In vitro selections for imidazole-modified RNA catalysts

by

Cara Lena Florance
B.S., Iona College, 2008

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Chemistry and Biochemistry
2014

The thesis entitled:
In vitro selections for imidazole-modified RNA catalysts
written by Cara L. Florance
has been approved for the Department of Chemistry and Biochemistry.

_____
Dr. Bruce Eaton

_____
Dr. Daniel Feldheim

_____
Dr. Robert Kuchta

_____
Dr. Michael Stowell

_____
Dr. Hubert Yin

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

1. ABSTRACT

Florance, Cara Lena (Ph.D., Biochemistry)

In vitro selections for imidazole-modified RNA catalysts

Thesis directed by Professor Bruce Eaton

In vitro selection from a random RNA library allows for the identification of single sequences capable of performing a desired task. This process has yielded RNA sequences that function as specific and efficient ligands and catalysts for many targets and reactions. This thesis details the development and application of advanced methods for in vitro selection of complex RNA catalysts. Imidazole modifications were used to impart additional functionality to the RNA scaffold in the experiments described herein. The first chapter details a new type of selection that can be used to identify aptamer-like sequences that can not only bind a protein, but also catalyze the addition of a tethered reactive molecule to target residues, forming a covalently bound RNA/protein complex. These sequences are termed reactamers. This method has isolated a sequence that forms a target protein-dependent, denaturing-resistant complex. The presence of a covalent bond, however, has yet to be confirmed. Recommended optimization strategies for future reactamer selections are discussed. The second chapter describes a selection to identify RNA capable of cleaving an amide bond. This experiment did not yield catalysts with the intended activity, however an improved scheme prompted by results during the selection is described. The last chapter details experiments performed on a sequence that can self-circularize, discovered during the

amide cleavage selection. In addition, it appears this sequence forms a four-way junction in the middle of the circle, creating a pinched figure-eight-like structure. It is hypothesized that this junction is a phosphotriester bond. Together, these three reports highlight not only the power of *in vitro* selections, but also the catalytic versatility of modified RNA.

CONTENTS

1. Abstract iii

2. Introduction 1

3. Reactamer Selection 6

3.1 Materials and Methods 33

4. RNA protease Selection 41

4.1 Materials and Methods 55

5. Non-linear RNA 61

5.1 Materials and Methods 82

LIST OF TABLES

3.1 Selection Progression Table 1 18

3.2 Selection Progression Table 2 19

5.1 Circle Isolates 66

LIST OF FIGURES

2.1 Imidazole Uridine Triphosphate 2

3.1 Kinetic diagram of reactamer binding and catalysis 7
3.2 Crystal Structure of SOD1 G93A 11
3.3 α-methacrylamide construct 13
3.4 An acid-catalyzed reaction mechanism for SOD1-RNA adduct formation . . 14
3.5 A base-catalyzed reaction mechanism for SOD1-RNA adduct formation . . . 15
3.6 Reactamer Selection Scheme 20
3.7 Alternate protein-RNA conjugation chemistries 22
3.8 Photocleavable linker to select for methacrylamide usage 23
3.9 PCR curves showing enrichment of SOD and methacrylamide dependent RNA 25
3.10 PCR curves showing enrichment of UV light-dependent sequences 25
3.11 SOD-dependent RNA gel shift 27
3.12 PCR curves showing final round's selection sample and controls 29

4.1 RNA as a serine protease mimic 43
4.2 Tripeptide construct 45
4.3 Original selection scheme 47
4.4 Gel mobility comparison of the 12mer-ligated product and circularized RNA 48
4.5 Modified selection scheme 50
4.6 Recommended selection scheme 53

5.1 Hypothesized model of RNA circle formation by T4 DNA Ligase 64
5.2 Circle formation catalyzed by T4 DNA Ligase 65
5.3 Gel of initial discovery of self-circularizing RNA 68
5.4 Magnesium inhibition and Imidazole-UTP dependence 70
5.5 pH profile of c6 catalysis 71
5.6 Figure-eight structure hypothesis development 72
5.7 Loop-dependent RT-PCR experiment explanation and results gel 74
5.8 Loop-dependent RT-PCR experiment description 74
5.9 Loop-dependent RT-PCR experiment description 75
5.10 Loop-dependent RT-PCR experiment description 75
5.11 Loop-dependent RT-PCR experiment description 76
5.12 A structure of a phosphotriester bond 79
5.13 A structure of a phosphorimidazolide bond to a backbone phosphate 80

2. INTRODUCTION

In the 1980s, Cech and Altman independently discovered that ribonucleic acid (RNA) can catalyze reactions.[43,63] This finding was pivotal to understanding the role of RNA in present and ancient biology; however, through a process called *in vitro* selection, researchers soon found that RNA is capable of performing an incredible scope of reactions including high affinity binding and catalysis that currently includes the Diels-Alder reaction[67], Michael addition[60] and aldol[23] and amide bond formation.[78]

The four nucleotides- adenosine, cytidine, guanosine, and uridine- polymerized in random succession make up the sequence of an RNA molecule. This sequence can reproducibly fold into a three-dimensional structure through interactions between the nucleotides. This structure can then bind other molecules discriminatingly or act as a catalyst. Discovery of RNA sequences that can bind to a specific target or catalyze a reaction, however, is quite empirical and due to the vast amount of sequence possibilities, must be attempted through combinatorial means. *In vitro* selections allow the identification of nucleic acids with desired functions, culled from a starting library containing over 10^{14} unique sequences.

To capitalize on the functional abilities of RNA, this iterative, cyclic process takes advantage of RNA's information storage capacity, allowing polymerase enzymes to amplify sequences which successfully complete the required task. The output of one selection cycle is then used as the input for the next cycle, thereby enriching the pool with active sequences. Once the pool is sufficiently enriched, the winning RNA molecules are sequenced and the individual sequences, called isolates, are studied for activity.

Nucleic acids have an inherent disadvantage as catalysts compared to proteins when

considering the number and variety of functional groups. This discrepancy may explain why proteins perform the majority of catalytic tasks in current biology. While proteins possess several types of hydrophobic, acidic, nucleophilic and other types of side chains, most nucleic acids are limited to modestly decorated purines and pyrimidines. These allow for functional traits such as hydrogen bonding, metal coordination and π-stacking, but are limited with respect to acid/base catalysis, nucleophillicity, and positively charged residues. Our lab incorporates modifications into the sequence during transcription to supply a nucleic acid library with more useful functional groups. The modification is typically through an amide linkage to C5 of uridine.[19,66,73]

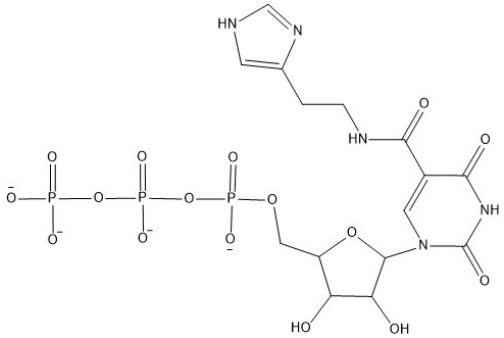


Fig. 2.1: Imidazole Uridine Triphosphate

In all of the following experiments, an imidazole-modified uridine was used (Fig. 2.1). This molecule can not only participate in acid-base catalysis at neutral pH, which was required in the catalytic mechanisms we projected to occur during the following selections, but can also provide additional metal-coordination sites, substrate recognition and structural functions. T7 RNA polymerase incorporates this modification into the growing strand with minimal bias and very little loss of yield.[74]

In selections, modifications may facilitate the discovery of active sequences. Not only can they impart functionality that was absent in the native nucleic acid, but they can also decrease the sequence space necessary for a specific reaction by replacing a non-ideal innate functional group, therefore allowing an increased number of structural presentations of an

active motif in a finite library. For example, in the HDV antigenomic ribozyme, a cytosine's N^3 may undergo significant pKa perturbation to act as the base catalyst for 2'-hydroxyl attack of the phosphate backbone, then act as the acid catalyst to protonate the leaving 5'-hydroxy anion during the phosphoester cleavage reaction. A non-bridging phosphate oxygen is responsible for the increased pKa of the cytosine's N^3 through hydrogen bonding with the exocyclic N^4 and subsequent imino tautomer formation.[21,54] The necessary sequence space to position this oxygen is not insignificant. If the RNA instead could have evolved with a general acid-base catalyst with a near-neutral pKa incorporated into the strand, such as imidazole, the structure and therefore sequence space required to perturb the catalytic nitrogen's pKa would not have been needed. Strikingly, mutation of the HDV ribozyme's active cytosine to uracil, which inactivates the catalyst, can be rescued with imidazole buffer.[54] In addition, most nucleic acid catalysts require divalent cations, typically magnesium, for appreciable activity. But, if physiological applications are desired, this dependence is often detrimental to the their rate of reaction, given the low concentration of cellular magnesium.[69] Precedence suggests outfitting nucleic acids with modifications to serve in place of the divalent cations, as a Lewis acid for example, can circumvent this trend and endow catalytic activity where none was possible in the unmodified form.[33] In the case of selections, the chances of identifying an active sequence should greatly increase as the necessary motif size decreases. In fact, several demanding selections have succeeded using modifications, but failed using a native library.[67,78]

Interestingly, naturally modified nucleotides can be found in modern biology in almost all classes of RNA. These conserved modifications are added post transcriptionally by dedicated enzymes and are thought to reduce conformational dynamics,[2] reduce the number of alternate folding pathways,[32] and to discriminatingly decode codons in tRNA.[48] As the pseudouridine modification may have direct catalytic roles in biological RNA,[3,28,51,72] one cannot help but wonder if the modifications are relics of reactive ancestors.

The process of *in vitro* selection, with and without modifications, has discovered thousands of nucleic acids that can tightly and specifically bind proteins, some of which have been used as therapeutics[52] or in biomarker arrays.[26] In addition, the field of catalytic synthetic nucleic acids has produced a fascinating and exciting repertoire of reactions as a result of this powerful tool. This thesis further explores the versatility of RNA through selections for activity towards several challenging reactions.

Though there are many examples of RNA-catalyzed small molecule conjugation chemistry, we sought to identify sequences that can perform a coupling reaction between an exogenous small molecule and a protein. This blends the themes of specific reactivity seen in catalytic RNA and target recognition seen in aptamers. To describe the sequences resulting from this combination of fields, the term reactamer was coined. The first chapter describes a proof of concept selection which outlines the first procedure tested for identification of these catalysts. Specifically, imidazole-modified RNA were challenged to catalyze the addition of a tethered methacrylamide moiety to a mutant of the protein Superoxide Dismutase 1 (SOD1). Seminal selections such as these, where the RNA are tasked to perform multiple and difficult reactions can fail to yield highly active catalysts without substantial optimization. Though a sequence from this selection has been identified to form a SOD-dependent gel-shifted complex in denaturing conditions, it is not known yet if a covalent bond is being formed. In and of itself, the developed protocol can serve as a foundation on which to base further reactamer selections.

The second chapter recounts a selection preformed in collaboration with Thermo Fisher Scientific to isolate imidazole-modified RNA that can cleave a tethered tripeptide. This was also a proof of concept selection, with the main goal of identifying sequences that could discriminatingly cleave phosphorylated proteins. This experiment unfortunately did not result in active RNA catalysts; however, it resulted in valuable lessons and techniques that would prove useful if this desirable reaction was attempted again. Cleavage of an amide

bond by RNA is a sought after reaction and several other labs have unsuccessfully tried using unmodified RNA.[4,10,17] In addition to a detailed account of the techniques developed in this selection protocol, this chapter also discusses the projected benefits of further nucleotide modifications for this challenging reaction.

The third chapter reports an unintended yet fascinating byproduct of the RNA-protease selection, RNA that can self-circularize and also possibly form a phosphotriester bond, creating a figure-eight-like secondary structure. In addition, these sequences can be ligated with T4 DNA ligase with good efficiency to form true circles. As the circular RNA field has blossomed in the past year, a simple, high yield synthetic route towards circle ladder markers, standards or even mechanistic insight to their biological formation would be useful to the field.

3. METHOD AND APPLICATION OF THE SELECTION OF RNA SEQUENCES THAT CAN CATALYZE THE COVALENT ADDITION OF A TETHERED REACTIVE MOLECULE TO A TARGET PROTEIN

RNA have demonstrated the ability to perform many conjugation reactions between small molecules in solution. Expanding on this concept, two selections from different labs have successfully identified RNA that can form a covalent bond to a protein. Gold and colleagues used a 5-iodouracil modification to crosslink bound RNA to the HIV-1 Rev protein upon UV irradiation. Iterative cycles of this produced not only high affinity binders, but also several sequences that could form a complex with the protein without UV irradiation that could withstand denaturing gel conditions.[36] Baskerville and Bartel identified unmodified RNA that could form a covalent 5'phosphoamide bond with the amino terminus of the BIV-1 Tat peptide.[7] Both of these selections began with a biased pool of RNA that had an affinity for the target protein. We sought to identify RNA from a completely random library that could not only bind a protein, but catalyze the covalent addition of a tethered reactive group (Fig. 3.1).

A molecule that forms a covalent bond to a specific target has great use as a therapeutic. Covalent inhibitors make up a small but powerful class of drugs, which include aspirin, omeprazole (Prilosec), and penicillin. Many take advantage of the catalytic mechanism of an enzyme to form a covalent bond, while others target general nucleophilic residues and rely on non-covalent interaction with the surrounding protein to properly align the reactive groups. This alignment which forms the covalent bond contributes to the specificity of the

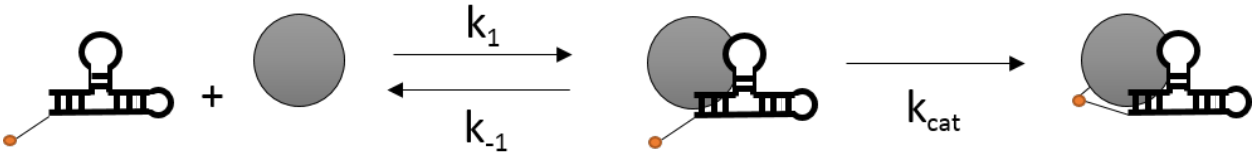


Fig. 3.1: Kinetic diagram of reactamer binding and catalysis. RNA and the target protein bind to form a RNA-protein complex. The rate is expressed as the term k_1. The disassociation rate of the complex is expressed as the constant k_{-1}. Once bound, the RNA can catalyze the addition of the tethered reactive group to the protein forming a covalent bond, expressed by the rate constant k_{cat}. Though not pictured, the reversal of some types of covalent bonds may not be negligible in all chemical environments, and therefore should be considered. The turnover rate in which the organism replaces the targeted protein as a whole may also be considered a factor in this component.[62]

reaction as unintended binding sites for the drug may not have a similarly placed nucleophile and therefore would not form a covalent bond. However, due to the irreversible nature of the bond, off target effects are highly undesired. Due to this, despite their prevalence among top selling drugs, the search for new covalent therapeutics is not proportional to their predecessors' effectiveness.[62] The hesitancy is not without warrant as non-specific reactions could have grave consequences, such as inhibition of vital pathways or creation of an antigenic site on benign proteins, called haptenization. As with all new drugs, these factors would need to be assessed for the individual lead, however the ability to select for or against these types of interactions without knowledge of the mechanism involved would be a powerful tool in drug discovery and optimization. During an *in vitro* selection experiment, a nucleic acid is selected based on it's ability to perform binding or catalysis. Unless a bias is added, there is no force dictating how these tasks are performed and the results are not limited to a compound library or a decorated substrate analog. In addition, successful binders can be reselected to avoid unwanted or dangerous off-target interactions, tolerate modifications that improve pharmacokinetics or degradation, or even overcome a drug resistant target mutation, for example.

Currently, many drug discovery efforts begin by screening a large library of compounds

for the ability to affect a target, either as an antagonist or agonist depending on the desired phenotype. Once molecules are identified, they can be further tested for unwanted interactions and chemically modified to fine tune their activity. Typical compound libraries are limited in terms of chemical diversity, when compared to vast amount of chemical space possible, due to experimental, financial, and storage restrictions. Researchers, however, can circumvent this limitation by selecting or synthesizing libraries based on known substrate mimics of their target, utilizing *in silico* structure docking programs to identify appropriate scaffolds, or using libraries with chemical subspace similar to previously identified biologically active molecules.[44,77] High throughput screening is typically used to assay each separate compound. However, it is costly to establish and run the automation and informatics systems necessary to perform the thousands, sometimes millions,[1] of assays required to test a library. Recent advances have continued to make this process faster and less expensive per run,[1] however one pot libraries with identifying sequences as the active compound, such as *in vitro* selection pools, relieve the need (and associated cost) for compartmentalizing each library component. Modified nucleic acids can reproducibly fold into three dimensional scaffolds and present functional groups that can interact with a target molecule. A facilely synthesized modified library can contain upwards of 10^{15} nucleic acids each offering a unique representation of the available chemical space. In addition, based on the proven ability of nucleic acids to catalyze reactions, the library can be appended with reactive groups that facilitate covalent modification of the target.

The *in vitro* selection experiment described herein was designed to discover RNA that can covalently attach a methacrylamide to a protein. The RNA has the ability to be both the binder and the identification code, which greatly simplifies the creation of the library and removes the need for typical small molecule deconvolution strategies. Like several covalent therapeutics,[62,75] an acrylamide derivative will be used as the electrophile to form the covalent bond with a target nucleophile. The RNAs take the place of the small molecule scaffold

and supply the recognition and electrophile orientation components. Since the RNAs are being tasked to both bind a target and form a covalent bond, the portmanteau "reactamer" was coined to describe a "reactive aptamer." A mutant of the protein Superoxide Dismutase 1 was chosen as the proof of concept target for this type of selection. This protein is implicated in causing a subset of amyotrophic lateral sclerosis cases, a hallmark of which is high molecular weight oligomers and aggregates of the mutant protein. Mouse model studies have shown that mutant SOD-directed antibody therapy may block the formation of the oligomers and delay disease progression.[27] Reactamers towards this target may confer the same benefit. By testing reactamer efficacy in the same mouse models and cell lines that the antibody therapies were tested in, this field offered a method to compare reactamers to another high molecular weight therapeutic in it's initial stages to begin to realize the benefits and shortcomings of this new technology.

Superoxide dismutase 1 (SOD1) is a vital cytosolic obligate-dimer metalloprotein which catalyzes the dismutation of the superoxide anion into oxygen and hydrogen peroxide. Mutants of SOD1 are implicated in causing amyotrophic lateral sclerosis (ALS), more widely known as Lou Gehrig's Disease.[58] The first symptoms of the disease, with average onset occuring at age 50-60, are muscle weakness and failure caused by progressive motor neuron death. The patient experiences gradual paralysis and 3-5 years after the onset of symptoms, passes typically from respiratory failure. Approximately 10% of ALS cases are inherited, and about 20% (2% total) are caused by mutant SOD1. Over 150 mutations in SOD1 have been identified in ALS patients. Though it is not known how mSOD1 causes the disease, it is believed to be due to a gain of function in only the affected cells, despite ubiquitous expression in the patient's tissues.[8,50,61] As with many neurodegenerative diseases, late stage ALS patients' neurons contain highly ubiquinated aggregates of the mutant protein. Though it is now known whether these deposits are contributing to the disease or a result of it, several in vivo and mouse model studies concerning antibody therapy targeted at the pre-aggregated

protein have resulted in blocking aggregation and delaying onset of symptoms, respectively, the latter possibly by blocking aggregation or other aberrant interactions.[25,27] This disease model offers a way to test and compare reactamer efficacy after proven *in vitro*.

SOD1 G93A as proof of concept target: The mutant SOD1 G93A was chosen as the proof of concept for this project. In addition to its status as a therapeutic[27,34] or diagnostic target[79], availability of mouse models,[18] and numerous reports concerning this mutant in the literature, SOD1 G93A has several advantageous characteristics with respect to this type of selection which would help shift the burden of proof on the RNA and are described below.

Stability: WT SOD1 is an incredibly stable protein, having a melting temperature near 95^oC. G93A is the most stable mutant, and can withstand long incubation times at 37^oC.[64] As the first round of a selection is essentially 10^{14} single molecule experiments, considering almost every RNA present is a unique sequence, the stability of the target is critical as it allows the RNA longer incubation times with a uniform target population. In order for a sequence to be selected, it must not only encounter the appropriate target residue on the protein, but also be sampling the necessary kinetic energy that allows for catalysis to occur. Since in this first round, one cannot rely on the average activity of the ensemble to carry a single sequence forward, we tend to use long ($\sim$ 20 hour) incubation times to ensure each RNA molecule can collide with the target while also having the time to sample the necessary energy to perform a reaction, if capable. As winning isolates grow in number due to the amplification steps, incubation times can then be dropped for use as a selection pressure for increased rates of reaction, where if only a fraction of a certain population meet these criteria they will still be carried forward. As a new selection protocol under development, removing the concern of target stability proved beneficial.

Free Thiols as Michael Donors: SOD1 contains four reduced cysteines which are the most reactive natural amino acid residue towards α-β-unsaturated carbonyls (Fig.3.2). Two on each SOD monomer, these thiols can be modified by maleimides in solution,[6] suggesting

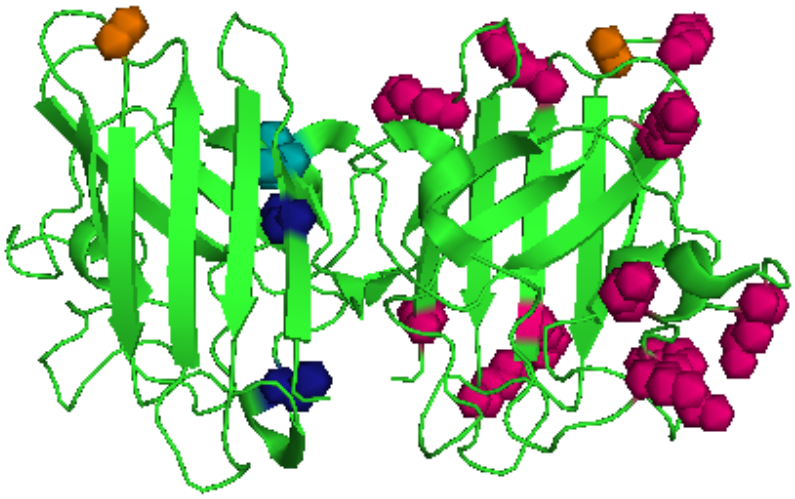


Fig. 3.2: A crystal structure of SOD1 G93A [24] highlighting the most probable target residues. Free cysteines, the most reactive amino acid towards α-β-unsaturated carbonyls, are colored in dark blue on the left monomer, while the two cysteines in a disulfide bond are colored in teal. On the right monomer, lysines are colored in pink. The site of the G93A mutation, though not a likely target residue, is highlighted in orange.(Image created in PyMol: PDB 3GZO)

the RNA should be able to access the groups with the tethered methacrylamide. The dimer also contains twenty two surface exposed lysines, which although less reactive nucleophiles than thiols towards the Michael acceptor, are another possible methacrylamide adduct site.

RNA reactive functional group choice: The α-methacrylamide is available commercially under the name *Acrydite*^{TM} and is attached to the 5' phosphate of an oligo through a hexane linker (Fig. 3.3). Though still an activated double bond, the methyl on the α-carbon decreases its reactivity due to steric hindrance. As a use for these types of catalytic RNA is in diagnostic assays or therapeutics, a high rate of nonspecific reactivity is unacceptable.

Several attempts to form an adduct with reduced glutathione failed, which suggested this compound has a low uncatalyzed reaction rate. This is in agreement with the literature.[14,47] In addition, acrylamide sustituents are common in small molecule covalent inhibitors of the EGFR family, and have lower allergenic and toxic effects than their more reactive counterparts.[38] Therefore, we initially chose the methacrylamide construct, as opposed to a more reactive derivative.

The short linker between the methacrylamide and DNA may limit the pool to structures that can position the 5' end of the nucleic acid near it's active site. In past selections in our lab, a longer hydrophilic PEG_{2000} polymer was used to attach functional groups to the 5' end. This allows the group to access active sites distant from its tether point thereby putting less constraints on the finite pool. This practice was forgone in favor of the commercially available construct so immediate focus could be placed on the selection scheme itself as opposed to the synthesis of the new 12mer.

RNA modification choice: The RNA in this selection contained two modifications that differ from natural RNA. 2'-Fluoro-2'-deoxycytidine was used to impart nuclease resistance if the RNA was to be introduced to biological fluids. 5'-Imidazole-uridine was incorporated to facilitate acid-base catalysis at neutral pH. Envisioned mechanisms can utilize the imidazole as a Lewis acid (Fig.3.4) where the imidazolium interacts with the carbonyl oxygen, making the double bond a more potent electrophile. Alternately, the imidazole can act as a base (Fig.3.5) to deprotonate the cysteine, making it a more potent nucleophile. Therefore, appending the library with this group to specifically position for one or both of these functions should increase the chances of identifying a catalyst for the overall reaction.

Pre-Selection controls: The initial selection scheme involved a gel partitioning step to purify the RNA/SOD covalent complex, however a sample of complex could not be isolated. Attempts included 24 hour incubation at 37^oC pH 9 to create the active thiolate species with a high μM concentration of methacrylamide-12mer. In addition, several catalysts were

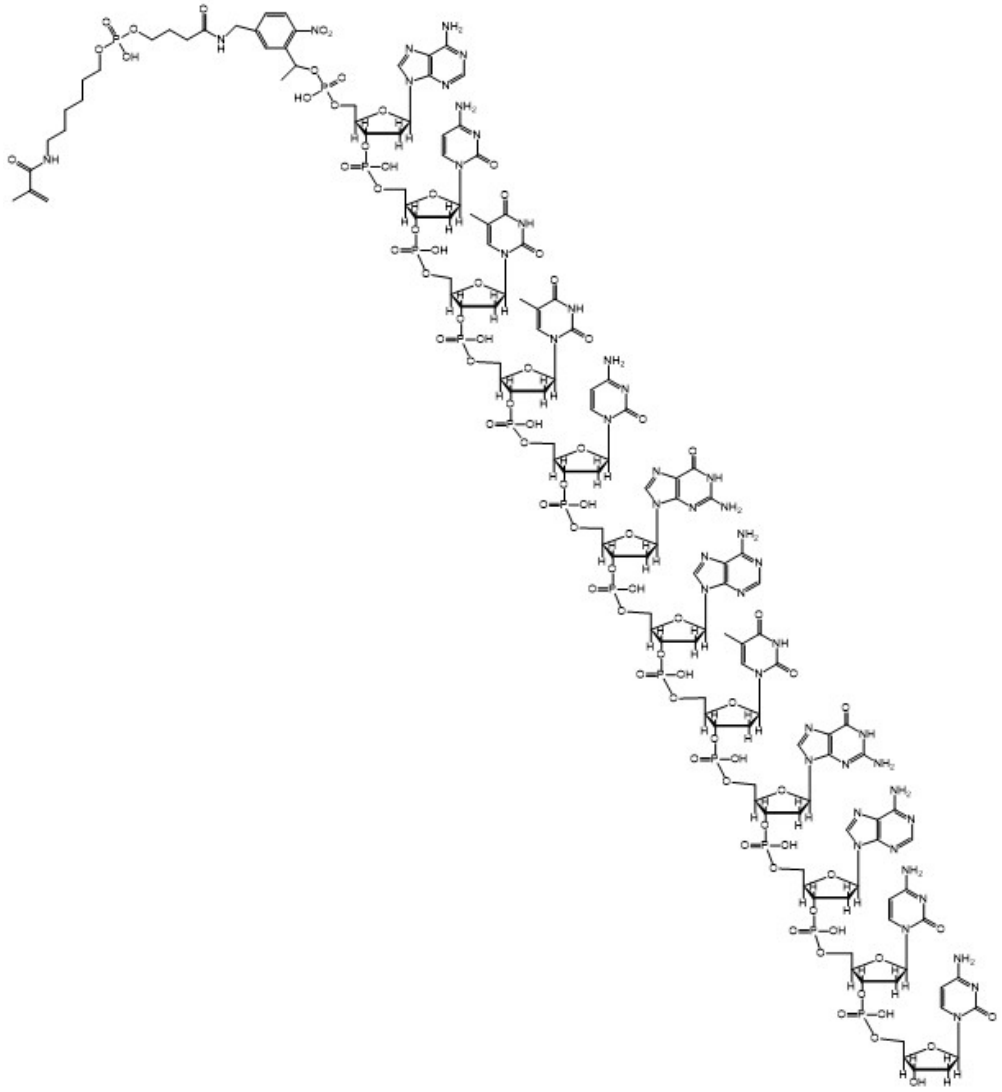


Fig. 3.3: The commercially available *Acrydite*^{TM} seen here for scale attached to the DNA 12mer with a photocleavable linker. In the selection, this functionalized 12mer is ligated to the 5' end of the RNA. This construct is referred to herein as the α-methacrylamide moiety.

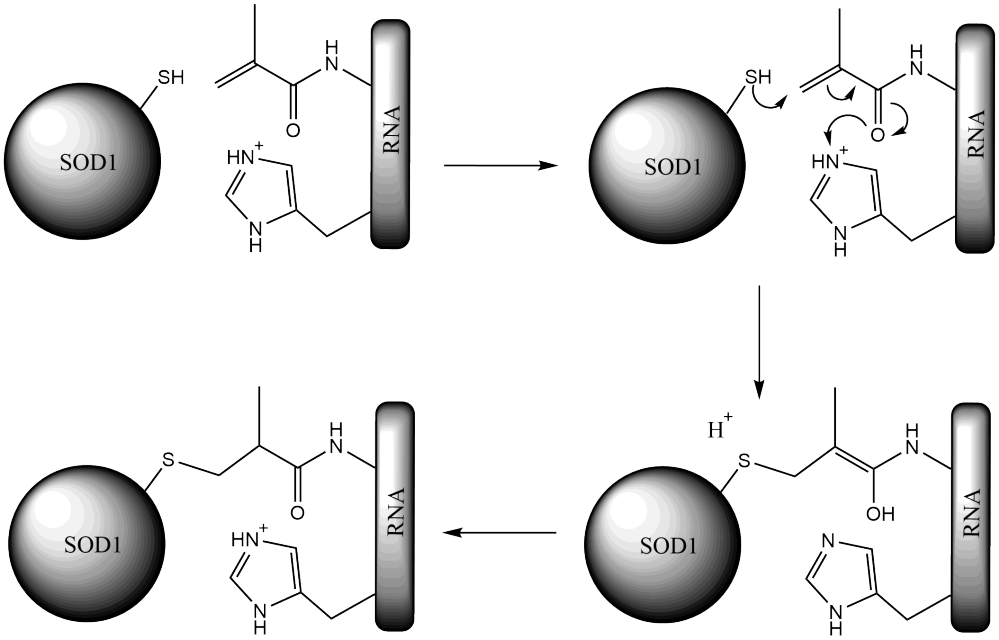


Fig. 3.4: A possible acid-catalyzed reaction mechanism for SOD1-RNA adduct formation. The protonated imidazolium cation acts as a lewis acid, activating the methacrylamide. This increases the electrophilicity of the C=C double bond, making it prone for attack by a protein nucleophile.

introduced, such as sulphonated triphenylphosphine, n-butylamine, and diethylamine.[47] The controls were analyzed by denaturing polyacrylamide gels, but no adducts were seen. This is likely due to the low reactivity of the methacrylamide combined with the sterically crowded location of the accessible cysteines. Though a different selection route had to be devised, these controls suggests the uncatalyzed reaction between the methacrylamide and the protein during the selection would be negligible.

During the course of these controls, it was found that 1 uM RNA would form soluble aggregates when incubated with greater than 2 uM SOD1 G93A. These aggregates could be visualized on a 6% denaturing polyacrylamide gel with SYBR Green staining. These species were independent of RNA modification. Aggregates were not visible after heating for 5 minutes at 80^oC in 40%v/v formamide or by simply lowering the SOD1 G93A concentration. We did not discern whether SOD was present in the aggregates. Whether this is a general property of mutant SOD1 or an artifact of the conditions is not known; however, it was an

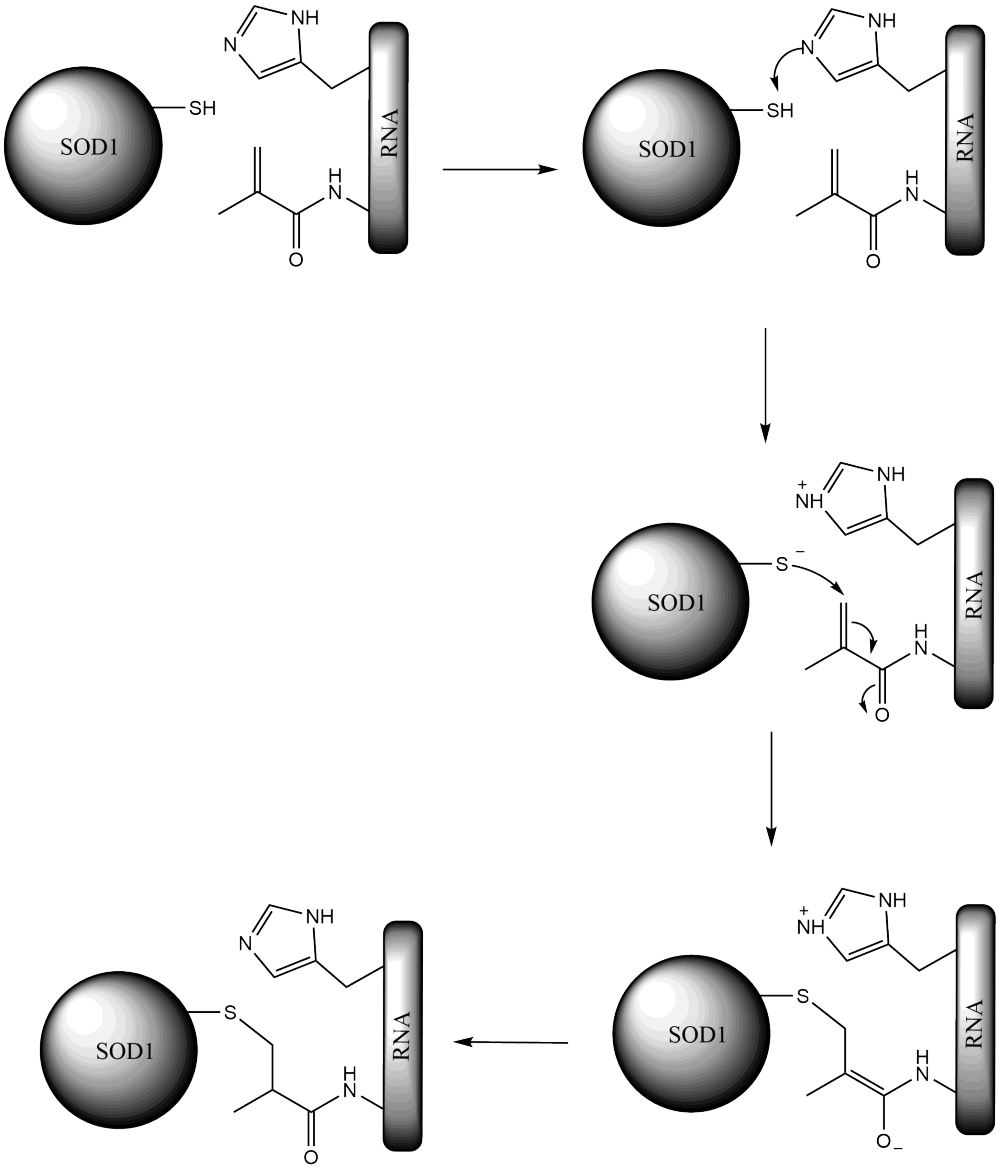


Fig. 3.5: A possible base catalyzed reaction mechanism for SOD1-RNA adduct formation. The imidazole modification on the RNA abstracts a proton from the free cysteine residue, forming the thiolate anion. Though not shown, the basicity of the imidazole could be increased by polarization of the ring by a backbone phosphate. The nucleophilic thiolate anion attacks the activated C=C bond tethered to the RNA, forming an enolate intermediate, and ending in formation of a covalent adduct.

unexpected result and may warrant further study considering soluble aggregates are thought to appear in early stage ALS.[61]

Necessity of biotin tag: This selection aims to identify reactamers, and therefore to separate these sequences from strictly high affinity binders, the physical partitioning method should be as strong as the desired bond. The target protein should be capable of binding a solid support to enable partitioning. Several long washing and enzymatic steps were used to select for covalent bonds to the protein, and if the protein-to-solid support bond has an off rate or can be destabilized, this would greatly reduce the stringency of those steps. We chose an amine reactive NHS-PEG_4-biotin to tag the protein, as the streptavidin-biotin bond has an incredibly stable interaction, a fast association rate, a femtomolar K_D, and a wide range of products to support altered protocols. This bond could also withstand the washing and enzymatic steps which the RNA-to-target bond also needed to survive. A disadvantage of this system, however, is that N-hydroxysuccinimide also reacts with an intended target amino acid of this selection, lysine. To decrease the negative effects of this issue, the biotin:SOD ratio was kept less than one using empirically determined reagent concentrations. The ratio was measured using a fluorescent biotin quantitation kit (Pierce) after excess biotin was removed with size exclusion chromatography. Although there is the chance there is a single most reactive lysine on SOD that was removed from the selection from this biotinylated protocol, it is likely there existed a distribution of biotin adduct sites.

Selection scheme and optimization: *Summary:* To begin each selection round, RNA attached to the methacrylamide moiety, termed maRNA, was diluted in selection buffer, thermally denatured then refolded using the protocol described in the methods section. Except where noted (Table 3.1), the maRNA concentration was 500 nM. Biotinylated SOD1 G93A was added to the maRNA and the sample was incubated at 37^oC. Streptavidin beads were used to isolate SOD and any bound maRNA. The beads were thoroughly washed to remove unbound maRNA. The transcripts were reverse transcribed, then the cDNA was

removed and subjected to PCR. An aliquot of the resulting dsDNA was used as the template for *in vitro* transcription. Those RNA sequences were ligated to the methacrylamide and purified before entry into the next round of selection (Fig.3.6). Over the course of the selection, various background populations appeared which required altering the selection scheme. The following is a description of the evolving experiment to select for reactamers.

Utilizing reverse transcription as a selection pressure: SuperScript III lacks RNaseH activity, unlike most reverse transcription enzymes, and therefore leaves the copied RNA strand intact to create a duplex with the nascent cDNA. In this selection, reverse transcription was not only used to prepare cDNA for amplification, but also as added selection pressure for a covalent bond. SuperScript is an incredibly processive polymerase,[70] and it is the formation and stability of this duplex that strips away RNA structure and aptamer-like interactions the maRNA may have with the target protein (Fig.3.6). In a traditional aptamer selection, this would remove the RNA from the target; however in this case, maRNA bound covalently will still be attached despite the removal of the RNA's base contacts. Therefore, strictly aptamer sequences, the presumed largest source of background, will not be favored in the selection.

For the first eleven rounds of the selection, reverse transcription was performed after the biotinylated SOD/RNA complexes were bound to the bead. Late in the selection, a control that did not include SOD1 G93A was shown to yield similar amounts of cDNA as the sample that did include the protein when measured by real time PCR. Analyses by gel electrophoresis confirmed that both samples contained full length PCR product. This indicated that sequences were in the pool that could bind to the beads and remain bound after reverse transcription to survive the selection steps. To decrease this population, reverse transcription was performed in solution *before* bead binding so that the folded active structure of the single stranded RNA would be abolished prior to exposure to the beads (Fig. 3.6).

Both of these reverse transcription approaches reduced background binding in this type

Round	[RNA]	[SOD]	Incubation Time	Gel partitioning	Bead Binding Time	RT vs. Bead Binding	cDNA Release Method
1	2 uM	500 nM	23 hours	no	40 min	after	heat
2	1 uM	500 nM	21.5 hours	no	30 min	after	heat
3	500 nM	250 nM	17.5 hours	no	40 min	after	heat
4	500 nM	250 nM	17 hours	no	40 min	after	heat
5	500 nM	200 nM	12 hours	no	30 min	after	heat
6	500 nM	200 nM	12 hours	no	30 min	after	heat
7	500 nM	100 nM	12.5 hours	no	30 min	after	heat
8	500 nM	200 nM	12 hours	no	15 min	after	PC/UV
9	500 nM	200 nM	18 hours	no	10 min	after	PC/UV
10	500 nM	200 nM	14 hours	no	15 min	after	PC/UV
11	500 nM	200 nM	10 hours	no	15 min	after	PC/UV
12	500 nM	200 nM	6.5 hours	no	6 min	before	PC/UV
13	500 nM	200 nM	5 hours 15 min	no	6 min	before	PC/UV
14	500 nM	200 nM	3 hours	no	6 min	before	PC/UV
15	500 nM	200 nM	1 hour	no	6 min	before	PC/UV
16	500 nM	200 nM	1 hour	no	6 min	before	PC/UV
17	500 nM	200 nM	20 min	no	6 min	before	PC/UV
18	500 nM	200 nM	2 hours	yes	10 min	after	PC/UV
19	500 nM	200 nM	1 hour	yes	10 min	after	PC/UV
20	500 nM	200 nM	1 hour	yes	10 min	after	PC/UV
21	500 nM	200 nM	20 min	yes	10 min	after	PC/UV
22	500 nM	200 nM	20 min	yes	10 min	after	PC/UV
23	500 nM	200 nM	20 min	yes	10 min	after	PC/UV
24	500 nM	200 nM	20 min	yes	10 min	after	PC/UV

Tab. 3.1: Selection Progression Table 1. This table describes the procedural specifications of each round. The "Incubation Time" column refers to the amount of time the RNA pool was incubated with SOD prior to partitioning. "Bead binding time" refers to the time the RNA/SOD pool was bound to the streptavidin beads. The time was reduced as it became evident RNA could react with the beads. The "RT vs Bead Binding" column refers to the order in which they occurred during that round. "After" denotes that reverse transcription occurred after bead binding. "Before" means that the RT reaction was in solution and the cDNA/SOD complexes were then bound to beads. In the last column, "heat" shows that the cDNA were melted from their bead-bound duplex, while PC/UV means the were photocleaved with UV light.

	Cycle of Amplification Detection								
Round	Selection Sample	(-) SOD	(-) MA	(-) light	(-) PC	Blank	+/- SOD difference	NB page	Notes
1	14	-	-	-	-	26	-	(8) 28-30	
2	12	-	-	-	-	>24	-	(8) 31-32	
3	17.5	23	-	-	-	26	5.5	(8) 38	
4	11	19	-	-	-	>24	8	(8) 41	Increased washing stringency
5	13.5	18.5	-	-	-	>23	5	(8) 47-48	
6	17.5	19.5	-	-	-	>24	2	(8) 68	Tween-20 added to washes
7	11.5	14	-	-	-	20	2.5	(8) 70	
8	18	22	-	-	-	23	4	(8) 74	PC-linker added
9	14.5	19.5	-	-	-	25	5	(8) 78	
10	15.5	19.5	-	-	-	>23	4	(8) 79	
11	16.5	21	-	-	-	26	4.5	(8) 80	
12	15.5	18.5	-	-	-	>24	3	(8) 90	RT before bead binding
13	14.5	20	-	-	-	>23	5.5	(8) 91	
14	14.5	20.5	-	-	-	>22	6	(8) 92	
15	14	18	-	19	-	>22	4	(8) 93	
16	12	20.5	19	18	-	21	8.5	(8) 94	
17	14.5	22	20	20	17	24.5	7.5	(8) 96	Preliminary Sequencing
18	19.5	26	-	-	-	26	6.5	(9) 4-5	Began gel partitioning
19	20.5	25	-	-	-	25	4.4	(9) 5	
20	17.5	26	-	-	-	26	8.5	(9) 6	
21	19.5	26	-	-	-	24	6.5	(9) 6	
22	17.5	26	-	-	-	26	8.5	(9) 7	
23	18.5	26	-	-	-	26	7.5	(9) 7	
24	18	24	24	24	-	24	6	(9) 12	Sequenced

Tab. 3.2: Selection Progression Table 2. This table shows the progress of the selection as analyzed by the cycle of PCR amplification. The selection sample refers to the (+)SOD, (+)MA, and where applicable (+)PC and (+)UV sample. (-)MA controls were RNA pools with the PEG_{18}-12mer ligated while the (-)PC control had the methacrylamide-12mer with no photocleavable linker. (-)light controls had the MA-PC-12mer ligated and were given SOD, but were not UV irradiated to free cDNA from the beads. The blank is all PCR reagents with no template sample added. +/-SOD difference shows the cycles of separation between those two samples, a larger number representing a larger dependence of the pool on SOD. "NB page" gives the notebook and page number(s) for each round for reference.

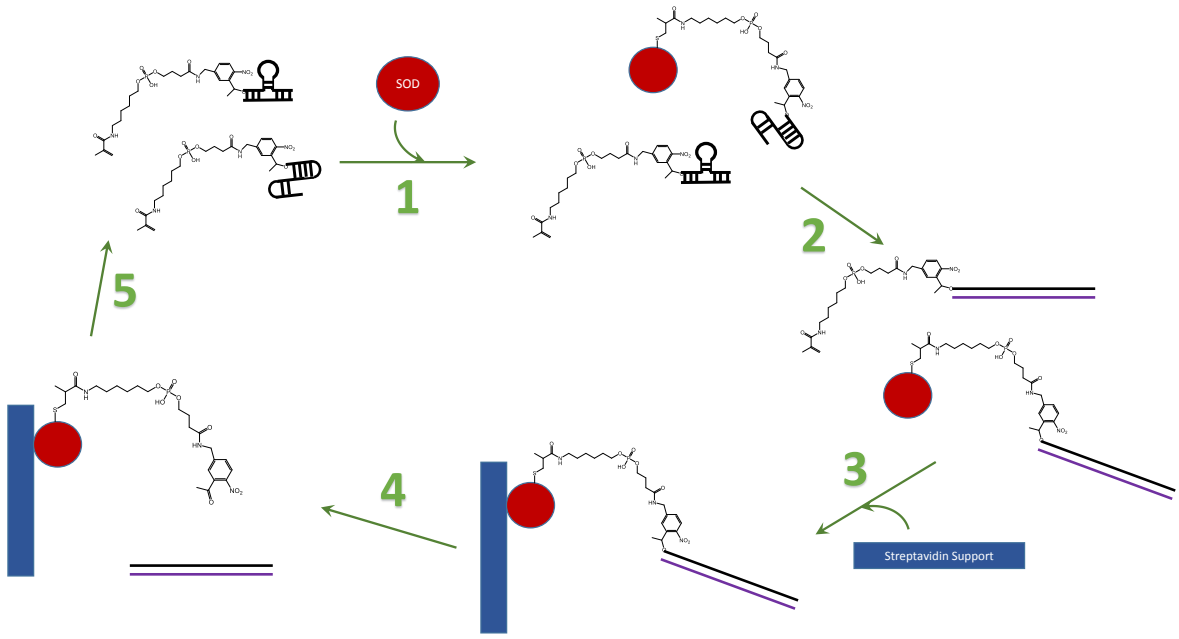


Fig. 3.6: A schematic of a reactamer selection cycle. Step 1: Biotinylated target protein is introduced to folded, ligated RNA and allowed to react. Step 2: RNA are reverse transcribed to inhibit reaction with the streptavidin solid support. Step 3: Biotinylated protein-RNA complex is bound to streptavidin and stringently washed. Step 4: Exposure to UV light cleaves the nitrobenzene linker and releases RNA bound to the protein by the methacrylamide. Step 5: RNA is PCR amplified, transcribed, ligated, then folded to enter the next round.

of selection. However further (-)SOD control experiments identical to that described above indicated enrichment of sequences that could now survive the new partitioning scheme. Incubating this reverse transcribed pool with biotin-saturated streptavidin beads (so as not to capture the RNA/SOD complexes) prior to incubation with ligand-free streptavidin beads greatly decreased this population presumably through active sequences reacting with the first exposure to the beads. However it did not remain effective after several rounds suggesting a new population arose that could overcome the added pressure. This would indicate that rotation between techniques every few rounds would help prevent method-specific background populations from arising.

In addition to the aforementioned controls for bead reactivity, other control samples were included that did not have the ligated methacrylamide. These were assayed by real time-PCR and gel analysis after partitioning, as described before, to compare relative amounts of cDNA surviving partitioning. Assuming unbiased reverse transcription and amplification of each sample, these showed that the pool was not utilizing the methacrylamide due to the result of the (+) and (-) methacrylamide samples having similar amounts of RNA being retained on the beads.

Using a photocleavable linker to select for use of the methacrylamide: Originally, the cDNA was being melted from the bead-bound RNA to gather it for PCR amplification. Though the maRNA pool was showing enrichment towards SOD binding when compared to a (-)SOD control, the removal of the methacrylamide did not affect the results, suggesting the maRNA were using an alternate Michael acceptor, different conjugation chemistry that did not require the methacrylamide, or some unknown unanticipated background (Fig.3.7).

We decided to insert a photocleavable linker between the methacrylamide and RNA to specifically release RNA/cDNA heteroduplexes that were bound by the methacrylamide (Fig.3.8). It must be noted that with the introduction of the photocleavable linker, a new 12mer was used in a successful attempt to increase ligation yield. The six 5' bases of the

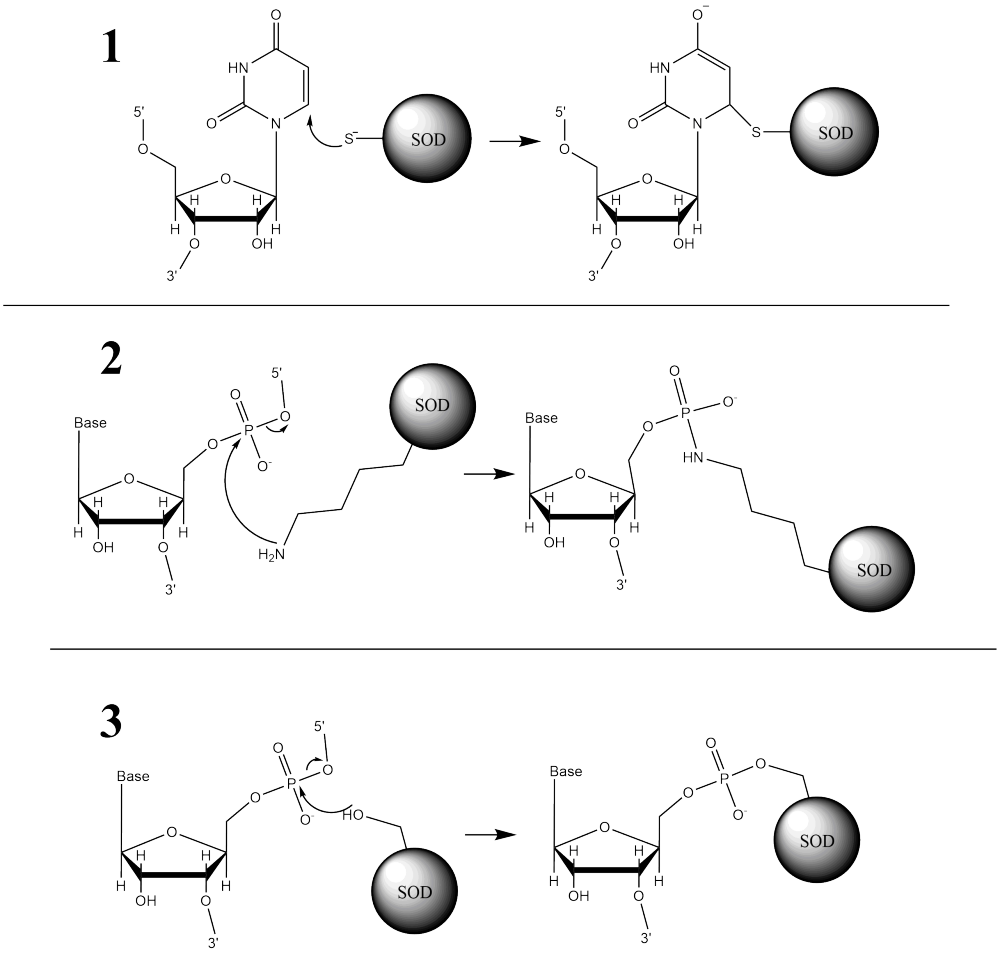


Fig. 3.7: Alternate protein-RNA conjugation chemistries. Although the intended reaction is methacrylamide adduct formation with a protein nucleophile, there is no doubt RNA exist in the pool that can catalyze conjugation reactions that do not involve the methacrylamide. (1) The thiolate anion of a cysteine can attack the C4-C5 double bond on uridine (or cytidine) in a similar reaction to the methacrylamide adduct. Though not pictured, the electrons on the formed oxyanion return to the carbonyl carbon after the enolate intermediate. This reaction resembles an intermediate in thymidylate synthase and cytosine-methyltransferase reactions.[65,80] (2) The RNA can also catalyze the reaction between a basic lysine and a bridging phosphodiester with the 5' (or 3') strand as a leaving group. If this reaction occurred in the constant regions with enough bases left to make a stable duplex with the primer, that RNA sequence would still be selected despite sequence loss. A phosphoamide bond like this was formed in a selection which sought to identify RNA that could covalently react with a peptide.[7] It is also found as a lysyl-AMP intermediate in many members of the nucleotidyltransferase superfamily, including T4 DNA ligase.[59] (3) A serine hydroxyanion can attack the backbone phosphate in a similar manner to scheme 2, but form a phosphodiester bond. To prevent these mechanisms from dominating the pool, a photocleavable linker was added to specifically select for methacrylamide usage.

12mer were scrambled to avoid a possible homodimer formation that may have been decreasing the effective reagent concentrations during ligation. This may have come at a cost as some sequences may have had base pairs to the 12mer to position the methacrylamide for the reaction. Conversely, this may have put pressure on sequences to act independently of the DNA 12mer, putting more emphasis on recognition and binding of the methacrylamide to drive the reaction and less on base-pair derived positioning of the electrophile.

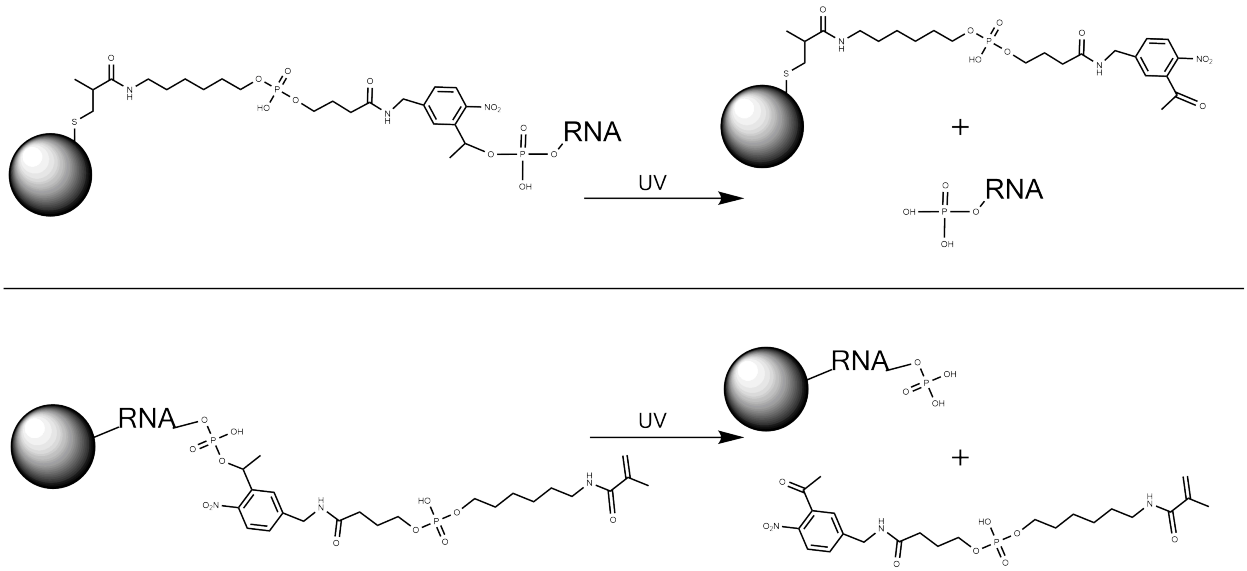


Fig. 3.8: Inserting a photocleavable linker between the methacrylamide and the RNA sequence allows the specific release of RNA that are bound through their methacrylamide. If a sequence reacts with the target without using the methacrylamide (bottom panel) it would remain bound to the target and solid support after photocleavage. If the RNA was bound through the methacrylamide, photocleavage would release the RNA sequence into solution where it can be gathered for reverse transcription (top panel).

Despite this change, this more specific technique eventually allowed the maRNA sequences that utilized the methacrylamide to dominate the selection. However, with a control sample that had an attached inert PEG_{18}-12mer instead of the methacrylamide-photocleavable-12mer, we detected a small population of RNA that could also be released from the beads in a light-dependent manner. This is shown in Figures 3.9 and 3.10 through real-time PCR amplification curves. Though described briefly before, these graphs depict SYBR Green intercalation-induced fluorescence which is proportional to the amount of ds-

DNA present. Once the amount of fluorescence passes the detection threshold during PCR, signal can be seen on the graph. The earlier a sample begins to show amplification on the graph, the more DNA was present in the initial sample. For reference, if PCR was 100% efficient, two samples amplifying 3.2 cycles apart would mean they had about a 10x different amount of DNA in the original sample. As this method detects all DNA, and not just that of the appropriate length, analytical gels must be run to show that the amplification curves were in fact tracing the intended PCR product. In all PCR samples described, full length product was amplified. Samples traced in figures 3.9 and 3.10 were all performed with the same PCR master mix and amplified together. The data are just split for simplicity. The amplification of the PEG_{18}-12mer (sample (+)SOD (-)MA (-)PC relative to the methacrylamide-photocleavable-12mer suggests that a population that used alternate conjugation chemistry persisted, and did so because they contained a sequence that could cleave in response to UV/near-UV light. The (+)SOD (+)MA (-)PC sample also suggest a similar mechanism as these molecules lacked a PC moiety. When comparing the light (Fig.3.9) to the dark (Fig.3.10), it is evident a light dependent, PC independent reaction is taking place. It is not known at the time the mechanism by which this occurred.

Preliminary Sequencing Though electrophoretic mobility shift assays could not detect covalent addition of the RNA to the protein at this time, controls analyzed by real time-PCR suggested enrichment of a biotinylated-protein-dependent and methacrylamide-dependent population in the pool that could be released from bead-bound streptavidin upon exposure to UV light.

We posited that sequencing may reveal a conserved motif that, though at this time may be inefficient at completing the desired reaction, could be chemically mutagenized and improved to the point that detection could be confirmed. After 18 rounds, the pool was inserted into plasmids, cloned and sequenced.

Isolate analysis and addition of a gel partitioning step: Four sequences were cho-

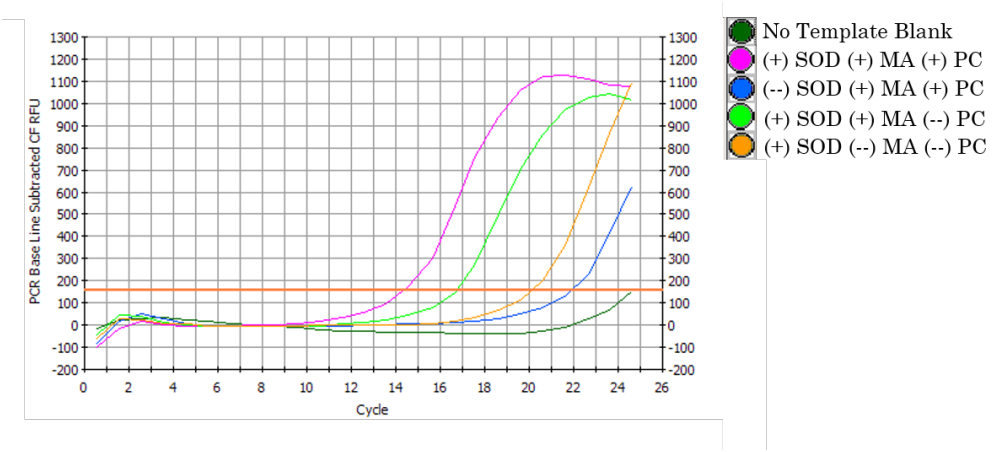


Fig. 3.9: Round 24 PCR amplification curves showing enrichment of SOD- and methacrylamide-dependent RNA (Pink trace). (MA=methacrylamide PC=photocleavable linker) All samples were exposed to UV light to release the cDNA/RNA duplex from the beads. The blue trace shows RNA that survived the partitioning without SOD. The 8 cycles of separation between the (+) and (-) SOD samples suggests a strong dependence on the presence of the protein. The green trace shows amplification of RNA that did not have the photocleavable linker but did have the MA, and was still exposed to UV light. The orange trace is a sample with no MA and no PC but was incubated with SOD showing dependence on the methacrylamide. The amplification of the green sample relative the the pink and orange suggests there is a population that can form a stable complex with SOD dependent on the acrylamide, but can be freed from the beads without the photocleavable linker.

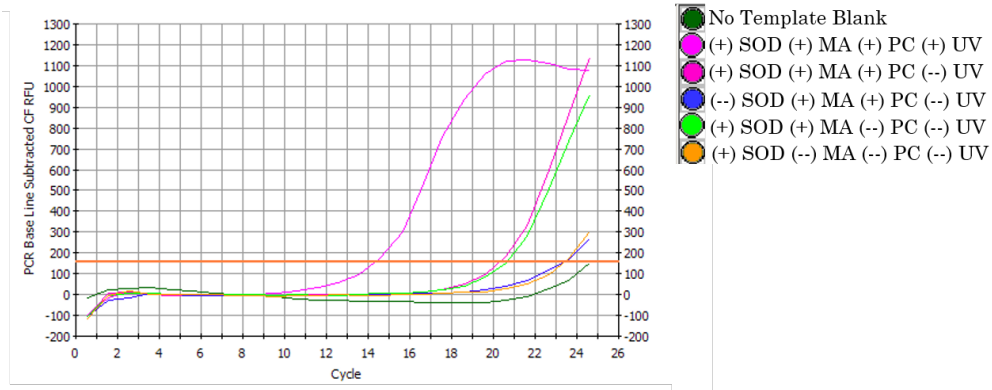


Fig. 3.10: Round 24 PCR curves showing enrichment of UV light-dependence. Cleavage of the photolabile linker is require for the RNA to be freed from the beads and amplified. This graph shows the selection sample (light pink) for reference, while all other traces are samples that were not irradiated. These control for alternate methods of release from the beads. Based on this graph it seems there is a small population that can be released from the beads without UV light. Both enriched traces (dark pink and green) are SOD dependent.

sen after convergence analysis and a denaturing electrophoretic mobility shift assay (EMSA) was performed to test if enough sequences in the pool could covalently interact with SOD1 G93A to visualize with SYBR Green. In one isolate, number 7, a positive gel shift was detected in SOD1 G93A containing samples. The shifted signal could be diminished when the sample was exposed to the UV LED prior to running on the gel. No shift was seen if the methacrylamide-12mer was not ligated. These experiments suggest the maRNA was connected through the methacrylamide, however only a small amount of the input RNA were able to react (Fig.3.11). Visualization of single stranded nucleic acid by SYBR Green staining is limited as its fluorescence is greatest upon binding double stranded nucleic acids. This technique is sufficient for typical analytical procedures, but it is not ideal for low level detection. However, based on the appearance of signal when adding SOD1 G93A and the disappearance of signal after UV irradiation, the band was tentatively assigned to the SOD-maRNA complex.

Based on this assumption, several more rounds were conducted with the addition of a gel-partitioning step in an attempt to enrich the RNA in the observed band to facilitate further characterization. To partition, after incubation with SOD, the maRNA were incubated with biotinylated SOD1 G93A, then heated in 40% formamide at 75^oC for 4 minutes, and ran on a 6% 8M urea polyacylamide gel. The area corresponding to where the aforementioned band appeared was excised, eluted, then RNA-SOD complexes were bound to streptavidin coated beads and the round proceeded as before. This partitioning addition abolished background signal to where all controls amplified with the no template control around PCR cycle 26 (due to primer dimer and low-level contamination)(Fig.3.12). Unfortunately, after several rounds of this procedure and reduction of the incubation time to 20 minutes, while still maintaining a dominance of SOD-dependent signal, the analytical EMSA of a 14 hour incubation still indicated only ∼1-5% activity. In addition, despite several successive identical selection rounds, the amplification cycle of the (+)SOD pool did not move, suggesting no further

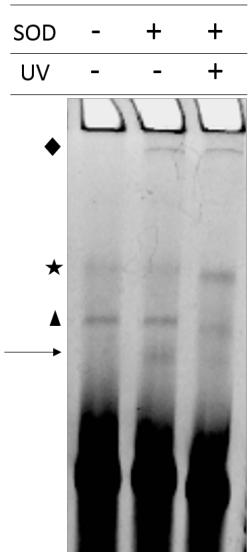


Fig. 3.11: A 6% polyacrylamide gel with 8M urea stained with SYBR Green. The middle lane shows an SOD-dependent gel shift marked by an arrow, and is presumably the RNA-SOD complex. The presence of this shift in the denaturing conditions of the gel suggest a very stable interaction. Upon exposure to UV light, third lane, the intensity of this band decreases, which is consistent with photocleavage of the RNA from the protein. The bands marked with an asterisk and triangle are artifacts that appear after the ligation reaction. The bands marked by the diamond are SOD dependent, but not sequence dependent and appeared in Round 1 controls. This band is not selection-derived.

enrichment could be gained from the current experimental conditions.

One possible reason is that the most reactive conformation of the protein could be unfavored. SOD1 normally exists as a dimer, but the monomer is thought to be biologically relevant and associated with disease.[57] In the monomeric form, the free thiols on the protein become more accessible. If the majority of active maRNA can only target the monomer and the rate of dimer disassociation is low, this may have limited the amount of SOD-maRNA product possible in the alloted time, explaining the stagnation of enrichment. On the other hand, given that the starting number of sequences in this selection, 10^{14}, was an incredibly small sampling of the possible 4^{50} unique sequences with this library, we may have just exhausted the abilities of the current pool. This selection not only required RNA to bind a protein, but also catalyze the formation of a covalent bond. If the effective motifs for these functions were not well represented in the initial library, this could explain the stagnation of

enrichment. Just as mutations in biology can confer a selective advantage to organisms, so can they during *in vitro* selections.

Pool Mutagenesis: In order to introduce more variability in an attempt to access a modified active site, folded structure, or reaction pathway, for example, the pool was mutagenized using an error-prone PCR protocol tailored for mutagenisis of an evolved pool.[12] In brief, Taq polymerase was used with 7 mM $MgCl_2$ and 0.5 mM $MnCl_2$ which introduces a higher rate of mutations than the already error-prone polymerase does naturally. This protocol retains the complexity of the current pool and introduces about 3 mutations per 100 bases. The selection process was restarted with the mutangenized round 22 pool. After several cycles of selection, the SOD-containing pool regained enrichment over the negative control, however it again stagnated at the same point as the selection did before. This suggested either the mutagenesis did not alter the sequences enough to improve them, or that another reagent or step, such as SOD dimer disassociation or methacrylamide reactivity, was limiting the progress.

Sequencing: Due to the failure of the mutagenesis branch to enrich the observed band, it was decided to sequence a past round that had not been mutagenized. An active sequence can, in the future, be chemically mutangenized, a process that can introduce more mutations than enzymatic mutagenesis, and then be re-selected for an increased reaction rate. This is beneficial when a sequence, by fold or mechanism for example, is prevented from reaching an active state due to a kinetic boundary.[68] For example, if only a small portion of the isolate with which the gel shift was seen was folded correctly to perform the binding and catalysis reactions, mutagenesis may destabilize the inactive fold or stabilize the active fold thereby increasing the amount of reactamers in the sample.

Round 24 was chosen for sequencing. This round had undergone several partitioning steps containing gel purification which seemed to greatly reduce background. Controls for SOD dependence, methacrylamide usage, bead binding, and light dependence were performed

for this round, all of which amplified with the no template blank suggesting a very low background population capable of surviving the selection steps (Fig.3.12). The presence of SOD G93A is crucial for selection, indicated by the separation between the (+) and (-) SOD samples. To test for use of the methacrylamide, shown (-)MA in the figure, a PEG_{18}-12mer was ligated to the pool instead of the methacrylamide-photocleavable linker moiety. The PEG_{18} occupies a similar amount of space, but is highly unlikely to take the place of the methacrylamide in a reaction, if needed. The 12mer was included in the control in case it forms an integral part of the structure or catalytic site of a methacrylamide-independent reaction, which would cause its absence to falsely indicate a dependence on the methacrylamide. This control also amplified with the blank, indicating a low number of methacrylamide-independent SOD-binding sequences. UV-induced cleavage dependence was also measured by forgoing the UV irradiation step for samples identical to Fig.3.12 (not shown) and all samples amplified with the blank. This also suggests the sequences being amplified in the selection sample are bound through the methacrylamide.

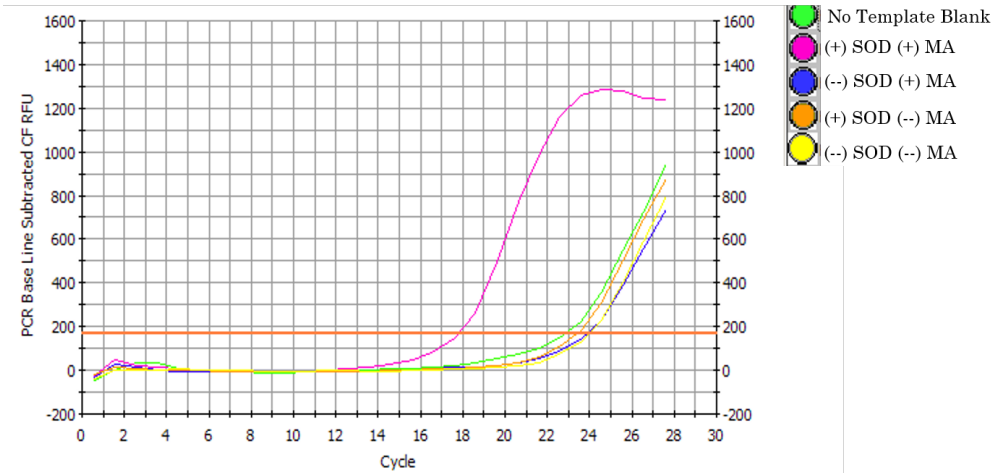


Fig. 3.12: PCR curves showing Round 24, the final round, selection sample and controls. MA=methacrylamide moiety. The (-)MA control RNA were ligated to an inert PEG_{18}-12mer. All samples were exposed to UV light to photocleave the RNA from beads. This graph shows there is a strong dependence on SOD and a requirement of the methacrylamide moiety. Due to the late amplification of the selection control, it also shows that only a small fraction of the RNA can perform the selected reaction during the 20 minute incubation time

The dsDNA resulting from this round was sequencing on an Illumina MiSeq by Dr. Jim Huntley at the BioFronteirs Institute at the University of Colorado, Boulder. Deep sequencing was chosen over the low throughput method used previously to gain better insight into the motifs represented in the final pool. Typically, our lab clones and sequences 96 isolates. This is an incredibly small sample of the diversity still present in evolved pools.[5] Because of this, we may be missing important motifs that are not only under-represented in the small sample, but also seeing results biased from the plasmid ligation step. In addition, the cost of next generation sequencing has greatly decreased making it comparable in cost to previously used methods. The obtained sequences were subject to typical processing.[9,15] Only reads with Phred scores higher than 20 were studied, resulting in approximately 8.5 million sequences. Repeat sequences were quantified then grouped as a single read. This resulting data was then ordered according to sequence abundance.

Initially, the first 1500 most populated sequences were analyzed with Daughter of Sequence Alignment (DOSA). This is a multiple sequence alignment program that aligns nucleic acid base pattern motifs, but allows for gaps between the conserved regions, as is often seen in *in vitro* selection results. It also allows slight mismatches in a pattern dictated by a user-input threshold values. Sequences that contain a consensus pattern are grouped into families. The output of the program presents two different alignments: a consensus sequence alignment which groups reads according to the presence of a single consensus sequence, and DOSA, which stresses the presence of multiple motifs when grouping into families. Though DOSA is well-suited for analyzing selected motifs, it cannot process the millions of sequences gathered from the NGS run. To ameliorate this issue, several other groups of 1500 were taken from throughout the sequencing data and processed by DOSA. All groupings contained similar families which suggests the motifs seen in the 1500 most populated sequences were pervasive.

Two sequences were chosen to study, 1-16180, the most abundant sequence, and 176-285,

a member of a highly populated family. Neither yielded a gel shift after incubation with SOD1 G93A. Further isolates from this pool have not been tested, instead efforts have been refocused on isolate 7 from the earlier sequencing analysis to characterize its interaction with SOD. It is not known why several of the selected sequences have no appreciable activity, nor why the final pool failed to converge on an active motif. This work can serve as a general method to conduct a selection experiment for reactamers, however there are several key points in need of improvement.

The random region of the library was 50 nucleotides in length. Though this is approximately the randomized length many modified aptamer selections are performed with, this may not provide enough sequence space to identify isolates that can both bind and catalyze the conjugation reaction. Future reactamer selections may benefit from a longer random region or the presence a pre-selected target binding sequence to focus the use of the random region on catalysis.

The commercially available methacrylamide moiety used is also a key issue to improve. In addition to the aforementioned constraints placed on the library by the short linker portion of the molecule, the low reactivity of the methacrylamide may also have adversely affected the outcome of the selection. After an RNA binds the target protein, the formation of the covalent bond must occur before the complex disassociates. If the RNA could not catalyze the reaction during the lifetime of the RNA/SOD complex due to the low reactivity of the methacrylamide, the contacts that align the reactive groups would be lost and the covalent bond would not form. This scenario may be a possible reason why only a small amount of isolate 7 was gel shifted in the presence of SOD. Future experiments with isolate 7 could attempt the reaction with an acrylamide or possibly a vinyl ketone moiety to observe a higher covalent complex yield, though recognition of the reactive group may be lost or reduced. For future reactamer selections, the use of a longer inert linker should facilitate the selection of a greater variety of structures by allowing the tethered functional group to access non-proximal

active sites. In addition, a more reactive group or even several different types of reactive molecules in parallel could be attempted to increase the chances of finding reactamers for specific targets. Even though cysteines are frequently used for covalent adducts, SOD may have been better targeted through a different residue and corresponding reactive group.

The potential for this new technology is high. This study sets the groundwork for future selections by offering methods to direct the library towards covalent bond formation, use of the tethered functional group, and importantly several low background partitioning steps. This work also reveals the interesting types of background populations that can arise from the implementation of the new method. With the improvement of more active and diverse functional groups, better tethering schemes, and early-selection execution of background-decreasing maneuvers, the definitive isolation, and therefore study, of reactamers can be realized.

3.1 Materials and Methods

Target Protein Preparation SOD1 G93A Cu/Zn was obtained from Giotto BioTech (Sesto Fiorentino, Italy). To biotinylate, the protein was mixed in a 1:40 molar ratio with excess biotin-PEG_4-NHS (*Pierce*) in 50 mM phosphate buffer pH 7 for 30 minutes at room temperature. Afterwhich, the excess linker was removed with Micro Bio-Spin P-6 columns (*BioRad*). The biotin to protein ratio was determined with the Fluorescence Biotin Quantitation Kit ($ThermoScientific$) and a Packard Fusion Microplate reader.

Library Preparation The random library was produced on a 394 ABI DNA synthesizer. The random region was produced by mixing the four phosphoramidites in a 3:3:2.4:2 ratio of dA:dC:dT:dG.[76] This ratio inversely corresponds to their coupling reactivities, which should result in a 1:1:1:1 distribution in the randomized portion. The DNA synthesized was the complement to the desired RNA sense strand. The RNA sequence was 5'-GGGAGACACGAGAAACGAGCAGCCA-50N-AGACAGAACCGCAACACGGAC-3'. The library was cleaved from the solid support and deprotected in 30% ammonium hydroxide. Full length ssDNA was gel purified from 6% polyacrylamide gels with 8M urea. 2 nanomoles of the purified product underwent 2-cycle PCR with the forward primer 5'-TAATACGACTCACTATAGGGAGACACGAGAAACGAGCAGCCA-3' and reverse primer 5'-GTCCGTGTTGCGGTTCTGTCT-3' using KOD XL polymerase. Due to the large number of legions towards the 5' end of long chemically synthesized DNA, the T7 RNA polymerase promoter sequence was added to the library during primer extension.

Primer Extension with KOD XL Polymerase

1x KOD Buffer

5 mM $MgCl_2$

250 μM dNTPs

1 μM primers

0.2 μM DNA template

0.0125 U/μl KOD XL polymerase (Novagen)

Cycle: 2x(95^oC 20 seconds, 53^oC 15 seconds, 72^oC 1 minute)

50 μl samples in 200 μl thin walled PCR tubes

The native-gel purified dsDNA was transcribed with T7 RNA polymerase according to the following protocol.

Transcription modified for 2'Fluoro-NTP Incorporation

1 mM ATP

1 mM GTP

1 mM imidazole-UTP

2.5 mM 2'F-CTP

30 mM GMP

10 mM DTT

0.5 μM dsDNA template

1x T7 R&DNA polymerase buffer (*Epicentre*)

5 U/μl T7 R&DNA polymerase (*Epicentre*)

37^oC for at least 5 hours

First Round: 1 mL

Following Rounds: 100-200 μl

Inclusion of excess GMP during transcription creates a 5'monophosphate on the transcript which is required for the next ligation step. T7 R&DNA polymerase (*Epicentre*) is a Y639F mutant T7 RNA polymerase that allows more efficient incorporation of 2'fluoro NTPs. The resulting RNA was desalted or gel purified then ligated to a DNA 12mer conjugated to specified functional groups. For the first several rounds, the 12mer sequence 5'-CTGATCATGACC-3' (Integrated DNA Technologies) was used when there was no photocleavable linker attached. For the rounds containing a photocleavable linker, the 12mer sequence was 5'-ACTTCGATGACC-3'(Integrated DNA Technologies).

Ligation

2.5 μM RNA (5'monophosphate)

5 μM DNA 12mer

5 μM DNA 24mer bridge

Ligase Reacton Buffer (*Invitrogen*) (50 mM Tris-HCl (pH 7.6) 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% w/v PEG-8000)

2 mM ATP

10 mM DTT

0.1 U/μl T4 DNA Ligase (*Invitrogen*)

37^oC for at least 12 hours

Notes: Nucleic acids and 1/4 of the buffer were mixed, then heated at 75^oC and cooled to room temperature over 10 minutes. The remaining buffer, water and T4 DNA ligase was added. A typical reaction used 500 pm RNA.

The ligated product was gel purified on 6% denaturing polyacrylamide gels. Due to the methacrylamide moiety's possible reactivity with polymerization agents in the forming gel,

all gels used to purify these ligation products were made at least 12 hours in advance to allow ample time to fully polymerize. In addition, all gels were pre-run for at least 15 minutes to remove unreacted ammonium persulfate. Ligated RNA were passively eluted from crushed gel slices, then desalted on 10 kDa MWCO Amicon ultracentrifugal filters at 9,000 rpm.

Reactamer Selection For the first round, 1.2 nanomoles ligated RNA were used, for all remaining rounds 20-100 picomoles per sample were used. All RNA samples were diluted to 500 nM, unless otherwise state (Table 1), with selection buffer (10 mM MOPS pH 7.2, 150 mM NaCl, 5 mM KCl, 2 mM $CaCl_2$, 1 mM $MgCl_2$) and heated in thin-walled PCR tubes in $\leq$50 ul aliquots according to the procedure below. This process denatures the RNA pool, then slowly brings the sample down to the incubation temperature to encourage reproducible folding of the RNA's tertiary structure.

Folding Protocol

Lid 105^oC

95^oC for 30 seconds

Decrease $80^oC - 38^oC$ at 2^o/minute

37^oC for $\geq$ 15 minutes

After folding of the RNA, biotinylated SOD1 G93A was added to the selection sample and select control samples. In all rounds except the first, one selection sample, which received the SOD, and one negative control, which received no SOD, were included. Some rounds included additional controls which are stated in the selection progression chart (Table 1). All incubations were performed in solution at 37^oC. Upon the end of the incubation time, samples were processed one of several ways.

MyOne Streptavidin C1 Dynabeads® (*Invitrogen*) were added to the RNA/protein solution to bind the biotinylated SOD and any attached RNA. Binding was performed in 10 mM

Tris-HCl pH 7.5, 1 mM EDTA, 1 M NaCl, and 0.005% Tween-20 (B&W buffer) to dilute the RNA to 10-30 nM. This diluted concentration greatly decreases non-specific binding to the beads. The samples was incubated at 37^{o}C for 5-30 min (See Table 1) shaking at 1100 rpm. The time was decreased over the selection as it became evident RNA could react with the beads. The beads were thoroughly washed 3x with 150 μl B&W buffer. Remaining RNA was reverse transcribed using SuperScript III which lacks an active RNaseH subunit, therefore leaving the RNA/cDNA duplex intact. Not only does this step create a cDNA strand from which to preform PCR, but the polymerization of the cDNA along the RNA also acts to abolish the structure and contacts required to bind the protein or beads. This assists in selecting for RNA that are covalently attached to the target or solid support since it is improbable for a sequence to tightly bind a target in both the single stranded and double stranded forms.

Reverse Transcription

1x First Strand Buffer (*Invitrogen*) (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$)

4 μM 3' primer

250 μM dNTPs

10 mM DTT

10 U SuperScript III (*Invitrogen*)

RNA/beads

12.5-25 μl sample volume

50^oC for 1 hour

Note: If beads were used, they were resuspended in solution by pipetting every 10-15 minutes.

The reverse transcription reaction supernatant was removed from the beads and discarded, as the SOD-RNA/DNA duplexes of covalently bound RNA should still be attached to the beads. The beads were washed 3x with 150 μl B&W buffer, then 800 μl B&W buffer was added and was shook at 1100 rpm at 37^{o}C for 10-20 minutes, discarding the supernatant. Starting in round 13, this 3x washing and 1x waiting and shaking was repeated with 50 mM phosphate buffer pH 7.5 with 0.005% Tween-20, then with 10 mM Tris-HCl pH 7 with 0.005% Tween-20 (PC buffer).

In later rounds, it was observed through (-)SOD controls, analyed by real time PCR, that some RNA sequences were able to react with the beads and were not dependent on SOD. To decrease this population, RNA was reverse transcribed in solution before bead binding under the assumption RNA in a duplex would have lower catalytic abilities than single stranded RNA. After the incubation with SOD, samples were buffer exchanged into 5 mM Tris-HCl pH 8 on Bio-gel P-6 spin columns ($Bio-Rad$), vacuum concentrated, then carried into reverse transcription as described above. The reaction was then bound to streptavidin-beads and washed also as described above.

Initial Sequencing PCR product from round 17 was inserted into the pCR^{TM}2.1-TOPO® vector and submitted to SeqWright for transformation, cloning, colonly picking, and sequencing. Four isolates were chosen to study from the results. These were PCR amplified off the sequenced plasmid, transcribed, ligated to the methacrylamide-photocleavable 12mer and purified. The isolates, 30 picomoles at 1 uM, were incubated with 500 nM SOD G93A in selection buffer at 37^{o}C overnight. 40% v/v formamide with 0.1% w/v bromophenol blue was added to each sample. They were heated at 75^{o}C for 5 minutes then ran on a 6% 8M urea polyacrylamide gel at 350V 20W for 50 minutes. Bands were visualized with SYBR Green staining. One sequence, isolate 7 revealed an SOD-dependent gel shift. This experiment was repeated with just isolate 7 and the shift proved to be repeatable. To see the affect of UV light on the complex, since photocleavage should release the RNA from the

protein if it is bound through the methacrylamide, the sample was incubated as above then illuminated for 10 minutes with a 6W 365 nm UV lamp before the addition of formamide and subsequent gel analysis. The sequence of the random region of isolate 7 is 5'-TCG TGC TGA AGT AAC AAT CAG GTT GGT GGA TCA GCG TTG GAT CAT CAC CT-3'.

Gel partitioning To increase selection pressure in later rounds, a gel partitioning step was added based on the appearance of a SOD-dependent gel shift with an isolate from preliminary sequencing data. In a selection round after incubation with SOD, 40% v/v formamide with 0.1% w/v bromophenol blue was added to each sample. They were heated at 75^{o}C for 5 minutes then ran on a 6% 8M urea polyacrylamide gel at 350V 20W for 50 minutes, including a 10 bp DNA ladder for reference. Bands were stained with 0.1x SYBR Green solution and visualized with a 473 nm SHG blue laser on an FLA-5100 scanner ($Fujifilm$). The area corresponding to where the isolate's gel shift was seen was excised in all samples with the back of a 1000 μl pipette tip. The gel was crushed and eluted in 400 μl 0.3 M NaAcetate and 1 mM EDTA overnight. The supernatant was filtered from the gel on a 0.2 μm Spin-X filter ($Corning$), then captured with 10 μl streptavidin beads, washed and reverse transcribed as described above.

Release of RNA from the beads In Rounds 1-7, these beads were heated at 95^oC for 5 minutes to melt the cDNA from the solid support. Immediately following, the supernatant was removed, then carried onto PCR. In Rounds 8 and on, the beads were resuspended in 80μl PC buffer added to white flat well 96-well plates and illuminated with a near UV LED source ($SallyHansen$) that fit snuggly over the top of the well holding the beads. This was done for 10 minutes, mixing the beads every 3 minutes. The mixture was then removed, and 38 μl of the supernatant was carried onto PCR, the rest was archived.

PCR

1x KOD Buffer (*Novagen*) (20 mM Tris-HCl pH 7.5, 8 mM $MgCl_2$, 7.5 mM DTT, 50 ug/ml BSA)

5 mM $MgCl_2$

250 μM dNTPs

1 μM primers

1% 1000x SYBR® Green (*LifeTechnologies*)

DNA template, 38μl, unknown concentration

0.0125 U/μl KOD polymerase (*Novagen*)

Cycle: 95^oC 20 seconds, 53^oC 15 seconds, 72^oC 25 seconds

50 μl samples in 200 μl thin-walled PCR tubes

Amplification of each sample was monitored using the fluorescence of SYBR Green and MyIQ^{TM} Single-Color Real-Time PCR Detection System. PCR products were analyzed by PAGE. 10-20 μl of the amplified dsDNA in the selection sample was transcribed and entered into the new selection round.

Gel Shift Assays To analyze the progress of the selection, a gel shift assay was used to detect complexes of RNA and SOD that could withstand denaturing gel conditions. Typically 30 picomoles of maRNA were refolded in 30 μl selection buffer. 500 nM SOD G93A was added, then incubated at 37^{o}C overnight in the dark. 20 μl formamide with 0.5% w/v bromophenol blue was added, the samples were heated at 75^{o}C for 5 minutes, then run on a 6% 8M urea polyacrylamide gel at 350V 20W for 50 minutes. RNA were visualized on a FLA-5100 gel scanner (*Fujifilm*) with a 473 nm SHG blue laser after soaking for 15 min in a 0.1% SYBR Green solution.

4. RNA PROTEASE SELECTION

In and of itself, an RNA that could cleave an amide bond would be a welcomed addition to the field of nucleic acid catalysis due to the demonstration of hydrolyzing such a stable and biologically relevant bond. At neutral pH and ambient temperature, an internal amide bond of a peptide has an uncatalyzed half-life around 600 years.[56] Moreover, a protease able to be engineered for specific substrates has astounding utility in medicine, biotechnology, and analytical biochemistry.[55] Several protease therapeutics are currently available, however all retain their natural substrate specificity.[16] In the future, proteases could be engineered to specifically cleave problem proteins and abolish their toxic function. This however would require more expansive specificity than natural proteases possess, and the field is far from achieving this objective. A more pragmatic current goal would be to engineer proteases for proteomics, namely for analysis of posttranslational modifications. Many of these modifications impart a significant perturbation in structure or chemical environment which can serve as a handle by which an engineered protease can differentiate between the modified and unmodified forms. In an early step towards this goal, the protease subtilisin BPN' has been mutated to preferentially cleave phosphotyrosine over tyrosine, the enzyme's natural target.[42] There is, however, no further altered sequence or structural dependence beyond that of the wild type enzyme. To begin to have further substrate specificity with respect to the surrounding protein, in order to have control over which substrates are cleaved, it may be beneficial to start *de novo* from a combinatorial library considering the proteolytic catalytic triads in question have continuously been selected by Darwinian evolution.

Members of the serine protease superfamily can be grouped according to structure and

substrate specificity; however all members, despite low sequence conservation, create their active site by aligning three types of amino acid residues to cleave a peptide bond. In a prime example of convergent evolution, several other enzymes in addition to proteases employ incredibly similar catalytic triads including β-lactamases, esterases, and acylases.[11] Inspired by the numerous protein scaffolds capable of properly orienting these functional groups to hydrolyze bonds, we sought to utilize *in vitro* selection to identify imidazole-modified RNA sequences capable of peptide cleavage.

Despite RNA's impressive repertoire of reactions, including both biological and *in vitro*-discovered sequences that can catalyze the formation of an amide bond,[78] the isolation of nucleic acids that can perform the reverse reaction continues to elude researchers.[4,10,17] Though the following experiments also failed to isolate protease-active sequences, it is clear several labs are concurrently working on this goal and public comparison of the different techniques may lead to a more efficient route to success.

As a proof of concept that imidazole modified-RNA can hydrolyze a peptide bond, we sought to initially identify sequences that could cleave a tethered biotinylated tripeptide, threonine-proline-tyrosine, which was modeled after the activation loop of c-JUN N-terminal kinase. After the proof of concept stage, the tripeptide would be synthesized with a phosphorylated residue and sequences would be selected that could cleave that over the non-phosphorylated form. Afterwards, the best sequences would be tested and possibly reselected for activity towards the tripeptide when part of the whole protein. This project was a collaboration with Thermo Fisher Scientific in Lafayette, CO. Dr. Robert Kaiser, the Director of Chemistry R&D, synthesized the peptide-containing molecules mentioned herein, and offered valuable insight on planning and problem-solving during the experiment.

RNA modification choice: The imidazole functionality was chosen because of histidine's crucial role in the serine protease active site (Figure 4.1). With a near neutral pKa, imidazole can take on both acidic and basic properties at physiological pH. In the catalytic

triad of serine proteases, an aspartic acid begins to deprotonate the acidic nitrogen on histidine, which in turn increases the basicity of histidine's other nitrogen. This nitrogen then deprotonates a proximal serine which then becomes a nucleophile and attacks the carbonyl of the amide bond to be cleaved. The forming negative charge on the carbonyl is supported by an oxyanion hole of positively charged amino acid residues, and the first step of the reaction ends in the collapse of the tetrahedral intermediate, which releases the N-terminus of the scissile bond, and forms of a covalent enzyme-peptide intermediate. The process then essentially repeats using water as the nucleophile to cleave the formed ester, releasing the carboxy-terminated peptide and thereby restoring the catalyst.[31]

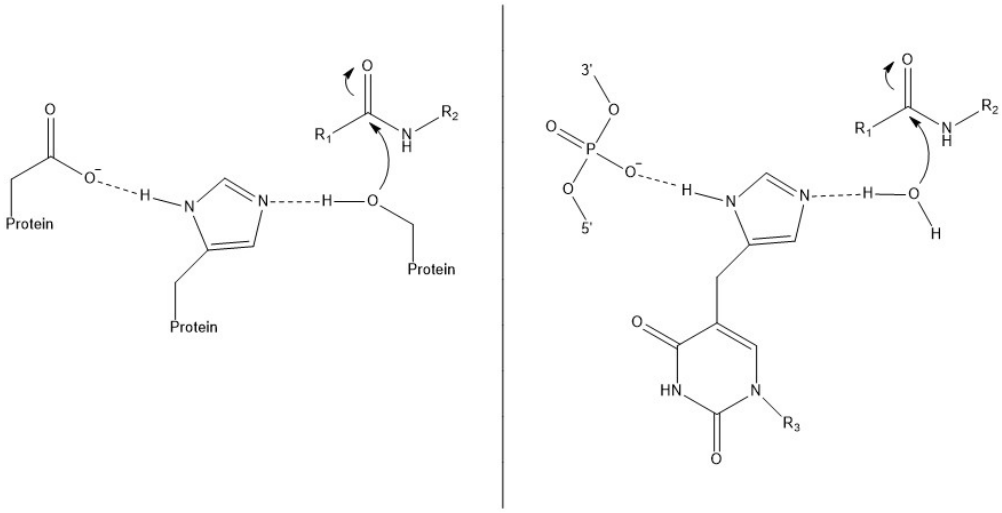


Fig. 4.1: Left: The catalytic triad of serine proteases. Right: Functional groups on the RNA can act similarly to the catalytic triad

In our system, RNA will be outfitted with the imidazole to equip them with a more active base at physiological pH. We envisaged the phosphate backbone could serve as the acidic residue with which to polarize the imidazole, while water or the 2'OH of the ribose ring could serve as the nucleophile for the initial carbonyl attack after activation, the former forgoing the covalent intermediate step. Additionally, protonated imidazoles or pKa-perturbed nucleotides could act as an oxyanion hole to support the forming tetrahedral intermediate. In retrospect, future attempts may greatly benefit from a stronger nucleophile, or even one

that can be positioned more precisely. It is probably no coincidence that biology evolved the use of serine instead of water for the initial reaction in this class of proteases.

Peptide construct design: A biologically relevant tripeptide, threonine-proline-tyrosine (TPY), was chosen for this selection. *In vivo* it is found on the solvent exposed activation loop in c-Jun N-terminal kinases (JNK) and can be phosphorylated on both the threonine and tyrosine residues.[81] In the context of the selection, this peptide had several attractive attributes. The tyrosine residue could increase the likelihood of RNA interaction with the peptide, and the proline could provide rigidity and decrease the number of peptide conformers possible. In addition, the activation loop is accessible to other kinases, which suggests the RNA would also be able to easily access it if winning sequences were tested on JNK, as opposed to the free tripeptide.

The biotin-dPEG$_4$-TPY-dPEG$_{36}$-DNA 12mer was synthesized by Dr. Robert Kaiser from Thermo Fisher Scientific and was ligated to each round's pool with T4 DNA Ligase. (Figure 4.2) The tripeptide was located between the biotin and the RNA sequence, and therefore cleavage of it would result in the loss of the biotin and consequent loss of streptavidin binding ability, which was used as the basis of the partitioning step. The long polyethylene glycol linker was chosen to facilitate active site access to the TPY residues. The construct was ligated to the 5' end of the RNA, and if the active site was on the opposite side of the folded RNA from the 5' end, a shorter linker may have restricted the pool to accommodating structures. Conversely, the linker also needed to be short enough and RNA concentration low enough that an *in trans* reaction would be unlikely so that protease-active RNA would only free their own biotin. This linker length approximates an effective concentration of 500 uM peptide to the RNA and has been used in prior selections with similar partitioning steps.[67,78] Discrete PEG (dPEG) refers to an exact amount of ethylene glycol units, in contrast to typical PEGs which are denoted with the average molecular weight, i.e. PEG$_{2000}$, and have a polydispersed population. Discrete PEGs were used to facilitate mass spectrometry

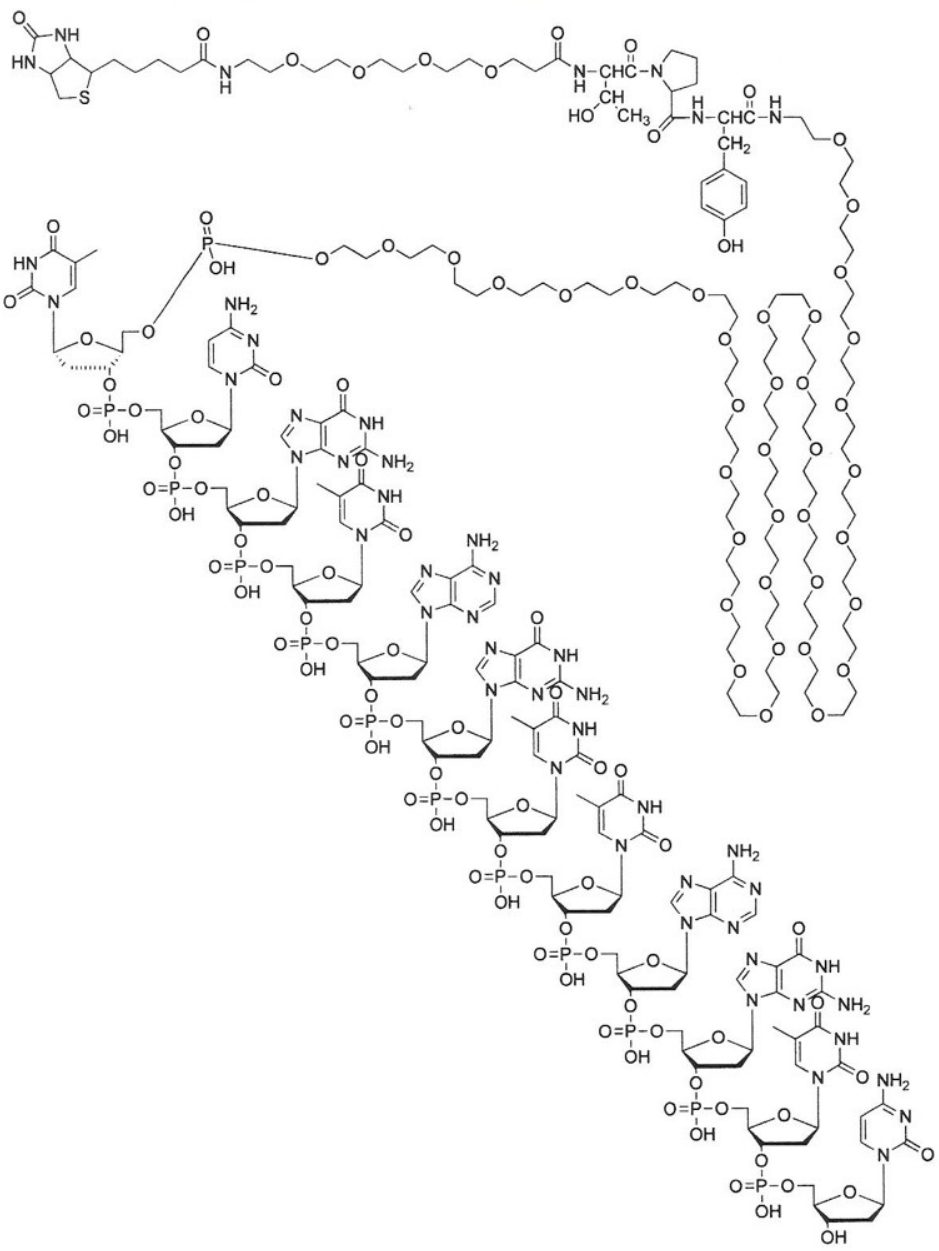


Fig. 4.2: Biotin-dPEG$_4$-TPY-dPEG$_{36}$-DNA 12mer constuct synthesized by Dr. Robert Kaiser from Thermo Fisher Scientific

analysis post selection.

All rounds contained a pool with the biotin-dPEG$_4$-TPY-dPEG$_{36}$-DNA 12mer construct ligated to the 5' end, and a "no peptide" control pool with a biotin-dPEG$_{40}$-DNA 12mer ligated. In later rounds, persistent background populations necessitated a more stringent purification step than PAGE excision, and therefore the biotin tag was used as such. During these ligations, 10 mM L-Threonine was included in the reaction to inhibit premature cleavage of the peptide and after it became available, 1 mM dPEG$_4$-TPY-dPEG$_4$ was also included as a competitor. Due to the limited amount of free peptide available, competitor concentrations could not be pushed higher than 1 mM. This may not have been sufficient to fully inhibit the sequences during an overnight incubation.

Selection Scheme: An imidazole-modified library with 1.5x10^{14} unique sequences was used, containing a N$_{100}$ random region flanked by 24 nucleotide constant priming regions. After ligated RNA were gel purified and refolded by thermal denaturation, they were incubated in selection buffer (1mM MOPS pH 7.25, 100mM NaCl, 100mM KCl, 1mM $MgCl_2$, 500uM $CaCl_2$, 10uM $NiCl_2$ and 10uM $ZnCl_2$) for 20 hours to cleave the peptide. If successful, the sequences would be released from the biotin and would not be pulled down with streptavidin-coupled (SA) beads. With this scheme, DNA- and RNA-cleaving sequences would also be selected, so gel partitioning of the SA bead supernatant was added as nucleic acid cleavage products would have greater mobility than the peptide cleavage products (Fig.4.3). A few rounds into this strategy, a dominating background population appeared which was eventually discovered to be circularly ligated RNA. The formation of the structures appeared to be dependent on T4 DNA Ligase. It became apparent that they were inadvertently selected for by the gel partitioning step, as in 6% denaturing gels, they ran where the peptide-cleaved RNA complex would run (Fig.4.4). These are further described in Chapter 3.

To reduce this population, several techniques were tried including simply partitioning from an 8% PAGE (where circular nucleic acids are shifted away from the area of interest

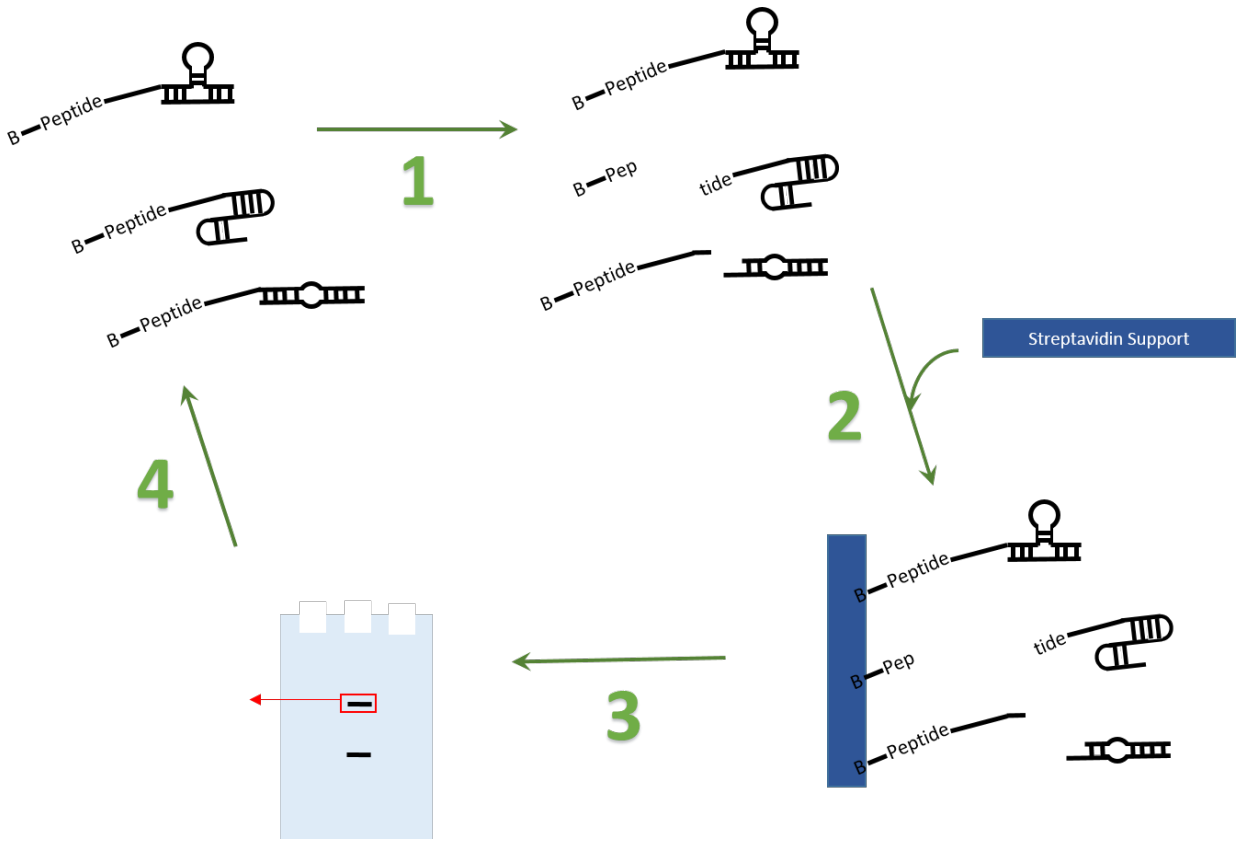


Fig. 4.3: A schematic of the original selection scheme for the RNA protease selection. (1) Folded RNA ligated to the peptide are incubated in selection buffer to cleave. (2) Support-bound streptavidin is added to bind sequences that are still attached to the biotin. Sequences that could perform some type of cleavage, i.e. peptide or phosphodiester cleavage, are left in solution. (3) The supernatant is run on a denaturing gel where different cleavage products can be separated according to length. The area corresponding to peptide cleavage is excised, eluted, reverse transcribed, PCR amplified, transcribed, ligated, then used as input for the next round.

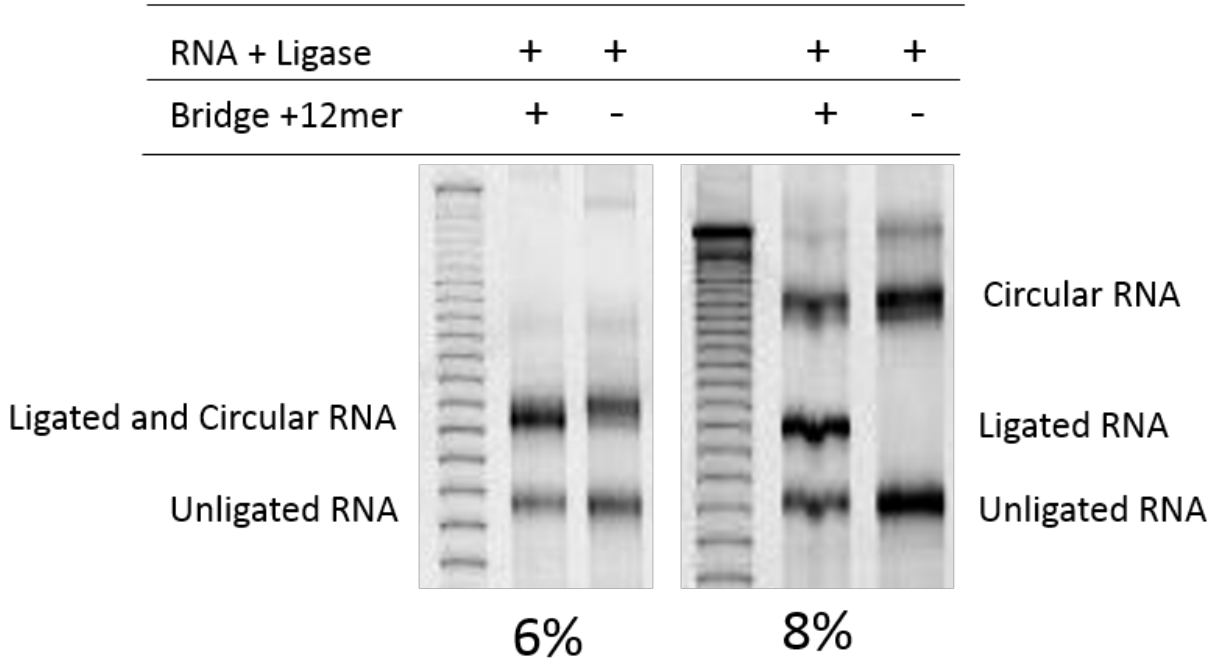


Fig. 4.4: Gel mobility comparison of the 12mer-ligated product and circularized RNA. These are denaturing gels of 6% polyacrylamide on the left and 8% on the right comparing the mobilities of the intended product to purify, the 12mer-ligated transcript, verses the circular RNA. The middle lane in each shows a full ligation reaction which included RNA, T4 DNA Ligase, 12mer, and bridge. The right-most lane in each gel shows a ligation reaction in which only RNA and ligase were added. The respective aliquots run on each gel were from the same reaction. In the 6% gel, ligated RNA and circular RNA run with nearly identical mobilities. In the 8% gel, however, the circular RNA exhibits decreased mobility, while the ligated species mobility does not change. Differences between the mobility of the circular species in different gel percentages is characteristic of non-linear nucleic acids. In the selection, the area corresponding to about where the ligated RNA ran was being excised from the gel to purify peptide-cleaved product. Since the circularized RNA ran at this same point, and most likely had larger numbers than amide-hydrolysis catalysts at this point, they were selected for and began to dominate the pool.

due to aberrant gel mobility) and extending the 3' primer during PCR to change the sequence at the site of ligation. All attempts worked to suppress their numbers for a few rounds, but the circular population proved incredibly resilient and would adapt to the added selection pressure. It is not known if this was done through polymerase mutations, or by possessing multiple reaction sites to perform an alternate reaction. The latter is evident in the sequenced circularizable pool described in chapter three. Not only could the sequences be ligated by T4 DNA Ligase to form circles, but some of them could also catalyze a different reaction to form non-linear RNA that possessed gel mobilities compliant to the selection constraints.

Eventually the selection was restarted from round two, before the circularizable population appeared. To ensure contending RNA had the TPY-construct ligated, as the gel purification method used before led to circle enrichment, the RNA pool was bound to SA-beads in the presence of L-Threonine to inhibit cleavage, and washed in denaturing conditions to remove sequences not bound through the biotin-streptavidin bond. The bead-bound sequences were then transfered into selection buffer, refolded, and incubated without competitor to cleave the peptide. Sequences freed from the beads, presumably by cleavage, were then run on a denaturing polyacrylamide gel to separate different cleavage products (i.e. peptide vs RNA cleavage would have difference gel mobilities). The area corresponding to the cleaved peptide product was excised. The RNA was passively eluted, reverse transcribed and PCR amplified (Fig.4.5). After one round of this, a self-cleaving ribozyme population appeared in the partitioning gels. This was an interesting result as the self-cleaving ribozymes ran on the gel about 3cm further than the area being excised. This may be due to the possibility that the amount of ribozyme RNA running towards the outer edge of the Gaussian shaped band could still be greater than the amount of RNA cleaving the peptide, and therefore were the species being selected. The following rounds showed an abundance of a ribozyme population that self-cleaved in only the peptide-containing pool, but not in the "no peptide" control, suggesting use of the peptide as a cofactor or for allostery. Several rounds later, the

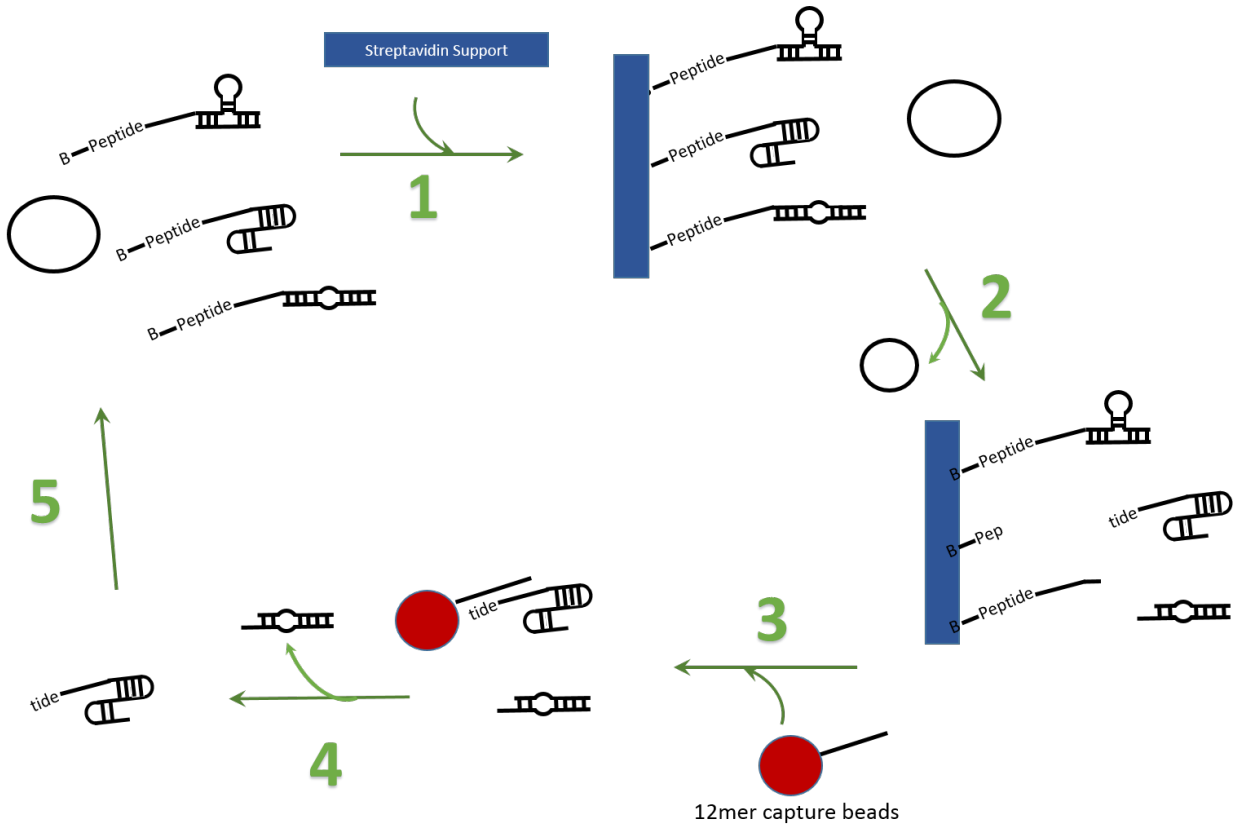


Fig. 4.5: Due to the dominant populations of circular RNA followed by phosphodiester-cleaving RNA in the pools, a modified selection scheme was developed to select against them. (1) To ensure sequences carried forward were ligated to the 12mer, after ligation, biotin-peptide-functionalized RNA were purified on streptavidin beads in the presence of peptide competitor and stringently washed. This should remove the population that could not form the biotin/streptavidin bond. (2) Bead-bound RNA were then thermally denatured and folded, and tasked to cleave the peptide in competitor free buffer. Cleavage would free the RNA from the bead. (3) Primer capture beads that bind the 12mer sequence were introduced to ensure only sequences that did not cleave off the 12mer would be selected. (4) The supernatant is removed to isolate sequences that cleaved beyond the 12mer. A gel partitioning step was also used in some rounds at this point. (5) The purified RNA were reverse transcribed, PCR amplified, transcribed, then ligated to use as input for the next round.

ribozymes were equal in abundance in both samples. At this time, the $dPEG_4$-TPY-$dPEG_4$ was synthesized and used in addition to L-Threonine to inhibit premature cleavage.

Fascinatingly, the selection soon developed a phosphodiester-cleaving ribozyme population that was inhibited by the presence of the free peptide. This became evident when a counter-selection step was added. The self-cleaving ribozymes were performing their reaction in selection buffer, but not during the ligation step as evident in analytical gels. This suggested a metal or pH dependence that was supported in selection buffer, but not in ligation bufer. To get the self-cleaving RNA to cleave, but also not lose peptide-cleaving RNA, after ligated sequences were bound to streptavidin beads, they were introduced to selection buffer with 1 mM $dPEG_4$-TPY-$dPEG_4$. This competitor is usually absent when sequences are exposed to selection buffer, as this is when they are intended to cleave the peptide. The supernatant from this step was saved, then the remaining steps in the normal incubation and partitioning were performed. When the samples were run on an analytical gel, there were sequences that self-cleaved during the "competitor with selection buffer" step and during the normal selection buffer only step suggesting sequences that did not cleave in the presence of competitor but did in its absence were either inhibited by the competitor or did not have enough time to cleave during the first step. This was repeated, but the time and number of "competitor with selection buffer" steps was increased, followed by the remaining incubation and partitioning steps. This revealed that a population of phosphodiester-cleaving ribozymes freed themselves from the beads during the initial incubation in the selection buffer with competitor, but after that supernatant was removed, and the process was repeated, no additional cleavage products were visible on the analytical gel. Interestingly, after adding competitor-free selection buffer, a population of RNA appeared in the supernatant that could self-cleave. This suggests inhibition by the exogenous peptide. Though this was an interesting result, it was not studied further.

To decrease the ribozyme population that was now dominating the selection, the step that

was being used to concentrate the bead supernatant was enlisted for partitioning (Fig.4.6). The technique, primer capture, uses a streptavidin-bead-bound biotinylated primer complementary to the RNA to pull those specific sequences out of solution after annealing. The sequences can then be melted from the beads into a smaller volume or different buffer with little yield loss. Traditionally the 3' primer is used for this step, which binds to the end opposite the peptide in our system. For the self-cleaving ribozymes to be freed from the beads but also carried on in the selection, they needed to cleave at the 5' end of the RNA and leave enough of the constant region in tact in order for the 5' primer to anneal during PCR. Since the cleaved transcripts were shorter than full length, they were probably cleaving the RNA and not the DNA 12mer. So, the sequence used to primer capture the RNA after release from the beads was changed to the complement of the 12mer. The first three bases of the RNA were also included so that the duplex was more stable at room temperature. In the first round it was used, this method reduced the ribozyme population so that it was no longer detectable on SYBR stained gels. After seven rounds, no new discernible population arose, and unfortunately no enrichment was seen for the peptide containing sample over the no peptide control.

Discussion: Due to the constant struggle against dominating background reactions, we decided to restart the selection and preemptively employ the new techniques; however, the improved scheme yielded no appropriate length PCR product. Though disconcerting, these peptide hydrolysis selections have been plagued with thwarting background reactions,[10,17] and a selection scheme that has no appreciable starting background population could be valuable. Based on continued failures to isolate active sequences in several labs, it can be assumed that the necessary structures are exceedingly rare, difficult to select due to polymerase bias, or non-existent in the given experimental conditions. As labs move forward towards this goal with the isolation of sequences that can catalyze ester and aromatic amide hydrolysis,[10] and others that form a covalent RNA-adduct to a transition-state analog of

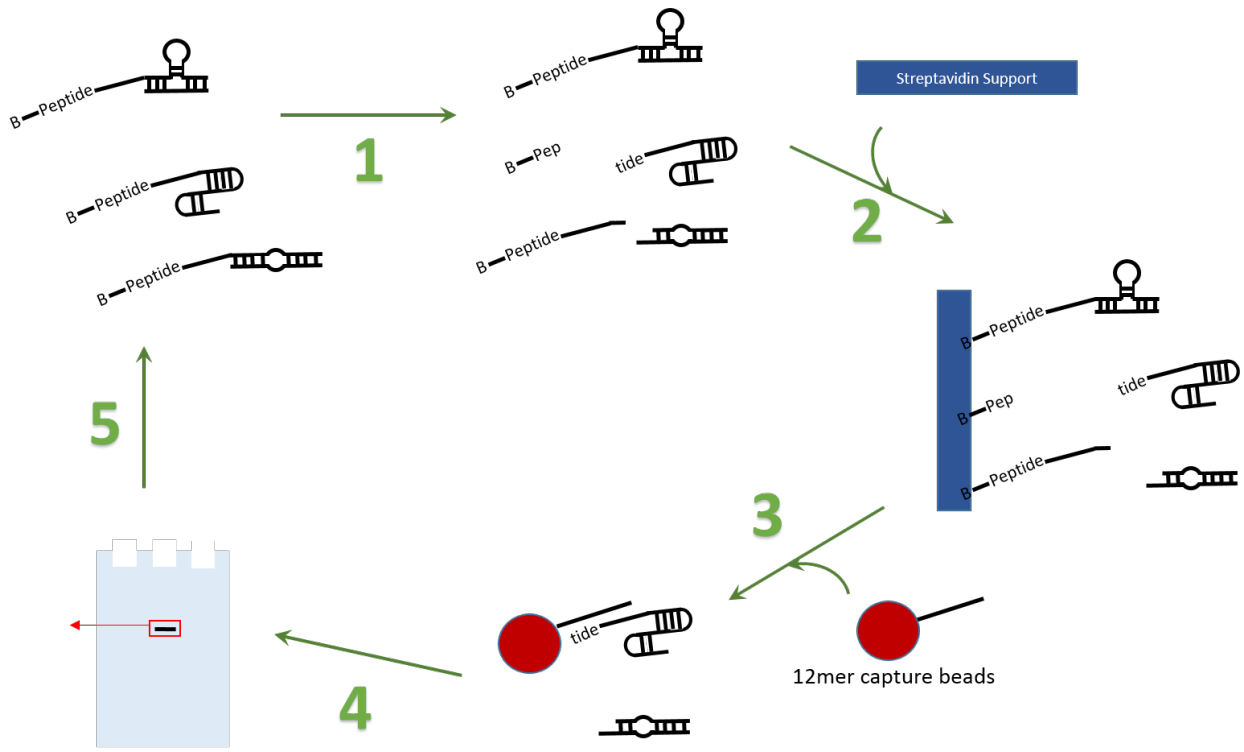


Fig. 4.6: Based on results in the previous selections, this figure depicts a recommended selection scheme. (1) Folded RNA are incubated to perform a cleavage reaction. (2) Streptavidin beads are added to bind sequences that did not perform a cleave that would remove the biotin bound to their 5' end. Sequences that did free the biotin remain in solution. (3) 12mer-capture beads are introduced to select sequences that did not remove the 12mer during the cleavage reaction. (4) The supernatant can then be run on a gel to separate the desired cleavage product from other background reactions. This step may be omitted in early rounds when the non-ideal elution yield from the gel may lose low copy number sequences. (5) Sequences are enzymatically amplified, transcribed, and ligated in preparation for the next round.

peptide hydrolysis,[4] it may be beneficial to start the aforementioned stringent selection scheme with a mutagenized sequence or bias library that already has some type of related activity or affinity towards the substrate. Alternately, modifications such as serine mimics can be added in addition to the imidazole (as the 2' ribose hydroxyl nor water seem to be readily taking its place as the nucleophile). It may also be beneficial to emulate a different proteolytic enzyme. Metalloproteinases employ a zinc ion, a glutamate residue and a water molecule to hydrolyze the amide bond, while aspartyl proteases appear to just use water and two conserved aspartate residues. These examples suggest an RNA carboxy modification may be advantageous, in addition to the cocktail of metals we already include.

Even though this selection was unsuccessful with respect to its goal, the constant rise and fall of background populations was an intriguing testament to the amount of diversity still possible in evolved pools. From the circularizable RNA being able to adapt to a new ligation sequence, to the self-cleaving ribozymes that effortlessly developed an interaction with the peptide that flowed from dependence, to independence, to inhibition based on the selection criteria, it begs to question whether these sequences were present all along, or beneficiaries of polymerase mutations. As researchers learn how a pool can evolve to increase or change its catalytic activity, well-timed interventions, such as intentional mutagenesis or new selection pressures, may be used to guide pools towards more complex and difficult reactions, such as peptide hydrolysis, with a higher success rate.

4.1 Materials and Methods

Library Preparation DNA was chemically synthesized as described in Experimental Chapter One. 2.5 nanomoles of ssDNA were extended in the primer-extension reaction. The complement of the RNA was synthesized. The constant sequence of the RNA library was 5'-GGGAGGAAACCACAGCCGAAGAAAU-CAAGCAGGCAGGCACGCAGAG-3. The forward primer with T7 promoter was 5'-TAATACGACTCACTATAGGGAGGAAACCACA GCCGAAGAAAU-3 and the reverse primer was 5'-CTCTGCGTGCCTGCCTGCTTG-3'.

Primer Extension with Taq Polymerase

10 mM Tris-HCl pH 8.3

50 mM KCl

3 mM $MgCl_2$

250 μM dNTPs

2 μM primers

0.5 μM DNA template

0.025 U/ul Taq polymerase (*NewEnglandBiosciences*)

Cycle: 2x(95^{o}C for 1 min, 54^{o}C for 1 min, 72^{o}C for 5 min) followed by 15 minutes at 72^{o}C

50 μl samples in 200 μl thin walled PCR tubes

The dsDNA was gel purified resulting in 1.5 nanomoles, all of which was used for *in vitro* transcription with recombinant T7 RNA Polymerase for the initial library. 25% of the PCR reaction was used for transcription in all subsequent rounds.

Transcription

1 mM ATP

1 mM GTP

1 mM CTP

1 mM imidazole-UTP

30 mM GMP

10 mM DTT

0.5 μM dsDNA template

1x T7 RNA polymerase buffer (40 mM Tris-HCl (pH 8), 10 mM NaCl, 8 mM $MgCl_2$)

2 U/μl T7 RDNA polymerase

37^oC for at least 2 hours

First Round: 3 mL

Following Rounds: 100 μl

The transcripts were purified from denaturing 6% PAGE, eluted from the gel pieces in 0.3 M sodium acetate, 1 mM EDTA pH 7 overnight, then dialyzed into dH_2O. The RNA was concentrated by a centrifugal evaporator then ligated to the biotin-$dPEG_4$-Thr-Pro-Tyr-$dPEG_{36}$-DNA 12mer (5'-TCGTAGTTAGTC-3') in the first round using T4 DNA ligase. Subsequent rounds included a ligation of biotin-$dPEG_{40}$-DNA 12mer for use as a negative control.

Ligation

1 μM RNA (5'monophosphate)

2 μM DNA 12mer

3 μM DNA 24mer bridge

1x Ligase Buffer (*Invitrogen*) (50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% (w/v) PEG-8000)

2 mM ATP

10 mM DTT

10 mL L-Threonine or 1 mM $dPEG_4$-TPY-$dPEG_4$

0.075 U/μl T4 DNA Ligase (*Invitrogen*)

37^oC for at least 12 hours

Notes: Nucleic acids and 1/4 of the buffer were mixed, then heated at 75^oC and cooled to room temperature over 10 minutes. The remaining buffer, water, peptide cleavage inhibitors, and T4 DNA ligase was added. A typical reaction used 200- 500 pm RNA.

10 mM L-Threonine or 1 mM $dPEG_4$-TPY-$dPEG_4$ was included to presumably inhibit premature cleavage of the tripeptide. The ligated pool was either purified from 6% or 8% PAGE or directly bound to streptavidin-coupled beads and thoroughly washed.

Amplification After partitioning, described below, selected RNA were reverse transcribed in solution.

Reverse Transcription

1x First Strand Buffer (*Invitrogen*) (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$)

3 μM 3' primer

unknown RNA concentration

250 μM dNTPs

10 mM DTT

10 U/μl SuperScript III (*Invitrogen*)

12.5-25 μl sample volume

First three reagents were mixed heated at 75^{o}C for 5 minutes then cooled to anneal primers. The remaining reagents were added, then incubated at 50^oC for 1 hour

The entire reverse transcription reaction was then added straight into PCR at 25% (v/v).

PCR

1x KOD Buffer (*Novagen*) (20 mM Tris-HCl pH 7.5, 8 mM $MgCl_2$, 7.5 mM DTT, 50 ug/ml BSA)

5 mM $MgCl_2$

250 μM dNTPs

1 μM primers

1% 1000x SYBR® Green (*LifeTechnologies*)

DNA template, 38μl, unknown concentration

0.0125 U/μl KOD polymerase (*Novagen*)

Cycle: 94^{o}C for 30 s, 58^{o}C for 25 s, 72^{o}C for 45 s

50 μl samples in 200 μl thin-walled PCR tubes

After amplification leveled off as measured by real-time monitoring of SYBR green fluorescence, the reactions were incubated 72^{o}C for 5 minutes. PCR products were analyzed by PAGE and the selection sample was carried on to the next round.

Original Selection Scheme: Gel purified ligated RNA, both the peptide-containing sample and the no-peptide control, were suspended in selection buffer (1mM MOPS pH 7.25, 100mM NaCl, 100mM KCl, 1mM $MgCl_2$, 500uM $CaCl_2$, 10uM $NiCl_2$ and 10uM $ZnCl_2$), heated at 78^{o}C for 8 minutes then passively cooled to room temperature. Incubation to cleave the attached substrate occurred at room temperature. The initial incubation lasted for 20 hours. The time was reduced conservatively in subsequent rounds. Post incubation, streptavidin-coupled beads (Dynabeads® MyOne^{TM} Streptavidin C1, Life Technologies) were added in 10x excess to the pools to bind inactive, still biotinylated sequences. Additional free streptavidin was added to the supernatant to bind any residual biotinylated constructs. The supernatant was then run on a denaturing 6% polyacrylamide gel. The cleaved peptide product was predicted to have a similar mobility to the fully ligated product, as only a biotin-$dPEG_4$ and an unknown amount of the peptide would be missing (no additional charge and a small fraction of the total mass) so the area corresponding to where the fully ligated product and a small amount below that was excised from the gel. The gel slices were crushed, eluted in 0.3 M sodium acetate and 1 mM EDTA, buffer exchanged by 3' primer capture, and subjected to RT-PCR as described above. Once it was discovered that circularized RNA co-migrated with the ligated population and was being enriched by the gel partitioning step, gel purification of the ligated pool was replaced with binding the crude ligation reaction to streptavidin beads in the presence of free peptide competitor. They were washed in 40% formamide, 60% 10 mM Tris-HCl pH 7.5, 1 M NaCl and 1 mM EDTA to remove unbound and unligated sequences. The beads with bound RNA were suspended in selection buffer and heated at 78^{o}C for 8 minutes, then passively cooled to room temperature to refold the RNA. After the alloted incubation time, the supernatant was collected, 3' primer captured to

concentrate, then run on an 8% denaturing polyacrylamide gel. The area of the gel described above was excised, and processed in the same manner for amplification.

5. TENTATIVE IDENTIFICATION OF A STABLE PHOSPHOTRIESTER IN A DUALLY CATALYTIC RNA SEQUENCE

Recently, circular exonic RNA were found to be abundantly expressed in tissues of higher order organisms.[49] Though several examples of these were known before, they were previously thought to be a result of splicing errors.[35] With conserved tissue-specific and developmental-stage-specific expression profiles,[49] and the discovery that they can bind complementary miRNA,[30] it is becoming clear that circular RNA have distinct roles in the regulatory processes of the cell. The mechanism of their formation, however is not clear. One proposed synthetic route to a specific class of circular RNA is through a spliceosomal interaction inferred from the presence of a consensus recruitment sequence for U1 snRNA.[13] The formation of circular RNA appears to be regulated,[49] however it is not known how the spliceosome differentiates between this role and it's known linear splicing role.

The spliceosome, itself has been found to have a new ability recently,[71] which may be an archaic trace of past functions. In a spliceosomal protein-free reaction, snRNA U2 and U6 were shown to create an RNA-X species in the presence of a branch point consensus sequence and absence of a 5' splice site. This reaction was dependent on the same sequence space necessary for the normal splicing reaction in cells. The RNA-X species has a four way junction thought to be a phosphotriester bond.[71,72] Biological RNA are constantly proving to be a catalytic and regulatory force in the cell. Just as several known in vivo reactions of RNA can also be observed emerging from random library space[41,53], so too may these recently observed phenomena. This chapter describes an imidazole-modified RNA sequence

that can self-circularize. There is also evidence to suggest a phosphotriester bond may be forming, creating a figure-eight-like secondary structure. This sequence requires exogenous imidazole-UTP for activity and is inhibited by magnesium ions. This is a fascinating example of metal-independent catalytic activity, a rarity with nucleic acids,[33] and is a testament to the resourcefulness of the modified polymer.

With the discovery that circular RNA is biologically important and ubiquitous,[49] this system may give mechanistic insight into its formation. If c6 can in fact form a stable phosphotriester bond, this sequence could also be a valuable contribution to understanding the formation and stability of the one thought to occur in RNA-X.

This chapter begins by recounting the discovery of RNA that could be circularized by T4 DNA Ligase. Though this is a fascinating reaction alone, it was soon discovered some of the sequences could self-catalyze their own formation into non-linear RNA in the absence of enzyme. This distinct reaction is the focus of the remaining chapter.

Identification of imidazole-modified RNA sequences that can be ligated by T4 DNA Ligase: As described in chapter two, an anomalous species appeared in a selection pool which was revealed to be circular RNA, the ligation of which was catalyzed by T4 DNA Ligase. The product had abnormally increased migration in decreasing percentage polyacrylamide gels when compared to DNA of known length. This is typical of non-linear species (See Fig. 4.3 in Chapter 2). The circular RNA could also be reverse transcribed to yield full length cDNA with a primer binding to the 5' end, suggesting the ends were connected. In addition to circles, end-to-end ligation products of dimers, trimers and tetramers were observed in lower amounts.

T4 DNA ligase is an ATP dependent enzyme which joins single stranded backbone nicks in double stranded DNA. In solution, the enzyme binds ATP and forms a covalent bond to AMP though the epsilon amino group of an active site lysine, releasing pyrophosphate. The enzyme then binds the nicked dsDNA, and transfers the AMP to the obligate 5' monophosphate of

the nicked strand in an intermediate step to form a 5'AppN moiety. A conformation change occurs and the enzyme then catalyzes the attack of the opposing strand's 3'OH to the penultimate phosphate to join the strands, releasing AMP.[46,59] T4 DNA ligase efficiently and quickly ligates nicked DNA, but it also exhibits some promiscuity towards RNA, though at a significantly decreased rate. A sample of 25% RNA/75% DNA duplex can be ligated within 12-15 hours at 37^{o}C, compared to 10 min for dsDNA at 25^{o}C. Our lab uses this flexibility to attach functional groups to the 5' end of RNA through a functionalized DNA 12mer and a DNA 24mer splint sequence that brings the 12mer and 5' end of the RNA together to be ligated. This technique was being used to functionalize this RNA library, and it became apparent that circularization of a fraction of the pool was occurring during the ligation step after five rounds of selection. This has not been seen in other selections employing the same technique and is believed to come as a result of the specific gel partitioning step with this length library.

In this step, an area of a polyacrylamide gel was excised after electrophoresis to purify the intended target population, however, the circularized RNA also ran at this same point. The circular RNA proved to be either more abundant or easier to amplify than the intended target, and soon dominated the selection. The only requirements for circularization were the RNA, T4 DNA Ligase, and typical buffer components. Since T4 requires a duplex to bind, we hypothesized that the active RNA contained a region complementary to each end that served as the splint sequence, bringing the two ends together to look like a backbone nick (Fig. 5.1). Ligation of a nicked RNA duplex by T4 DNA ligase occurs at a very low rate, and it therefore came as a surprise that this reaction proceeded at a faster rate than the 75% DNA duplex typically used in our experiments. To gain further insight in the mechanism of action, we decided to sequence the pool. Base pairing around the ligation site is crucial for catalysis by T4 DNA Ligase, and sequencing should reveal if this is a factor in circularizing the RNA. In addition, the extent to which the sequences mimic the natural substrate of

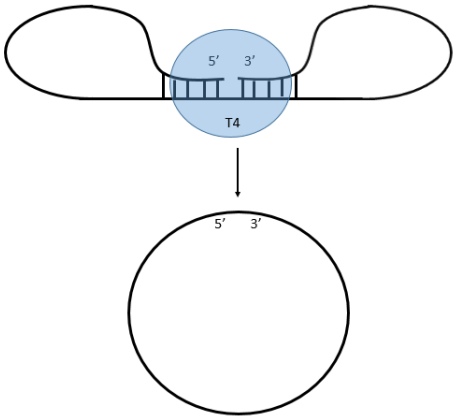


Fig. 5.1: T4 DNA Ligase requires a duplex to ligate nucleic acid. We hypothesized that the RNA had an internal sequence complementary to the ends of the primers, creating a duplex. The ligase could then bind and ligate the ends to form a circle.

T4 may provide clues regarding the role of the RNA's participation in the reaction. If the substrate is foreign to the enzyme, the RNA may be taking an active role in the initiation or completion of the reaction.

During the selection, there was no intended pressure for circularization, only the gel partitioning step which originally led to their appearance. This step, though, in no way *required* circularization to happen and hypothetically, there could be populations in the pool with other activities that also ran at that point in the gel. A round where the circularization yield was the highest was chosen, and a single round was performed to further enrich the pool. For the partitioning step, circularized RNA was excised from a gel, then reverse transcribed using a primer which bound to the 5' end of the RNA. Since reverse transcription proceeds in the 5' to 3' direction, only transcripts that had a 3' end connected to the 5' end would have template from which to copy. The cDNA was amplified with PCR then ligated into vectors, transvected, cloned and sequenced. To our surprise the sequences did in fact contain short primer binding sites in the randomized region that would bring the two ends together for ligation; however, there was a $T_N G_N$ insertion between the two internal primer binding sites (IPBS) in the large majority of sequences (Table 5.1). The other isolates had a different

insertion sequence, and only one did not have any. Interestingly all of the sequences except one had the IPBS near the 3' end of the randomized region. The reason for this is still not clear. The isolates were sorted into families based on their insertion sequences as there was not a significant amount of alignment patterns in the remaining sequence space.

The modest rate of the reaction and the gap between the primer binding sites suggests a supporting role for the RNA in catalysis in addition to the enzyme, as normal dsRNA is a poor substrate for T4 DNA Ligase and gapped ends are not substrates at all.[40] But, due to the lack of any discernible widely conserved patterns beyond the IPBS, this role was difficult to determine. The absence of larger patterns could suggest the rest of the RNA played little role in the enzyme-catalyzed ligation reaction or that many sequences are capable of the necessary catalytic assistance. Due to the IPBS only binding a few bases of each primer, this interaction alone should not be stable at the 37^{o}C temperature the ligation is performed at. The loops flanking the IPBS may have a structure which forces the ends to bind, increasing the effective melting temperature of that interaction. In addition, a structure could be possible that could contort the A-form IPBS RNA helix to mimic a B-form helix, which is the natural substrate of T4 DNA Ligase. This may explain the increased rate when compared to ligation of typical RNA helices by T4 DNA Ligase.

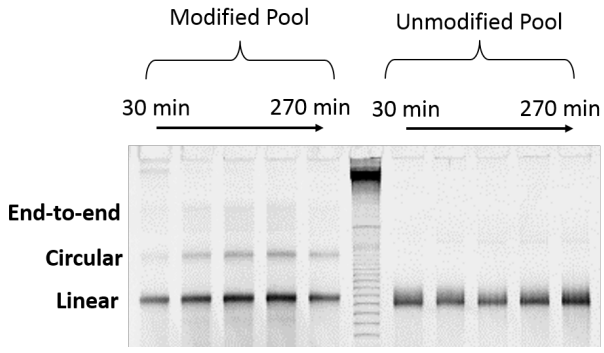
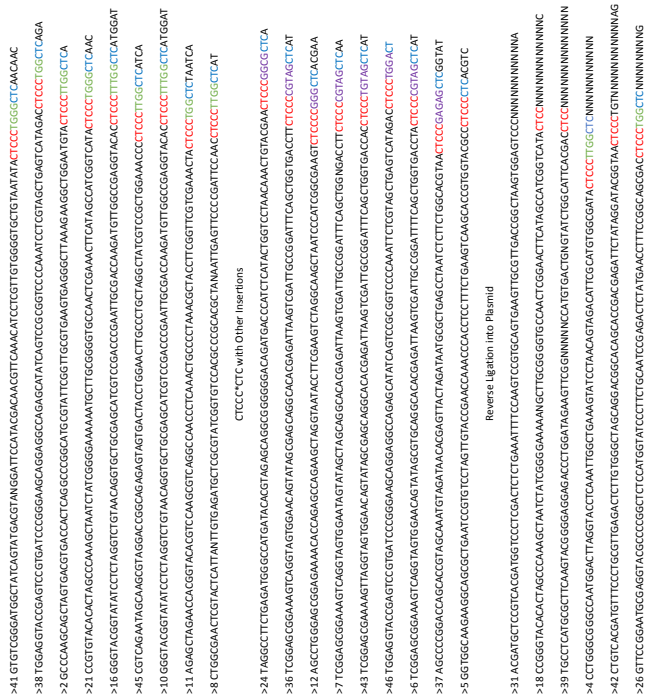


Fig. 5.2: Time course of T4 DNA Ligase creating circular RNA. To the left of the ladder is the imidazole-modified pool. To the right is the unmodified pool. Circular products are marked, as are the end-to-end ligated molecules.

Discovery of self-circularizing imidazole-modified RNA: The gap between the



Tab. 5.1: Select isolates from circularizable RNA sequencing results. Conserved regions are shown in color. Constant regions are not shown. The red sequence is complementary to the 5' end of the RNA while the blue sequence is complementary to the 3' end of the RNA. The green lettering in the top group highlights the most abundant insertion sequence, T_NG_N. The purple lettering in the middle group shows other examples of observed insertion sequences. The last group contains several isolates that were sequenced as the antisense strand. Fascinatingly, yet still inexplicable, all isolate obtained from this sequencing experiment that had inserted into the plasmid with the antisense strand ligating to the plasmid sense strand (giving the reverse complement of the RNA in the sequencing results), yielded poor reads in the first part of the random region that was read. The primer site before this (not shown) was read with no issues.

presumed IPBS in the sequencing data was an interesting result due to the fact a gapped substrate cannot be ligated by T4 DNA Ligase. The enzyme can bind and activate the 5' monophosphate with AMP with low efficiency, however it cannot catalyze the joining of the strands.[40] In the majority of sequences inserted between the primer binding sites, an imidazole-modified base was present. Since the ligase is unable to join the strands after adenylation when a gap is present, the RNA, itself, may be deprotonating the 3' hydroxyl with the imidazole and positioning it for attack of the 5' phosphate. To test this hypothesis, an adenosine-phosphate-phosphate-guanosine (AppG) was synthesized.[37] This dinucleotide can be incorporated into the 5' end of an enzymatically synthesized RNA.[29] Two of the sequences which could be ligated in the highest yields were synthesized each with a 5' triphosphate-G, 5' monophosphate-G, and 5' AppG.

Surprisingly, both isolates were able to form material, *without* T4 DNA Ligase, that had the anomalous gel migration patterns of non-linear species. However, one sequence, c26, was capable of forming it with only the AppG end, while the other isolate, c6, was only capable of forming the circular species with the mono- and triphosphate 5' termini (Fig.5.3). These species, however, ran slightly differently in gels compared to their T4-ligated counterparts. This suggests the material formed in the absence of enzyme are not quite the same species, but possibly lariats, smaller circles, or another shape permutation. This was a fascinating result as there was no obvious selection pressure to self-catalyze these reactions since the ligase enzyme seemed to do so efficiently. This demonstrates that any kinetic advantage a sequence has can greatly facilitate its selection.

In this case, c26 may have been able to complete the ligation reaction after T4 DNA Ligase adenylated the 5' end, as the enzyme itself is conformationally hindered to complete the reaction due to the gapped substrate. Isolate c26 was not studied further, however additional experiments could be performed to determine the role of the interstitial sequence of the IPBS. Mutants of T4 DNA Ligase exist that cannot perform certain key steps in the

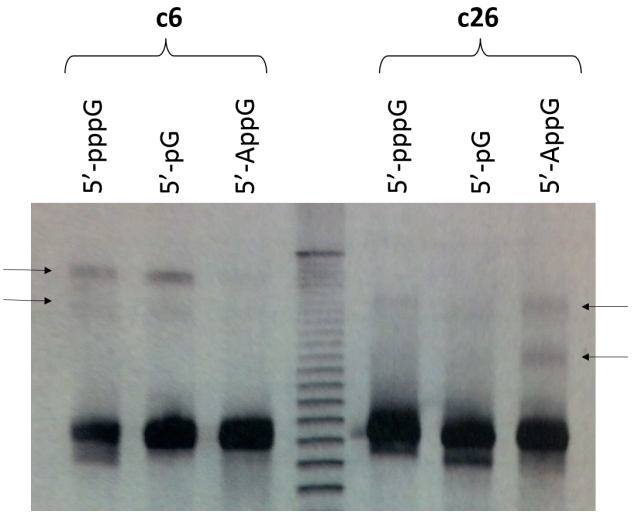


Fig. 5.3: 8% denaturing polyacrylamide gel showing products resulting from self-circularization. The 5' ends of the RNA were synthesized with either a triphosphate, a monophosphate, or an adenosine diphosphate. Sequence c6 could self-circularize with a pppG and pG, but not with an AppG end. Sequence c26 could only circularize with an AppG.

reaction, such as adenylation or ligation. The use of these could delineate the roles of the ligase and RNA along the reaction pathway. Since the yield of the self-catalyzed reaction was much lower than that of the enzyme catalyzed reaction, there may be an integral step or reagent missing from the experimental conditions. If the ligase offers an environment in which the RNA can more efficiently catalyze the joining of the strands, just as ribosomal and spliceosomal ribonucleoproteins work in concert to catalyze a reaction with faster rates than RNA alone,[53,71] this may explain the discrepancy in the yield.

Isolate c6 was chosen to study further. During the selection in which c6 was enriched, excess GMP in the transcription reaction created a 5'GMP on the transcript. However, as seen in Fig 5.3, c6 can also self-ligate with a 5'GTP to a lesser extent, but cannot ligate with a 5'AppG. Though the leaving group, i.e. the monophosphate or one or more 5' nucleotides, for this reaction is not known, these results suggest bulk in the binding site for the 5'end can hinder the reaction.

It was found that with this isolate, self-circularization was occurring during the transcrip-

tion reaction. T7 RNA polymerase was not required for circularization. Experiments were performed to identify which components of transcription were necessary for the reaction. Since the buffer used during transcription was 40 mM Tris-HCl pH 8, this was used for all further experiments. Initially the 8 mM $MgCl_2$ present in transcription was also included in the solution. By adding each individual component present in the transcription reaction, it was found that either 4 mM imidazole-UTP and GTP addition would result in the formation of the circular product, however the latter to a lesser extent. In experiments where single reagents were excluded, the absence of magnesium seemed to slightly increase the formation of the product. Combining these findings, it was discovered that magnesium inhibits the reaction with imidazole-UTP, while slightly facilitates that with GTP (Fig. 5.4). ATP and CTP were completely inactive (not shown). At this time, yields were great enough to reveal that the sequence could make two main non-linear products. The reaction with imidazole-UTP was chosen to study further as the yield was over five times greater than with GTP.

The necessity of the free imidazole-UTP was quite surprising as the strand itself contains many imidazole-uridines, however, the requirement of an exogenous nucleotide is also seen in group I catalytic introns and is therefore not a concept without precedence. It is not known at this time whether the free imidazole-UTP is being incorporated into the strand, or is utilized as a bound cofactor. An unmodified c6 sequence with exogenous imidazole-UTP could not form any products suggesting the body-imidazoles also play an important role in structure or catalysis. The use of free imidazole-dUTP showed no significant decrease in activity, while free UTP and pyridyl-UTP were inactive. The use of the imidazole as a reactive group is supported by a pH profile of the reaction, where the buffer, 40 mM Tris HCl typically at pH 8 for transcription, was varied from 6.0 to 9.5 (Fig.5.5) revealing a sharp appearance of activity at pH 7.2, an optimum pH at 8.2, and retained activity through 9.2. Though the exact pKa's of the imidazole nitrogens in the context of the nucleotide is not known, assuming the value

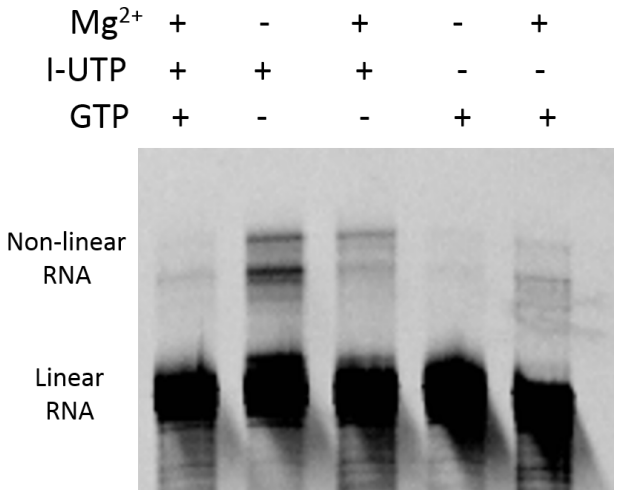


Fig. 5.4: 8% denaturing polyacrylamide gel showing c6 RNA incubated overnight in 40 mM Tris-HCl (pH 8) and 4 mM of the listed nucleotides and 8 mM $MgCl_2$. Two main non-linear products are formed. This gel shows a dependence of non-linear RNA formation on imidazole-UTP, and inhibition by magnesium when I-UTP is present. Exogenous GTP allows the sequence to form the circular structures, though to a lesser extent than I-UTP. Unlike I-UTP, however, the addition of magnesium seems to assist the structures' formation.

is near histidine's, the pH variation is possibly titrating the basic nitrogen, suggesting the active form of the nucleotide is the neutral tautomer of imidazole. The decreasing activity after pH 8.2 suggests another ionizable group is also required for activity. Due to the proven ability of ribozymes to significantly perturb the pKa of a residue, though, no conclusions regarding specific identification of either of the functional groups at play can be made.

We shifted our focus to deconvoluting the non-linear structure of c6. As mentioned before, the self-catalyzed circular species behaved slightly differently in gels than did the circular species made by T4 DNA Ligase. Since c6 had no activity in ligation buffer, nor towards an AppG appended 5' end, we assumed that T4 was the only source of catalysis with respect to circle formation during ligation and that the product of this reaction should resemble the product of T4's biological reaction, 3'-to-5' ligated nucleic acid. Therefore, the T4-ligated c6 was assumed to be a true circle, and self-circularized c6 a circular-like shape, which was thought to be a lariat at the time.

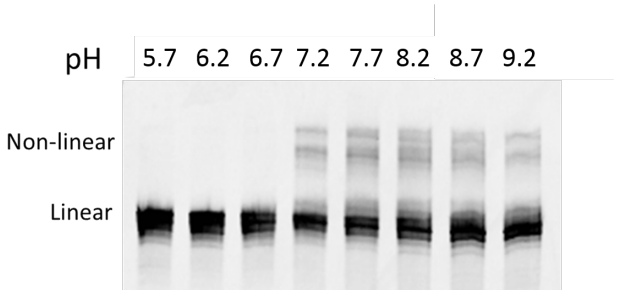


Fig. 5.5: 8% denaturing polyacrylamide gel of the pH dependence of c6.

Isolate c6 was capable of forming two main circular species when visualized in gels. Both of these products could be excised, purified, and heated in 75% formamide for 5 minutes at 80^{o}C then re-run on an 8M urea gel with no change in mobility or appearance of linear product, demonstrating their stability. This also suggests that this product contains a covalent bond, as opposed to a stable non-covalent structure. For the following experiments, the highest running circular species was studied in it's pure form.

Other evidence had already suggested the self-catalyzed population of circular RNA were not truly circles, but possibly another type of non-linear RNA. Analysis of the products of RT-PCR during the selection revealed several shorter than full length DNA, ranging from 40 to 100 bases (full length product is 147 bases). Performing this experiment with the separate isolates revealed that individual sequences were responsible for each of these PCR products. Isolate c6 gave rise to an approximately 80 base amplicon. RT-PCR of the purified non-linear product of c6 revealed that this truncate resulted from it, and not the linear species. The PCR product was excised, purified, and sequenced revealing a 68 base deletion from the random region beginning at base 58 and ending just before the start of the 3' constant region (Fig. 5.6).

This was an interesting result as we believed this amplicon was encoded by a lariat RNA with nucleotides near the 3' and 5' ends ligated together (demonstrated by full length reverse transcription from the 5' constant region). Though it is possible the 68 base region was being excised, this would result in a much shorter circular product that most likely would not run

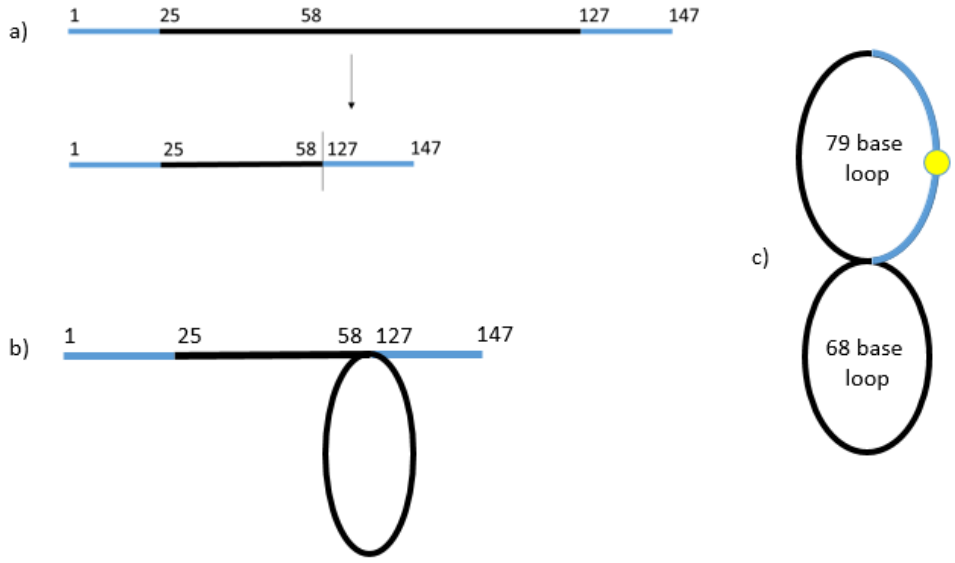


Fig. 5.6: (a) Sequencing results showed a 68 base deletion in RT-PCR products from a purified c6 circular species. (b) Over time, a purified c6 circle can degrade back to it's linear form suggesting the 68 base portion was not excised, but just couldn't be accessed by the reverse transcriptase. (c) c6 circles are a substrate for DBR1, suggesting there is a 5'-2' possibly branched phosphodiester bond present. RT-PCR experiments imply this bond joins the ends of the transcript together (yellow circle). The exact nature of the bond which joins base G58/C59 to U126/C127 is not known.

near the full length ligase-made circular RNA. In addition, over time, the purified species could degrade back to full length linear RNA suggesting the transcript remains intact in it's circular form. This would suggest the RNA is connected as a circle at the ends, but has a "pinch" in the middle connecting approximately G58 to C127 (Fig.5.6). This obstruction prevented the reverse transcriptase from accessing the sequence that was deleted in the final amplicon. The 3' primer used in this experiment would anneal right on the pinch site and possibly guide the polymerase straight over the looped sequence. In contrast, when the polymerase is primed before the pinch site, RT-PCR results imply it can choose to read through or bypass the unusual bond, resulting in cDNA with corresponding lengths. Bartel and colleagues reported a similar phenomenon with SuperScript II jumping a branched template.[70] While this selection and experiments were performed with SuperScript III, a

more thermostable variant, this template jumping ability appears to also apply.

Additional experiments were performed to test the hypothesized figure-eight-like secondary structure of c6. Nucleic acid catalyzed end-to-end ligation of RNA often results in unnatural 2'-5' linkages due to the greater nucleophilicity of the 2'hydroxyl.[20,22,45] To test if this bond was present in the non-linear c6 species, debranching enzyme (DBR1) was received as a gift from Dr. Thomas Menees at UMKC. This enzyme specifically cleaves 2' to 5' branch points in lariat introns. Purified circular c6 was treated with the enzyme, and branched and debranched products were probed with circle/loop-dependent RT-PCR experiments. Primers were made that bound to sites within the template sequence, depicted in Figure 5.7. Primer pairs that abut on the sequence were used for reverse transcription and PCR, and the products were analyzed by gel electrophoresis. Since the primers are designed to be extended away from each other, the reverse transcriptase needs to make it around the structure, almost back to it's starting primer location in order to polymerize the cDNA for the binding site of the forward primer in the pair. If this does not happen and the polymerase cannot complete a circle, the forward primer cannot bind during PCR and therefore there will be no PCR product. This is contrary to typical RT-PCR, where primer sites are located on opposite ends. This experiment requires some type of circular structure to yield any PCR product, and requires end-to-end circularization to yield full length product. If the structures have an accessible 2' to 5' linkage which makes them non-linear, this site may be a substrate for DBR1. By comparing the PCR products resulting from untreated and DBR1-treated purified c6 non-linear RNA, linkage site information can be gathered. Two primer pairs were initially used. One set binds in the "top" loop of the figure with the putative 2' to 5' linkage. The other set binds to the bottom loop. These experiments are described in Figures 5.7-5.11.

These results suggest the formation of a bond that can block polymerase processivity though the full RNA structure. Though non-covalent interactions that can cause this are

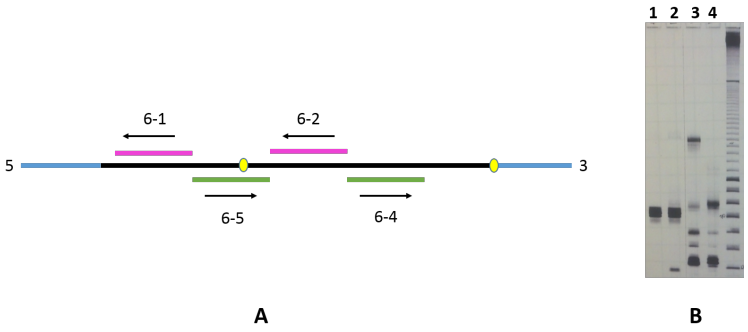


Fig. 5.7: (A) The sites of primer binding used in the structure probing experiment. Reverse primers are pink and forward primers are colored green. Constant regions are blue, and the proposed sites of the "pinch" are marked by yellow circles. (B) An 8% denaturing PAGE of PCR products from the structure probing priming experiment. Each lane is described in Figures 5.8-5.11. Lanes 1 and 3 result from RT-PCR of the purified top-most circular product, and lanes 2 and 4 result from the DBR1 treated purified circular product. The primers used in lanes 1 and 2 are thought to bind to the lower 68 base loop in the structure and the primers used in lanes 3 and 4 are thought to bind to the upper 79 base loop containing the end-to-end ligation site. The site of 5' to 2' end ligation is termed the branch point, and the site of the presumed "pinch" is called the pinch point in figures 5.8-5.11.

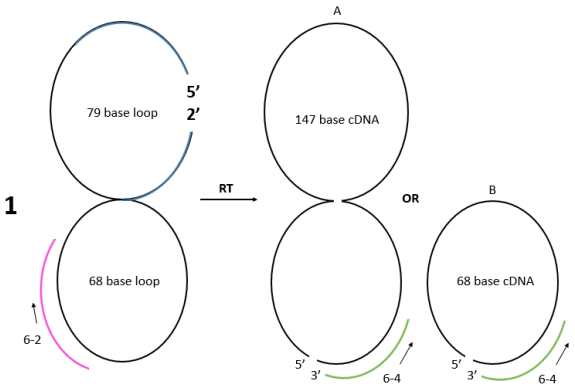


Fig. 5.8: In lane 1, the purified circular transcript was reverse transcribed with the primer 6-2 and PCR amplified with primers 6-2 and 6-4. The figure above illustrates the binding sites of the primers followed by possible RT products. If the reverse transcriptase can traverse the pinch site twice (A), full length PCR product should be visible, however there is only a very small amount on the gel. If, by the nature of the pinch point, one side is more difficult to be read through by the polymerase, it may stall and abort the strand. There is no full length product when priming from within the 68 base loop. Alternately, the polymerase may bypass the pinch junction and stay within the original loop (B). This would create a 68 base cDNA and PCR product, which is in fact seen on the gel.

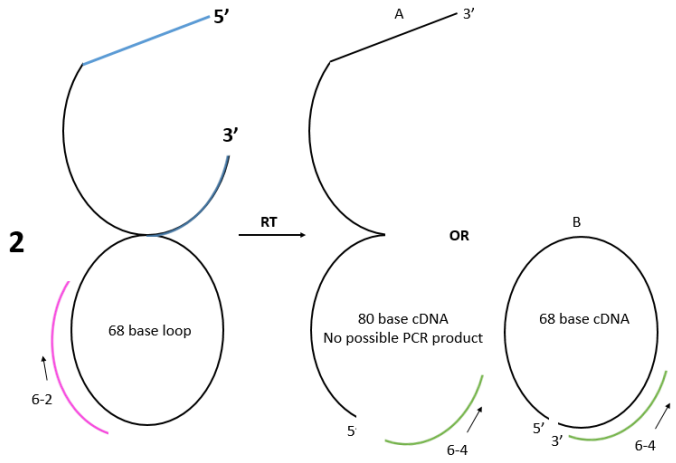


Fig. 5.9: In lane 2, the purified circular transcript was debranched, then reverse transcribed with the primer 6-2 and PCR amplified with primers 6-2 and 6-4. The figure above illustrates the binding sites of the primers followed by possible RT products. No PCR product can be made if the polymerase traverses the pinch point but the branch point is no longer intact (A). If the polymerase bypasses the pinch point, a 68 base cDNA and PCR product can be made. This is seen on the gel.

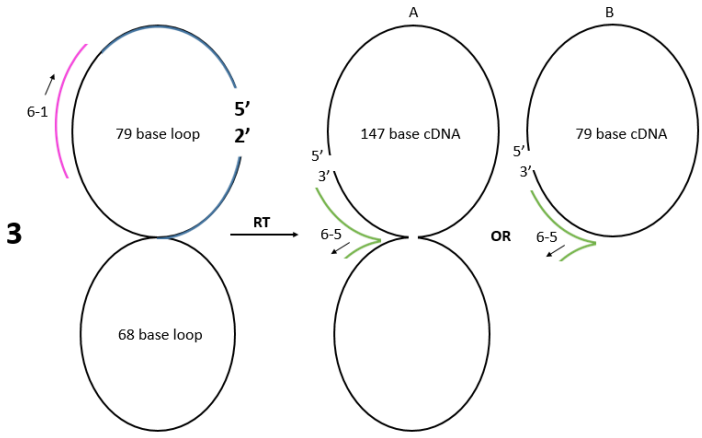


Fig. 5.10: In lane 3, the purified circular transcript was reverse transcribed with the primer 6-1 and PCR amplified with primers 6-1 and 6-5. The figure above illustrates the binding sites of the primers and possible RT products. Since the primers abut, PCR product should only occur if the reverse transcriptase can complete the full length cDNA, or complete one of the smaller loops to create the complement to the forward primer. In one possible scenario (A), the full length RT product can be obtained by the polymerase passing through the pinch point at least once, as enough cDNA may be complementary to 6-2 if the second time it encounters the pinch point it aborts. Surprisingly, full length PCR product can be seen in the gel. The second possible RT product (B) should be very difficult to amplify as the 3' of the primer is not complementary to this cDNA, however there is a PCR product that corresponds to this length.

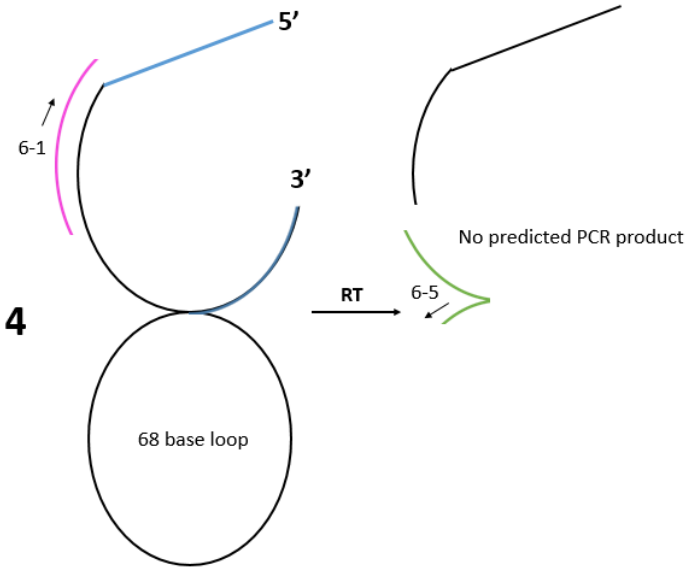


Fig. 5.11: In lane 4, the purified circular transcript was debranched with DBR1, then was reverse transcribed with the primer 6-1 and PCR amplified with primers 6-1 and 6-5. The figure above illustrates the binding sites of the primers and possible RT products. No full length cDNA and therefore no PCR product should be possible with this primer pair when the template is debranched. This is seen in the analytical gel. However, there is a dark band between 75 and 80 bases that corresponds to a lighter band in the analogous branched sample. This product cannot be explained by the current model.

probably possible, the fact that the purified product analyzed can be run on gels in denaturing conditions without appearance of linear product suggests this is a covalent bond being formed. In addition, it is clear this structure is a substrate for DBR1 based on the vastly different PCR products in a treated and untreated sample. The DBR1 experiment also strengthens the case for the formation of the covalent bond as the primer pair that bound to the lower loop's resulting PCR products were not affected by DBR1 treatment.

Based on the PCR products resulting from the primers used in the selection that bound on either end, if self-circularization was c6's only function, the sequence information would not make it beyond reverse transcription due to the 68 base deletion in the cDNA when the 3' primer is used. This sequence most likely only survived the selection because T4 DNA Ligase could form true circles with it, which result in full length cDNA. The experiments described in Figures 5.7-5.11 seem to support both the existence of a 5' to 2' branch point between the ends of c6, based on the DBR1 experiments, and the presence of another point within the circle that can block normal reverse transcription. The mechanism by which two different types of bonds can form is unknown, however dually catalytic RNA have been discovered before.[45] This RNA can catalyze ligation and cleavage reactions at different sites along its sequence, presumably made possible by two active conformations. In addition, the necessary catalytic sequences for these reactions are only a few nucleotides long.

Using similar RT-PCR experiments described earlier, the structure was probed further. Several new primer pairs were introduced that bound to different parts of the c6 sequence. The circular construct was reverse transcribed and PCR amplified, and the DNA products were analyzed by gel electrophoresis for patterns regarding termination sites and the ability of the primer pairs to produce full length product. Since reverse transcriptases are known to pause at difficult sites in the template, then either proceed or abort,[70] the reverse transcription reaction was allowed to proceed even longer to give the polymerase ample time to copy the strands. Astonishingly, termination sites could be traced to six AGC or AGT

triads along the sequence including the AGC site seen in the shorter experiment. All of these sites seemed to be bound to the same base as before, approximately C127. This experiment suggests the pinch between G58 and C59 seen previously, probed by the shorter reverse transcription time, may be the dominant species and the additional termination sites seen in this experiment are also pinched structures that occur less frequently. In fact, when the purified circular species was run on a higher percent gel, the once seemingly single band split into multiple bands, indicating multiple species were present during the experiment.

The only known four-way branch point in natural RNA is the RNA X species detected when U2 and U6 snRNA, the catalytic core of the splicesosome, are incubated with a consensus branch site oligonucleotide without splicososomal protein and 5' splice site. This forms a very low yield phosphotriester bond between the 3' phosphate of A53 of U6 and A21 2' hydroxyl of the branch site oligo (Fig.5.12).[71] Fascinatingly, an AGC triad is required for this reaction, which parallels the results from this c6 sequence. Though more research needs to be done to confirm the hypothesis, if c6 can in fact form a phosphotriester at the proposed pinch site it may serve as a stable, relatively high yield tool to study this phenomenon. Work is being done to isolate the four-way junction by nuclease digestion in preparation for mass spectrometry analysis.

Exactly how c6 is able to catalyze this type of reaction not only once but at multiple consensus sites across the sequence is unknown. There are eight AGC/T sites in c6 and six of them are seen in the structure probing experiments. There does not seem to be sequence similarity surrounding the AGC/T sites which would lead them to interact with the same catalytic core sequence. Several different tertiary structures may be possible with this sequence, however, and their distribution in the ensemble of folded c6 may directly relate to the distribution of figure-eight-like structures seen in the last experiment. The ability to form these different products, though, may have led to the sequence's domination of the selection population. While early partitioning steps selected for sequences that could

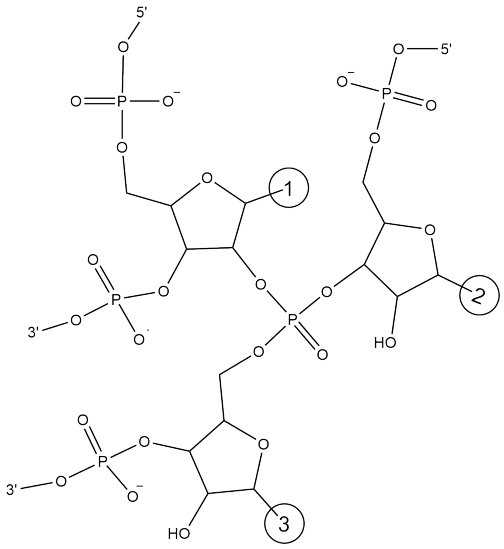


Fig. 5.12: A structure of a phosphotriester bond where the 2' hydroxyl of nucleobase 1 is bound to the phosphate between nucleobases 2 and 3.

circularize, as the gel percentage was changed to shift this population away from the area being excised, the ability of c6 to form a pinch site allowed it to have a gel mobility that still corresponded with the excision site, and therefore, continue to be selected.

The exact leaving group of RNA X is not known, but the authors suggest an H_2O molecule may serve as such.[71] In the case of this four-way junction, similar questions remain. The requirement of imidazole-UTP to make this structure, however, does offer an interesting synthetic route to test. A phosphorimidazolide bond can activate nucleotides to spontaneously polymerize in the presence of divalent cations.[37] They are facilely synthesized with triphenylphosphine, dipyridyl disulfide, triethylamine, imidazole and the mono-, di-, or triphosphate nucleotide of choice.

If the c6 sequence can catalyze a similar reaction where imidazole-U126 or the exogenous imidazole-UTP forms a phosphorimidazolide bond between G58 and C59 (Fig.5.13), the 2'OH of C127 may attack the phosphorimidazole, causing the imidazole to leave, forming a phosphotriester bond. The structure of c6 would be used to align the reactive groups, forgoing the need for a divalent cation. Alternately, the "pinch" site may be the stable

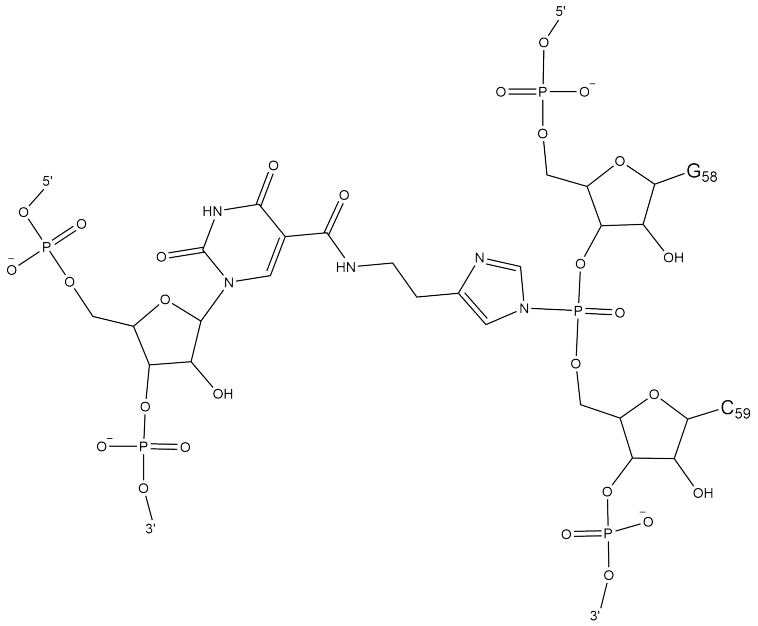


Fig. 5.13: A structure of a phosphorimidazolide bond that could be forming at the pinch site in c6. Imidazole modified U126, which is proximal to the proposed site of bond formation, can react with the backbone phosphate between G58 and C59. This could be the source of the stable pinched structure, or could serve as an intermediate, priming the phosphate for 2'hydroxyl attack to form a phosphotriester. Though not shown, the required exogenous imidazole-UTP could also be serving this intermediate role.

phosphorimidazolide to imidazole-U126 and the presence of magnesium may catalyze the cleavage reaction which was seen as inhibition in the assays. Additional future work should involve mutagenesis of these key residues and radiolabeling of the exogenous imidazole-UTP to determine if it is incorporated at any point in the reaction mechanism.

5.1 Materials and Methods

Circle enrichment from protease selection Since there was no selection pressure *specifically* for circles during the protease selection that resulted in their appearance, a final round was performed to enrich them. Round 17 from the protease selection was chosen as over half of the pool could be circularize in an overnight incubation with T4 DNA Ligase. After ligation, the non-linear transcripts were purified from an 6% gel, which separates them from not only their unligated linear predecessors but also the linear end-to-end ligated species. The RNA was eluted form the gel and primer captured with 3' primer-conjugated beads to concentrate and buffer exchange. They were then reverse transcribed at 50^oC with SuperScript III (Invitrogen) according to the protocol described in Chapter 2 with an 18 base primer that binds to the 5' end of the transcript and three bases of the 3' end. Its sequence was 5'-TGGTTTCCTCCCCTCTGC-3' The extra three bases of the 3' constant region were used so that the melting temperature of the reverse primer was high enough to bind at 50^oC, but still allow for enough sequence remaining in the 5' constant region in the cDNA for the forward primer to bind. The cDNA were amplified by PCR using the usual forward and reverse primers for the library. The forward primer with T7 promoter was 5'-TAATACGACTCACTATAGGGAGGAAACCACAGCCGAAGAAAU-3 and the reverse primer was 5'-CTCTGCGTGCCTGCCTGCTTG-3'. Both were purchased from Integrated DNA Technologies. The resulting dsDNA was inserted into the pCR^{TM}2.1-TOPO® vector with topoisomerase using the TOPO® TA Cloning® Kit(Life Technologies) and submitted to SeqWright for transformation, cloning, colony picking, and sequencing. Resulting sequences were initially analyzed by DOSA, and further grouped by visual analysis. These can be seen in Table 5.1

Select isolates were amplified through PCR from the plasmids, then transcribed and gel purified. Initial analysis of the isolates consisted of testing their circularization efficiency when catalyzed by T4 DNA Ligase according to the following protocol.

Circularizing Ligation

1 μM compliant RNA (5'monophosphate)

1x Ligase Buffer (*Invitrogen*) (50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% (w/v) PEG-8000)

2 mM ATP

10 mM DTT

0.075 U/μl T4 DNA Ligase (*Invitrogen*)

37^oC for at least 12 hours

Ligation products were analyzed by gel electrophoresis. Abnormal gel behavior was observed by comparing mobility in 6% vs 8% polyacrylamide gels.

To test the hypothesis that the RNA could be catalyzing at least a part of the ligation due to the presence of the gapped doubled stranded substrate, isolate c6 and c26 were chosen to study due to their high yields of circularized products. The random region sequence of c6 is 5'-TCG GAG CGG AAA GTC AGG TAG TGG AAC AGT ATA GCG TGC AGG CAC ACG AGA TTA AGT CGA TTG CCG GAT TTT CAG CTG GTG ACC TAC TCC CCG TAG CTC AT-3. The random region sequence of c26 is 5'-GTT CCG GAA TGC GAG GTA CGC CCC GGC TCT CCA TGG TAT CCC TTC TGC AAT CCG AGA CTC TAT GAA CCT TTC CGG CAG CGA CCT CCC TGG CTC NNN NNN NNG-3' Since the natural ligaton with T4 DNA Ligase proceeds via a 5' AMP adduct to the nicked double stranded nucleic acid, an AppG mimic of this moiety was synthesized according to a protocol first described by Leslie Orgel.[37]

The detailed synthetic procedure was obtained from the McManus Lab's (UCSF) online protocol database. In brief, a mixure of 5'-AMP in DMF was added to a stirred solution of triphenylphosphine, dipyridyldisulfide, triethylamine, and imidazole in DMF. The formed adenosine phosphorimidazolide (ImpA) was precipitated in sodium perchlorate, acetone and

anhydrous ethyl ether and pelleted. The precipitate was washed with acetone and ethyl ether, then dried under vacuum. The ImpA was added to 5'-GMP in MOPS in the presence of magnesium to form the AppG dinucleotide.[37] The product was analyzed by phosphorous-31 NMR.

The AppG was incorporated into the 5' end of the c6 and c26 by in vitro transription with T7 RNA polymerase,[29] using 30 mM AppG and 1 mM NTPs. The same sequences were also transcribed using 1 mM NTPs to form a 5' GTP, and with 1 mM NTPs and 30 mM GMP to form the 5'-GMP. The transcription protocol is outlined in chapter 2.

After transcription, the replicons were purified with a 6% polyacrylamide gel. Upon analysis in an 8% gel, species with decreased mobility were seen. Upon analysis in a 6% gel, these species demonstrated the abnormal mobility that is characteristic of non-linear RNA by shifting back near the linear species.

Experiments identifying the necessary components for non-linear RNA formation The non-linear RNA was being formed sometime between transcription and gel purification. To discover at what point the reaction was happening, the c6 transcript was incubated in all the conditions it had come into contact with such as full transcription mix, gel running buffer, and elution buffer. The transcription mix was the only condition that yielded gel shifted material and was therefore the conditions in which the reaction took place.

Each component of the transcription reaction was singly removed during overnight incubations at 37^{o}C with 1 uM c6 RNA (10 ul reactions). Upon 8% denaturing gel analysis, it was shown that no one component could abrogate the gel shift in its absence. Performing these same incubation conditions with only one component at a time using their concentrations used in transcription, in addition to the 40 mM Tris-HCl pH 8 buffer, revealed that the imidazole-UTP and GTP could, separately, both induce non-linear RNA formation. To test metal dependence, 1 uM c6 was incubated with imidazole-UTP and GTP with and without the magnesium chloride or with and without sodium chloride. The absence of sodium had no

effect on the reaction. However, magnesium had an inhibitory effect on the reaction with the imidazole and a stimulatory effect on the reaction with GTP. These incubation experiments were also performed, in the absence of magnesium, with 4 mM UTP, imidazole-dUTP, and pyridyl-UTP and analyzed by 8% gel electrophoresis.

pH dependence assay To determine if c6 has a pH dependence, possibly brought on by activation of a titratable functional group, the pH of the Tris-HCl buffer was varied. In 10 ul reactions, 1 uM RNA were mixed with 4 mM imidazole-UTP in 40 mM Tris-HCl of pH 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, and 9.5. These were incubated overnight at 37^oC then analyzed on 8% denaturing polyacrylamide gels.

Sequencing of c6 PCR products When PCR amplified with the forward and reverse primers used in the selection, purified non-linear product of c6 would result in an approximately 90 base amplicon. This was purified from a 10% gel and submitted to SeqWright for sequencing in both the forward and reverse direction. Of note, when the amplicon was gel purified, it became evident it was actually made of about three individual amplicons, one of these was purified and submitted to sequencing. The sequencing chromatagrams were analyzed using FinchTV 1.3.1 software.

Loop-dependent RT-PCR experiments The additional primer sequences used in these experiments are as follows:

6-1: 5'-TCCCACTACCTGACTTTC-3'

6-2: 5'-GACTTAATCTCGTGTGCC-3'

6-3: 5'-ATGAGCTACGGGGAGTAG-3'

6-4: 5'-GATTGCCGGATTTTCAGC-3'

6-4: 5'-CAGTATAGCGTGCAGGCA-3'

Purified non-linear RNA from the top-most product of c6 was treated with debranching enzyme (DBR1) as follows. 1 ug of non-linear RNA was incubated with 1 ul recombinant DBR1 (obtained from Thomas Menees, UMKC) in 50 mM Tris-HCl pH 7.0, 4 mM $MnCl_2$,

2.5 mM DTT, 25 mM NaCl, 0.01% Triton-X-100, 0.1 mM EDTA and 0.15% glycerol. This was incubated at 30^{o}C for 1 hour.[39] The reaction was buffer exchanged into 10 mM Tris-HCl pH 7.5 using Bio-Gel P6 size exclusion columns (Bio-Rad), then added to the following reverse transcription reactions.

For each reaction, the reverse primer was added to the untreated or DBR1-treated purified c6 non-linear product in First Strand buffer (Invitrogen) and heated at 75^{o}C for 3 min, then passively cooled to room temperature to anneal the primer. The remaining reagents of the reverse transcription reaction, described in chapter 2, were added to a total volume of 12 ul each and incubated at 50^{o}C for 1 hour. 6 ul of these reactions were added to 50 ul PCR reactions, described in chapter 2, with their respective forward primer and amplified with KOD XL polymerase. Aliquots of the reaction were analyzed by denaturing gel electrophoresis.

BIBLIOGRAPHY

[1] Jeremy J Agresti, Eugene Antipov, Adam R Abate, Keunho Ahn, Amy C Rowat, Jean-Christophe Baret, Manuel Marquez, Alexander M Klibanov, Andrew D Griffiths, and David a Weitz. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(9):4004–9, March 2010.

[2] Paul F Agris. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO reports*, 9(7):629–35, July 2008.

[3] Iraj K Ali, Laura Lancaster, Jason Feinberg, Simpson Joseph, and Harry F Noller. Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Molecular cell*, 23(6):865–74, September 2006.

[4] Sandeep Ameta and Andres Jäschke. An RNA catalyst that reacts with a mechanistic inhibitor of serine proteases. *Chemical Science*, 4(3):957, 2013.

[5] Sandeep Ameta, Marie-Luise Winz, Christopher Previti, and Andres Jäschke. Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic acids research*, 42(2):1303–10, January 2014.

[6] Jared R Auclair, Kristin J Boggio, Gregory a Petsko, Dagmar Ringe, and Jeffrey N Agar. Strategies for stabilizing superoxide dismutase (SOD1), the protein destabilized in the most common form of familial amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50):21394–9, December 2010.

[7] Scott Baskerville and David P Bartel. A ribozyme that ligates RNA to protein. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14):9154–9, July 2002.

[8] Joseph S Beckman, Alvaro G Estévez, John P Crow, and Luis Barbeito. Superoxide dismutase and the death of motoneurons in ALS. *Trends in Neurosciences*, 24(11):S15–S20, November 2001.

[9] Daniel Blankenberg, Assaf Gordon, Gregory Von Kuster, Nathan Coraor, James Taylor, and Anton Nekrutenko. Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)*, 26(14):1783–5, July 2010.

[10] Benjamin M Brandsen, Anthony R Hesser, Marissa a Castner, Madhavaiah Chandra, and Scott K Silverman. DNA-catalyzed hydrolysis of esters and aromatic amides. *Journal of the American Chemical Society*, 135(43):16014–7, October 2013.

[11] Andrew R Buller and Craig a Townsend. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8):E653–61, February 2013.

[12] R C Cadwell and G F Joyce. Randomization of genes by PCR mutagenesis. *Genome Research*, 2(1):28–33, August 1992.

[13] B Capel, a Swain, S Nicolis, a Hacker, M Walter, P Koopman, P Goodfellow, and R Lovell-Badge. Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell*, 73(5):1019–30, June 1993.

[14] Justin W. Chan, Charles E. Hoyle, Andrew B. Lowe, and Mark Bowman. Nucleophile-Initiated Thiol-Michael Reactions: Effect of Organocatalyst, Thiol, and Ene. *Macromolecules*, 43(15):6381–6388, August 2010.

[15] Peter J a Cock, Christopher J Fields, Naohisa Goto, Michael L Heuer, and Peter M Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, 38(6):1767–71, April 2010.

[16] Charles S Craik, Michael J Page, and Edwin L Madison. Proteases as therapeutics. *The Biochemical journal*, 435(1):1–16, April 2011.

[17] Xiaochang Dai, Alain De Mesmaeker, and Gerald F Joyce. Cleavage of an Amide Bond by a Ribozyme. *Science*, 267:237–240, 1995.

[18] Mauro C. DalCanto and Mark E. Gurney. Neuropathological changes in two lines of mice carrying a transgene for mutant human Cu,Zn SOD, and in mice overexpressing wild type human SOD: a model of familial amyotrophic lateral sclerosis (FALS). *Brain Research*, 676(1):25–40, April 1995.

[19] B E Eaton. The joys of in vitro selection: chemically dressing oligonucleotides to satiate protein targets. *Current opinion in chemical biology*, 1(1):10–16, 1997.

[20] Eric Ekland, Jack Szostak, and David Bartel. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science*, 269(5222):364–370, 1995.

[21] Adrian R FerreDAmare, Kaihong Zhou, and Jennifer A Doudna. Crystal structure of a hepatitis delta virus ribozyme. *Nature*, 395:567–574, 1998.

[22] Amber Flynn-Charlebois, Tracey K Prior, Kelly A Hoadley, and Scott K Silverman. In Vitro Evolution of an RNA-Cleaving DNA Enzyme into an RNA Ligase Switches the Selectivity from 3 $'$ - 5 $'$ to 2 $'$ - 5 $'$. *J. Am. Chem. Soc.*, 125:5346–5350, 2003.

[23] Stefan Fusz, Alexander Eisenführ, Seergazhi G Srivatsan, Alexander Heckel, and Michael Famulok. A ribozyme for the aldol reaction. *Chemistry & biology*, 12(8):941–950, August 2005.

[24] Ahmad Galaleldeen, Richard W Strange, Lisa J Whitson, Svetlana V Antonyuk, Narendra Narayana, Alexander B Taylor, Jonathan P Schuermann, Stephen P Holloway, S Samar Hasnain, and P John Hart. Structural and biophysical properties of metal free pathogenic SOD1 mutants A4V and G93A. *Archives of biochemistry and biophysics*, 492(1-2):40–7, December 2009.

[25] Ghanashyam D Ghadge, John D Pavlovic, Sujatha P Koduvayur, Brian K Kay, and Raymond P Roos. Single chain variable fragment antibodies block aggregation and toxicity induced by familial ALS-linked mutant forms of SOD1. *Neurobiology of disease*, 56:74–8, August 2013.

[26] Larry Gold, Deborah Ayers, Jennifer Bertino, Christopher Bock, Ashley Bock, Edward N Brody, Jeff Carter, Andrew B Dalby, Bruce E Eaton, Tim Fitzwater, Dylan Flather, Ashley Forbes, Trudi Foreman, Cate Fowler, Bharat Gawande, Meredith Goss, Magda Gunn, Shashi Gupta, Dennis Halladay, Jim Heil, Joe Heilig, Brian Hicke, Gregory Husar, Nebojsa Janjic, Thale Jarvis, Susan Jennings, Evaldas Katilius, Tracy R Keeney, Nancy Kim, Tad H Koch, Stephan Kraemer, Luke Kroiss, Ngan Le, Daniel Levine, Wes Lindsey, Bridget Lollo, Wes Mayfield, Mike Mehan, Robert Mehler, Sally K Nelson, Michele Nelson, Dan Nieuwlandt, Malti Nikrad, Urs Ochsner, Rachel M Ostroff, Matt Otis, Thomas Parker, Steve Pietrasiewicz, Daniel I Resnicow, John Rohloff, Glenn Sanders, Sarah Sattin, Daniel Schneider, Britta Singer, Martin Stanton, Alana Sterkel, Alex Stewart, Suzanne Stratford, Jonathan D Vaught, Mike Vrkljan, Jeffrey J Walker, Mike Watrobka, Sheela Waugh, Allison Weiss, Sheri K Wilcox, Alexey Wolfson, Steven K Wolk, Chi Zhang, and Dom Zichi. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS one*, 5(12):e15004, January 2010.

[27] Francois Gros-Louis, Geneviève Soucy, Roxanne Larivière, and Jean-Pierre Julien. Intracerebroventricular infusion of monoclonal antibody or its derived Fab fragment

against misfolded forms of SOD1 mutant delays mortality in a mouse model of ALS. *Journal of neurochemistry*, 113(5):1188–99, June 2010.

[28] Nancy S Gutgsell, Murray P Deutscher, and James Ofengand. The pseudouridine synthase RluD is required for normal ribosome assembly and function in Escherichia coli. *RNA*, 11:1141–1152, 2005.

[29] Alicia J Hager and Jack W Szostak. Isolation of novel ribozymes that ligate AMP-activated RNA substrates pool RNA. *Chemistry & biology*, 4:607–617, 1997.

[30] Thomas B Hansen, Trine I Jensen, Bettina H Clausen, Jesper B Bramsen, Bente Finsen, Christian K Damgaard, and Jø rgen Kjems. Natural RNA circles function as efficient microRNA sponges. *Nature*, 495(7441):384–8, March 2013.

[31] Lizbeth Hedstrom. Serine protease mechanism and specificity. *Chemical reviews*, 102(12):4501–24, December 2002.

[32] Mark Helm. Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic acids research*, 34(2):721–33, January 2006.

[33] Marcel Hollenstein, Christopher J Hipolito, Curtis H Lam, and David M Perrin. A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M2+). *Nucleic acids research*, 37(5):1638–49, April 2009.

[34] Peng Huang, Li Feng, Elizabeth A Oldham, Michael J Keating, and William Plunkett. Superoxide dismutase as a target for the selective killing of cancer cells. *Nature*, 407(September), 2000.

[35] William R Jeck, Jessica a Sorrentino, Kai Wang, Michael K Slevin, Christin E Burd,

Jinze Liu, William F Marzluff, and Norman E Sharpless. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA*, 19(2):141–57, February 2013.

[36] K B Jensen, B L Atkinson, M C Willis, T H Koch, and L Gold. Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. *Proceedings of the National Academy of Sciences of the United States of America*, 92(26):12220–4, December 1995.

[37] G. F. Joyce, T. Inoue, and L. E. Orgel. Non-enzymatic Template-directed Synthesis on RNA Random Copolymers. *J. Mol. Biol.*, 176:279–306, 1984.

[38] Amit S Kalgutkar and Deepak K Dalvie. Drug discovery for a new generation of covalent drugs. *Expert opinion on drug discovery*, 7(7):561–81, July 2012.

[39] M Fahad Khalid, Masad J Damha, Stewart Shuman, and Beate Schwer. Structure-function analysis of yeast RNA debranching enzyme (Dbr1), a manganese-dependent phosphodiesterase. *Nucleic acids research*, 33(19):6349–60, January 2005.

[40] Joohoon Kim and Milan Mrksich. Profiling the selectivity of DNA ligases in an array format with mass spectrometry. *Nucleic acids research*, 38(1):e2, January 2010.

[41] Robin D Knight and Laura F Landweber. Rhyme or reason : RNA-arginine genetic code interactions and the. *Chemistry & biology*, 5(9):R215–R220, 1998.

[42] Zachary a Knight, Jennifer L Garrison, Karina Chan, David S King, and Kevan M Shokat. A remodelled protease that cleaves phosphotyrosine substrates. *Journal of the American Chemical Society*, 129(38):11672–3, September 2007.

[43] K Kruger, P J Grabowski, a J Zaug, J Sands, D E Gottschling, and T R Cech. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, 31(1):147–157, 1982.

[44] Hugo Lachance, Stefan Wetzel, Kamal Kumar, and Herbert Waldmann. Charting, navigating, and populating natural product chemical space for drug discovery. *Journal of medicinal chemistry*, 55(13):5989–6001, July 2012.

[45] L. F. Landweber and I. D. Pokrovskaya. Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 96(1):173–8, January 1999.

[46] I. R. Lehman. DNA ligase: structure, mechanism, and function. *Science (New York, N.Y.)*, 186(4166):790–7, November 1974.

[47] Guang-Zhao Li, Rajan K. Randev, Alexander H. Soeriyadi, Gregory Rees, Cyrille Boyer, Zhen Tong, Thomas P. Davis, C. Remzi Becer, and David M. Haddleton. Investigation into thiol-(meth)acrylate Michael addition reactions using amine and phosphine catalysts. *Polymer Chemistry*, 1(8):1196, 2010.

[48] Debabrata Mandal, Caroline Köhrer, Dan Su, Susan P Russell, Kady Krivos, Colette M Castleberry, Paul Blum, Patrick a Limbach, Dieter Söll, and Uttam L RajBhandary. Agmatidine, a modified cytidine in the anticodon of archaeal tRNA(Ile), base pairs with adenosine but not with guanosine. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7):2872–7, March 2010.

[49] Sebastian Memczak, Marvin Jens, Antigoni Elefsinioti, Francesca Torti, Janna Krueger, Agnieszka Rybak, Luisa Maier, Sebastian D Mackowiak, Lea H Gregersen, Mathias Munschauer, Alexander Loewer, Ulrike Ziebold, Markus Landthaler, Christine Kocks, Ferdinand le Noble, and Nikolaus Rajewsky. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–8, March 2013.

[50] Brett M. Morrison and John H. Morrison. Amyotrophic lateral sclerosis associated

with mutations in superoxide dismutase: a putative mechanism of degeneration. *Brain Research Reviews*, 29(1):121–135, January 1999.

[51] Meredith I Newby and Nancy L Greenbaum. Sculpting of the spliceosomal branch site recognition motif by a conserved pseudouridine. *Nature structural biology*, 9(12):958–65, December 2002.

[52] Eugene W M Ng, David T Shima, Perry Calias, Emmett T Cunningham, David R Guyer, and Anthony P Adamis. Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. *Nature reviews. Drug discovery*, 5(2):123–32, February 2006.

[53] Poul Nissen, Jeffrey Hansen, Nenad Ban, Peter Moore, and Thomas Steitz. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science*, 289(5481):920–930, August 2000.

[54] Anne Perrotta, Ihung Shih, and Michael Been. Imidazole Rescue of a Cytosine Mutation in a Self-Cleaving Ribozyme. *Science*, 286(5437):123–126, October 1999.

[55] Mark Pogson, George Georgiou, and Brent L Iverson. Engineering next generation proteases. *Current opinion in biotechnology*, 20(4):390–7, August 2009.

[56] Anna Radzicka and Richard Wolfenden. Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases. *J. Am. Chem. Soc.*, 118:6105–6109, 1996.

[57] Rishi Rakhit, John P Crow, James R Lepock, Leslie H Kondejewski, Neil R Cashman, and Avijit Chakrabartty. Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *The Journal of biological chemistry*, 279(15):15499–504, April 2004.

[58] Daniel Rosen, Teepu Siddique, David Patterson, Denise Figlewicz, Peter Sapp, Afif Hentati, Deirdre Donaldson, Jum Goto, Jeremiah ORegan, HanXiang Deng, Zohra Rahmani, Aldis Krizus, Diane McKenna, Annarueber Cayabyab, Sandra Gaston, Ralph Berger, Rudolph Tanzi, John Halperin, Brian Herzfeldt, Raymond VandenBergh, WuYen Hung, Thomas Bird, Gang Deng, Donald Mulder, Celestine Smyth, Nigel Laing, Edwin Soriano, Margaret Pericak, Jonathan Haines, Guy Rouleau, James Gusella, Robert Horvitz, and Robert Brown. © 19 9 3 Nature Publishing Group. *Nature*, 362:59–62, 1993.

[59] R Rossi, a Montecucco, G Ciarrocchi, and G Biamonti. Functional characterization of the T4 DNA ligase: a new insight into the mechanism of action. *Nucleic acids research*, 25(11):2106–13, June 1997.

[60] G Sengle, a Eisenführ, P S Arora, J S Nowick, and M Famulok. Novel RNA catalysts for the Michael reaction. *Chemistry & biology*, 8(5):459–473, 2001.

[61] Bryan F Shaw and Joan Selverstone Valentine. How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends in biochemical sciences*, 32(2):78–85, March 2007.

[62] Juswinder Singh, Russell C Petter, Thomas a Baillie, and Adrian Whitty. The resurgence of covalent drugs. *Nature reviews. Drug discovery*, 10(4):307–17, April 2011.

[63] B C Stark, R Kole, E J Bowman, and S Altman. Ribonuclease P: an enzyme with an essential RNA component. *Proceedings of the National Academy of Sciences of the United States of America*, 75(8):3717–3721, 1978.

[64] P B Stathopulos, J A O Rumfeldt, G A Scholz, R A Irani, H E Frey, R A Hallewell, J R Lepock, and E M Meiering. Cu Zn superoxide dismutase mutants associated with

amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *PNAS*, 100(12):7021–7026, 2003.

[65] Robert M Stroud and Janet S Finer-moore. Current Topics Conformational Dynamics along an Enzymatic Reaction Pathway : Thymidylate Synthase , " the Movie " †. *Biochemistry*, 42(2):239–47, 2003.

[66] Theodore M. Tarasow and Bruce E. Eaton. Dressed for success: Realizing the catalytic potential of RNA. *Biopolymers*, 48(1):29–37, 1998.

[67] TM Tarasow, SL Tarasow, and BE Eaton. RNA catalysed carbon carbon bond formation. *Nature*, 389(SEPTEMBER):2–5, 1997.

[68] D. K. Treiber. Kinetic Intermediates Trapped by Native Interactions in RNA Folding. *Science*, 279(5358):1943–1946, March 1998.

[69] J Trepanier, J E Tanner, R L Momparler, O N L Le, F Alvarez, and C Alfieri. Cleavage of intracellular hepatitis C RNA in the virus core protein coding region by deoxyribozymes. *Journal of viral hepatitis*, 13(2):131–8, February 2006.

[70] T Tuschl, P a Sharp, and D P Bartel. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *The EMBO journal*, 17(9):2637–50, May 1998.

[71] S Valadkhan and J L Manley. Splicing-related catalysis by protein-free snRNAs. *Nature*, 413(6857):701–7, October 2001.

[72] Saba Valadkhan and James L Manley. Characterization of the catalytic activity of U2 and U6 snRNAs. *RNA*, 9:892–904, 2003.

[73] Jonathan D Vaught, Chris Bock, Jeff Carter, Tim Fitzwater, Matt Otis, Dan Schneider, Justin Rolando, Sheela Waugh, Sheri K Wilcox, and Bruce E Eaton. Expanding the

chemistry of DNA for in vitro selection. *Journal of the American Chemical Society*, 132(12):4141–51, March 2010.

[74] Jonathan D Vaught, Torin Dewey, and Bruce E Eaton. T7 RNA polymerase transcription with 5-position modified UTP derivatives. *Journal of the American Chemical Society*, 126(36):11231–11237, 2004.

[75] Annette O Walter, Robert Tjin Tham Sjin, Henry J Haringsma, Kadoaki Ohashi, Jing Sun, Kwangho Lee, Aleksandr Dubrovskiy, Matthew Labenski, Zhendong Zhu, Zhigang Wang, Michael Sheets, Thia St Martin, Russell Karp, Dan van Kalken, Prasoon Chaturvedi, Deqiang Niu, Mariana Nacht, Russell C Petter, William Westlin, Kevin Lin, Sarah Jaw-Tsai, Mitch Raponi, Terry Van Dyke, Jeff Etter, Zoe Weaver, William Pao, Juswinder Singh, Andrew D Simmons, Thomas C Harding, and Andrew Allen. Discovery of a mutant-selective covalent inhibitor of EGFR that overcomes T790M-mediated resistance in NSCLC. *Cancer discovery*, 3(12):1404–15, December 2013.

[76] B Ward and T Juehne. Combinatorial library diversity: probability assessment of library populations. *Nucleic acids research*, 26(4):879–86, February 1998.

[77] Matthew E Welsch, Scott a Snyder, and Brent R Stockwell. Privileged scaffolds for library design and drug discovery. *Current opinion in chemical biology*, 14(3):347–61, June 2010.

[78] T W Wiegand, R C Janssen, and B E Eaton. Selection of RNA amide synthases. *Chemistry & biology*, 4(9):675–683, 1997.

[79] Leah Winer, Dushyanth Srinivasan, Seung Chun, David Lacomis, Matthew Jaffa, Anne Fagan, David M Holtzman, Ed Wancewicz, C Frank Bennett, Robert Bowser, Merit Cudkowicz, and Timothy M Miller. SOD1 in cerebral spinal fluid as a pharmacodynamic

marker for antisense oligonucleotide therapy. *JAMA neurology*, 70(2):201–7, February 2013.

[80] J C Wu and D V Santi. Kinetic and catalytic mechanism of HhaI methyltransferase. *The Journal of biological chemistry*, 262(10):4778–86, April 1987.

[81] X Xie, Y Gu, T Fox, J T Coll, M a Fleming, W Markland, P R Caron, K P Wilson, and M S Su. Crystal structure of JNK3: a kinase implicated in neuronal apoptosis. *Structure (London, England : 1993)*, 6(8):983–91, August 1998.