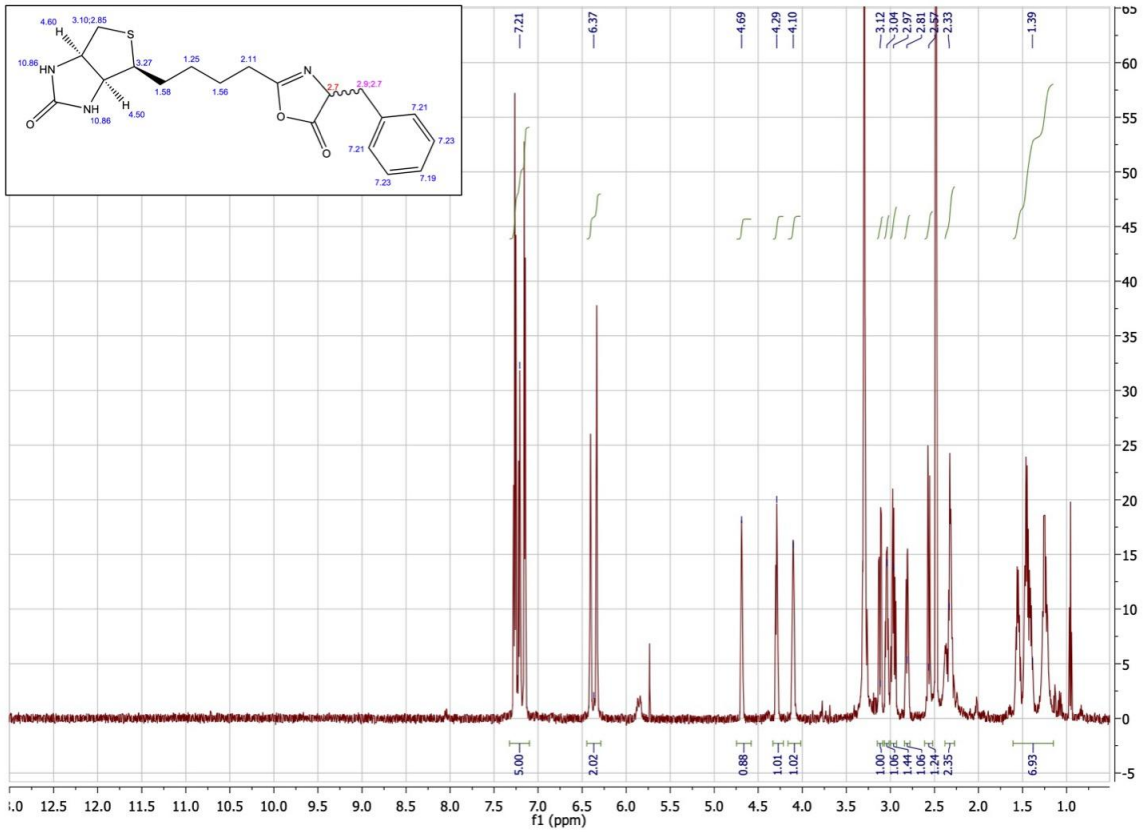

Figure 2.2. Proton nuclear magnetic resonance (H^1 NMR) spectrum of biotinyl-Phe-5(4*H*)-oxazolone (BFO) in deuterated dimethyl sulfoxide (DMSO). Peak centers labeled at top. Integrated values shown at bottom. Inset: structure of BFO with predicted H^1 NMR shifts generated by ChemDraw software.

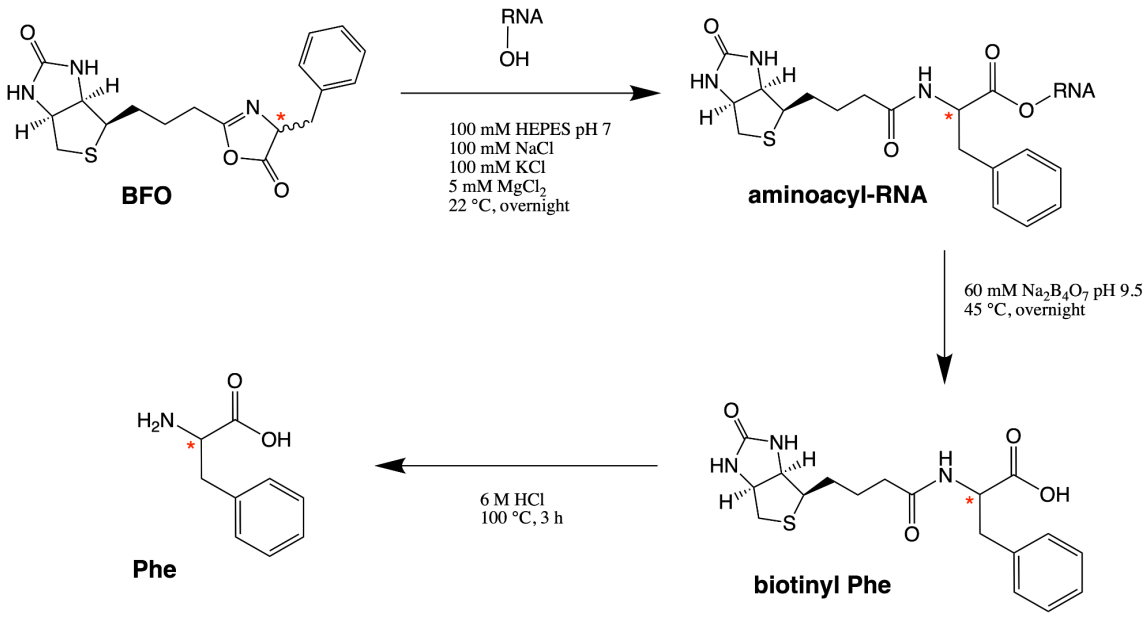

Figure 2.3. Preparation of phenylalanine for chiral separation. The analyzed chiral center is indicated with a red asterisk.

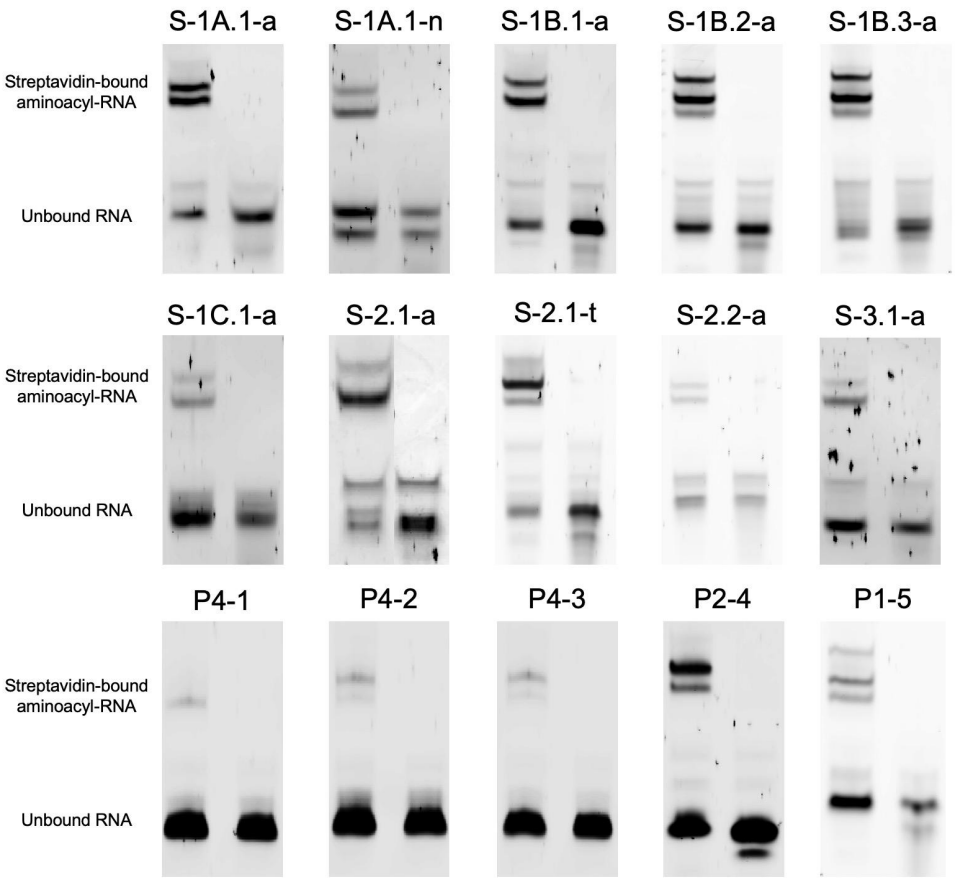

Figure 2.4. Confirmation of reactivity of ribozymes with BFO and product hydrolysis. Electrophoretic mobility shift assays (EMSA) of self-aminoacylated ribozymes after reaction, incubated with streptavidin before (left lane) and after (right lane) selective hydrolysis of the aminoacyl ester bond (incubation overnight at 45 °C in 60 mM $Na_2B_4O_7$ (pH 9.5)). 1-μL sample aliquots were saved before and after hydrolysis. These were incubated with 2 μM streptavidin for 5 min at room temperature, mixed with 1 μL 6X native loading dye, and loaded without heating into 8% 29:1 mono:bis-acrylamide gels cast and run with the Bio-Rad minigel system. The gels were run at 300 V for 15 min, stained with SYBR Gold (Thermo Fisher Scientific), and visualized with an Amersham Typhoon scanner.

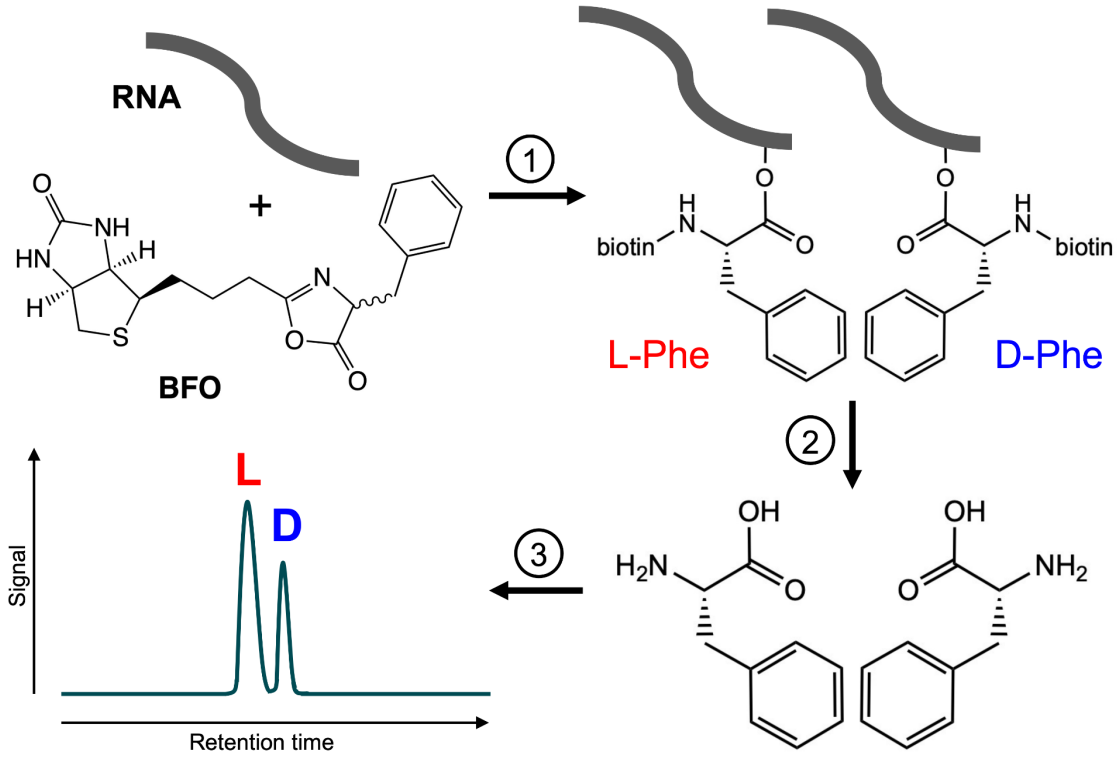

Figure 2.5. Method used to determine stereoselectivity of ribozymes. Each ribozyme is incubated with racemic biotinyl-Phe-5(4*H*)-oxazolone (BFO). Upon reaction (1), an ester bond is formed at an internal 2' hydroxyl to yield aminoacyl-RNA of unknown isomeric ratio. (2) The ester is selectively hydrolyzed, yielding biotinylated phenylalanine, which is separated from the RNA by size exclusion chromatography. The amide bond is then cleaved to remove the biotin moiety. (3) The resulting mixture of L- and D-phenylalanine is separated by HPLC on a chiral column, the peak identities verified by mass spectrometry, and the enantiomeric ratio determined by peak integration.

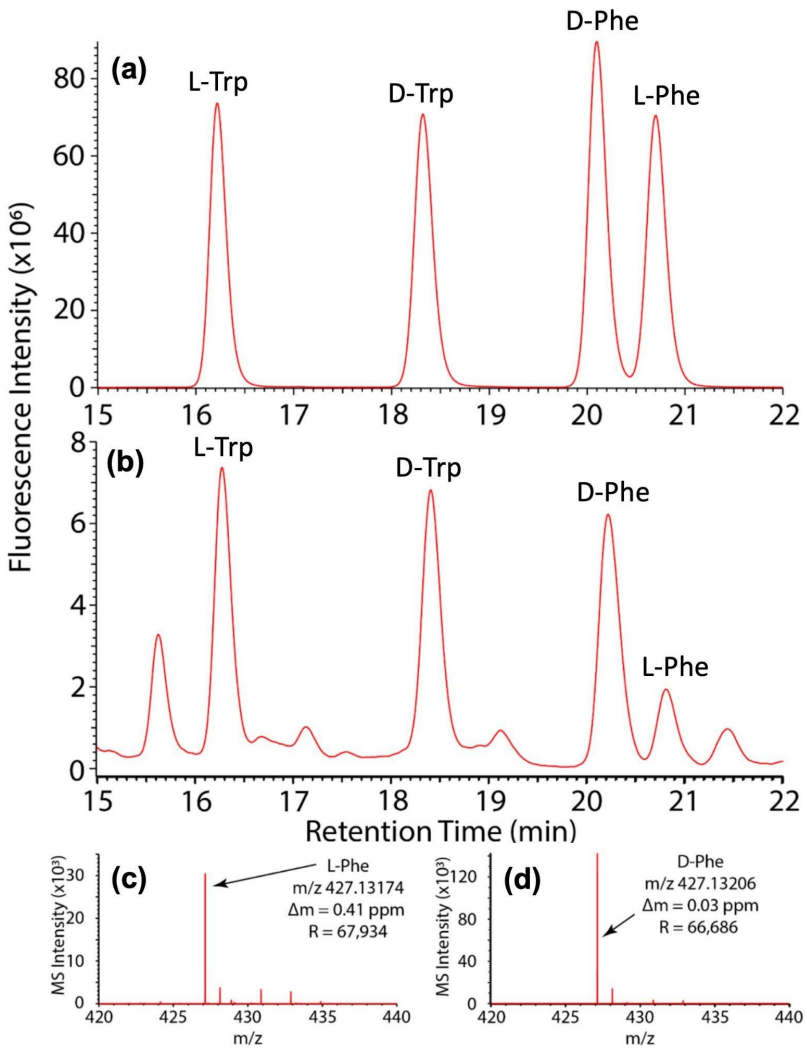

Figure 2.6. Determination of L-Phe and D-Phe concentrations by chiral LC-MS. (a-b) Example LC-MS analysis for S-1A.1-a, [BFO] = 1 mM. Comparison of target analyte detection in a combined 10 µM standard (a) and the sample (b). Racemic Trp was used as an internal standard during ion exchange chromatography. An excess of D-Phe over L-Phe was observed in this sample. (c-d) Mass spectra of L-Phe (c) and D-Phe (d) peaks detected in the sample from (b).

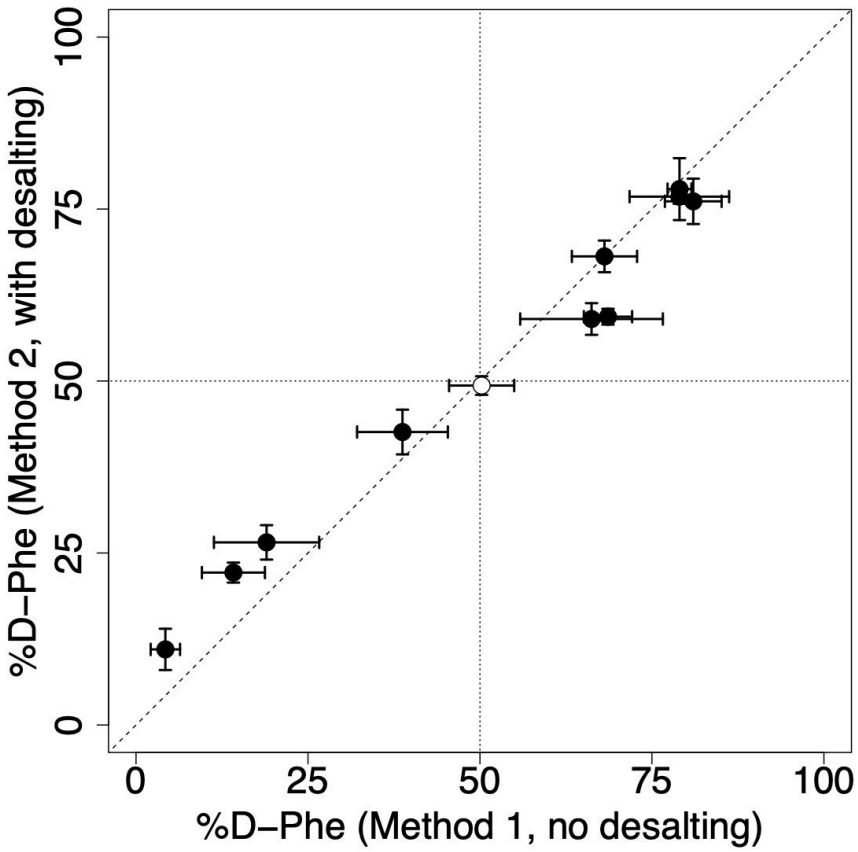

Figure 2.7. Validation of chiral chromatography procedure. Ultrahigh performance liquid chromatography-mass spectrometry (UHPLC-MS) data used for enantioselectivity analyses were obtained by method 1 (6 M HCl, no desalting or derivatization) and validated by method 2 (acid vapor, desalting column, and derivatized by *o*-pthaldialdehyde/*N*-acetyl-L-cysteine). Each point represents the mean of experimental triplicates. Error bars indicate standard error of the mean. Ribozymes were incubated with 1 mM BFO. The negative control (open circle) shows the result of 500 nM BFO incubated overnight in 60 mM $Na_2B_4O_7$ (pH 9.5). The identity line is shown as a dashed line. The slope of the linear best-fit was not significantly different than 1 ($p = 0.16$).

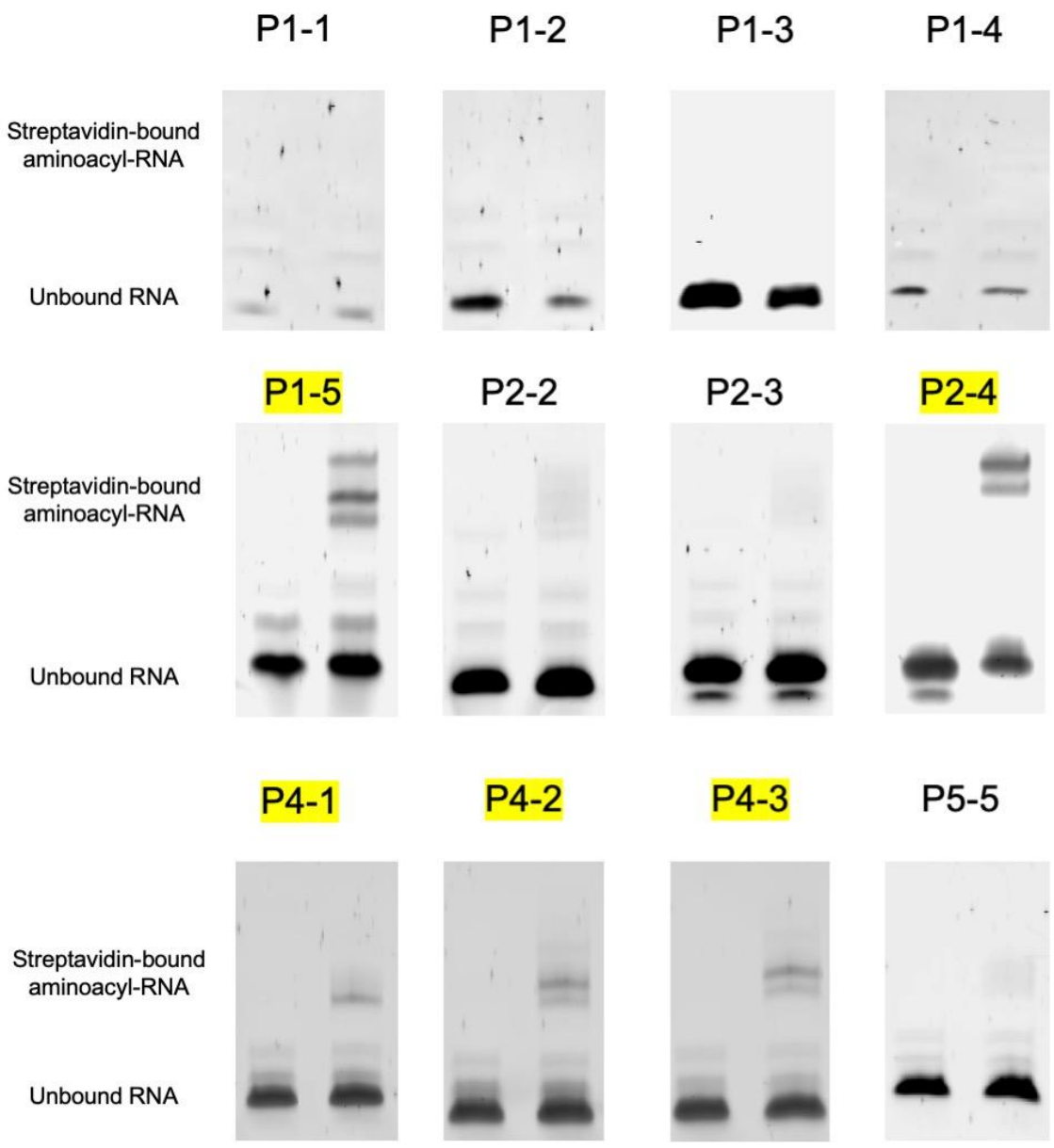

Figure 2.8. Electrophoretic mobility shift assay (EMSA) of aminoacyl-RNA without (left lane) or with (right lane) streptavidin for candidate ribozyme sequences. Based on these results, five sequences were chosen for enantioselectivity analysis (highlighted yellow).

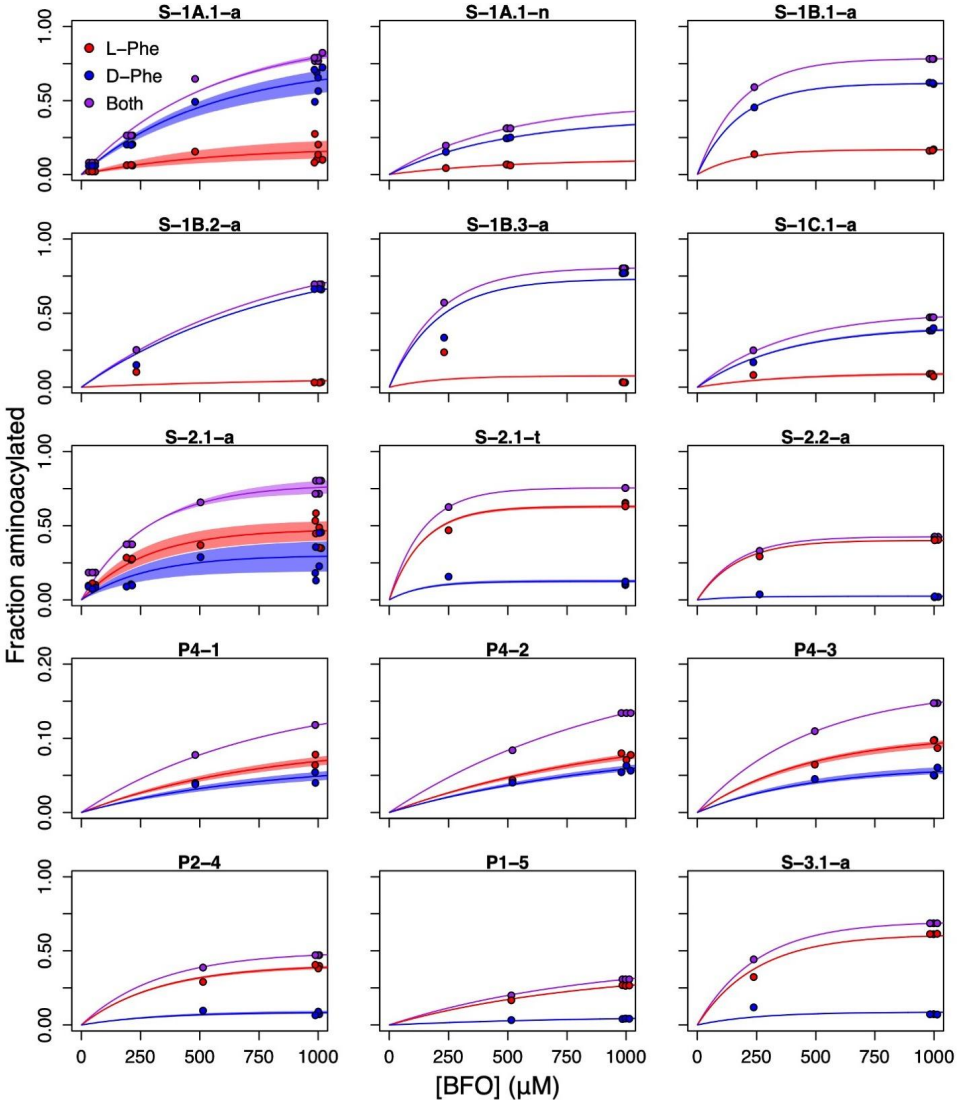

Figure 2.9. Determination of reaction rates for L- and D-products for all 15 ribozymes analyzed in this chapter. For each ribozyme, the fraction of RNA aminoacylated by L- and D-BFO was measured at multiple concentrations of BFO and fit to the pseudo-first order rate law as described, to estimate the enantioselectivity index $i_e = \log_2(k_D/k_L)$. Each point represents one experimental replicate. Replicates are jittered horizontally for visibility. Lines show the best fits to the rate law. Shaded areas indicate 95% confidence intervals from bootstrapping analysis.

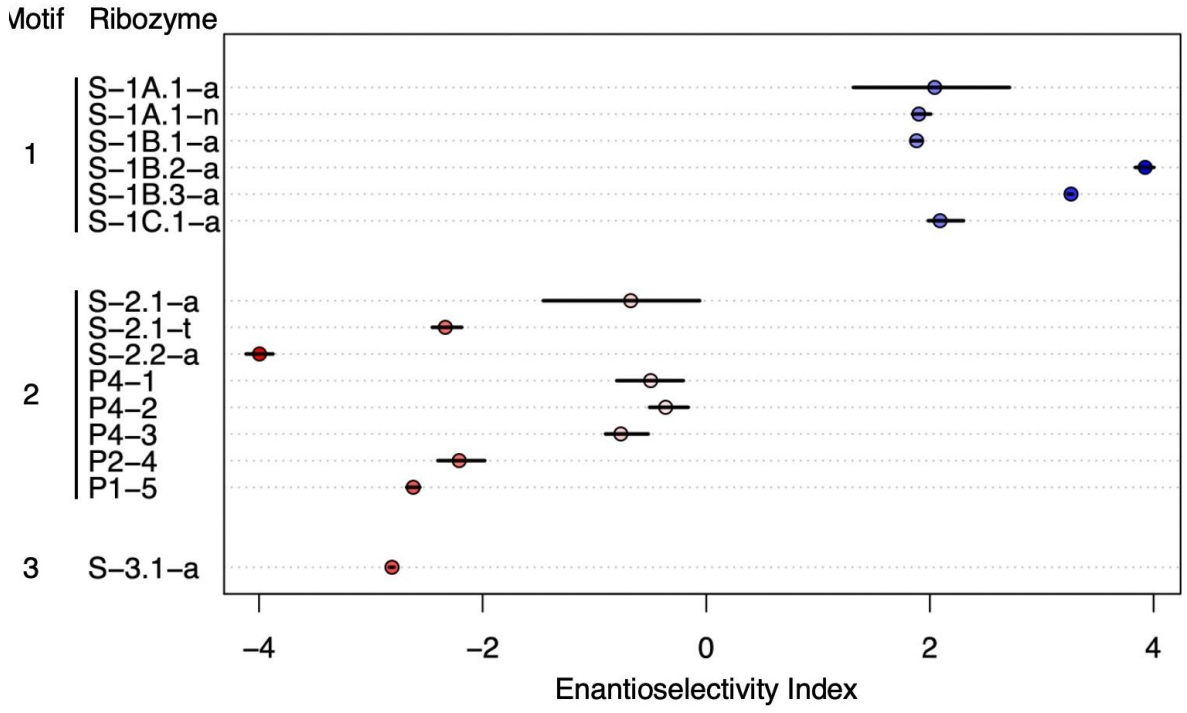

Figure 2.10. Enantioselectivity indices (i_e) for all 15 ribozymes analyzed in this chapter. Points indicate the median i_e and error bars indicate 95% confidence intervals determined by bootstrapping.

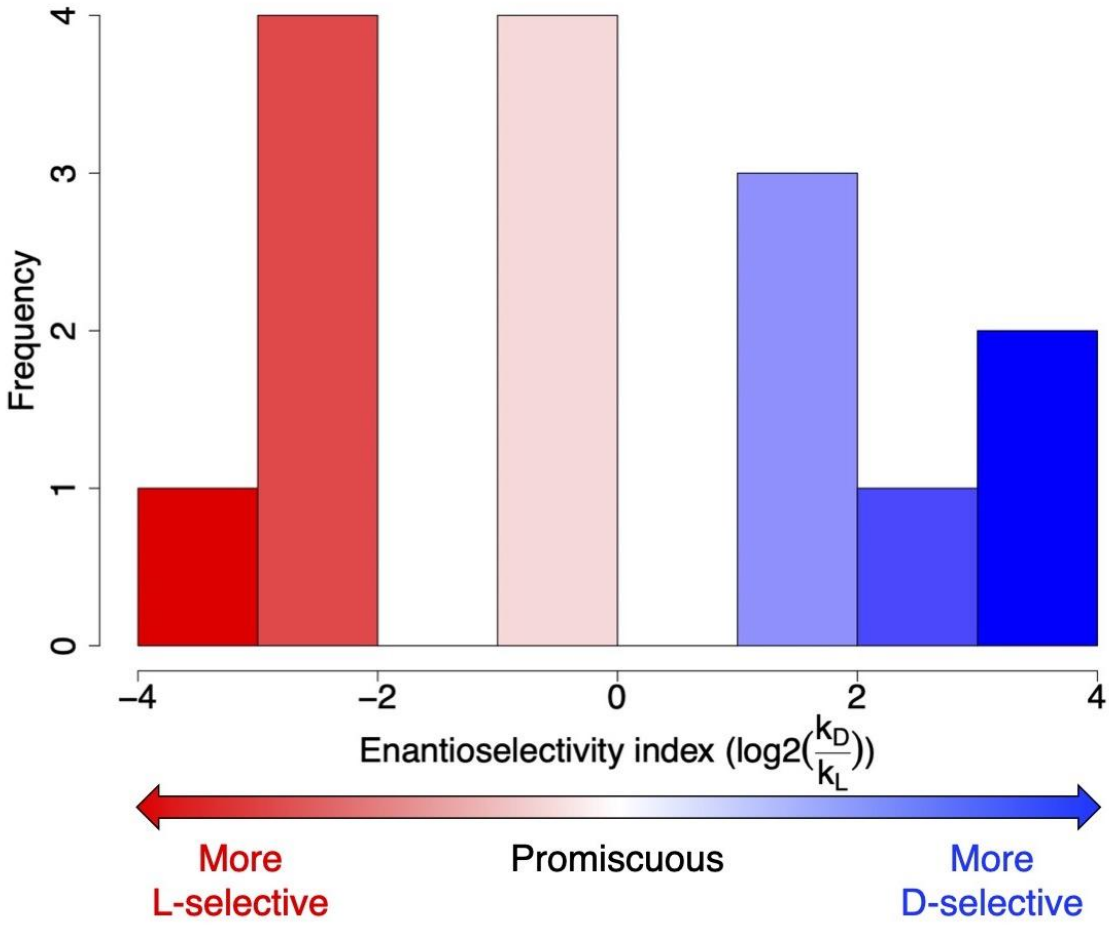

Figure 2.11. Distribution of enantioselectivity indices for the ribozymes analyzed in this chapter. We found a range of positive and negative values, with nearly half (6/15) preferring D-Phe and the other half preferring L-Phe.

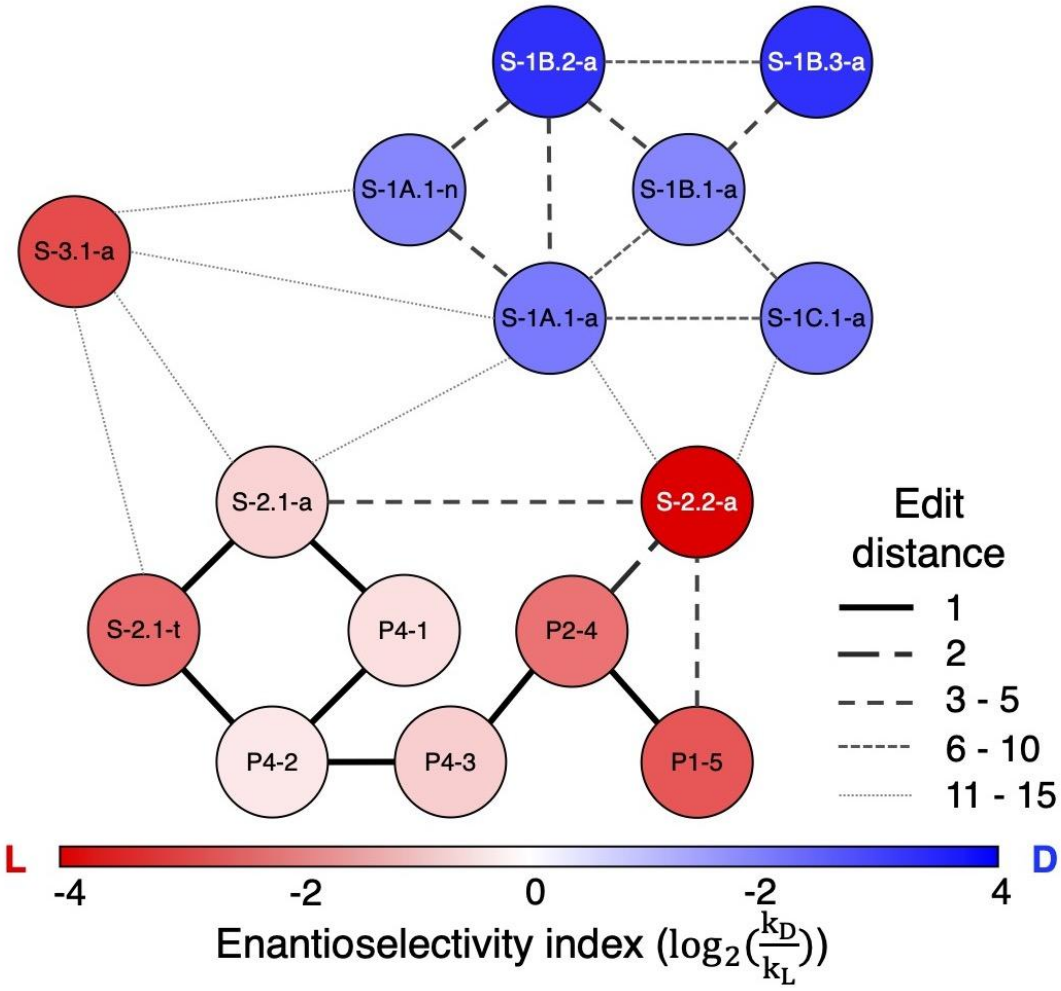

Figure 2.12. "Enantioselectivity landscape" of ribozymes analyzed in this chapter. The enantioselectivity index is shown as a heat map, with each ribozyme represented by a labeled circle. The Levenshtein edit distance (number of substitutions, insertions, or deletions) between sequences is indicated by the line strength (see legend). Direction of preference is conserved among related ribozymes.

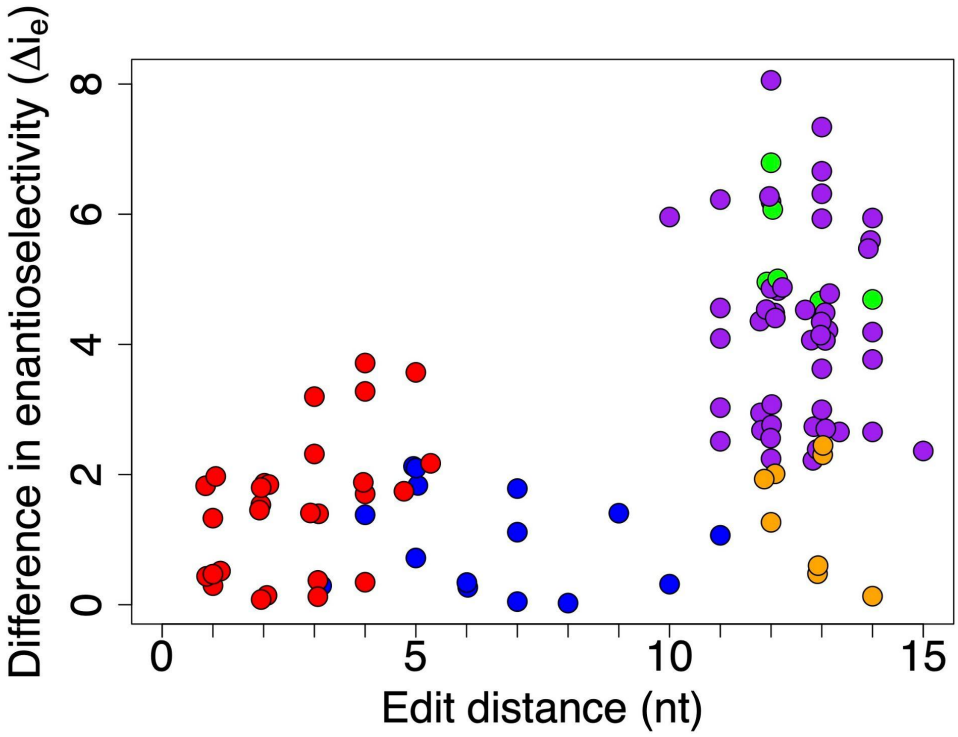

Figure 2.13. Difference in enantioselectivity vs. edit distance for each pair of ribozymes. Blue: between ribozymes within Motif 1; red: between ribozymes within Motif 2; purple: between ribozymes of Motifs 1 and 2; green: between ribozymes of Motifs 1 and 3; orange: between ribozymes of Motifs 2 and 3. Across motifs, evolutionary proximity correlates with shared enantiomeric preference (Spearman's rank correlation, $p < 0.001$), but the statistical significance of this relationship does not hold within motifs for these data.

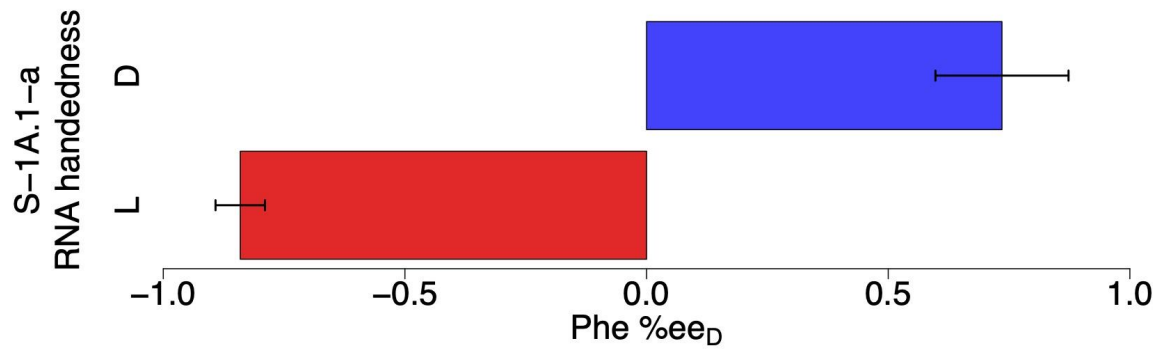

Figure 2.14. Stereoselectivity of ribozymes is dependent on the aminoacyl α-carbon. L- or D-RNA stereoisomers of ribozyme S-1A.1-a were incubated with 1 mM BFO and the ratio of D- to L-Phe in the aminacyl-RNA product was determined as described. Shown is the percent enantiomeric excess of the D-product (%ee$_D$). Bars indicate the mean of triplicates; error bars indicate the standard error of the mean. The %ee$_D$ of each was not significantly different from the inverse of the other (t-test, $p = 0.29$), indicating chiral reversibility of the reaction and supporting the hypothesis that the enantioselectivity of the ribozymes is dependent on the α-carbon, rather than other chiral centers in BFO.

Table 2.1. List of ribozymes analyzed in this chapter. Sequence motifs were described in Pressman et al. 2019. Enantioselectivity index $i_e = \log_2(k_D/k_L)$ determined by best fit to the first-order rate law as described in the Methods; 95% confidence intervals were determined by bootstrapping. Catalytic rate k and maximum aminoacylation fraction A were determined by kinetic sequencing (k-Seq)[41,42,60]; 95% confidence intervals are given in parentheses.

Name	Source	DNA sequence of variable region	Motif	i_e	k ($M^{-1}min^{-1}$)	A
S-1A.1-a	Pressman et al. 2019	CTACTTCAAAC AATCGGTCTG	1	2.07 (1.32, 2.77)	118.06 (20.17, 181.83)	0.79 (0.45, 1.00)
S-1A.1-n	"	CTCTTCAAACA ATCGGTCTTC	1	1.90 (1.84, 2.01)	122.46 (47.98, 227.55)	0.27 (0.22, 0.46)
S-1B.1-a	"	CCACACTTCAA GCAATCGGTC	1	1.88 (1.83, 1.92)	87.49 (34.90, 196.82)	0.99 (0.62, 1.00)
S-1B.2-a	"	CCGCTTCAAGC AATCGGTCGC	1	3.92 (3.84, 4.00)	136.11 (48.96, 279.25)	0.29 (0.21, 0.50)
S-1B.3-a	"	CCGAGTTTCAA GCAATCGGTC	1	3.26 (3.24, 3.28)	320.34 (112.17, 592.19)	0.61 (0.48, 0.86)
S-1C.1-a	"	CTCTTCAATAAT CGGTTGCGT	1	2.09 (1.99, 2.30)	103.94 (13.09, 503.88)	0.41 (0.14, 1.00)
S-2.1-a	"	ATTACCCTGGTC ATCGAGTGA	2	-0.67 (-1.40, -0.07)	1,656.27 (146.41, 4,616.20)	0.55 (0.37, 0.80)
S-2.1-t	"	ATTACCCTGGTC ATCGAGTGT	2	-2.33 (-2.45, -2.19)	1,860.97 (431.66, 5,167.17)	0.58 (0.40, 0.81)

S-2.2-a	"	ATTCACCTAGGTCATCGGGTG	2	-4.00 (-4.12, -3.88)	2,114.88 (880.72, 5,739.42)	0.31 (0.21, 0.42)
S-3.1-a	"	AAGTTTGCTAATAGTCGCAAG	3	-2.81 (-2.83, -2.79)	65.88 (17.87, 309.25)	0.66 (0.23, 1.00)
P4-1	This work	ATTACCTTGGTCATCGAGTGA	2	-0.50 (-0.80, -0.21)	108.30 (20.69, 2,219.98)	0.02 (0.01, 0.09)
P4-2	"	ATTACCTTGGTCATCGAGTGT	2	-0.36 (-0.51, -0.16)	49.597 (2.788, 1,428.20)	0.10 (0.03, 0.99)
P4-3	"	ATTACCTAGGTCATCGAGTGT	2	-0.76 (-0.90, -0.52)	550.48 (212.40, 2,646.26)	0.44 (0.19, 1.00)
P2-4	"	ATTCACCTAGGTCATCGAGTGT	2	-2.21 (-2.40, -1.98)	378.20 (172.18, 612.16)	0.47 (0.40, 0.61)
P1-5	"	ATTCACCTAGGTCATCGAGTTT	2	-2.62 (-2.68, -2.56)	222.78 (33.21, 673.91)	0.51 (0.26, 1.00)

2.8. Supplementary Text

Determination of enantioselectivity index from the integrated rate law

The concentration of D-BFO ([D]) and L-BFO ([L]) are assumed to be equal and constant, since BFO is synthesized in a racemic mixture and [BFO] >> [RNA]. At time $t = 0$, let the relative concentration of unmodified RNA ([R]) be equal to 1, and the concentration of aminoacylated RNA is 0. Aminoacylated RNA may be modified by D-BFO, giving product *RD*, or by L-BFO, giving product *LD*. The concentration constraints are:

(1) $$[L] = [D] = [B]$$

(2) $$R + [RL] + [RD] = 1$$

The chemical system consists of two irreversible reactions:

(3) $$R + L - k_L \rightarrow [RL]$$

(4) $$R + D - k_D \rightarrow [RD]$$

where k_L and k_D are the rate constants for the reaction with each respective stereoisomer. The change in [R] is given by the following differential equation:

(5) $$\frac{dR}{dt} = - k_L RB - k_D RB$$

the analytical solution to which is

(6) $$R(t) = Ae^{-B(k_L+k_D)t}$$

where *A* represents the maximum fraction of aminoacylated RNA. Focusing on one stereoisomer, the change in L-aminoacyl-RNA is given by the following differential equation:

(7) $$\frac{d[RL]}{dt} = RBk_L$$

Substituting (6) for R gives

(8) $$\frac{d[RL]}{dt} = ABk_L e^{-B(k_L+k_D)t}$$

the analytical solution to which, given the initial conditions, is

(9) $$[RL] = \frac{Ak_L}{k_L+k_D}(1 - e^{-B(k_L+k_D)t})$$

Similarly for D-aminoacyl-RNA,

(10) $$[RD] = \frac{Ak_D}{k_L+k_D}(1 - e^{-B(k_L+k_D)t})$$

Note that $k_D + k_L$ can be estimated from fitting of [*RL*]+[*RD*] vs. time to a pseudo-first-order rate equation. Note that the ratio k_D/k_L can be estimated from the ratio of amplitudes upon fitting [*RL*] and [*RD*] to equations 9 and 10, respectively, given data from multiple [BFO], or from direct calculation given a single [BFO], and that the ratio [*RD*]/[*RL*] = k_D/k_L is independent of time.

2.9. Acknowledgements

We thank Yang Li and Xiangning Huang for assistance with HPLC-MS and Ziwei Liu, Donna Blackmond, John Sutherland, and Robert Pascal for technical discussions. NMR was performed in the MRL Shared Experimental Facilities, which are supported by the MRSEC Program of the NSF under Award No. DMR 1720256; a member of the NSF-funded Materials Research Facilities Network (www.mrfn.org). Funding from NASA (NNX16AJ32G and 80NSSC21K0595), the Simons Foundation Collaboration on the Origin of Life, and the UCSB Mellichamp Graduate Fellowship in Systems Biology and Bioengineering are acknowledged.

Concluding Remarks

The work from our research group that has been summed up in this dissertation ultimately sought to probe interesting and important questions related to theoretical evolution and abiogenesis. Our findings are briefly recounted here. In Chapter 1 we examined the effect of environmental conditions on fitness landscapes. The results suggested that a dynamic environment increases the proportion of genetic space accessible for evolution to optimize function, and enables evolution between specialized variants for a given function (Figure 1.18). In Chapter 2 we tested whether chiral transfer from the RNA World to proteins would deterministically lead to a left-handed (L-)protein world. We documented the first self-aminoacylating ribozymes selective for right-handed (D-)amino acids, and found evidence that the bifurcation point between an L- or D-protein world may have been a random event (Figures 2.10-12).

While the impetus for our work was academic interest in fundamental biological questions about the distant past or yet-unreachable worlds – questions that may not be answered in our lifetimes – we acknowledge the potential applications to current and pressing efforts to improve the material condition of humans on Earth. Some of these were briefly covered in the Introduction. The chief beneficiary of a thorough understanding of the origin and early evolution of life will be the field of synthetic biology, which seems inexorably linked to the fate of humanity as it seeks to transform the way we produce food, fuel, and medicine. Knowledge of the forces governing fitness landscapes will empower more efficient directed evolution, and also enable prediction and prevention of emergent human pathogens[30,31,99]. Substrate-specific self-aminoacylating ribozymes are likely essential

components of a minimal translation system[100]. Furthermore, D-selective self-aminoacylating ribozymes could be used to construct D-proteins, which hold promise as potential hyperstable therapeutics or industrial enzymes[101].

The tangible benefits of this research may be years or decades away, but the expansion of the human-explored region of the infinite, *n*-dimensional continuous space of possible knowledge is of immediate benefit to the collaborative pursuit of science. We hope that the incremental contributions of the work presented in this dissertation will at least be presently of interest to some, and ultimately of service to many.

Bibliography

1. Joyce, G. F. RNA evolution and the origins of life. *Nature* **338**, 217–224 (1989).
2. Crick, F. H. C. The origin of the genetic code. *Journal of Molecular Biology* **38**, 367–379 (1968).
3. Orgel, L. E. Evolution of the genetic apparatus. *Journal of Molecular Biology* **38**, 381–393 (1968).
4. Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M. & Saxinger, W. C. On the Fundamental Nature and Evolution of the Genetic Code. *Cold Spring Harb Symp Quant Biol* **31**, 723–736 (1966).
5. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science (New York, N.Y.)* **332**, 209–12 (2011).
6. Chumachenko, N. V., Novikov, Y. & Yarus, M. Rapid and Simple Ribozymic Aminoacylation Using Three Conserved Nucleotides. *J. Am. Chem. Soc.* **131**, 5257–5263 (2009).
7. Zaug, A. J. & Cech, T. R. The intervening sequence RNA of Tetrahymena is an enzyme. *Science* **231**, 470–475 (1986).
8. Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68–74 (1994).
9. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561 (1970).
10. Cech, T. R. The Ribosome Is a Ribozyme. *Science* **289**, 878–879 (2000).
11. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *PNAS* **84**, 4398–4402

(1987).

12. Pressman, A., Blanco, C. & Chen, I. A. The RNA World as a Model System to Study the Origin of Life. *Current Biology* **25**, R953–R963 (2015).
13. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* **409**, 387 (2001).
14. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
15. Duve, C. de. The second genetic code. *Nature* **333**, 117 (1988).
16. Schimmel, P. R. & Söll, D. Aminoacyl-tRNA Synthetases: General Features and Recognition of Transfer RNAs. *Annu. Rev. Biochem.* **48**, 601–648 (1979).
17. Wen, K. & Orgel, L. E. The Specificity of Peptide Chain Extension by N-Carboxyanhydrides. *Orig Life Evol Biosph* **31**, 241–248 (2001).
18. Biron, J.-P., Parkes, A. L., Pascal, R. & Sutherland, J. D. Expeditious, Potentially Primordial, Aminoacylation of Nucleotides. *Angewandte Chemie International Edition* **44**, 6731–6734 (2005).
19. Liu, Z., Rigger, L., Rossi, J.-C., Sutherland, J. D. & Pascal, R. Mixed Anhydride Intermediates in the Reaction of 5(4H)-Oxazolones with Phosphate Esters and Nucleotides. *Chemistry – A European Journal* **22**, 14940–14949 (2016).
20. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97–159 (1931).
21. Kimura, M. Evolutionary Rate at the Molecular Level. *Nature* **217**, 624 (1968).
22. Kimura, M. The Neutral Theory of Molecular Evolution. *Scientific American* **241**, 98–129 (1979).
23. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *Journal of Theoretical Biology* **141**,

211–245 (1989).

24. Aita, T. & Husimi, Y. Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscapeII. Effect of small non-additivity. *J Math Biol* **41**, 207–231 (2000).
25. Li, C., Qian, W., Maclean, C. J. & Zhang, J. The fitness landscape of a tRNA gene. *Science* **352**, 837–840 (2016).
26. Puchta, O. *et al.* Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840–844 (2016).
27. Athavale, S. S., Spicer, B. & Chen, I. A. Experimental fitness landscapes to understand the molecular evolution of RNA-based life. *Current Opinion in Chemical Biology* **22**, 35–39 (2014).
28. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proceedings of the National Academy of Sciences* **110**, 14984–14989 (2013).
29. Pitt, J. N. & Ferré-D'Amaré, A. R. Rapid Construction of Empirical RNA Fitness Landscapes. *Science* **330**, 376–379 (2010).
30. He, X. & Liu, L. Toward a prospective molecular evolution. *Science* **352**, 769–770 (2016).
31. de Visser, J. A. G. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics* **15**, 480–490 (2014).
32. Joyce, G. F. Amplification, mutation and selection of catalytic RNA. *Gene* **82**, 83–87 (1989).
33. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA

ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).

34. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818 (1990).
35. McCall, M. J., Hendry, P. & Jennings, P. A. Minimal sequence requirements for ribozyme activity. *Proceedings of the National Academy of Sciences* **89**, 5710–5714 (1992).
36. Shippy, R., Lockner, R., Farnsworth, M. & Hampel, A. The Hairpin Ribozyme: Discovery, Mechanism, and Development for Gene Therapy. *MB* **12**, 117–130 (1999).
37. Hammann, C. & Lilley, D. M. J. Folding and Activity of the Hammerhead Ribozyme. *ChemBioChem* **3**, 690–700 (2002).
38. Hamasaki, K., Killian, J., Cho, J. & Rando, R. R. Minimal RNA Constructs That Specifically Bind Aminoglycoside Antibiotics with High Affinities. *Biochemistry* **37**, 656–663 (1998).
39. T. Le, T., Chumphukam, O. & G. Cass, A. E. Determination of minimal sequence for binding of an aptamer. A comparison of truncation and hybridization inhibition methods. *RSC Advances* **4**, 47227–47233 (2014).
40. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566 (1992).
41. Pressman, A. D. *et al.* Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* **141**, 6213–6223 (2019).
42. Janzen, E., Shen, Y., Liu, Z., Blanco, C. & Chen, I. A. *Error minimization and specificity could emerge in a genetic code as by-products of prebiotic evolution*. 2021.05.14.444235

https://www.biorxiv.org/content/10.1101/2021.05.14.444235v1 (2021) doi:10.1101/2021.05.14.444235.

43. Deamer David, Singaram Sara, Rajamani Sudha, Kompanichenko Vladimir, & Guggenheim Stephen. Self-assembly processes in the prebiotic environment. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 1809–1818 (2006).
44. Bonner, W. A. Chirality and life. *Origins Life Evol Biosphere* **25**, 175–190 (1995).
45. Blackmond, D. G. The Origin of Biological Homochirality. *Cold Spring Harb Perspect Biol* **2**, (2010).
46. Glavin, D. P., Burton, A. S., Elsila, J. E., Aponte, J. C. & Dworkin, J. P. The Search for Chiral Asymmetry as a Potential Biosignature in our Solar System. *Chem. Rev.* **120**, 4660–4689 (2020).
47. Blout, E. R., Doty, P. & Yang, J. T. Polypeptides. XII. The Optical Rotation and Configurational Stability of α-Helices [1]. *J. Am. Chem. Soc.* **79**, 749–750 (1957).
48. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science* **332**, 1193–1196 (2011).
49. Flynn, K. M., Cooper, T. F., Moore, F. B.-G. & Cooper, V. S. The Environment Affects Epistatic Interactions to Alter the Topology of an Empirical Fitness Landscape. *PLOS Genetics* **9**, e1003426 (2013).
50. Wright, S. (University of C. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress of Genetics* **1**, 356–366 (1932).

51. Fragata, I., Blanckaert, A., Dias Louro, M. A., Liberles, D. A. & Bank, C. Evolution in the light of fitness landscape theory. *Trends in Ecology & Evolution* **34**, 69–82 (2019).

52. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53–59 (2008).

53. Modi, A., Vai, S., Caramelli, D. & Lari, M. The Illumina Sequencing Protocol and the NovaSeq 6000 System. in *Bacterial Pangenomics: Methods and Protocols* (eds. Mengoni, A., Bacci, G. & Fondi, M.) 15–42 (Springer US, 2021). doi:10.1007/978-1-0716-1099-2_2.

54. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-Throughput Sequencing Technologies. *Molecular Cell* **58**, 586–597 (2015).

55. Aalto, A. P. & Pasquinelli, A. E. Small non-coding RNAs mount a silent revolution in gene expression. *Current Opinion in Cell Biology* **24**, 333–340 (2012).

56. Mutschler, H., Wochner, A. & Holliger, P. Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry* **7**, 502–508 (2015).

57. Liu, Z. *et al.* Harnessing chemical energy for the activation and joining of prebiotic building blocks. *Nature Chemistry* **12**, 1023–1028 (2020).

58. Willcott, M. R. MestRe Nova. *J. Am. Chem. Soc.* **131**, 13180–13180 (2009).

59. Blanco, C., Verbanic, S., Seelig, B. & Chen, I. A. EasyDIVER: A Pipeline for Assembling and Counting High-Throughput Sequencing Data from In Vitro Evolution of Nucleic Acids or Peptides. *J Mol Evol* **88**, 477–481 (2020).

60. Shen, Y., Pressman, A., Janzen, E. & Chen, I. A. Kinetic sequencing (k-Seq) as a massively parallel assay for ribozyme kinetics: utility and critical parameters. *Nucleic Acids Research* **49**, e67 (2021).

61. AbouHaidar, M. G. & Ivanov, I. G. Non-Enzymatic RNA Hydrolysis Promoted by the Combined Catalytic Activity of Buffers and Magnesium Ions. *Zeitschrift für Naturforschung C* **54**, 542–548 (1999).
62. Salditt, A. *et al.* Thermal Habitat for RNA Amplification and Accumulation. *Phys. Rev. Lett.* **125**, 048104 (2020).
63. O'Sullivan, C. T. Newton's law of cooling—A critical assessment. *American Journal of Physics* **58**, 956–960 (1990).
64. Connell, J. H. Diversity in Tropical Rain Forests and Coral Reefs. *Science* **199**, 1302–1310 (1978).
65. Kohlberger, M. & Gadermaier, G. SELEX: Critical factors and optimization strategies for successful aptamer selection. *Biotechnology and Applied Biochemistry* **n/a**,.
66. Plückthun, A., Schaffitzel, C., Hanes, J. & Jermutus, L. In vitro selection and evolution of proteins. in *Advances in Protein Chemistry* vol. 55 367–403 (Academic Press, 2001).
67. Nei, M., Maruyama, T. & Chakraborty, R. The Bottleneck Effect and Genetic Variability in Populations. *Evolution* **29**, 1–10 (1975).
68. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Research* **31**, 3429–3431 (2003).
69. Cline, D. B. On the physical origin of the homochirality of life. *European Review* **13**, 49–59 (2005).
70. Bernstein, M. P., Dworkin, J. P., Sandford, S. A., Cooper, G. W. & Allamandola, L. J. Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues. *Nature* **416**, 401–403 (2002).
71. Schwieterman, E. W. *et al.* Exoplanet Biosignatures: A Review of Remotely Detectable

Signs of Life. *Astrobiology* **18**, 663–708 (2018).

72. Cronin, J. R. & Pizzarello, S. Amino acid enantiomer excesses in meteorites: Origin and significance. *Advances in Space Research* **23**, 293–299 (1999).
73. Soares, T. A., Lins, R. D., Longo, R., Garratt, R. & Ferreira, R. Plural Origins of Molecular Homochirality in Our Biota Part II. The Relative Stabilities of Homochiral and Mixed Oligoribotides and Peptides. *Zeitschrift für Naturforschung C* **52**, 89–96 (1997).
74. Blackmond, D. G. The Origin of Biological Homochirality. *Cold Spring Harb Perspect Biol* **11**, a032540 (2019).
75. Tamura, K. Origin of amino acid homochirality: Relationship with the RNA world and origin of tRNA aminoacylation. *Biosystems* **92**, 91–98 (2008).
76. Chela‐Flores, J. The origin of chirality in protein amino acids. *Chirality* **6**, 165–168 (1994).
77. Noorduin, W. L. *et al.* Complete chiral symmetry breaking of an amino acid derivative directed by circularly polarized light. *Nature Chemistry* **1**, 729–732 (2009).
78. Woese, C. R. The Present Status of the Genetic Code. in *Progress in Nucleic Acid Research and Molecular Biology* (eds. Davidson, J. N. & Cohn, W. E.) vol. 7 107–172 (Academic Press, 1967).
79. Hein, J. E., Tse, E. & Blackmond, D. G. A route to enantiopure RNA precursors from nearly racemic starting materials. *Nat Chem* **3**, 704–706 (2011).
80. Yu, J., Jones, A. X., Legnani, L. & Blackmond, D. G. Prebiotic access to enantioenriched glyceraldehyde mediated by peptides. *Chem Sci* **12**, 6350–6354 (2021).
81. Murakami, H., Saito, H. & Suga, H. A Versatile tRNA Aminoacylation Catalyst Based

on RNA. *Chemistry & Biology* **10**, 655–662 (2003).

82. Suga, H., Hayashi, G. & Terasaka, N. The RNA origin of transfer RNA aminoacylation and beyond. *Philos Trans R Soc Lond B Biol Sci* **366**, 2959–2964 (2011).
83. Tamura, K. & Schimmel, P. Chiral-Selective Aminoacylation of an RNA Minihelix. *Science* **305**, 1253–1253 (2004).
84. Tamura, K. & Schimmel, P. R. Chiral-selective aminoacylation of an RNA minihelix: Mechanistic features and chiral suppression. *PNAS* **103**, 13750–13752 (2006).
85. Danger, G. *et al.* 5(4H)-Oxazolones as Intermediates in the Carbodiimide- and Cyanamide-Promoted Peptide Activations in Aqueous Solution. *Angewandte Chemie International Edition* **52**, 611–614 (2013).
86. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat Protoc* **6**, 779–790 (2011).
87. Saito, H., Kourouklis, D. & Suga, H. An in vitro evolved precursor tRNA with aminoacylation activity. *The EMBO Journal* **20**, 1797–1806 (2001).
88. Ahmed, Y. L. & Ficner, R. RNA synthesis and purification for structural studies. *RNA Biology* **11**, 427–432 (2014).
89. Barfod, E. T. & Cech, T. R. Deletion of nonconserved helices near the 3' end of the rRNA intron of Tetrahymena thermophila alters self-splicing but not core catalytic activity. *Genes Dev.* **2**, 652–663 (1988).
90. Rio, D. C., Ares, M., Hannon, G. J. & Nilsen, T. W. Ethanol Precipitation of RNA and the Use of Carriers. *Cold Spring Harb Protoc* **2010**, pdb.prot5440 (2010).
91. Evans, M. E., Clark, W. C., Zheng, G. & Pan, T. Determination of tRNA aminoacylation levels by high-throughput sequencing. *Nucleic Acids Research* **45**, e133–e133 (2017).

92. Simkus, D. N. *et al.* Methodologies for Analyzing Soluble Organic Compounds in Extraterrestrial Samples: Amino Acids, Amines, Monocarboxylic Acids, Aldehydes, and Ketones. *Life* **9**, 47 (2019).

93. Glavin, D. P., Callahan, M. P., Dworkin, J. P. & Elsila, J. E. The effects of parent body processes on amino acids in carbonaceous chondrites. *Meteoritics & Planetary Science* **45**, 1948–1972 (2010).

94. Burton, A. S., Stern, J. C., Elsila, J. E., Glavin, D. P. & Dworkin, J. P. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem. Soc. Rev.* **41**, 5459–5472 (2012).

95. Beaufils, D., Danger, G., Boiteau, L., Rossi, J.-C. & Pascal, R. Diastereoselectivity in prebiotically relevant 5(4 H)-oxazolone-mediated peptide couplings. *Chemical Communications* **50**, 3100–3102 (2014).

96. Takayama, T. *et al.* A novel approach for LC-MS/MS-based chiral metabolomics fingerprinting and chiral metabolomics extraction using a pair of enantiomers of chiral derivatization reagents. *Analytica Chimica Acta* **898**, 73–84 (2015).

97. Tanwar, S. & Bhushan, R. Enantioresolution of Amino Acids: A Decade's Perspective, Prospects and Challenges. *Chromatographia* **78**, 1113–1134 (2015).

98. Joyce, G. F. Non-enzymatic template-directed synthesis of RNA copolymers. *Origins Life Evol Biosphere* **14**, 613–620 (1984).

99. Acevedo, A., Brodsky, L. & Andino, R. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature* **505**, 686–690 (2014).

100. Luisi, P. L., Ferri, F. & Stano, P. Approaches to semi-synthetic minimal cells: a review. *Naturwissenschaften* **93**, 1–13 (2006).

101. Aileen Funke, S. & Willbold, D. Mirror image phage display—a method to generate d - peptide ligands for use in diagnostic or therapeutical applications. *Molecular BioSystems* **5**, 783–786 (2009).