Title
In vivo selection and in vivo evolution of a trans-splicing ribozyme

Permalink
https://escholarship.org/uc/item/4r26j4rt

Author
Olson, Karen Elizabeth

Publication Date
2012

Peer reviewed|Thesis/dissertation
In vivo selection and in vivo evolution of a trans-splicing ribozyme

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Chemistry

by

Karen E. Olson

Committee in charge:
Professor Ulrich Müller, Chair
Professor Thomas Hermann
Professor Simpson Joseph
Professor Stanley Opella
Professor Lorraine Pillus

2012
The Dissertation of Karen E. Olson is approved, and is acceptable in quality and form for publication on microform and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

Chair

University of California, San Diego

2012
Table of Contents

Signature Page ......................................................................................... iii

Table of Contents ...................................................................................... iv

List of Figures ............................................................................................. vi

List of Tables ............................................................................................... ix

Acknowledgements .................................................................................... x

Vita .............................................................................................................. xii

Abstract of the Dissertation ....................................................................... xiii

Chapter 1  Introduction .............................................................................. 1
  1.1 Discovery of the *Tetrahymena thermophila* catalytic RNA ................ 1
  1.2 Diversity of group I intron ribozymes .............................................. 3
  1.3 Determining the structure ................................................................. 6
  1.4 Group I intron *trans*-splicing ......................................................... 10
  1.5 Improvements in ribozyme efficiency through sequence optimization .... 15
  1.6 Gene therapy by mRNA repair ......................................................... 18
  1.7 Limitations of current *trans*-splicing group I introns ....................... 19
  1.8 Goals of this thesis ........................................................................... 21
  1.9 References .......................................................................................... 23

Chapter 2  Computational prediction of efficient splice sites for *trans*-splicing ribozymes ......................................................... 29
  2.1 Abstract .............................................................................................. 29
  2.2 Introduction .......................................................................................... 30
  2.3 Results .................................................................................................. 33
2.4 Discussion.................................................................................................................. 55
2.5 Materials and Methods............................................................................................. 60
2.6 Acknowledgements................................................................................................. 67
2.7 Supporting Information........................................................................................... 68
2.8 References............................................................................................................... 72

Chapter 3  An in vivo selection method to optimize trans-splicing ribozymes..... 79

3.1 Abstract................................................................................................................... 79
3.2 Introduction............................................................................................................. 80
3.3 Results.................................................................................................................... 83
3.4 Discussion.............................................................................................................. 98
3.5 Materials and Methods........................................................................................ 101
3.6 Acknowledgements.............................................................................................. 107
3.7 Supporting Information......................................................................................... 109
3.8 References............................................................................................................ 113

Chapter 4  Evolution of group I intron internal sequences for high trans-splicing efficiency in vivo ................................................................. 117

4.1 Introduction............................................................................................................ 117
4.2 Results................................................................................................................... 121
4.3 Discussion............................................................................................................. 132
4.4 Acknowledgements............................................................................................. 133
4.5 References............................................................................................................ 134

Chapter 5  Future Directions..................................................................................... 136

Appendix ...................................................................................................................... 145
List of Figures

**Figure 1.1:** Intron removal from a primary transcript. ........................................ 2

**Figure 1.2:** Comparison of host and group I intron phylogeny. ......................... 6

**Figure 1.3:** *Tetrahymena thermophila* group I intron secondary structure........ 8

**Figure 1.4:** Intron core structure overlap. .......................................................... 9

**Figure 1.5:** *Cis* and *trans*-splicing diagrams. .............................................. 11

**Figure 2.1:** Interactions within and between the mRNA substrate and the ribozyme, and their treatment during computation ......................... 36

**Figure 2.2:** Computed values of $\Delta G_{\text{bind}}$ for splice sites on CAT mRNA .... 38

**Figure 2.3:** *Trans*-splicing efficiency measured on 18 splice sites on CAT mRNA .................................................................................................................. 41

**Figure 2.4:** *Trans*-splicing efficiency with short, 13-nucleotide substrates .... 44

**Figure 2.5:** Energetic contributions to the computed binding free energy on CAT mRNA .............................................................................................................. 48

**Figure 2.6:** Comparison between *trans*-splicing efficiency and *trans*-tagging results .............................................................................................................. 50

**Figure 2.7:** Transcription efficiencies of ribozymes analyzed in this study ...... 53

**Figure S2.1:** Energetic contributions to $\Delta G_{\text{bind}}$ from the three molecular events for the short, 13-nucleotide substrates .............................................. 70
Figure S2.2: Correlation between $trans$-splicing efficiencies experimentally determined with CAT mRNA substrate and the three individual energetic contributions to $\Delta G_{\text{bind}}$ ................................................................. 71

Figure 3.1: Secondary structure of the $trans$-splicing ribozyme construct used in this study, and its change during $trans$-splicing ....................... 81

Figure 3.2: Design of the pool of EGSs for the $in$ $vivo$ selection ............... 85

Figure 3.3: Schematic for the work flow of the $in$ $vivo$ selection procedure .... 88

Figure 3.4: Enrichment of sub-pools in the ribozyme library, during 5 rounds of $in$ $vivo$ selection ................................................................. 90

Figure 3.5: Correlation between the measurements of bacterial growth with two methods ............................................................................. 94

Figure 3.6: Predicted secondary structures .............................................. 96

Figure 3.7: Correlation between two methods measuring $trans$-splicing efficiency ........................................................................... 97

Figure S3.1: Predicted secondary structures between EGSs identified an earlier selection’ and the mRNA target site ..................................... 112

Figure S3.2: Library plasmid for the $in$ $vivo$ selection of ribozymes that can repair the mRNA of the chloramphenicol acetyltransferase (cat) gene . 113

Figure 4.1: Steps of group I intron $cis$-splicing ........................................ 118

Figure 4.2: Conversion of the $cis$-splicing group I intron into a $trans$-splicing intron ............................................................................. 132
Figure 4.3: Plasmid map of the substrate and ribozyme used for each round of the selection………………………………………………………………………………….122

Figure 4.4: Average number of mutations per sequence over the 21 rounds of evolution……………………………………………………………………………..124

Figure 4.5: Doubling times of the best selected ribozyme (R21C24) and knockout variants in E. coli…………………………………………………………….127

Figure 4.6: P6b and P8 stem loop mutations and their effects on the in vivo activity of the ribozyme…………………………………………………………….129

Figure 4.7: Schematic of protein pull-down by biotinylated RNA………………..130

Figure 4.8: Silver stained SDS PAGE of proteins separated through the pull down assay ………………………………………………………………………….131

Figure 4.9: Proposed role of 30S ribosomal protein S1 in increasing the trans-splicing efficiency of the evolved R21C24 ribozyme ……………………..133

Figure 5.1: Summary of the results from the previous in vivo selection and in vivo evolution ……………………………………………………………………137

Figure 5.2: Schematic of additional group I intron splicing reactions…………..141
List of Tables

**Table 2.1:** Results of the *trans*-tagging assay .......................................................... 43

**Table S2.1:** Experimentally tested splice sites on CAT mRNA .................. 68

**Table S2.2:** Energetic contributions to $\Delta G_{\text{bind}}$ for the splice sites tested on CAT mRNA and on the short, 13-nucleotide substrates .......................... 69

**Table 3.1:** Extended Guide Sequences (EGSs) that were used in this study 92

**Table S3.1:** Splice sites identified using the *trans*-tagging assay .............. 109

**Table S3.2:** Number of clones identified from each sub-pool per round ...... 109

**Table S3.3:** PCR primer sequences ................................................................. 110-111

**Table 4.1:** Concentration of chloramphenicol ($\mu$g/mL) used in the selection plates by round .......................................................... 125

**Table 4.2:** Results of mass spectrometry analysis ................................. 132
Acknowledgements

There have been many individuals responsible for inspiring, encouraging, and challenging me toward a PhD in chemistry. From my early public school science teachers through college professors who gave me the space to perform my own experiments. I owe a debt of gratitude to those who took extra time or quietly acknowledged and fostered my affinity for science.

The Müller, Toor, Joseph, Yang and Tezcan lab members are thanked for sharing supplies, instruments, lending a hand and offering helpful advice over the years. Thank you for sharing the joy and frustrations of research with me.

The support from my family and friends over the last five years is greatly appreciated. From coffee breaks to happy hours, phone conversations to care packages, I would not have made it here without your encouragement.

Lastly, but certainly not least, I would like to thank my advisor Ulrich Müller. It has been a rewarding adventure starting both a new lab and new research project with you. Your combination of experience, guidance, passion and patience is rare. Thank you for teaching and challenging me while allowing me to grow as a scientist.

Chapter 2, in full, is a reprint of the material as it appears in RNA, Meluzzi, D., K.E. Olson, G.F. Dolan, G. Ayra, and U.F. Müller. RNA, 2012. 18(3): p. 590-602. The dissertation author was the second author on this paper.
Chapter 3, in full, is a reprint of the material as it appears in RNA, Olson, K.E. and U.F. Müller. RNA, 2012. 18(3): p. 581-9. The dissertation author was the first author on this paper.

Chapter 4, in part, is currently being prepared for submission for publication, Olson, K.E. and U.F. Müller. The dissertation author was the first author on this paper.

This research was funded in part by the National Institutes of Health Hemoglobin and Blood Protein Chemistry training grant.
Vita

University of Wisconsin, La Crosse
2006 Bachelors of Science in Chemistry (ACS Certified)

University of California, San Diego
2008 Master of Science in Chemistry
2012 Doctor of Philosophy in Chemistry

PUBLICATIONS


Abstract of the Dissertation

*In vivo* selection and *in vivo* evolution of a *trans*-splicing ribozyme

by

Karen E. Olson

Doctor of Philosophy in Chemistry

University of California, San Diego 2012

Professor Ulrich Müller, Chair

Group I introns are catalytic RNAs (ribozymes) that are capable of self-splicing out of primary transcripts. Through two consecutive transesterification reactions, the intron is removed and flanking exons are ligated. The group I intron sequence evolved for *cis*-splicing in nature. However, sequence modifications allow this ribozyme to catalyze a *trans*-splicing reaction as well. If the 5'-portion of the intron is truncated so that it no longer contains an exon, the ribozyme can then bind in *trans* at its 5'-end to an RNA substrate, cleave the substrate and replace the 3'-terminal portion of the substrate with its own 3'-exon. When the splice site is upstream of a mutation in a messenger RNA (mRNA) sequence, the transfer of the 3'-exon can replace and repair the downstream portion of the mRNA. Thus, this reaction can be utilized as a type of gene therapy on the mRNA level. Repairing mRNA offers the benefit of maintaining natural gene regulation and the ability to target gain-of-function mutations that gene therapy via gene transfer does not. Although the idea of utilizing a group I intron for gene therapy is not new, to date no gene therapy trials have used this *trans*-splicing
reaction. This is in part due to the low efficiency of the ribozyme under clinically relevant conditions.

This thesis represents an effort to increase trans-splicing efficiency of the *Tetrahymena thermophila* group I intron toward the development of a gene therapy. To find ribozyme variants with increased trans-splicing efficiency an *in vivo* selection procedure was developed in which the most efficient ribozymes are selected from millions of ribozyme variants. Ribozymes with an improved 5'-terminus were selected from a library with $9 \times 10^6$ ribozyme variants. Ribozymes with an improved internal sequence resulted from 21 rounds of evolution, which included mutagenesis and recombination. The *in vivo* evolution produced a highly efficient repair ribozyme that appears to be recruiting a cellular protein to increase activity. Future experiments that establish an analogous selection strategy in mammalian cells may make it possible to develop trans-splicing ribozymes for treating genetic diseases.
Chapter 1

Introduction

1.1 Discovery of the *Tetrahymena thermophila* catalytic RNA

In the late 1970s and early 1980s the laboratory of Tom Cech was studying the transcription of ribosomal DNA (rDNA) in the model organism, *Tetrahymena thermophila*. In this era, before the polymerase chain reaction (PCR) had been invented, it was desirable to study the genes of organisms that allowed extraction of a large quantity of DNA. Not only do *Tetrahymena* carry their rDNA on a small chromosome separate from the genomic DNA, in an organelle termed the macronucleus, they offer the ability to purify both the rDNA and the proteins responsible for ribosomal RNA (rRNA) transcription separately from the majority of the cell (1). These features made members of the genus *Tetrahymena* an obvious choice for studying the method of and proteins necessary for rRNA transcription.

In 1977, Phil Sharp and Richard Roberts independently discovered intervening sequences in genes of higher organisms (2, 3). Therefore, it was already known that transcription of DNA led to a primary transcript including coding exons and non-coding introns which underwent processing to form the mature RNA sequence (Fig. 1.1). The gene encoding 26S rRNA of *Tetrahymena thermophila* was known to contain an approximately 400 nucleotide intron. By 1980, the Cech lab found that the intron interrupting the 26S gene was removed
post-transcriptionally (4). This result was similar to that observed with *Tetrahymena pigmentosa* 6UM which also had an intron in the 26S RNA, but did not occur with the 8ALP strain of the same species, which was known to have a linear gene sequence without an intron (5).

Figure 1.1: Intron removal from a primary transcript. Introns are found in mRNA, rRNA, and tRNA sequences. During primary transcript processing, introns (red) are spliced out and the flanking exons (blue) are joined to give the coding sequence.

Cech became interested in finding the enzyme that was able to splice the intron from the pre-rRNA. What he did not know at the time was that there was no enzyme to be found, at least not in the traditional sense. He and his research associate Arthur Zaug, attempted to show there was a factor in the nuclear extracts that was necessary for splicing. They isolated pre-rRNA, radioactively labeled via transcription in nuclear extracts, and then incubated the pre-rRNA with either fresh nuclear extract or buffer alone. However, in both samples the intron was removed from the 26S rRNA. In fact, when they incubated heat denatured or even SDS treated and phenol extracted pre-rRNA with buffer alone, they still observed intron removal (6). Some of the intron splicing requirements
identified at this time included a guanosine cofactor including a 2'-hydroxyl, magnesium, and a buffer containing monovalent cations (in this case (NH₄)₂SO₄). However, the most important contribution from this work was likely the suggestion that splicing might take place without a protein enzyme, and that the RNA could form an active site capable of catalysis.

In 1982, the Cech lab reported that the intron RNA sequence was able to self-splice, without the need for proteins. For the first time, this RNA with catalytic activity, was called a ribozyme (7). The defining experiment involved transcribing the intronic RNA with a small portion of the rRNA flanking exon sequence \textit{in vitro}. The observation that splicing of \textit{in vitro} transcribed RNA occurred without having been in contact with \textit{Tetrahymena thermophila} nuclear extracts demonstrated that this RNA did not need a protein, and must therefore have inherent catalytic activity. Although this ribozyme does not participate in a multiple turnover reaction, the argument was made that it indeed was a catalyst as it was able to significantly lower the activation energy and favorability of the reaction. One year later, in 1983, the laboratory of Sidney Altman provided a second example of a ribozyme by determining that the RNA component of ribonuclease P held the catalytic activity (8). Cech and Altman shared the 1989 Nobel Prize in Chemistry for their discovery of catalytic RNAs.

1.2 Diversity of group I intron ribozymes

Not long after the first description of the catalytic properties of the \textit{Tetrahymena thermophila} (\textit{T. th}) intron, several other introns with similar activity
were found. These catalytic introns were later termed group I introns, to
differentiate them from group II introns, another class of self-splicing introns that
use a different splicing mechanism, have a different structure and remove the
intron in a lariat form (9). By 1990, 87 group I intron sequences had been
identified (10), in 1994, a total of 219 such sequences had been reported (11),
and by 2008, over 17,000 sequences of group I introns were validated and
categorized into subgroups (12). Although the first group I intron discovered was
in an rDNA gene, in the macronucleus of a single celled eukaryote, the currently
identified group I introns are very diverse. To date, group I introns have been
identified in mRNA, rRNA and tRNA sequences. These intron RNAs are part of
nuclear, chloroplast and mitochondrial genes, from eukaryotic, prokaryotic,
bacterial and even viral genomes. These introns are divided into 14 subgroups
based on sequence and structure homology. The T. th group I intron is of the
class IC1 and is one of more than 800 known introns in this subgroup.

It is not well understood why group I introns exist. They do not appear to
afford a specific benefit to the organism in which they reside (12), other than
perhaps removal of introns without the need of the complex spliceosome
machinery. Therefore, they may be considered parasitic genetic elements.
Although the reason group I introns persist in genomes is not known, their
presence in such a wide variety of genes and organisms can be explained
through their two methods of mobility. The majority of group I intron mobility is
likely due to the presence of a homing endonuclease gene (HEG) found in the
peripheral stem loop of the intron (13). These HEGs encode proteins that cleave
the DNA at an identical gene locus without a HEG, allowing recombination of the gene encoding the HEG and group I intron into a new site. These HEG sequences are then shortened, mutated, or lost over time, fixing the intron in the gene. Another potential method of intron mobility comes from the reversibility of the splicing reaction (14). After self-splicing, a liberated intron can reverse splice into new RNA locations and through reverse transcription and recombination, become integrated into a new host genome (15). However, the reverse splicing mechanism is likely the minor mode of group I intron mobility.

Because group I introns are both inherited vertically (parent to progeny) and gained through horizontal gene transfer (transfer between organisms) of the gene containing the intron (15), organisms of the same genus may have vastly different group I intron sequences, whereas more distantly related organisms may have closely related introns (Fig. 1.2). The availability of the known group I intron sequences has made it possible to analyze sequence conservation among group I intron ribozymes of the same subgroup and predict at which positions nucleotides must be absolutely conserved, may be co-varied, can be mutated or deleted all together (16, 17).
1.3 Determining the structure

Although the discovery of many additional group I introns occurred quickly, the structure elucidation did not. Identification of some structural elements in the secondary structure was immediate, but it took several more years to identify tertiary interactions and over twenty years to determine the crystal structure (18).

Initially, two independent groups were able to predict the ten paired regions formed in the intron sequence (19, 20). Their predictions involved analyzing the sequence homology of several introns found in the mitochondrial genomes of fungi. These papers helped to determine the conserved guanine
residues in the internal guide sequence (IGS) and at the 3'-splice site, and confirm that the intron sequence folded to bring the 5’ and 3’ ends in proximity for splicing.

By 1989, the P7 helix was identified as the binding site for the exogenous G that would initiate the first step in the reaction cascade as well as for the conserved ωG that would position the 3’-exon for the ligation step (21). By 1990 the first model of the catalytic core was generated by the Westhof lab after analyzing the alignment of 87 group I intron sequences (10). This model proposed a 3D structure of the group I intron core and was used to predict the hierarchal pathway of folding.

The well substantiated secondary structure, initial 3D model of the core nucleotides, and additional information on tertiary interactions led to a new representation of the secondary structure of group I introns that is still used today (22). A variation on this structure representation with the P4-P6 domains move to the right is seen below in figure 1.3.
Figure 1.3: *Tetrahymena thermophila* group I intron secondary structure. The 10 paired regions formed in the cis-splicing reaction are shown along with helix junctions and the guanosine binding site in the P7 helix. Adapted from (22) by placing the P4,5,6 domains on the right side.

In 1996, the Westhof lab refined their model to include the complete *T. th* ribozyme sequence (23). By including tertiary contacts as well as long range loop-loop interactions, additional data on Fe(II)-EDTA protection and further sequence comparisons, the first complete, three dimensional structure was available. By 2005, the first crystal structures were published for three different group I introns, *Tetrahymena thermophila*, *Azoarcus* and the *Staphylococcus aureus* bacteriophage Twort (18, 24-26). Although these introns came from different subgroups with vastly different primary sequences and intron lengths,
they showed that there was remarkable structural similarity in the catalytic core. Figure 1.4 below shows the overlap of the core structure of these three crystallized ribozymes. These structures also elucidated the metal binding site which helped provide the structural stability of the ribozyme’s core. The determination of the secondary and tertiary structure provided a better understanding of the ribozyme and paved the way for many future studies.

**Figure 1.4:** Intron core structure overlap. Alignment of *Azoarcus* (grey), *Tetrahymena* (gold) and Twort (green) intron core structures. The Azoarcus and Twort 5’-exons are shown in red, and the Azoarcus 3’-exon is shown in blue. The ωG is show as ball and stick to denote the active site. Adapted from (27).
1.4 Group I intron trans-splicing

In nature, group I introns evolved for self-splicing, also termed cis-splicing because the substrate and ribozyme lie on the same molecule (Fig. 1.5a). However, with a few modifications, the intron can be converted into a trans-acting ribozyme (Fig. 1.5b). To act in trans the 5’-exon and a portion of the ribozyme at the 5’-end are removed so that the ribozyme’s internal guide sequence (IGS) can base pair with a target RNA. The ribozyme can then transfer the 3’-exon to the target RNA in a single turnover reaction.

The ribozyme is also capable of a multiple turnover reaction in which a trans-splicing intron without a 3’-exon, binds in trans to a substrate and catalytically cleaves the substrate at the splice site. This multiple turnover reaction has been utilized for studying the kinetics of the cleavage step (28, 29). The ability to study cleavage kinetics has also made it possible to study the functional group requirements on the substrate and IGS for efficient P1 helix docking (30). When a portion of the 3’-exon is retained, both the cleavage and ligation steps can be monitored (31).

Although the majority of the catalytic core of group I introns is highly conserved, especially in structure, many of the peripheral sequences can be mutated. The sole sequence requirement in the 6 nucleotide IGS is a G-U wobble at the splice site. Therefore, the 5’-most nucleotide of the IGS must be a guanine and the substrate splice site must be a uridine which pairs with the guanine in the P1. The 3’-end of the intron is also a conserved guanine residue, termed the ωG,
but the 3'-exon can be changed seemingly freely, to accommodate trans-splicing repair.

**Figure 1.5:** Cis and trans-splicing diagrams. (a) Diagram of the natural cis-splicing reaction of the *Tetrahymena thermophila* group I intron. Through cis-splicing two transesterification reactions ligate the exons (red) and release the intron (black). (b) Diagram of the trans-splicing reaction. By removing the 5’-portion of the intron and 5’-exon, the intron and can pair in trans to the substrate (red) and transfer the 3’-exon (blue) to in the splicing reaction. The addition of an EGS (green) in the trans reaction increases specificity and efficiency.

To increase the ribozyme’s ability to select the correct target site in trans, an extended guide sequence (EGS) could be added on the 5’-end (green in Fig. 1.5b). This sequence was found to increase efficiency and decrease the likelihood that an incorrect target site would be chosen. These minimal requirements and the ability to modify the target recognition sequences and 3’-
exon have made the group I intron *trans*-splicing reaction an attractive foundation for developing a gene therapy.

An initial proof of principle reaction was performed in 1994 in which *E. coli* cells were transformed with a plasmid carrying the 5'-portion of a *lacZ* gene substrate and *trans*-splicing ribozyme which would complete the gene via 3'-exon ligation (32). Although repair was low, approximately 1%, the observation of β-galactosidase activity in these cells demonstrated *trans*-splicing repair could occur in frame, *in vivo*, and that functional proteins could be translated. This result indicated that transcription and translation were not too tightly coupled for a *trans*-acting ribozyme.

Further mRNA repair reactions focused on *trans*-splicing in mammalian cell lines (33) and splicing on disease gene targets (34-37). These ribozyme reactions were able to catalyze up to 50% of substrate repair and show that in frame repair was possible in mammalian cell lines. These initial studies were performed in the Sullenger lab and focused on *in vivo* repair of truncated or mutated substrates.

The first group I intron ribozyme *trans*-splicing study in mammalian cells was a reaction that sought to determine if *trans*-splicing could occur in mammalian cells (33). The substrate was a short portion of the *lacZ* gene containing the natural IGS recognition sequence found in the *Tetrahymena* intron, CCCUCU, followed by a spacer sequence and primer binding site. While *trans*-splicing did not repair, or complete the *lacZ* gene, it shortened the spacer, which generated a size difference when comparing the substrate and product
formed via reverse transcription and PCR. The substrate and ribozyme were carried on different plasmids, and were transformed in varying ratios, from 0.1:1 up to 10:1 ribozyme:substrate, into OST7-1 mouse fibroblast cells. Both the substrate and ribozyme had T7 promoters and the fibroblast cell line was engineered to constitutively express T7 RNA polymerase. When 20 µg of the ribozyme plasmid was transformed with 2 µg of the substrate plasmid, 49% of the substrate was trans-spliced to product, as determined by qRT-PCR.

Three more papers from the Sullenger lab followed, this time focusing on repair of the β-globin gene transcripts (34-36). In the first paper (34), accessible splice sites upstream of the most common mutation causing sickle cell anemia, an A to T mutation in the sixth codon of the β-globin gene, were mapped in vivo and in vitro. Ribozymes were then generated to trans-splice the γ-globin sequence in frame, converting mutant β-globin transcript into γ-globin RNA, which has been shown to reduce polymer formation of sickle hemoglobin. Trans-splicing was observed in vitro with transcribed substrates and in vivo using erythrocyte precursor cells from healthy patients and those with sickle cell anemia. The only quantified repair efficiency was with in vitro splicing and was 8% determined via qcRT-PCR.

The second paper (36) combined trans-splicing ribozymes targeting two of the splice sites identified on the β-globin transcript in the previous paper. This study found that combining two trans-splicing ribozymes led to conversion of a greater percentage of substrate to product in vitro. Up to 25% of the substrate was trans-spliced with the 3'-exon tag using both ribozymes in concert.
Additionally, it was determined that adding the ribozyme targeting the second splice site, after a long incubation with the first, increased splicing. This effect was attributed to the second ribozyme being able to target substrate molecules with alternate folds that allowed access to the second splice site, but not the first, and was seen no matter which ribozyme was initially chosen.

In the third paper from the Sullenger lab (35), again splicing on the β-globin gene, trans-splicing efficiency was improved by adding an antisense EGS sequence. Efforts were made to optimize the P10 interaction between the EGS and 3'-tail transferred in the splicing reaction as well as the EGS length. In this study, ribozymes expressed under control of the T7 promoter, this time in HEK293 cells expressing T7 RNA polymerase, accomplished up to 56% conversion of the substrate to product as determined by quantitative competitive RT-PCR (qcRT-PCR) with high ribozyme expression levels. However, the goal of this paper was to increase splicing efficiency of ribozymes expressed by pol II in vivo. Initially, the ribozymes expressed via pol II, without an EGS, were not active in vivo. Addition of an EGS containing at least a three nucleotide P1 extension, engineered P10 interactions and additional antisense pairing to the substrate were active and provided almost 10% in vivo repair by converting sickle β-globin transcripts to γ-globin mRNA.

Another splicing study on a different mutant substrate was able to assess both the percentage of substrate repaired in vivo and the activity of the protein translated from the repaired mRNA (37). For this in vivo repair HEK293 cells, stably expressing mutated canine skeletal muscle chloride ion channel mRNA,
were transformed with the trans-splicing repair ribozyme. The ribozyme targeted a splice site determined by mapping accessible splice sites, and the engineering EGS contained a three nucleotide P1 extension, a P10 helix and 33-nucleotide antisense pairing to the substrate. Although the ribozyme in this system was only able to repair a modest 1.2% of the mutant substrate, as determined by qRT-PCR, it was still enough to be able to significantly improve the ion channel function in 18% of the cells analyzed.

The group I intron trans-splicing reaction has also been utilized to specifically reprogram mRNA to kill target cells. In human cells, the transcripts of mutated tumor suppressor genes, such as p53, have been repaired to restore the natural ability to induce apoptosis in vivo (38). Similarly, mRNAs typically up-regulated in cancer, such as hTERT, (39, 40) have been targeted by trans-splicing a cytotoxic gene to the mRNA transcript, for use as a cancer therapeutic. Viral RNAs have also been targeted, adding trans-splicing group I introns to the repertoire of RNA therapeutics together with siRNA, miRNA and other trans-acting catalytic RNAs (41, 42).

1.5 Improvements in ribozyme efficiency through sequence optimization

Through rational design, selections, and evolutions, the ribozyme sequence has been altered in attempts to improve specificity, splicing efficiency and to better understand the ribozyme. One method of increasing the trans-splicing efficiency is to optimize the target mRNA splice site. Although the only sequence requirement of an mRNA is a uridine at the splice junction, the
accessibility of the chosen splice site can be very important for the efficiency of the reaction. A method for trans-tagging of accessible splice sites was developed that allows identification of accessible target sites in vitro, and even in vivo (43). In this assay a ribozyme library is generated with a randomized IGS that maintains the guanosine residue at the splice site (GNNNNN) and that has an arbitrary 3'-exon. This ribozyme library is then able to trans-tag its 3'-exon to any accessible uridine in the substrate sequence. Reverse transcription, cloning and sequencing can identify the accessible sites at which trans-splicing occurred.

Initial selections performed in vitro utilized the reversibility of either the first or second transesterification reaction in the natural splicing cascade (44-46). In these selections, an RNA substrate is ligated to the ribozyme at the position of the natural 5' or 3' exon. The ligated sequence enables the identification and selection of ribozymes, from a pool of variants, which are able to efficiently perform the reverse reaction. Placing a new selection pressure on the ribozyme during selection or randomization of different ribozyme domains allowed the evolution of ribozyme sequences with different metal preferences (45) or minimizations and replacements for a portion of the structural P4-P6 domain (46, 47). Although this method of selection yielded ribozymes with faster reverse-splicing kinetics, it did not often lead to ribozymes with improved forward reaction rates (46).

In vivo selections have also been utilized to select more efficient ribozymes. A self-splicing ribozyme inserted into an antibiotic resistance gene was mutagenized and selected in E. coli for efficient activity in a new exon
sequence context (48). In this one-round selection, it was determined that the P1 helix should not be too strong, or too weak, for efficient self-splicing. Another *in vivo* selection was performed in yeast cells using repair of auxotrophy genes for selection (49). This selection included 13 trans-splicing ribozyme sequences with different designed EGS lengths ranging from 0 to 764 nucleotides. Although this study was able to determine that the EGS with 293 nucleotides of complementarity to the substrate was the most efficient, this result was more attributed to serendipity than an optimal and universal best EGS length.

Aside from using selections to improve the ribozyme, rational design has also been attempted extensively. The 5’-EGS and 3’-exon have been the primary targets for rational design techniques with the goal of optimizing both the pairing with the substrate in the cleavage step, (also termed the P1 extension) and stabilization of the 3’-exon in the ligation step (or P10 helix). However, best design principles have remained inconsistent and will not work for every substrate or splice site. In a recent paper, Kohler et al. optimized the EGS sequence of the *Tetrahymena* ribozyme for *trans*-splicing the β-galactosidase α-peptide sequence onto the chloramphenicol acetyltransferase transcript (50). However, when they attempted to utilize the same EGS structures for targeting the cucumber mosaic virus or HIV RNA, only 2% of the β-galactosidase activity identified in when targeting the *cat* gene was observed. Further attempts to establish design rules for efficient EGS sequences on different species and subclasses of group I introns have not yielded a set of principles that hold for all substrates and splice site sequences (51, 52).
1.6 Gene therapy by mRNA repair

Early gene therapy trials sought to cure diseases associated with genetic mutations by inserting a new, correct gene sequence into the genome. An early trial utilized a lentiviral vector for gene transfer and repair of X-linked severe combined immunodeficiency. However, the site of integration was not well defined and in at least one patient oncogenes were activated resulting in leukemia (53). Although the use of different viral vectors, such as the adenovirus or adeno-associated virus vectors, can overcome this problem through more defined integration (54), these gene therapies do not offer all of the benefits of repair on the mRNA level.

It is currently impossible to excise and replace a mutated, disease-causing gene, and therefore gene transfer occurs at a different location in the genome. Because the replacement gene will be inserted in a different context, it will likely not retain normal gene regulation which may result in unanticipated consequences. Additionally, simple insertion of the correct gene sequence into the genome will not be an effective gene therapy for diseases resulting from gain-of-function mutations, such as in sickle cell anemia which results from a mutation causing misfolded proteins that aggregate and deform red blood cells (55). Repair of gene products on the mRNA level can maintain proper gene regulation, can be effective in repair of gain-of-function mutations, and if therapy is consistently applied, can remove the need for integration into the genome.
There are many known genetic diseases, but not all are good targets for mRNA repair. To be a good target, there must be a single, known gene causing the disease. The specific mutation must be identified so an appropriate splice site can be selected and the 3’-exon can be engineered for repair. It is convenient if the disease affects organs that are easily accessible for delivery techniques. For blood diseases such as sickle cell anemia or hemophilia gene therapy may be able to be administered by injection. For diseases affecting the lungs such as cystic fibrosis, the primary effects may be alleviated by delivering a trans-splicing ribozyme gene therapy via inhalation. Diseases of the muscles and connective tissue, such as muscular dystrophy or osteogenesis imperfecta are harder to target due to difficulty in specifically and thoroughly targeting these tissues. Lastly, in the best target genes, a small repair of the mRNA will lead to a significant improvement of symptoms. For example, in hemophilia B, restoration of clotting factor IX activity to 1% of normal levels is enough to upgrade the patient from the severe hemophilia classification (56). As more efficient ribozymes are generated, diseases which require higher repair can be targeted as well.

1.7 Limitations of current trans-splicing group I introns

Despite the advancements in understanding the group I intron structure and function, these introns have not yet been developed into a viable gene therapy. The limiting factors include low efficiency of target gene repair at clinically relevant ribozyme expression levels, unknown issues with possible off-
target splicing, and determining an appropriate method for delivery. We aimed to further improve trans-splicing efficiency in a cellular environment without the need for artificially high ribozyme expression levels.

Although in vitro selections offer the possibility of sampling on the order of 1,000 times more sequences than in vivo selections, they often do not accurately represent the cellular environment. In contrast, in vivo selections, afford the ability to select inside a cell, but do not allow high pool numbers, and to date the largest in vivo selection included only 200,000 group I intron variants (48).

The trans-tagging assay developed to identify accessible splice sites has been useful as a current first step in optimizing a trans-splicing ribozyme. However, many of the initial in vivo repair reactions did not utilize this method, and instead kept the natural IGS target site of CCCUCU (32, 33). It is likely these initial mRNA repair studies could have been improved upon via determination of more accessible splice junctions.

Rational design has primarily been utilized for efforts of optimization of the P1 extension and P10 pairing. Although rational design methods and analysis of optimal EGS lengths have been shown to improve trans-splicing efficiency on a chloramphenicol acetyltransferase target, utilizing the same optimized base pair formation on cucumber mosaic virus or HIV target mRNA did not yield similarly efficient splicing (50). Rational design of ribozyme sequences often does not lead to the trans-splicing efficiency desired.

Group I introns evolved in nature to cis-splice in a different cellular environment than is present during in vivo trans-splicing for mRNA repair. It is
possible that there is a different optimal internal sequence that will facilitate trans-splicing. To increase the in vivo splicing efficiency, there must be a coordinated effort to choose an accessible splice site, optimize the EGS sequences and P10 structures, and evolve the internal sequence in the desired repair environment.

1.8 Goals of this thesis

To date, no group I intron ribozyme has been employed for mRNA gene therapy clinical trials. Improvements must be made to increase both the specificity of trans-splicing as well as the efficiency of repair. The following chapters discuss methods for improving trans-splicing ribozyme activity toward the development of a successful gene therapy.

Chapter 2: Computational prediction of efficient splice sites for trans-splicing ribozymes

Although a functional method for selection of trans-splicing products has previously been utilized (43), this method may not always select the best splice sites most frequently. A computational method is described in Chapter 2 that takes into account the accessibility of the substrate mRNA, the accessibility of the specific IGS sequence and the stability of the P1 helix to determine the optimal splice sites on an mRNA. This method for splice site selection has been shown to find more efficient splice site targets than the previously available method with less effort.
Chapter 3: An *in vivo* selection method to optimize *trans*-splicing ribozymes

After selecting an accessible splice site on the mRNA, the second determinant of splicing efficiency is the structure formed between the EGS and both the substrate and 3'-exon. While rational design has extensively been employed in the development of EGS sequences and structures, it has not produced clear design rules that can reliably predict efficient EGSs. To date *in vivo* repair methods have shown up to 50% repair, but this level was only achieved by utilizing an artificially high expression level in a T7 driven system. Chapter 3 discusses a novel *in vivo* selection procedure to optimize the EGS sequence of a *trans*-splicing ribozyme. This system focuses on the repair of a mutated antibiotic resistance gene and is able to select the best ribozyme sequences in a cellular environment.

Chapter 4: Evolution of group I intron internal sequences for high *trans*-splicing efficiency *in vivo*

Group I introns evolved in nature to self-splice and ligate the specific 5' and 3' exon sequences that flank the intron. However, for mRNA repair, the ribozymes must be converted into efficient *trans*-splicing ribozymes that form IGS and EGS base pairing with new mRNA substrate sequences. Because this intron was not evolved for *trans*-splicing, there may be better and more efficient internal ribozyme sequences for mRNA repair reactions. Chapter 4 expands on the selection method designed in chapter 3 and includes mutagenesis of the internal
ribozyme sequence via mutagenic PCR and recombination. Over 21 rounds of selection, the antibiotic selection pressure was increased to select a ribozyme with much higher efficiency than the parent sequence. Interestingly, the much increased efficiency seems to be associated with the evolved ribozyme recruiting a cellular protein.

1.9 References


Chapter 2

Computational prediction of efficient splice sites for *trans*-splicing ribozymes

2.1 Abstract

Group I introns have been engineered into *trans*-splicing ribozymes capable of replacing the 3'-terminal portion of an external mRNA with their own 3'-exon. Although this design makes *trans*-splicing ribozymes potentially useful for therapeutic application, their *trans*-splicing efficiency is usually too low for medical use. One factor that strongly influences *trans*-splicing efficiency is the position of the target splice site on the mRNA substrate. Viable splice sites are currently determined using a biochemical *trans*-tagging assay. Here, we propose a rapid and inexpensive alternative approach to identify efficient splice sites. This approach involves the computation of the binding free energies between ribozyme and mRNA substrate. We found that the computed binding free energies correlate well with the *trans*-splicing efficiency experimentally determined at 18 different splice sites on the mRNA of chloramphenicol acetyltransferase. In contrast, our results from the *trans*-tagging assay correlate less well with measured *trans*-splicing efficiency. The computed free energy components suggest that splice site efficiency depends on the following secondary structure rearrangements: hybridization of the ribozyme’s internal guide sequence (IGS) with mRNA substrate (most important), unfolding of
substrate proximal to the splice site, and release of the IGS from the 3'-exon (least important). The proposed computational approach can also be extended to fulfill additional design requirements of efficient trans-splicing ribozymes, such as the optimization of 3'-exon and extended guide sequences.

2.2 Introduction

Group I introns are catalytic RNAs (ribozymes) that excise themselves from primary transcripts (1). These ribozymes have been well characterized both biochemically and structurally, with five high-resolution crystal structures solved to date (2-5). Furthermore, it has been possible to convert the Tetrahymena cis-splicing group I intron into an artificial trans-splicing ribozyme that can modify the sequence of an arbitrary mRNA substrate. This was achieved by removing the 5'-portion of the wild-type Tetrahymena group I intron, thus exposing the intron's internal guide sequence (IGS) at the 5'-terminus (6, 7). The resulting ribozyme can hybridize its IGS to a complementary target site on a substrate mRNA, forming a helix equivalent to the P1 duplex in the wild-type intron (8, 9). The ribozyme then catalyzes cleavage of the substrate at the target site and transfer of the ribozyme 3'-exon to the remaining 5'-portion of the substrate (10).

The ability to replace a portion of a substrate mRNA with the 3'-exon carried by the ribozyme has created potential roles for trans-splicing ribozymes in therapeutic applications (11). These ribozymes could be used to treat genetic disorders by repairing the sequence of a mutated mRNA (7, 12, 13), to selectively kill cancer cells by splicing a sequence that encodes a toxic peptide
into cancer-specific mRNAs (14-16), and to kill virally infected cells with the same strategy (16-18).

However, technical difficulties have so far prevented the medical use of trans-splicing ribozymes. One problem is the localized and efficient delivery of therapeutic RNAs to the affected tissues (19, 20), though solutions involving viral vectors and modified Salmonella strains appear to be promising (21, 22). Another problem of trans-splicing ribozymes is their low efficiency in cells. Here, trans-splicing efficiency denotes the fraction of target mRNA that is converted to product by the ribozymes. Trans-splicing efficiency is typically 10% or less in cells (11). Although efficiencies of up to 50% have been reported (23), the necessary intracellular ribozyme concentrations appear too high to be acceptable in a clinical setting. Therefore, an outstanding task for the development of trans-splicing ribozymes is to increase their efficiency, allowing a sufficient fraction of mRNA substrate to be trans-spliced at low ribozyme concentrations.

The trans-splicing efficiency varies strongly with the location of the splice site within the mRNA substrate (24, 25). In particular, secondary structures within the substrate can render potential target sites inaccessible to the ribozyme. The trans-splicing efficiency may also vary between splice sites owing to different stabilities of the P1 duplex formed between ribozyme and substrate, and to different base-pairing interactions between the ribozyme IGS and the ribozyme 3'-exon. Thus, the identification of efficient splice sites, i.e., those sites that mediate a high trans-splicing efficiency, is a non-trivial problem in the design of trans-splicing ribozymes.
An elegant experimental method, known as the trans-tagging assay, has been developed to identify efficient splice sites (26). In this assay, the mRNA substrate is incubated in vitro with a pool of trans-splicing ribozymes whose IGS is randomized. This pool of ribozymes is thus able to recognize any potential splice site on the substrate. Ribozymes targeting efficient splice sites are able to trans-splice at those sites. The resulting products are detected by reverse transcription, sub-cloning, and sequencing. The sequences from several clones are then used to deduce the positions of efficient splice sites (12, 26, 27). This trans-tagging assay has been successfully applied—without explicitly reported limitations—to uncover efficient splice sites on various mRNA substrates (12, 28-35).

The problem of identifying efficient splice sites for trans-splicing ribozymes is in concept similar to that of finding accessible or optimal binding sites on mRNA substrates targeted by antisense DNA oligonucleotides (36), siRNAs (37, 38), and other non-protein coding RNAs (39). In these cases, several computational methods have been proposed and found to be effective at predicting the relevant target sites (40-46). One recurring theme is the use of RNA secondary structure prediction algorithms to calculate the free energy changes, ΔG_{bind}, associated with binding an oligonucleotide to each possible target site on the mRNA substrate. Target sites with strongly negative values of ΔG_{bind} are then predicted to be the most accessible sites on the substrate (44, 47-50).

Here, we have adapted this approach of computing ΔG_{bind} to the problem
of finding efficient splice sites for *trans*-splicing ribozymes. The mRNA of chloramphenicol acetyltransferase (CAT) was used as a model substrate for both computations and experiments. In particular, we computed $\Delta G_{\text{bind}}$ for all 186 splice sites on this substrate and we experimentally determined the *trans*-splicing efficiency *in vitro* for 18 of those splice sites. Furthermore, to map the accessible splice sites on CAT mRNA, we carried out the *trans*-tagging assay, which differs substantially from direct measurements of *trans*-splicing efficiency. Surprisingly, the experimental *trans*-splicing efficiencies at the 18 tested splice sites were found to correlate better with the computed $\Delta G_{\text{bind}}$ values than with the results of the *trans*-tagging assay. These results suggest that the proposed calculation of $\Delta G_{\text{bind}}$ could be used to predict efficient splice sites for *trans*-splicing ribozymes more quickly, cheaply, and accurately than with the presently used experimental methods.

### 2.3 Results

**Calculation of $\Delta G_{\text{bind}}$ for *trans*-splicing ribozymes**

We defined $\Delta G_{\text{bind}}$ as the free energy change associated with the binding of the ribozyme’s IGS to a given splice site on the mRNA substrate. Our hypothesis was that a very negative $\Delta G_{\text{bind}}$ corresponds to a high *trans*-splicing efficiency. To calculate $\Delta G_{\text{bind}}$, the binding process was modeled as consisting of three idealized molecular events that involve only local changes in RNA secondary structure: the unfolding of the target site on the substrate, the release of the IGS on the ribozyme, and the hybridization between target site and IGS to
form the P1 duplex (Fig. 2.1A). The corresponding component free energy changes were denoted by $\Delta G_{\text{unfold-target}}$, $\Delta G_{\text{release-IGS}}$, and $\Delta G_{\text{hybrid}}$, respectively. By "release" of the IGS we mean the breaking up of any base pairs that may form between the IGS and the 3'-exon. This release is necessary to make the IGS available for binding to the target site. We thus computed $\Delta G_{\text{bind}}$ by summing the above components, i.e., $\Delta G_{\text{bind}} = \Delta G_{\text{unfold-target}} + \Delta G_{\text{release-IGS}} + \Delta G_{\text{hybrid}}$. These components can in turn be computed using several available RNA folding algorithms, which predict the possible secondary structures for a given RNA sequence as well as the free energy changes of those structures relative to the unfolded state (51).

To compute the components of $\Delta G_{\text{bind}}$ for trans-splicing ribozymes that interact with a long mRNA substrate, we adapted the ribozyme and mRNA sequences before submission to RNA folding algorithms. The 3D model of the Tetrahymena ribozyme (52) suggests that before pairing to an mRNA target site, the IGS is sterically allowed to base pair with the 3'-exon, but does not base pair with the body of the folded intron (Fig. 2.1B). Thus, to calculate $\Delta G_{\text{release-IGS}}$, we used a ribozyme sequence in which the intron portion downstream of the IGS was replaced by a five-nucleotide linker region (Fig. 2.1C). This shortened sequence allowed the IGS to interact with the 3'-exon, while permitting rapid computation of $\Delta G_{\text{release-IGS}}$ by RNA folding algorithms. To choose an appropriate linker sequence, we noted that in the Tetrahymena group I intron, the IGS is followed by three unpaired adenosines, and the 3'-exon is preceded by one unpaired guanosine (53). Although these nucleotides seem natural candidates
for inclusion in the linker, the 3D structure of the ribozyme shows that these nucleotides are conformationally prohibited from pairing with the substrate and from engaging in π-stacking interactions with the IGS or the 3'-exon (Fig. 2.1B). Therefore, we chose the linker sequence to consist of five unspecified nucleotides, which are effectively ignored by RNA folding algorithms in the calculation of ΔG_{release-IGS} and ΔG_{hybrid}. Varying the length of the linker from four to seven unspecified nucleotides did not significantly change the computed ΔG values (data not shown).

To predict ΔG_{unfold-target} from the local secondary structure around each splice site on a long mRNA substrate, we directed the RNA folding algorithms to analyze subsets, or windows, of the entire substrate sequence. Thus, only the sequence within each window containing a given splice site was used to calculate ΔG_{unfold-target} for that site. For the 678-nucleotide sequence of our model mRNA (see below), the resulting values of ΔG_{unfold-target} were found to vary by more than 1 kcal/mol over different window sizes (data not shown). This variation may be due to the somewhat arbitrary exclusion of base pairs involving nucleotides outside of a given window. To minimize the influence of window size on the predicted ΔG_{bind}, we computed ΔG_{unfold-target} using window sizes of 100, 200, 300, 400, 500, and 600 nucleotides, and then we averaged the resulting values of ΔG_{bind}. 
Figure 2.1: Interactions within and between the mRNA substrate and the ribozyme, and their treatment during computation. (A) Schematic of the substrate binding step in trans-splicing. The mRNA substrate is illustrated as a hairpin structure and the body of the ribozyme is illustrated as a grey oval. Unfolding of the target site on the mRNA (first step) and release of the internal guide sequence (IGS) (second step) allow the mRNA target site and the IGS to hybridize (third step). (B) The 3D structure of the Tetrahymena ribozyme (52) indicates that the AAA linker (black) is not stacked on the P1 duplex. Additionally, the 3'-exon (grey) is close enough to the IGS (black) for interactions if the mRNA (grey) is absent. (C) To facilitate computational treatment of the binding process, the body of the ribozyme was replaced by a linker sequence joining the IGS to the 3'-exon. This linker does not include the natural AAA linker but consists of five unspecified nucleotides (N), which are ignored in the calculation of base-pairing and π-stacking interactions.
Splice sites chosen on a model mRNA substrate, CAT mRNA

As a model mRNA substrate for our computational and experimental investigations of trans-splicing efficiency we chose the mRNA of chloramphenicol acetyltransferase (CAT). This mRNA sequence is 678 nucleotides long and contains 186 uridines downstream of the AUG translation start codon. Each of these uridines represents a potential splice site for trans-splicing ribozymes (54). We thus computed ΔG_{bind} for each of these splice sites. Figure 2.2A plots the resulting ΔG_{bind} values as a function of splice site position relative to the adenine of the AUG start codon. These values are distributed over the range from 0 to −6.5 kcal/mol, with most values closer to 0 kcal/mol (Fig. 2.2B). In particular, only seven out of the 186 available splice sites had ΔG_{bind} < −4.0 kcal/mol, suggesting that this value of ΔG_{bind} could be used as threshold to predict efficient splice sites.

To determine experimentally whether the computed values of ΔG_{bind} can be used to identify efficient splice sites, we chose a set of 18 different target splice sites on CAT mRNA. To reveal possible correlations between trans-splicing efficiency and ΔG_{bind}, as well as between trans-splicing efficiency and position on mRNA substrate, these 18 sites were chosen to cover the full range of ΔG_{bind} values (Fig. 2.2B) and were distributed over the length of the CAT mRNA (Fig. 2.2A). Two splice sites with ΔG_{bind} slightly below −4.0 kcal/mol were omitted in order to cover the wide range of the predicted values for ΔG_{bind} with the available resources.
Figure 2.2: Computed values of $\Delta G_{\text{bind}}$ for splice sites on CAT mRNA. (A) plot of $\Delta G_{\text{bind}}$ against splice site position relative to the adenosine (position 1) of the AUG transcription start codon. Only splice sites downstream of this codon were mapped by the trans-tagging assay in this study. Circles with or without diamond inside indicate splice sites chosen for experimental determination of trans-splicing efficiency. Diamonds indicate splice sites found among 66 product sequences obtained by trans-tagging assay. (B) Distribution of the 186 splice sites (grey bars) of CAT mRNA over the computed values of $\Delta G_{\text{bind}}$. Black bars represent the 18 splice sites that were chosen for experimental testing. The topmost bar represents 61 splice sites and is shown truncated.
Experimental trans-splicing efficiencies on CAT mRNA

To measure experimentally the trans-splicing efficiency of the 18 chosen splice sites, we generated 18 trans-splicing ribozymes that contained different IGSs, but were otherwise identical. Each IGS was designed to target a particular one of the chosen splice sites (Table S2.1). Each of the 18 ribozymes was incubated with 5'-radiolabeled CAT mRNA, and the reaction products were analyzed by polyacrylamide gel electrophoresis and autoradiography (Fig. 2.3A). As a measure of trans-splicing efficiency, we quantified the fraction of radiolabeled substrates converted to trans-splicing products after 4 hours of incubation. Bands consistent with the expected trans-splicing products were observed for nine of the 18 tested splice sites, namely for splice sites at uridines 83, 87, 97, 197, 222, 258, 350, 405, and 448. Product fractions for these splice sites ranged from 0.4% of the substrate at uridine 222, to 14.3% at uridine 258 (Table S2.1). Nine of the 18 splice sites did not yield quantifiable products.

The experimental trans-splicing efficiencies were found to agree with the computed ΔG_{bind} values (Fig. 2.3B). In particular, the efficiency of splice sites with ΔG_{bind} > −4 kcal/mol was relatively low, reaching at most 2.2%. On the other hand, the efficiency of splice sites with ΔG_{bind} < −4 kcal/mol increased steadily with increasingly more negative values of ΔG_{bind}, reaching 14.3% at ΔG_{bind} = −6.1 kcal/mol. Only the five most efficient splice sites had a computed ΔG_{bind} < −4.0 kcal/mol, again suggesting that efficient splice sites could be predicted by comparing the computed ΔG_{bind} to the empirical threshold of ΔG_{bind} ≈ −4.0 kcal/mol.
To test the identity of the observed trans-splicing products, we reverse transcribed, cloned, and sequenced the products of all 9 reactions that had resulted in quantifiable bands on polyacrylamide gels (sequence data not shown). The expected product sequences were obtained for 8 of the 9 tested splice sites. Product sequences for splice site 350, which did not yield the expected product, indicated that the corresponding ribozyme had spliced primarily at uridine 307. Therefore, the actual trans-splicing efficiency at splice site 350 may be lower than its measured trans-splicing efficiency (2.2%), possibly improving further the correlation between computed $\Delta G_{\text{bind}}$ and trans-splicing efficiency seen in Fig. 2.3B. In summary, the sequencing results confirmed all but one of the nine products observed in the trans-splicing reactions.
Figure 2.3: Trans-splicing efficiency measured on 18 splice sites on CAT mRNA. (A) Autoradiogram of reaction products from trans-splicing reactions with radiolabeled substrate. Lanes are labeled with the targeted splice sites, or with (-) where ribozyme was absent. RNA marker sizes in nucleotides are indicated. The unreacted substrate had a length of 678 nucleotides, whereas the length of reaction intermediates (i) and trans-splicing products (*) varied with the splice sites. Note that the ribozymes targeting splice sites 405 and 448 also spliced at each other's splice site. (B) Correlation between computed binding free energies and experimental trans-splicing efficiencies. Each diamond represents a specific splice site. Splice sites 131, 240, 369 have $\Delta G_{\text{bind}} \geq 0$ kcal/mol and are represented by a white diamond at the origin. The data points near or over the $\Delta G_{\text{bind}}$-axis around the value of $-3$ kcal/mol correspond, from left to right, to splice sites 222, 518, 273, 346, 325, 378, and 551. The thick grey line represents a least-mean-squares exponential fit, $y = A \exp(Bx)$, to the data points, with a coefficient of determination $R^2 = 0.87$. Alternatively, a linear fit (not shown) to the same data points yields a correlation coefficient $R = -0.75$, with a probability $p = 0.00033$ that the values are not correlated. Horizontal error bars are standard deviations over three to six different window sizes (100 to 600 nt). Vertical error bars are standard deviations from three independent experiments.
Experimental trans-splicing efficiency with unstructured RNA substrates

To confirm that all of the 18 ribozymes employed in experiments with CAT mRNA were catalytically active, we measured their trans-splicing efficiency on short substrates that were designed to prevent formation of secondary structure. Hence, this experiment also allowed us to examine the effects of substrate secondary structure on trans-splicing efficiency. Each short substrate consisted of a 13-nucleotide sequence containing one of the 18 splice sites already targeted on CAT mRNA. The single-stranded conformation predicted by RNA folding algorithms for these substrates effectively abolishes the hypothesized $\Delta G_{\text{unfold-target}}$ component of $\Delta G_{\text{bind}}$. The reaction conditions were the same as with the CAT mRNA, except that the incubation time was 30 times shorter and the ribozyme concentration was 15 times lower. The reaction products were then analyzed by polyacrylamide gel electrophoresis and autoradiography (Fig. 2.4A). The trans-splicing product fractions ranged from $0.16 \pm 0.07\%$, for splice site 369, to $29 \pm 7\%$, for splice site 240 (Table S2.1), confirming that all ribozymes were indeed active in the absence of substrate secondary structure. Because the short substrates required considerably shorter reaction times and lower ribozyme concentrations to yield product fractions comparable to those obtained with CAT mRNA, our results reinforce the notion that substrate folding can significantly reduce trans-splicing efficiency.
For almost all ribozymes, the expected trans-splicing product of 85 nucleotides (8 nucleotides of cleaved substrate plus 77 nucleotides of 3'-exon) was accompanied by two other products, whose lengths were approximately 80 and 90 nucleotides (Fig. 2.4A). The additional products most likely resulted from one or more of the possible side reactions in which group I intron ribozymes are known to engage (55-57). These side reactions depend on the same process modeled in this study, i.e., the binding of ribozyme’s IGS to substrate target site. Nevertheless, because the identities of the side products were not determined experimentally, and because our focus was on the correct trans-splicing product, we took the fraction of the 85-nt product alone as a measure of trans-splicing efficiency for splice sites on the short substrates. The ensuing qualitative correlation of computed $\Delta G_{\text{bind}}$ with experimentally measured efficiency (see below) was not strongly affected when the sum of all three product fractions was used as a measure of trans-splicing efficiency (data not shown).

<table>
<thead>
<tr>
<th>Count</th>
<th>Position of the splice site</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>97</td>
</tr>
<tr>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>8</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>99</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>2</td>
<td>18, 83, 87, 166, 177</td>
</tr>
<tr>
<td>1</td>
<td>14, 65, 73, 84, 110, 131, 153, 175, 179, 180, 197, 356, 448, 499, 573</td>
</tr>
</tbody>
</table>

1 Number of times that a splice site was identified among 66 product sequences.
2 Position of a uridine relative to the adenine of the start codon in CAT mRNA.
Figure 2.4: Trans-splicing efficiency with short, 13-nucleotide substrates. (A) Representative autoradiograms of products from trans-splicing reactions. Each 13-mer substrate contains one of the 18 splice sites that were targeted on CAT mRNA. Each group of lanes shows samples taken after 1, 8, 16, and 31 minutes of reaction and is labeled with the position of the targeted splice site. The positions of the unreacted substrate (s; 13 nt), reaction intermediate (i; 8 nt), and trans-splicing product (p; 85 nt) are indicated. The size of the product was confirmed by comparison with an 85-nt RNA marker in separate experiments (not shown). (B) Plot of experimental trans-splicing efficiency as a function of computed ΔG_{bind}. The thick grey line represents a least-mean-squares exponential fit to the data points, with a coefficient of determination $R^2 = 0.27$. Only splice site 240 (~3.9 kcal/mol; 29% reacted) results in a strong deviation from this trend. Alternatively, a linear fit (not shown) to the same data points yields a correlation coefficient $R = -0.51$, with a probability $p = 0.030$ that the values are not correlated. Error bars are standard deviations from three independent experiments.
To determine whether the two remaining energetic contributions, P1 duplex stability, modeled by ΔG_{hybrid}, and base-pairing interactions between IGS and 3'-exon, modeled by ΔG_{release-IGS}, were able to explain the observed variation in \textit{trans}-splicing efficiency for splice sites on the short substrates, we plotted the \textit{trans}-splicing efficiency measured for each splice site as a function of ΔG_{bind} computed for that site (Fig. 2.4B). This plot reveals a correlation between calculations and experiments that is less tight than the correlation for the full-length CAT mRNA, suggesting that factors other than P1 duplex stability and IGS/3'-exon interactions might influence \textit{trans}-splicing efficiency. One reason why these factors become important for unstructured substrates (Fig. 2.4B) but not for structured substrates (Fig. 2.3B) may be the faster \textit{trans}-splicing kinetics with short substrates. Several processes may play a stronger role at faster reaction kinetics, such as the initial folding of the ribozyme (58, 59) and later steps in the \textit{trans}-splicing process (60, 61), which are not captured by the thermodynamic model used in this study.

\textbf{Computed energetic contributions to \textit{trans}-splicing efficiency}

The good correlation observed between computed values of ΔG_{bind} and experimental \textit{trans}-splicing efficiencies for CAT mRNA supports the notion that the ribozyme-substrate binding process can be approximately modeled by assuming three idealized molecular events: unfolding of target site, release of the IGS, and IGS-target site hybridization. Here we examine the relative energetic contributions of these hypothetical molecular events to the observed efficiency of
a given splice site.

The thermodynamic model (Fig. 2.1A) suggests that a very negative hybridization energy, $\Delta G_{\text{hybrid}}$, would favor trans-splicing, whereas very positive energies of substrate unfolding, $\Delta G_{\text{unfold-target}}$, and of IGS release, $\Delta G_{\text{release-IGS}}$, would disfavor trans-splicing. Figure 2.5A and Table S2.2 show the three energy components of $\Delta G_{\text{bind}}$ for each of the experimentally tested splice sites on CAT mRNA. The average of $\Delta G_{\text{unfold-target}}$ over all splice sites is $1.8 \pm 0.8$ kcal/mol, the average of $\Delta G_{\text{release-IGS}}$ is $1.8 \pm 0.9$ kcal/mol, and the average of $\Delta G_{\text{hybrid}}$ is $-7.2 \pm 1.4$ kcal/mol. Thus, both the strongest energetic contribution (-7.2 kcal/mol) and the largest variation among splice sites ($\pm 1.4$ kcal/mol) come from the hybridization of IGS with the target sites. These results indicate that the free energy of IGS-target site hybridization is more important, on average, than the free energies of target unfolding and IGS release in modulating $\Delta G_{\text{bind}}$.

To assess the importance of the energetic components $\Delta G_{\text{unfold-target}}$, $\Delta G_{\text{release-IGS}}$, and $\Delta G_{\text{hybrid}}$ towards the correlation observed between experimental trans-splicing efficiency and $\Delta G_{\text{bind}}$ (Fig. 2.3B), we recomputed $\Delta G_{\text{bind}}$ by omitting each component in turn, and we calculated the coefficient of determination associated with a least-mean-squares exponential fit to the experimental trans-splicing efficiency as a function of computed $\Delta G_{\text{bind}}$. For splice sites on CAT mRNA, omitting any of the energetic components resulted in a poorer correlation and a consequent loss of a reliable threshold bounding values of $\Delta G_{\text{bind}}$ associated with efficient splice sites (Fig. 2.5, B-D). In particular, omission of $\Delta G_{\text{hybrid}}$ resulted in a non-negative $\Delta G_{\text{bind}}$ for all splice sites (Fig 2.5D), whereas
omission of either $\Delta G_{\text{unfold-target}}$ or $\Delta G_{\text{release-IGS}}$ resulted in the appearance of up to 6 false-positives, i.e., splice sites with $\Delta G_{\text{bind}} < -4$ kcal/mol but without any experimentally detected trans-splicing product (Fig. 2.5, B, C). A poor correlation also ensued when omitting pairs of $\Delta G$ components from $\Delta G_{\text{bind}}$ (Fig S2.2; Table S2.2). Comparing the coefficients of determination (Fig. 2.3B; Fig. 2.5, B-D) indicates that the importance of individual energetic contributions towards the observed correlation between trans-splicing efficiency and $\Delta G_{\text{bind}}$ follows the order $\Delta G_{\text{hybrid}} > \Delta G_{\text{unfold-target}} > \Delta G_{\text{release-IGS}}$, but these energetic contributions are all necessary to achieve a good correlation.

When the same analysis was performed for splice sites on the short substrates, we found that omitting $\Delta G_{\text{unfold-target}}$ made no significant difference because this component was always less than 0.4 kcal/mol, as expected (Fig. S2.1, A, B). Omitting either of the other components resulted in a worse correlation (Fig. S2.1, C, D). The values of $\Delta G_{\text{release-IGS}}$ for the 13-mers were identical to the values for corresponding splice sites on CAT mRNA, which was expected because $\Delta G_{\text{release-IGS}}$ does not depend on the substrate. Values of $\Delta G_{\text{hybrid}}$, however, differed by up to about 1 kcal/mol between corresponding splice sites on 13-mers and CAT mRNA. These small energetic differences arose from the different nucleotides flanking the target sites in the two contexts. Such nucleotides vary with splice site on CAT mRNA (Table S2.1) but do not vary with splice site on the short substrates (see Materials and Methods).
Figure 2.5: Energetic contributions to the computed binding free energy on CAT mRNA, from the three molecular events described in Fig. 2.1A. (A) The computed energetic contributions from target site unfolding ($\Delta G_{\text{unfold-target}}$, grey), ribozyme IGS release ($\Delta G_{\text{release-IGS}}$, white), and IGS-target site hybridization ($\Delta G_{\text{hybrid}}$, black) are shown for each tested splice site. Three of the 18 tested splice sites were omitted (131, 240, 369) because their computed energetic values were not among the strongest 10,000 interactions reported by INTEGRNA. Error bars are standard deviations of the energies calculated with three to six different window sizes as in Fig. 2.3. (B-D) Plots of trans-splicing efficiency as a function of computed $\Delta G_{\text{bind}}$ values when the contribution of (B) $\Delta G_{\text{unfold-target}}$, (C) $\Delta G_{\text{release-IGS}}$, and (D) $\Delta G_{\text{hybrid}}$ is omitted in turn. The thick gray lines in (B-D) represent exponential fits, as in Figures 2.3B and 2.4B, with coefficients of determination $R^2 = 0.38$, 0.57, and 0.076 for (B), (C), and (D), respectively. For additional details see Figure 2.3.
**Trans-tagging assay on CAT mRNA**

To determine whether the results of the *trans*-tagging assay also generate a good correlation with the experimentally determined *trans*-splicing efficiencies, we performed this assay using CAT mRNA as the substrate. The output of the *trans*-tagging assay is the number of times a given splice site is identified in a set of product sequences obtained from *in vitro* *trans*-splicing reactions with randomized IGSs (26). We obtained a total of 66 product sequences, which were consistent with *trans*-splicing at 25 different uridines among the 186 uridines that follow the start codon in the CAT mRNA sequence (Table 2.1). The splice sites detected most frequently were at uridines 97 and 33, with 14 and 12 occurrences, respectively. On the other hand, 15 of the 25 detected splice sites were found only once. The splice sites detected by the *trans*-tagging assay are represented by diamonds in Figure 2.2A.

When the experimentally determined *trans*-splicing efficiencies were plotted as a function of the splice site counts in the *trans*-tagging assay, no correlation was evident (Fig. 2.6). More strikingly, the assay did not find splice sites 258 and 405, even though these splice sites were more efficient than splice site 97, which was detected 14 times by the *trans*-splicing assay. Similarly, the assay found each of the splice sites 448 and 197 only once in 66 product sequences, even though these splice sites were two of the five most efficient ones among the 18 tested splice sites. Therefore, the number of times that a given splice site is found by the *trans*-tagging assay does not necessarily reflect the *trans*-splicing efficiency measured for that splice site.
Figure 2.6: Comparison between trans-splicing efficiency and trans-tagging results. (A) The measured trans-splicing efficiency for each splice site is plotted against the number of times that the splice site was found in 66 product sequences obtained with the trans-tagging assay (filled diamonds). Open diamonds denote splice sites that were not found by the trans-tagging assay. Note that eight of the 18 sequences are clustered at the origin of the plot.

Possible biases of the trans-tagging assay

Why did the trans-tagging results only partially reflect the measured splice site efficiency? One possible explanation is that the PCR step in the assay disfavors long RT-PCR products, a well-known phenomenon in quantitative PCR (62). In our trans-tagging experiments, this phenomenon may have caused the pattern seen in Figure 2.2A. Here, the forward PCR primer was designed to identify splice sites between positions 14 and 643 on CAT mRNA, but only 4 out of the 25 identified splice sites were found at positions beyond uridine 197 (Table 2.1). In contrast, the five most efficient splice sites (also with ΔG_{bind} below the threshold of −4 kcal/mol) were more evenly distributed over the CAT mRNA...
(positions 97, 197, 258, 405, and 448). We reasoned that if the PCR step indeed caused more splice sites to be identified near the forward PCR primer than elsewhere, then moving this primer farther downstream from the mRNA 5'-terminus should reveal a different pattern of detected splice sites, now crowded near the new position of the forward primer. Therefore, we repeated the trans-tagging experiment on CAT mRNA this time using a forward primer designed to identify only splice sites at positions 207 to 643. The resulting nine product sequences yielded eight new splice sites, those at uridines 248, 271, 273, 321, 350, 378, 384, and 405 (not shown in Figure 2.2A). None of these splice sites had been found among the previous set of 66 trans-tagging product sequences. The new splice sites were again crowded near the forward primer. These results suggest that the PCR step may cause the trans-tagging assay to miss efficient splice sites located 200 nucleotides or more downstream of the forward PCR primer.

A second possible explanation for a bias in the trans-tagging assay lies in the different rates at which ribozymes with different IGSs are transcribed by T7 RNA polymerase in vitro. The assay employs a pool of trans-splicing ribozymes whose IGS differs between ribozymes. Because the IGS represents the 5'-terminus of the transcript and because the transcription efficiency of T7 RNA polymerase is strongly dependent on the sequence of the first six nucleotides in the transcript (63, 64) we hypothesized that the difference in transcription efficiency achieved for different ribozymes may bias the results of the trans-tagging assay.
To test this second hypothesis we measured the transcription efficiency for the 18 studied ribozymes. Specifically, the ribozymes were individually transcribed in the presence of α-[³²P]-GTP, the transcription products were separated by denaturing polyacrylamide gel electrophoresis, and the transcription efficiency was measured by quantitating the relevant bands on autoradiograms. The measured transcription efficiency varied 8-fold among the different ribozymes (Fig. 2.7, Table S2.1). Notably, ribozyme 97 had the highest transcription efficiency, 5-fold higher than that of ribozyme 258. These measured differences in ribozyme amounts represent lower bounds for the actual differences affecting the trans-tagging assay. Specifically, in the individual transcription reactions, partial saturation may have been reached for the more efficient transcriptions but not for the less efficient ones. This partial saturation would have reduced the observed difference between efficient and less efficient transcriptions. On the other hand, in the trans-tagging assay, all ribozymes were transcribed together from equimolar amounts of templates under the same reaction conditions, leading to equal extents of product saturation for all ribozymes. Hence, the difference in the amounts of ribozymes 97 and 258 was likely greater than 5-fold in the trans-tagging assay. Therefore, the observed bias in transcription efficiency for ribozymes with different IGSs could explain why splice site 97 was found most frequently in the trans-tagging assay, while splice site 258, which was the most efficient in trans-splicing experiments with CAT mRNA, was absent in all 75 product sequences that we obtained with the trans-tagging assay.
Figure 2.7: Transcription efficiencies of ribozymes analyzed in this study. All values are relative to the transcription efficiency of the ribozyme targeting splice site 97. Transcription efficiencies were determined from band intensities of internally $[^{32}\text{P}]$-labeled ribozymes, which had been transcribed *in vitro* and separated by denaturing polyacrylamide gel electrophoresis. The targeted splice site is indicated at the bottom of each bar. Error bars are standard deviations from seven independent transcriptions.

A third potential source of bias for the outcome of the trans-tagging assay is the tendency of group I intron variants to lose their 3'-exon via side reactions known as 3'-specific hydrolysis and G-exchange (65-68). Indeed, a significant loss of 3'-exons was revealed during the separation of *in vitro* transcribed ribozymes on denaturing polyacrylamide gels (above). Specifically, for each band of a full-length ribozyme we detected a faster migrating band, which corresponded in size to ribozymes that had lost their 3'-exon. If the loss of 3'-exons during transcription and during trans-splicing depends on the IGS, then the pool of transcribed ribozymes used in the trans-tagging assay will be biased against specific splice sites.
To determine whether the loss of 3'-exon varied with IGS under in vitro transcription conditions, we quantitated the relevant bands on the same autoradiograms that were used to measure transcription efficiency. The resulting fractions of lost 3'-exon varied from 12 ± 4% to 39 ± 4%, with an average of 28 ± 7% over all 18 ribozymes (Table S2.1). This means that, among the studied ribozymes, between 88% and 61% of the transcribed ribozymes retained their 3'-exon, generating only a 1.4-fold bias for the trans-tagging assay. Therefore, 3'-exon loss during transcription was not a major contributor to the large bias observed in the trans-tagging assay.

The loss of 3'-exons could also have taken place during the trans-splicing reaction. Because these conditions were different from the transcription conditions, we also measured the loss of 3'-exons under trans-splicing conditions. Internally radiolabeled ribozymes with a 3'-exon were size purified by denaturing PAGE, then incubated under trans-splicing conditions, and the fraction of ribozymes that lost their 3'-exon was determined by a second denaturing PAGE and quantitation by phosphorimaging. The fraction of 3'-exons lost after four hours varied from 22 ± 4% to 65 ± 7% with an average of 40 ± 10% over all ribozymes (Table S2.1). Therefore, we estimate that, during the 1-hour incubation under trans-tagging conditions, between ~80% and ~95% of the ribozymes retained their 3'-exon, suggesting that the small proportion of ribozymes that lost their 3'-exon should not constitute a major bias in the trans-tagging assay.

In summary, we found that the results of the trans-tagging assay were
skewed by experimental biases. The strongest influences appeared to originate from a product size bias in the PCR step and from different efficiencies in the transcription of ribozymes, whereas a smaller influence came from the loss of 3'-exons by ribozymes during transcription and trans-splicing.

2.4 Discussion

We have developed and experimentally tested a computational approach to identify efficient splice sites for trans-splicing ribozymes. This approach is based on the computation of the free energy change, $\Delta G_{\text{bind}}$, for the hybridization and secondary structure rearrangements accompanying the first step of trans-splicing, i.e., the binding of IGS to substrate. The computed values of $\Delta G_{\text{bind}}$ were found to correlate well with experimentally determined trans-splicing efficiencies at 18 different splice sites on CAT mRNA. The correlation was significantly better than the correlation of experimental trans-splicing efficiencies with the results of the trans-tagging assay, suggesting that the proposed computational approach could provide an alternative solution to the identification of efficient splice sites for trans-splicing ribozymes.

In particular, our results suggest that a set of candidate efficient splice sites for trans-splicing ribozymes could be determined by selecting those sites with $\Delta G_{\text{bind}} < -4$ kcal/mol. Although other mRNAs may show different values for $\Delta G_{\text{bind}}$ that better discriminate between efficient and inefficient splice sites, a threshold of $\Delta G_{\text{bind}} < -4$ kcal/mol may provide a rough guideline for choosing candidate splice sites on other mRNA substrates. Our results also suggest that
the unfolding of substrate mRNA at the target site, the hybridization of the substrate to the ribozyme, and the release of IGS from its secondary interactions with the 3'-exon all contribute towards the efficiency of a given splice site on long mRNAs, albeit to varying degrees.

The above results may seem unsurprising at first, because computations similar to ours have already been applied to identify target sites for different types of RNA-binding molecules. For example, a thermodynamic cycle equivalent to our calculation of \( \Delta G_{\text{bind}} \) predicted antisense oligonucleotides with high binding affinity for rabbit \( \beta \)-globin and mouse tumor necrosis factor-\( \alpha \) mRNAs, achieving 60% accuracy and a significant correlation with experimental data (47). Analogous calculations included the concentration-dependent effects of oligonucleotide dimerization, predicting oligonucleotide-target affinities consistent with results from an RNase-H mapping assay on the AT\(_1\) receptor mRNA, and from a trans-tagging assay on sickle \( \beta \)-globin mRNA (48). More recently, a target site accessibility measure provided by the program RNA PLFOLD and related to our \( \Delta G_{\text{unfold-target}} \) was used to improve the prediction of effective siRNAs (44). A similar calculation also improved and simplified the prediction of effective microRNA targets in \textit{Drosophila melanogaster} tissue culture cells (69). Another study used a statistical sampling technique to calculate \( \Delta G \) values analogous to our \( \Delta G_{\text{bind}} \) (70). These values were found to correlate significantly with experimentally measured trans-cleavage activities of 15 hammerhead ribozymes targeting transcripts of human breast cancer resistance protein. Contrary to our results, this study found that \( \Delta G_{\text{unfold-target}} \) contributes more than \( \Delta G_{\text{hybrid}} \) toward
the observed correlation, underscoring the structural and mechanistic differences between trans-cleaving and trans-splicing ribozymes (71). Lastly, the program INTARNA, which we used to calculate ΔG\textsubscript{bind}, was more effective than other software in predicting 18 targets of small bacterial regulatory RNAs (50).

Despite the previous achievements by similar methods, it was not obvious at the outset whether our proposed computations of ΔG\textsubscript{bind} would correlate well with the efficiency of trans-splicing ribozymes, because these molecules are more complex than those investigated in most of the previous reports. Although Mathews et al. (48) found some agreement between their ΔG\textsubscript{bind} calculations and the counts of accessible splice sites reported by Lan et al. (12), this comparison considered only a short 70-nucleotide region of the mRNA substrate and did not directly measure trans-splicing efficiency. In contrast, our study individually tested the trans-splicing efficiency on 18 different splice sites, distributed over the full length of the CAT mRNA, and compared the experimental results to the computed values of ΔG\textsubscript{bind}. Our study also included the ribozyme 3'-exon into the calculations. The good correlation we observed between experiments and computation confirms that a calculation of ΔG\textsubscript{bind} based on established RNA folding algorithms can be used to identify efficient splice sites for trans-splicing ribozymes.

The scope of the present study was intentionally limited to providing a first assessment of the proposed calculation of ΔG\textsubscript{bind} as a tool for predicting efficient splice sites. Therefore, we carried out experiments using a single model mRNA substrate and we targeted a subset of the possible splice sites on this substrate.
On the other hand, the *trans*-tagging assay has already been successfully applied to map efficient splice sites on at least eight different mRNA substrates (12, 28-30, 32-35). Moreover, the proposed calculation of $\Delta G_{\text{bind}}$ relies on secondary structure prediction algorithms that are very fast but not always accurate because they do not take into account all possible RNA interactions and they are sensitive to errors in experimental energy parameters (51, 72). Thus, future experimental studies on additional mRNA substrates will show whether calculations of $\Delta G_{\text{bind}}$ provide a reliable means of predicting efficient splice sites.

The proposed method is attractive because the calculations of $\Delta G_{\text{bind}}$ do not require bench-work and can be performed in one afternoon. In contrast, the experimental assay, from the preparation of ribozymes and substrate mRNA to the analysis of *trans*-splicing product sequences, requires at least one week of work. Thus a computational approach, once proven reliable, will likely be cheaper and quicker than the experimental route. Moreover, the computational approach can be developed further to satisfy additional design requirements. For example, if the aim of the ribozymes is to repair a mutated mRNA sequence, then the sequence of the 3'-exon must be adjusted according to the targeted splice site (73). This adjustment, which is likely important based on our analysis of energetic contributions towards *trans*-splicing efficiency, would be difficult to achieve with the *trans*-tagging assay, whose pool of ribozymes presently carries the same 3'-exon sequence for all targeted splice sites. On the other hand, the proposed computational approach can easily be modified to adjust the 3'-exon sequence in accordance with the splice site.
Another design requirement might be to avoid trans-splicing ribozymes that react at multiple splice sites. This behavior was observed in the present study for ribozymes targeting splice sites 405 and 448 (Fig. 2.3A). Such cross-reactivity was likely caused by the similarity of the corresponding IGSs, which differ in only one out of six nucleotides (GGGGAA for splice site 405, versus GGGGAU for splice site 448). A computational approach can easily overcome this problem, as a simple comparison of the predicted IGSs suffices to detect and discard potentially cross-reactive ribozymes.

The computational approach could also accommodate the design of 5'-extended guide sequences (EGSs), which have been shown to strongly increase trans-splicing efficiency (11, 74). The design of these EGSs depends critically on the mRNA sequence immediately downstream of the splice site. If the trans-tagging assay were to be used for the simultaneous optimization of IGSs and EGSs then the sequence of the EGSs would have to co-vary with the randomized IGS. This would be a very laborious task because each sequence would have to be synthesized individually. Therefore, EGS optimization cannot be easily achieved with the trans-tagging assay. In contrast, our computational approach can be modified to include a splice site-specific EGS optimization, which would allow one not only to identify efficient splice sites on a given mRNA but also to maximize their efficiency.

In conclusion, the proposed computational approach, together with future algorithmic extensions aimed at optimizing 3'-exon and EGS, could facilitate the development of efficient trans-splicing ribozymes while elucidating the
interactions between *trans*-splicing ribozymes and their mRNA substrates.

2.5 Materials and Methods

**Prediction of ΔG_{bind}**. The values of the binding free energy change, ΔG_{bind}, and its components were initially computed using the Vienna RNA package (75). Specifically, we used RNAEVAL to calculate ΔG_{hybrid}, and RNAFOLD to calculate ΔG_{unfold-target} and ΔG_{release-IGS} using the partition function approach, which yields an ensemble average of ΔG values associated with all possible pseudo knot-free secondary structures achievable with a given RNA sequence (76). We then found that the same calculations can be carried out more conveniently using the INTARNA software ((50); [http://rna.informatik.uni-freiburg.de:8080/IntaRNA.jsp](http://rna.informatik.uni-freiburg.de:8080/IntaRNA.jsp)), which was developed to find putative target sites for bacterial small regulatory RNAs (sRNAs, typically 50-250 nt in length) on long mRNA sequences. This software takes as input two sequences: a short sRNA and a long target mRNA. The software outputs a list of possible base-pairing interactions between the two RNAs, together with the corresponding ΔG_{bind} = ΔG_{hybrid} + ΔG_{unfold-mRNA} + ΔG_{unfold-sRNA}, where ΔG_{hybrid} is the hybridization free energy component of the interaction, and ΔG_{unfold-mRNA/sRNA} is the cost in free energy for locally unfolding the region of mRNA/sRNA involved in the base-pairing interaction. To calculate ΔG_{bind} for the possible splice sites on CAT mRNA, the 678-nt sequence of this substrate was specified as the input mRNA sequence, and the shortened ribozyme sequence

GXXXXXNNNNNACGCACGTCAATTGGCCGCTGGATGGGGCCCCTGTGAAG
TGTTGCTGAGCAACGCGCTGGCGCGGCTCAGAGGCTTC was specified as the input sRNA sequence, where GXXXXX denotes a specific IGS, NNNNNN is a linker replacing the body of the ribozyme, and the rest of the sequence is the 77-nt 3'-exon used for the trans-tagging and trans-splicing experiments in this study. To calculate ΔG\textit{bind} for the short 13-mer substrates, the input sRNA sequence was the shortened ribozyme sequence shown above, and the input mRNA sequence was the 13-mer substrate sequence GGYYYYYUAAAAA, where YYYYYY is the reverse complement of the five IGS bases XXXXX. For all calculations, the maximum length of the hybridized region was seven, the upper energy threshold was 10 kcal/mol, the exact number of seed base pairs was three, the maximum number of reported sub-optimal interactions was 10,000, while all other parameters of the software had default values. Requesting 10,000 sub-optimal interactions was necessary to obtain ΔG\textit{bind} values for as many as possible of the least efficient splice sites tested in this study. However, requesting 40 sub-optimal interactions was sufficient to predict the five most efficient of the tested splice sites. Only interactions involving each IGS and the corresponding target site on the substrate were extracted from the program’s output. The above procedure was automated using a short Perl script, which is available from G.A. upon request. Values of ΔG\textit{bind} for CAT mRNA were computed using window sizes of 100, 200, 300, 400, 500, and 600 nucleotides, which cover uniformly the length of the substrate. Because each window size produced different outliers in the plot of experimental trans-splicing efficiency versus ΔG\textit{bind} (data not shown), we averaged ΔG\textit{bind} over all six window sizes,
thus obtaining a more consistent trend (Fig. 2.3B). Error bars of $\Delta G_{\text{bind}}$ in Fig 2.3B are standard deviations over the six window sizes. Splice sites whose $\Delta G_{\text{bind}}$ value was reported for less than three window sizes were not assessed for energetic contributions and are not shown in Fig. 2.3B. When no $\Delta G_{\text{bind}}$ values were returned for specific combinations of window size and splice site the value zero was used for $\Delta G_{\text{bind}}$ in Fig. 2.2. All calculations for the 13-mer substrates were done with a window size of 13 nucleotides.

**Ribozymes.** To prepare DNA templates for transcription of *trans*-splicing ribozymes with specific IGSs (Table S2.1), a DNA fragment for each such ribozyme was amplified by PCR, using forward primer GCGTAATACGACTCACTATAGXXXXXAAAAGTTATCAGGCATGCACC and reverse primer GAAGCCTCTGAGCCGCGCCAG (PR1), from a plasmid containing the Tetrahymena ribozyme gene linked to a 77-nt segment of the alpha-mannosidase gene as the 3'-exon (27). The above forward primer contains the promoter for T7 RNA polymerase, the desired IGS nucleotides GXXXXX, and the first 21 nucleotides downstream of the IGS in the Tetrahymena ribozyme sequence. All PCR products were cloned into the EcoRI / BamHI sites of pUC19 using appropriate PCR primers, and the ribozyme sequences were confirmed by sequencing. DNA templates for ribozyme transcription were obtained from these ribozyme-encoding plasmids by PCR amplification using forward primer GCGTAATACGACTCACTATAG and reverse primer PR1. To minimize the loss of 3'-exons, ribozyme transcriptions were carried out at 30°C for 20 - 30 minutes.
The transcripts were purified by denaturing polyacrylamide gel electrophoresis (PAGE), and eluted from gel slices. Ribozyme concentrations were determined from absorbance measurements at 260 nm using an extinction coefficient of 4.0 μM⁻¹cm⁻¹.

**Substrates.** The template for run-off transcription of CAT mRNA was amplified by PCR from plasmid pLysS (Novagen), in two consecutive PCRs. The first PCR used forward primer CAGGAGCTAAGGAAGCTAAAATG and reverse primer CGCCCCGCCCCTGCCACTCATC; the second PCR used forward primer GCGTAATACGACTCACTATAGCAGGAGCTAAGGAAGCTAAAATG and the same reverse primer. After transcription the full-length CAT mRNA was purified on Micro Bio-Spin 6 columns (Bio-Rad), dephosphorylated with Antarctic phosphatase (NEB), and radiolabeled at the 5'-terminus using γ-[³²P]-ATP (Perkin Elmer) and polynucleotide kinase (NEB). The radiolabeled substrate was purified by denaturing 5% PAGE and eluted from gel slices using RNA elution solution (20 - 50 mM MOPS/NaOH pH 7.0, 0.2% SDS, 300 mM NaCl). The 13-mer substrates were prepared as described previously (64) to obtain the sequences GGYYYYYUAAAAA, where YYYYY is the reverse complement of the five IGS bases XXXXX adjacent to the 5'-terminal G of the ribozyme. Templates for transcription were obtained by annealing the sense oligonucleotide GCTAATACGACTCACTATAG, which contains a T7 RNA polymerase promoter, with antisense DNA oligonucleotides TTTTAXXXXXCCTATAGTGAGTCGTAT TAGC. In vitro transcription was carried out at 37°C for one hour in the presence
of $\alpha-[^{32}\text{P}]-\text{UTP}$ (Perkin Elmer), using 500 nM DNA template and ~1 unit/μL of T7 RNA polymerase. The radiolabeled transcripts were purified by denaturing 20% PAGE.

**Trans-splicing reactions.** Ribozymes and substrates were mixed in a buffer containing 1 mM MgCl$_2$, 135 mM KCl, 50 mM MOPS/NaOH pH 7.0, 20 μM GTP, and 2 mM spermidine, and incubated at 37°C. Before mixing with the substrate, the ribozymes were pre-incubated for 10 min at 37°C in the absence of magnesium. In reactions with CAT mRNA substrate, ribozyme concentrations were 1.5 μM, substrate concentration was ~50 nM, and the incubation time was 4 hours. This reaction time was chosen as a measure for trans-splicing efficiency because after this time the signal for the most efficient splice site (258) began saturating, while the signals for most other splice sites were relatively weak or absent. In reactions with 13-mer substrates, ribozyme concentrations were 100 nM, substrate concentrations were ~10 nM, and the incubation time was 8 minutes. At this time none of the reaction products appeared saturated and all ribozyme-substrate pairs yielded bands sufficiently strong for quantitation. The samples were separated by denaturing 20% PAGE and visualized by phosphorimaging (Bio-Rad PMI). Band intensities in the resulting digital images were quantified using custom computer software that employs previously described curve-fitting procedures (77, 78). Trans-splicing efficiencies were calculated using the formula 

\[ \text{(trans-splicing efficiency)} = \frac{\text{(trans-splicing product)}}{\text{(trans-splicing product + 5'-cleavage product + unreacted substrate)}} \]
amount of each species was assumed proportional to the corresponding band intensity. Error bars were calculated as standard deviations over three independent experiments.

**Confirmation of specific trans-splicing products.** Samples from trans-splicing reactions that yielded visible bands on autoradiograms of polyacrylamide gels were reverse-transcribed using AMV reverse transcriptase (NEB) with reverse primer PR2. The RT products were amplified by PCR using the nested reverse primer PR3 and one of the forward primers PF1, CGGAATTTCATATGGGATAGTGTTCACCC, CGGAATTCCGTTGATATATCCCAATGGC, CGGAATTCTGGCCTATTTCCCTAAAGGG, and GCCTTTATTCACATTCTTGCC, which anneal at different positions on the CAT mRNA sequence. The PCR products were cloned into the plasmid pUC19 and sequenced.

**Trans-tagging assay.** The trans-tagging assay was carried out essentially as described (27). All ribozymes were synthesized by run-off in vitro transcription using T7 RNA polymerase, from DNA templates that were generated by PCR. Preparation of the ribozyme pool was as described above for ribozymes with specific IGSs, except that here the forward primer encoding the T7 promoter contained five randomized nucleotides at the positions denoted with X. The trans-tagging reactions contained 100 nM substrate, 10 nM ribozyme, 0.2 mM GTP, 5 mM MgCl$_2$, 50 mM MOPS/KOH pH 7.0, 135 mM KCl, and 2 mM spermidine. After the substrate and the ribozyme were pre-incubated in separate
tubes in the absence of MgCl₂ for 10 minutes at 37°C, all reagents were mixed and incubated for 1 hour at 37°C. Reaction products were reverse transcribed with AMV reverse transcriptase (NEB) using primer GAAGCCTCTGAGCCCG (PR2). The template RNA was hydrolyzed by incubation in 200 mM NaOH at 90°C for 10 min. RT products were amplified by PCR using nested reverse primer GCGGATCCTGCTCAGCAACACTTCACAGG (PR3) and forward primers CGAATTCCAGGAGCTAAAGGCTAAATG (PF1) or CGGAATTCCATTC TTGCCCAGCTGATGAATGC, cloned into pUC19, and sequenced. The resulting sequences were compared with the ribozyme 3'-exon sequence and the CAT mRNA sequence, to determine the positions of the splice sites on CAT mRNA.

**Ribozyme transcription efficiency and 3'-exon loss.** Ribozymes with specific IGSs (Table S2.1) were individually transcribed by T7 RNA polymerase in 20-μL reactions containing α-[³²P]-GTP (Perkin Elmer) from DNA templates that were prepared as described above. The reactions were carried out at 30°C for 20 min and were stopped by the addition of formamide loading buffer. The ribozymes were heat-denatured at 80°C for 2 min and separated by denaturing 5% PAGE. The resulting gels were analyzed as described above for the trans-splicing reactions. A total of seven independent reactions per ribozyme were carried out. To calculate the relative transcription efficiencies, the absolute intensity of the band corresponding to each intact ribozyme was divided by the absolute intensity of the band corresponding to ribozyme 97. Error bars are standard deviations of the relative transcription efficiencies. To calculate the
fraction of 3'-exon lost during transcription for each ribozyme, the quantitated intensity of the faster migrating band was divided by the sum of the intensities of this band and of the band for the intact ribozyme. Error bars are standard deviations of these fractions. Ribozymes were eluted from the gel slices and individually incubated without substrate under the same trans-splicing conditions described above. Samples taken at times 0 and 4 hours reaction time were analyzed by denaturing 5% PAGE and autoradiography as described above for trans-splicing reaction products. No loss of 3'-exon was seen at time 0. Fractions of lost 3'-exon were calculated as for the transcription reactions.

2.6 Acknowledgements

This work was supported by the National Institutes of Health [T32DK007233 to K.E.O; Hemoglobin and Blood Protein Chemistry training grant to E. Komives], the ARCS Foundation, San Diego Chapter [Scholarship to D.M.], and U.C. San Diego [to G.A.]. We thank the reviewers for their constructive comments, which resulted in several improvements to the manuscript.

Chapter 2, in full, is a reprint of the material as it appears in RNA, Meluzzi D., K.E. Olson, G.F. Dolan, G. Ayra, and U.F. Müller. RNA, 2012. 18(3): p. 590-602. The dissertation author was the second author on this paper.
Table S2.1: Experimentally tested splice sites on CAT mRNA.

<table>
<thead>
<tr>
<th>Position</th>
<th>IGS</th>
<th>Target site</th>
<th>Trans-tagging number</th>
<th>CAT mRNA $-\Delta G_{\text{bound}}$ (kcal/mol)</th>
<th>Product (%)</th>
<th>Short substrate $-\Delta G_{\text{bound}}$ (kcal/mol)</th>
<th>Product (%)</th>
<th>Ribozyme transcription (%)</th>
<th>RNA Transcription (%)</th>
<th>Trans-splicing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>83</td>
<td>GCUGAC</td>
<td>CAGUCA/GGUG</td>
<td>2</td>
<td>3.3±0.6</td>
<td>1.64±0.14</td>
<td>5.0</td>
<td>13±6</td>
<td>36±6</td>
<td>36±3</td>
<td>42±3</td>
</tr>
<tr>
<td>87</td>
<td>GCACGAC</td>
<td>CAGUCA/GUGCA</td>
<td>2</td>
<td>1.1±0.5</td>
<td>1.1±0.2</td>
<td>4.5</td>
<td>5.1±1.9</td>
<td>48±7</td>
<td>32±4</td>
<td>49±4</td>
</tr>
<tr>
<td>97</td>
<td>GGUGAC</td>
<td>AUGUACC/GAU</td>
<td>14</td>
<td>4.5±0.3</td>
<td>7.9±1.0</td>
<td>5.7</td>
<td>12.9±0.7</td>
<td>100</td>
<td>30±10</td>
<td>65±3</td>
</tr>
<tr>
<td>131</td>
<td>AGGCC</td>
<td>ACGUGGUGU</td>
<td>1</td>
<td>N/C</td>
<td>N/D</td>
<td>2.3</td>
<td>2.7±0.2</td>
<td>22±3</td>
<td>21±7</td>
<td>59±5</td>
</tr>
<tr>
<td>197</td>
<td>GGGCCG</td>
<td>GCCCGGCGGA</td>
<td>1</td>
<td>4.2±0.5</td>
<td>6.0±1.2</td>
<td>7.0</td>
<td>10.1±1.3</td>
<td>20±3</td>
<td>12±4</td>
<td>25±20</td>
</tr>
<tr>
<td>222</td>
<td>GGGGAA</td>
<td>AAUUCGGAU</td>
<td>0</td>
<td>2.6±0.09</td>
<td>0.42±0.05</td>
<td>4.8</td>
<td>9±6</td>
<td>43±8</td>
<td>18±2</td>
<td>39±5</td>
</tr>
<tr>
<td>240</td>
<td>GCCGUC</td>
<td>AACAGGGGA</td>
<td>0</td>
<td>N/C</td>
<td>N/D</td>
<td>3.9</td>
<td>29±7</td>
<td>17±3</td>
<td>28±6</td>
<td>22±4</td>
</tr>
<tr>
<td>258</td>
<td>GUGGCA</td>
<td>UAUGGAGUA</td>
<td>0</td>
<td>6.0±0.10</td>
<td>14.3±1.5</td>
<td>6.4</td>
<td>12±2</td>
<td>21±6</td>
<td>30±8</td>
<td>27±8</td>
</tr>
<tr>
<td>273</td>
<td>GACGCC</td>
<td>ACUAGGCUGA</td>
<td>0</td>
<td>2.9±0.2</td>
<td>N/D</td>
<td>5.4</td>
<td>4±2</td>
<td>70±10</td>
<td>21±2</td>
<td>47±4</td>
</tr>
<tr>
<td>325</td>
<td>GUGCAC</td>
<td>CAGUGAAUGA</td>
<td>0</td>
<td>3.4±0.11</td>
<td>N/D</td>
<td>4.7</td>
<td>1.0±0.2</td>
<td>13±3</td>
<td>36±3</td>
<td>42±7</td>
</tr>
<tr>
<td>346</td>
<td>GUGGC</td>
<td>CCGUCGUGU</td>
<td>0</td>
<td>3.1±0.4</td>
<td>N/D</td>
<td>6.1</td>
<td>20±8</td>
<td>54±12</td>
<td>32±5</td>
<td>37±4</td>
</tr>
<tr>
<td>350</td>
<td>GGAACG</td>
<td>CAGUUCUGA</td>
<td>0</td>
<td>2.5±0.13</td>
<td>2.2±0.7</td>
<td>4.0</td>
<td>0.4±0.5</td>
<td>70±20</td>
<td>31±2</td>
<td>48±3</td>
</tr>
<tr>
<td>369</td>
<td>GUCGUG</td>
<td>GCGAAGUGU</td>
<td>0</td>
<td>N/C</td>
<td>N/D</td>
<td>2.3</td>
<td>0.16±0.07</td>
<td>21±3</td>
<td>39±4</td>
<td>28±3</td>
</tr>
<tr>
<td>378</td>
<td>GCAUGC</td>
<td>UGCGGUGUA</td>
<td>0</td>
<td>3.6±0.5</td>
<td>N/D</td>
<td>6.9</td>
<td>29±5</td>
<td>15±3</td>
<td>36±9</td>
<td>24±5</td>
</tr>
<tr>
<td>405</td>
<td>GGAGGAA</td>
<td>AUUUCGGAUAA</td>
<td>0</td>
<td>5.32±0.12</td>
<td>9.3±1.6</td>
<td>6.3</td>
<td>21±4</td>
<td>50±9</td>
<td>30±10</td>
<td>65±7</td>
</tr>
<tr>
<td>448</td>
<td>GGAGGUA</td>
<td>AUUUCGGAU</td>
<td>0</td>
<td>5.4±0.3</td>
<td>10.8±2.3</td>
<td>6.3</td>
<td>25±10</td>
<td>50±10</td>
<td>25±6</td>
<td>60±2</td>
</tr>
<tr>
<td>518</td>
<td>GGUGUG</td>
<td>UUCCAGAGUGG</td>
<td>0</td>
<td>2.8±0.7</td>
<td>N/D</td>
<td>5.8</td>
<td>4±1.6</td>
<td>28±6</td>
<td>19±8</td>
<td>42±6</td>
</tr>
<tr>
<td>551</td>
<td>GCCACC</td>
<td>AAGGCGUGCA</td>
<td>0</td>
<td>3.7±0.3</td>
<td>N/D</td>
<td>6.8</td>
<td>14±1.6</td>
<td>60±20</td>
<td>24±7</td>
<td>35±3</td>
</tr>
</tbody>
</table>

3 Position of splice site uridine relative to the adenosine of the AUG translation start codon on CAT mRNA.
4 The underlined sequence corresponds to the nucleotides that base pair to the IGS. Note that all IGS-target sequence helices contain a G:U pair at the splice site.
5 Average and standard deviation over three to six window sizes (100, 200, 300, 400, 500, 600 nucleotides). N/C, $\Delta G_{\text{bound}}$ was not calculated by INTARNAs with at least three window sizes, because the interaction was not among the strongest 10,000.
6 Average and standard deviation over three independent reactions. N/D, product fractions could not be determined because product bands were not visible on autoradiograms.
7 Ribozyme transcription efficiency relative to the amount of the ribozyme 97% transcribed under identical conditions.
8 Fraction of ribozyme that lost its 3'-exon during 20 minutes of in vitro transcription at 30°C.
9 Fraction of ribozyme that lost its 3'-exon during 4 hours of incubation under trans-splicing conditions in the absence of mRNA substrate.
10 The sequences of subcloned RT-PCR products from reactions that targeted splice site 350 were consistent with trans-splicing primarily at splice site 307.
Table S2.2: Energetic contributions to $\Delta G_{\text{bind}}$ for the splice sites tested on CAT mRNA and on the short, 13-nucleotide substrates.

<table>
<thead>
<tr>
<th>Position</th>
<th>IGS</th>
<th>CAT mRNA (kcal/mol)</th>
<th>Short substrate (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\Delta G_{\text{unf}}$-target</td>
<td>$\Delta G_{\text{release}}$-IGS</td>
</tr>
<tr>
<td>83</td>
<td>GCUGAC</td>
<td>1.3±0.6</td>
<td>3.0±0.3</td>
</tr>
<tr>
<td>87</td>
<td>GGCAAC</td>
<td>2.9±0.5</td>
<td>2.2±0.2</td>
</tr>
<tr>
<td>97</td>
<td>GGGUAC</td>
<td>1.6±0.3</td>
<td>1.1±0.0</td>
</tr>
<tr>
<td>131</td>
<td>GAGGCC</td>
<td>N/C</td>
<td>N/C</td>
</tr>
<tr>
<td>197</td>
<td>GGCGGG</td>
<td>1.0±0.8</td>
<td>3.5±0.1</td>
</tr>
<tr>
<td>222</td>
<td>GGCGAA</td>
<td>2.3±0.1</td>
<td>2.3±0.1</td>
</tr>
<tr>
<td>240</td>
<td>GCGGUC</td>
<td>N/C</td>
<td>N/C</td>
</tr>
<tr>
<td>258</td>
<td>GUCCCA</td>
<td>0.3±0.1</td>
<td>1.8±0.0</td>
</tr>
<tr>
<td>273</td>
<td>GAAGG</td>
<td>1.9±0.2</td>
<td>1.6±0.0</td>
</tr>
<tr>
<td>325</td>
<td>GUUCA</td>
<td>1.3±0.1</td>
<td>0.4±0.0</td>
</tr>
<tr>
<td>346</td>
<td>GUCGCC</td>
<td>2.4±0.4</td>
<td>3.4±0.1</td>
</tr>
<tr>
<td>350</td>
<td>GAAAC</td>
<td>1.2±0.2</td>
<td>0.5±0.0</td>
</tr>
<tr>
<td>369</td>
<td>GUCUUG</td>
<td>N/C</td>
<td>N/C</td>
</tr>
<tr>
<td>378</td>
<td>GCAGC</td>
<td>2.7±0.5</td>
<td>1.8±0.0</td>
</tr>
<tr>
<td>405</td>
<td>GGGGAAG</td>
<td>1.1±0.1</td>
<td>1.3±0.0</td>
</tr>
<tr>
<td>448</td>
<td>GGGGAG</td>
<td>1.1±0.3</td>
<td>1.3±0.0</td>
</tr>
<tr>
<td>518</td>
<td>GGUGUG</td>
<td>3.0±1.0</td>
<td>1.3±0.3</td>
</tr>
<tr>
<td>551</td>
<td>GCCAC</td>
<td>2.9±0.5</td>
<td>2.3±0.0</td>
</tr>
</tbody>
</table>

1 Average and standard deviation over three to six window sizes (100, 200, 300, 400, 500, 600 nucleotides).
2 The computations for the short substrates used a single window including the full substrate sequence. Hence, no standard deviations are reported for these $\Delta G$ values.
3 Position of splice site uridine relative to the adenosine of the AUG translation start codon on CAT mRNA.
4 The computation of $\Delta G_{\text{release}}$-IGS did not make use of a window because the substrate was not involved. Hence, no standard deviations are reported for $\Delta G_{\text{release}}$-IGS.
5 N/C, $\Delta G$ was not calculated by INTA RNA for at least three window sizes, because the interaction was not among the strongest 10,000 queried from the program.
Figure S2.1: Energetic contributions to $\Delta G_{\text{bind}}$ from the three molecular events for the short, 13-nucleotide substrates. (A) The computed energetic contributions of substrate target site unfolding ($\Delta G_{\text{unfold-target}}$, grey), ribozyme IGS release ($\Delta G_{\text{release-IGS}}$, white), and IGS-target site hybridization ($\Delta G_{\text{hybrid}}$, black), are shown as a function of the splice sites. (B) Correlation of experimentally determined trans-splicing efficiency with $\Delta G_{\text{bind}}$ when $\Delta G_{\text{unfold-target}}$ is omitted. The thick gray line represents a least-mean-squares exponential fit with coefficient of determination $R^2 = 0.26$. An alternative linear fit (not shown) yields a correlation coefficient of $R = -0.51$, with a probability $p = 0.031$ that the values are not correlated. (C) Correlation when $\Delta G_{\text{release-IGS}}$ is omitted ($R^2 = 0.21; R = -0.54; p = 0.020$). (D) Correlation when $\Delta G_{\text{hybrid}}$ is omitted ($R^2 = 0.0079; R = 0.11; p = 0.68$). For additional details see Figure 3.
Figure S2.2: Correlation between trans-splicing efficiencies experimentally determined with CAT mRNA substrate and the three individual energetic contributions to ΔG_{bind}. (A) The correlation of trans-splicing efficiency with ΔG_{bind} serves as comparison. This is the same plot shown in Figure 3B. The thick gray line is a least-mean-squares exponential fit with coefficient of determination $R^2 = 0.87$. An alternative linear fit (not shown) yields a correlation coefficient of $R = -0.75$, with a probability $p = 0.00033$ that the values are not correlated. (B) Correlation of trans-splicing efficiency with ΔG_{unfold-target} ($R^2 = 0.064; R = -0.32; p = 0.20$). (C) Correlation of trans-splicing efficiency with ΔG_{release-IGS} ($R^2 = 0.0065; R = 0.099; p = 0.70$). (D) Correlation of trans-splicing efficiency with ΔG_{hybrid} ($R^2 = 0.14; R = -0.38; p = 0.12$). For additional details see Figure 3.
2.8 References


Chapter 3

An *in vivo* selection method to optimize *trans*-splicing ribozymes

3.1 Abstract

Group I intron ribozymes can repair mutated mRNAs by replacing the 3’-terminal portion of the mRNA with their own 3’-exon. This *trans*-splicing reaction has the potential to treat genetic disorders, and to selectively kill cancer cells or virus infected cells. However, these ribozymes have not yet been used in therapy, partially due to a low *in vivo* *trans*-splicing efficiency. Previous strategies to improve the *trans*-splicing efficiencies focused on designing and testing individual ribozyme constructs. Here we describe a method that selects the most efficient ribozymes from millions of ribozyme variants. This method uses an *in vivo* rescue assay where the mRNA of an inactivated antibiotic resistance gene is repaired by *trans*-splicing group I intron ribozymes. Bacterial cells that express efficient *trans*-splicing ribozymes are able to grow on medium containing the antibiotic, chloramphenicol. We randomized a 5’-terminal sequence of the *Tetrahymena thermophila* group I intron and screened a library with $9 \times 10^6$ ribozyme variants for the best *trans*-splicing activity. The resulting ribozymes showed increased *trans*-splicing efficiency and help the design of efficient *trans*-splicing ribozymes for different sequence contexts. This *in vivo* selection method can now be used to optimize any sequence in *trans*-splicing ribozymes.
3.2 Introduction

Group I intron ribozymes are catalytic RNAs (ribozymes) that can splice themselves out of primary transcripts, many of them without the help of proteins (1). In contrast to these naturally cis-splicing ribozymes, man-made variants can catalyze trans-splicing reactions (for a recent review see (2)). The trans-splicing group I introns in our study lack a 5'-exon, and their recognition sequence is modified such that the ribozyme can base pair to a target site on the substrate RNA. In this format, the ribozyme replaces the 3'-fragment of the substrate RNA downstream of the splice site, with its own 3'-exon (Fig. 3.1). This replacement can repair an mRNA if the mRNA carries a mutation in its 3'-fragment and the ribozyme 3'-exon is the sequence of the mRNA 3'-fragment without mutation (3). By being able to change the sequence of specific mRNAs, trans-splicing ribozymes take an important place in the toolbox of RNAs that can be used to alter the RNA inventory of a cell: The degradation of specific cellular RNAs can be affected by siRNAs, hammerhead ribozymes, HDV ribozymes, and the RNase P ribozyme. In contrast, trans-splicing ribozymes can alter the sequence of specific cellular RNAs. The alteration of RNA sequences in the cell could be used to repair mRNAs in the treatment of genetic disorders (3) and for the specific killing of cells that express viral RNAs (4, 5), and cancer cells by targeting their increased telomerase reverse transcriptase mRNA (6). Cell death is accomplished by a toxic peptide that is encoded in the transferred 3'-exon (5).

Current trans-splicing group I intron ribozymes are not efficient enough in
vivo for therapeutic applications. One way to increase the level of trans-splicing products in vivo is to raise the ribozyme expression levels, which can facilitate the repair of 10 - 50% of a specific mRNA in cells under ideal conditions (7-9). However, the high ribozyme expression levels necessary for these efficiencies are not applicable in a therapeutic setting. Our aim is to improve the in vivo repair efficiency with the aim that low ribozyme concentrations can achieve sufficient levels of mRNA repair.

Figure 3.1: Secondary structure of the trans-splicing ribozyme construct used in this study, and its change during trans-splicing. The ribozyme is shown in grey, with the duplexes labeled. Note that the P4-P6 domain is shown to the right of the catalytic core for clarity. The mRNA is shown in red, with the splice site marked by an arrowhead. The EGS is shown in green. The ribozyme 3'-exon is shown in blue, with the 3'-splice site marked with an arrowhead. The sequence of the EGS used in this figure is that of clone R3C7, which was a result of the in vivo selection. Secondary structures formed by the EGS are labeled as 5'-duplex, internal loop, P1ex (P1 extension), and P1 helix. The reaction from substrate (left) to product (right) is indicated by arrows.

Previous approaches to improve the trans-splicing efficiency were mostly based on designing and testing individual ribozyme constructs (9-11). In contrast, combinatorial approaches can test millions of ribozymes in parallel.
Combinatorial approaches were crucial for the development and optimization of many ribozymes that would not have been found by design (for review, see (12)). However, group I intron ribozymes present a challenge for combinatorial systems because the ribozymes separate themselves from their products, which prevents the traditional selection of the catalyst via the product. One elegant solution to this problem is to select these ribozymes based on their potential to catalyze the reverse reaction, thereby linking themselves with reaction products (13-15). However, the resulting ribozymes were inefficient in catalyzing the complete splicing process because they were selected to catalyze only one of the two catalytic steps (15).

To optimize trans-splicing ribozymes for the complete splicing reaction in a combinatorial approach we developed an in vivo selection method, which compartmentalizes the ribozyme with its substrate and product in cells. Combinatorial methods have also been used in vivo but did not employ the powerful combination with a ribozyme library of high complexity and multiple rounds of selection. Previous studies used fluorescence activated cell sorting (FACS) to select trans-splicing ribozymes that repair the mRNA of a fluorescent protein in mammalian cells (16), used histidine prototrophy to select trans-splicing ribozymes that activate a metabolic pathway in yeast cells (11), or used the resistance to the antibiotic kanamycin to detect excision of a cis-splicing ribozyme from the kanamycin nucleotidyl transferase mRNA in E. coli (17). In addition to the compartmentalization of ribozyme and substrate the in vivo format is advantageous because several ribozymes including the Tetrahymena group I
intron ribozyme show differences between in vitro and in vivo activity (18-20), and the ultimate goal for trans-splicing ribozymes is a high activity in vivo.

Here we report an in vivo selection method that selects the best trans-splicing ribozymes from millions of ribozyme variants. The method employs ribozyme libraries with high complexity, uses several rounds of enrichment, and is based on materials that exist in most molecular biology labs. A selection from millions of ribozyme variants can be completed by one person in less than three months. We validated this method by optimizing the ribozyme 5'-terminus for a given splice site in the mRNA of chloramphenicol acetyltransferase. The method yielded 5'-terminal sequences that mediated increased trans-splicing efficiency. The robust in vivo selection technique now makes it possible to optimize any sequence in trans-splicing ribozymes.

3.3 Results

Identification of efficient splice sites

Trans-splicing ribozymes can in principle target every uridine residue of a substrate RNA but the target sites need to be accessible. To determine accessible splice sites on our model mRNA substrate, the mRNA of chloramphenicol acetyltransferase (CAT), we first determined accessible splice sites by an established protocol, the trans-tagging assay (7). In this assay the mRNA is incubated with trans-splicing ribozymes that carry a randomized internal guide sequence (IGS). This enables the ribozyme population to splice on every accessible site on the mRNA. The sites at which trans-splicing occurred were
identified by RT-PCR, cloning and sequencing. Thirteen different splice sites were identified on cat mRNA (Tab. S1). We chose splice site 177 for all further experiments. The corresponding IGS for splice site 177 is 5’-GAAGGC-3’.

**Design of the ribozyme gene library**

In addition to targeting an accessible splice site, efficiently trans-splicing ribozymes require a 5'-terminal extension that forms specific secondary structures with the target mRNA (Fig. 3.1) (10). This 5'-terminal extension was termed the EGS (extended guide sequence). The secondary structures formed with the substrate mRNA are the P1 duplex, the P1 extension, an internal loop, a 5'-duplex, and a P10 duplex. However, several details on this structure were still unclear. First, it was unclear how long the single-stranded portions on each side of the internal loop should be. Second, the optimal sequence in the single-stranded portion of the internal loop was unknown. Third, it was unclear whether the first base pair of the P1 extension duplex should be a C:A base pair instead of a U:A base pair as suggested by one of our earlier selections (Fig. S3.1).

To address these questions we designed a library of ribozyme genes with six sub-pools (Fig. 3.2). Each of these sub-pools formed the 5'-duplex in a different register so that the selection could optimize the relative length of single-stranded nucleotides on each side of the internal loop. The 5'-duplex fixed the identity of eight nucleotides. The next nine nucleotides in the EGS were randomized. This allowed optimizing the length of the 5'-duplex because selected nucleotides could extend the 5'-duplex; it allowed optimizing the length of the P1
extension because selected nucleotides could extend the P1 extension duplex; and it allowed optimizing the sequence in the single-stranded portion of the EGS. The next two nucleotides were fixed because they were known to be involved in the minimum 3-base pair P1 extension duplex (10, 17). Finally, the last nucleotide of the EGS and the first nucleotide of the P1 extension was randomized to determine whether it should form a U:A or a C:A base pair with the mRNA.

Figure 3.2: Design of the pool of EGSs for the in vivo selection. Six sub-pools with six different registers (+6, +3, 0, -3, -6, -9) were generated to represent six different geometries of the internal loop with the mRNA (bottom). The positions of randomized nucleotides are denoted with N. Underlined characters show the position of duplexes between EGS and mRNA. All six sub-pools form identical P1 interactions at the IGS (GAAGGC), and at the P1 extension, which is constituted by the three base pairs adjacent to the mRNA splice site. The six sub-pools differ in the position of their eight base pair 5'-duplex on the mRNA, formed with nucleotides 3 to 10 of the EGS.

Generation of the plasmid library and the E. coli library

The library of ribozyme genes with partially randomized EGSs was generated by PCR and cloned into a library plasmid (Fig. S3.2). The ribozyme expression was driven by an attenuated version of the IPTG inducible trc promoter (21). The promoter for bacteriophage T7 RNA polymerase, which was
used in earlier studies (3), was not chosen because it inhibited cell growth. The down regulated trc promoter was chosen from three different variants of the trc promoter because it facilitated the best cell growth but expressed sufficient levels of ribozyme to show trans-splicing activity (data not shown).

The gene of the substrate mRNA (cat) was under the control of a constitutive promoter from its parent plasmid pLysS (Novagen). To allow the selection of mRNA repairing trans-splicing ribozymes we introduced a frame-shift deletion (DG322) into the cat mRNA. This mutation abolished the chloramphenicol resistance such that E. coli cells carrying a plasmid with this mutated gene did not form detectable colonies at a chloramphenicol concentration of 2 mg/mL on culture plates. Transcription termination sequences were inserted after the cat gene and after the ribozyme gene, to prevent read-through.

Each plasmid library was transformed into E. coli cells and plated on medium selecting for the presence of the library plasmid through ampicillin resistance. This resulted in at least 1.5 x 10^6 cells for each sub-pool. At this point the six sub-pools were mixed, resulting in the E. coli library that was used for the selection. To test the composition of this initial pool (round 0) we sequenced 50 of its clones. The sequences confirmed that each of the six sub-pools was represented equally, within statistical variation (Tab. S2).
Progress of the *in vivo* selection

To select those cells from the *E. coli* library that contained efficiently trans-splicing ribozymes we induced the expression of ribozymes by IPTG and plated the library on SOC agar plates that contained IPTG, and chloramphenicol at a concentration of 8 mg / mL. This chloramphenicol concentration allowed the growth of about 0.1% - 1% of the plated cells. Cells that formed colonies were assumed to contain ribozymes that efficiently repaired the inactivated *cat* mRNA to mediate resistance against chloramphenicol. The number of cells plated on the selection plates was 9 million, which contained 1.5 million cells from each sub-pool. Because this covered the sequence space \((4^{10})^{1.43}\)-fold we expected that 76% of all possible sequences were represented in this first round of the selection (see materials and methods for the calculation).

Five rounds of the selection were performed to eliminate false positives due to genomic mutations in *E. coli* or due to mutations in the library plasmid. After each selection step the library plasmids were extracted from the selected cells. The ribozyme genes were isolated by restriction digest, agarose gel purification, and PCR, and ligated back into fresh library plasmid (Fig. 3.3). In each selective step after the first round, at least 3-fold more cells encoding a ribozyme were plated than the maximum complexity of the pool at that stage. Therefore, at least 95% of the pool complexity was maintained in these later selection rounds.
Figure 3.3: Schematic for the work flow of the *in vivo* selection procedure. **SOC** plates refer to agar plates for bacterial growth, supplemented with the antibiotic ampicillin. Selection plates refer to agar plates for bacterial growth, supplemented with IPTG and the antibiotic chloramphenicol. More detailed descriptions are given in the text and in the materials and methods section.
Three rounds of selection were sufficient to select the best EGSs. This was concluded based on the sequences of 20 clones that were collected in each round. These sequences showed that the enrichment of sub-pools had saturated after 3 rounds of selection (Fig. 3.4) and that most clones were found as multiplicates after 3 rounds (Tab. 1). Therefore, we considered clones from round 3, 4, and 5 as the winners of the selection. Most of the winners contained EGSs from the sub-pools +6 (30/60) and +3 (22/60). In addition, a few isolates stemmed from sub-pools -3 (5/60) and -6 (3/60), while no isolates from sub-pools -9 and 0 were found. The clones were named by the selection round R in which they were first detected, and a successive clone number C, for example R3C7.

The early disappearance of sub-pools -9 and 0 may be due to the tendency of their respective EGSs to form self-structures in the absence of a substrate mRNA. Using mfold, we predicted the average energy for self-structures, of all 50 sequences obtained from the starting population (Table S2). Pools -9 and 0 showed the strongest self-structure energies (-3.5 ±0.6 kcal/mol and -5.1 ±0.8 kcal/mol, respectively). The other sub-pools had self-structure energies of -2.6 ±1.0 kcal/mol (+6), -3.1 ±0.8 kcal/mol (+3), -2.7 ±0.8 kcal/mol (-3), and -1.3 ±0.4 kcal/mol (-6). Such self-structure formation would prevent the IGS and the EGS from pairing with the substrate, thereby reducing the trans-splicing efficiency and causing the disappearance of sub-pools -9 and 0 from the population.
Figure 3.4: Enrichment of sub-pools in the ribozyme library, during 5 rounds of *in vivo* selection. 50 clones were sequenced from the initial pool (round 0) and 20 clones were sequenced from each selected pool (pools 1 - 5). The relative abundance of each sub-pool is shown as the size of the corresponding rectangle. Note that sub-pools -9 and 0 disappeared within two rounds of selection. The corresponding numerical values are given in Table S2.

**Trans-splicing efficiencies of selected EGSs *in vivo***

To test whether the selection succeeded in isolating ribozymes that mediate high *in vivo* activity we tested their *trans*-splicing efficiency *in vivo*. The eight most frequently selected sequences were chosen, as well as nine control sequences: four clones from the library before selection occurred, four clones that were designed based on published principles, and one clone without EGS (Tab. 1). Their *in vivo* *trans*-splicing efficiency was quantified by measuring the doubling times of *E. coli* cells expressing the chosen ribozymes and the inactivated *cat* gene in liquid medium containing chloramphenicol. All clones with
selected ribozymes showed doubling times below 100 minutes, whereas none of the clones with designed ribozymes showed doubling times shorter than 100 min (Tab. 1). Sequencing confirmed that these clones did not differ in their internal ribozyme sequence from the wild type sequence. These results showed that the selection was successful in isolating ribozymes with increased \textit{in vivo} activity.

The highest \textit{in vivo} activities stemmed from ribozymes of the sub-pools +3, -3, and -6. Interestingly, one pre-selected EGS (PS39) mediated good \textit{in vivo} activity. Most pre-selected and designed EGSs showed lower \textit{in vivo} activity than the ribozyme without EGS sequence (No EGS). This confirmed that non-optimized EGSs inhibited mRNA repair \textit{in vivo}, as it was found previously (10).
Table 3.1: Extended Guide Sequences (EGSs) that were used in this study. The first column shows the clone name. Names starting with an R label clones that were identified during the in vivo selection, which indicates the selection round R in which it appeared first, and its clone number C. Ribozymes from the R3C7 class are in bold, and ribozymes from the class R3C14 are in italics. The prefix "Des" indicates that the EGS was designed, based on published principles. The prefix "PS" indicates that the EGS was from the pre-selection pool (round 0). The second column shows the symmetry of the internal loop, i.e. the register between mRNA strand and EGS in the 5'-duplex relative to the register at the IGS. The third column shows the number of clones that were found during the in vivo selection, as the sum of rounds 3, 4, and 5. The fourth column shows the doubling time of cells expressing the ribozyme variant, in liquid medium, in the presence of chloramphenicol. N.d. indicates "not determined". Errors are standard errors of the means from at least three independent experiments. The fifth column shows the sequence of the EGSs. Bold characters show nucleotides that were predicted to participate in the 5'-duplex (left) or the P1 extension duplex (right).

<table>
<thead>
<tr>
<th>clone</th>
<th>loop sym.</th>
<th>clones found</th>
<th>doubling time / min</th>
<th>sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoEGS</td>
<td>-</td>
<td>-</td>
<td>115 ±27</td>
<td>-</td>
</tr>
<tr>
<td>R3C14</td>
<td>+6</td>
<td>19x</td>
<td>86 ±13</td>
<td>AAGCAAGAAGUGAAGUCGAUAUU</td>
</tr>
<tr>
<td>R3C7</td>
<td>+3</td>
<td>7x</td>
<td>63 ±3</td>
<td>AACGGCGAACUUACAAAUUU</td>
</tr>
<tr>
<td>R5C6</td>
<td>+3</td>
<td>7x</td>
<td>97 ±16</td>
<td>AACGGCAAGUAGUAAUUAU</td>
</tr>
<tr>
<td>R2C16</td>
<td>+3</td>
<td>4x</td>
<td>93 ±4</td>
<td>AACGGCAAGAAAACUAUU</td>
</tr>
<tr>
<td>R3C5</td>
<td>-3</td>
<td>4x</td>
<td>65 ±5</td>
<td>AAUACGGCGAGAAUACCAAAUU</td>
</tr>
<tr>
<td>R3C3</td>
<td>+6</td>
<td>3x</td>
<td>88 ±10</td>
<td>AAGCAAGAAGUAGUACCAAU</td>
</tr>
<tr>
<td>R3C4</td>
<td>+3</td>
<td>3x</td>
<td>65 ±5</td>
<td>AACGGGCAAGUAAGAAUAAU</td>
</tr>
<tr>
<td>R3C9</td>
<td>-6</td>
<td>3x</td>
<td>61 ±1</td>
<td>AAUUCAGAGGCAAGCAAGCAAGAU</td>
</tr>
<tr>
<td>R3C15</td>
<td>+6</td>
<td>2x</td>
<td>n.d.</td>
<td>AAGCAAGAAGUAGCAAGAU</td>
</tr>
<tr>
<td>R3C11</td>
<td>+6</td>
<td>2x</td>
<td>n.d.</td>
<td>AAGCAAGAAGUAGUAAUCAUA</td>
</tr>
<tr>
<td>R3C19</td>
<td>+6</td>
<td>1x</td>
<td>90 ±13</td>
<td>AAGCAAGAAGUACAAUUUU</td>
</tr>
<tr>
<td>R4C28</td>
<td>+6</td>
<td>1x</td>
<td>n.d.</td>
<td>AAGCAAGAAGUAGAACCAGU</td>
</tr>
<tr>
<td>R5C19</td>
<td>+6</td>
<td>1x</td>
<td>n.d.</td>
<td>AAGCAAGAAGUAAUAAUACCAU</td>
</tr>
<tr>
<td>R3C6</td>
<td>+6</td>
<td>1x</td>
<td>n.d.</td>
<td>AAGCAAGAAGUAGAUCGGAAAC</td>
</tr>
<tr>
<td>R3C12</td>
<td>+3</td>
<td>1x</td>
<td>n.d.</td>
<td>AAGCAGAGAAGUAGUAGAUA</td>
</tr>
<tr>
<td>R4C1</td>
<td>-6</td>
<td>1x</td>
<td>n.d.</td>
<td>AAUACGGAGAAAAAAEAU</td>
</tr>
<tr>
<td>Des1</td>
<td>+6</td>
<td>-</td>
<td>171 ±34</td>
<td>AAGCAAGAAGUAGAACCAGU</td>
</tr>
<tr>
<td>Des2</td>
<td>+6</td>
<td>-</td>
<td>113 ±24</td>
<td>AAGCAAGAAGUAGACUGACAGA</td>
</tr>
<tr>
<td>Des3</td>
<td>-1</td>
<td>-</td>
<td>123 ±26</td>
<td>AACAGGCGGGAGAUAUAAAGA</td>
</tr>
<tr>
<td>Des4</td>
<td>-1</td>
<td>-</td>
<td>131 ±8</td>
<td>AACAGGCGGGAGACGAUCGA</td>
</tr>
<tr>
<td>PS72</td>
<td>+6</td>
<td>-</td>
<td>167 ±12</td>
<td>AAGCAAGAAGUUCGUGUACAGA</td>
</tr>
<tr>
<td>PS45</td>
<td>+3</td>
<td>-</td>
<td>119 ±9</td>
<td>AAGCCGCCAGAAGUACGGA</td>
</tr>
<tr>
<td>PS39</td>
<td>0</td>
<td>-</td>
<td>68 ±4</td>
<td>AAGCGGCAGCGCGGUAAU</td>
</tr>
<tr>
<td>PS66</td>
<td>-9</td>
<td>-</td>
<td>128 ±6</td>
<td>AAGCAUUCAGUAGAUCGAAGA</td>
</tr>
</tbody>
</table>
We were surprised to find that the most frequently selected ribozyme (R3C14) did not mediate the fastest growth in liquid medium (Tab. 1). We hypothesized that this may be because the in vivo selection used growth on plates whereas the doubling times were determined in liquid culture, which differ in their requirements on the cells. The difference between growth on plates and in liquid medium is caused by the diffusion of nutrients and oxygen into the bacterial colony (22, 23), which limits the growth on plates but not in liquid culture. To test this hypothesis we measured the growth of all clones for growth on plates containing chloramphenicol that were selected at least three times in the selection. The growth was judged by the average diameter of their colonies (Fig. 3.5). The results showed that clone R3C14 grew better than any other clone on plates, which paralleled its frequency among the selected clones. This confirmed our hypothesis that the reason for the high enrichment of clone R3C14 during the selection was its superior growth on plates. While these clones are not truly artifacts (they do facilitate faster growth than any of the designed clones) one could define the winners of the selection as the four sequences facilitating the fastest growth in liquid medium, representing 17 out of 60 clones.
Figure 3.5: Correlation between the measurements of bacterial growth with two methods. The doubling time in liquid medium containing chloramphenicol and IPTG is shown in the y-axis, the diameter of bacterial colonies on LB agar plates containing IPTG and chloramphenicol is shown on the x-axis. The clone names are indicated. The two clusters are referred to in the text as the class with clone R3C14 (upper) and the class with clone R3C7 (lower).

The growth behaviors on plate and in liquid culture grouped the eight selected ribozymes into two classes (Fig. 3.5). One class contained clone R3C14, which grew, on average, better on plate but not as well in liquid culture. The other class contained R3C7, which grew better in liquid culture but not as well on plates. Because the only difference between these ribozymes lies in their EGSs we hypothesized that the secondary structures formed between the EGS and the substrate (Fig. 3.6A) or the EGS and the product (Fig. 3.6B) caused the different growth behavior of the two classes. In support of this hypothesis the
predicted secondary structures showed a systematic difference: The class that contained clone R3C7 formed an internal loop between substrate and EGS with 5 or 6 single-stranded nucleotides on the mRNA side but it did not form a conserved structure between product and EGS. In contrast, the class that contained clone R3C14 formed an internal loop between trans-splicing product and EGS with 5 or 6 single-stranded nucleotides on the mRNA side but not between substrate and EGS. If the length of 5-6 single-stranded nucleotides on the mRNA side of the internal loop represents an optimal geometry then the two classes of ribozymes would use two different strategies to increase trans-splicing efficiency: The class containing R3C7 may optimize the interactions for the first step of splicing, whereas the class containing R3C14 may optimize the interactions for the second step of splicing. This hypothesis will be tested in future studies, with additional selections using different splice sites on the cat mRNA.

Two additional observations were made for the predicted secondary structures between EGS and substrate or product. First, the length of the P1 extension was 3 base pairs for all selected clones. The four clones that had less than 3 base pairs (designed clones Des2 and Des4, and two clones from the pre-selection library, PS45 and PS66) showed poor in vivo repair efficiencies (Tab. 1). None of the selected clones formed a P1 extension longer than 3 base pairs although the randomized sequence would have allowed the formation of longer P1 extensions. This suggests that the optimal length of the P1 extension is exactly 3 base pairs, consistent with previously reported results (10). Second,
none of the selected EGSs formed a P10 helix, i.e. a duplex between the singlestranded region of the internal loop of the EGS with the 3'-exon. The two
designed clones that formed a P10 helix (Des2 and Des4) showed low transsplicing efficiency. Because the formation of a P10 helix increased the transsplicing efficiency in a different sequence context (9, 10) we assume that the beneficial effect of a P10 helix is dependent on the sequence context.

To test whether the growth rate of cells was a good measure for the transsplicing efficiency in vivo we measured the fraction of repaired mRNA in cells, by quantitative RT-PCR (Fig. 3.7). Four ribozyme variants were chosen to cover the range of doubling times mediated by the ribozyme variants. This included the two EGSs that promoted the fastest doubling time in liquid medium (R3C7, register
+3, and R3C9, register -6), one that facilitated medium growth (R3C19, register +6), and one pre-selected EGS that facilitated only slow growth (PS72, register +6). Growth in liquid culture was used for this correlation because the RNA samples for quantitative RT-PCR were prepared by growth in liquid medium. The quantitation of cat mRNA and trans-splicing products reported that between 0.4% and 3.8% of the mRNA was repaired by the trans-splicing ribozymes. The fraction of repaired mRNA and the doubling time of the cells were correlated well (r = 0.95). This suggests that the cell doubling time is a good measure for in vivo trans-splicing efficiency.

Figure 3.7: Correlation between two methods measuring in vivo trans-splicing efficiency. The percentage of repaired cat mRNA was determined by RT-PCR and is shown on the x-axis. The doubling time of E. coli cells expressing the respective ribozyme, in medium containing chloramphenicol, is shown on the y-axis. Error bars the standard errors of the mean from three quantitations of the doubling time and from six qRT-PCR experiments that determined the fraction of repaired mRNA.
The selection system can now be used to improve the efficiency of any \textit{trans}-splicing ribozyme. For example, \textit{trans}-excision ribozymes (25) or \textit{trans}-insertion ribozymes (26) could be optimized with this system. Additionally, in combination with mutagenic PCR, the \textit{in vivo} selection system can be converted to an \textit{in vivo} evolution system: After each round of selection new mutations can be introduced by mutagenic PCR (27) such that over multiple rounds, the population of evolving ribozymes successively accumulates beneficial mutations. This procedure would allow optimizing any sequence on the ribozyme, and the successive exploration of sequence space would facilitate the screening of far more functional sequences than can be tested in a selection experiment. We expect that such evolutions will facilitate even larger increases in \textit{trans}-splicing efficiency, and ultimately allow the application of \textit{trans}-splicing ribozymes for therapeutic applications.

### 3.4 Discussion

We developed an \textit{in vivo} selection procedure to optimize the efficiency of \textit{trans}-splicing ribozymes \textit{in vivo}. The method was validated by selecting ribozymes with improved EGSs from a library with 9 million ribozyme variants. The \textit{in vivo} selection method can now be used to optimize any sequence on \textit{trans}-splicing ribozymes.

The model organism \textit{E. coli} was chosen for the \textit{in vivo} selection for several reasons. First, \textit{E. coli} is used in most life-science laboratories and is easy to manipulate. Second, the transformation efficiency of \textit{E. coli} with commercially
available cells allows the easy generation of libraries with high complexity. While this study demonstrated the use of an *E. coli* library with a complexity of 9 million it is feasible for most laboratories to increase this complexity to about 100 million. Third, while the use of a eukaryotic system may seem an advantage because the eventual application of these ribozymes would be in eukaryotic cells, the limiting factor for trans-splicing in both systems appears to be the same, and the same principles for the design of these ribozymes apply to both systems (compare (9), (10), and this study). This suggests that the initial optimization of trans-splicing ribozymes in bacterial cells and later testing of fewer ribozyme constructs in mammalian cells is a promising route to develop efficient trans-splicing ribozymes for therapeutic purposes.

The therapeutic application of trans-splicing ribozymes requires them to have a high efficiency on disease-related mRNAs. However, in vivo selection procedures cannot directly optimize the ribozymes on disease-relevant mRNAs because the selection procedure requires selectable markers such as antibiotic resistance genes ((17) and this study), auxotrophy (11), or fluorescent proteins (16). A solution to this problem is that regardless of the broader sequence context, specific secondary structures between the trans-splicing ribozyme and the substrate mRNA are necessary for high trans-splicing efficiency (9, 10). Therefore, structural elements that increase trans-splicing efficiency can be identified on unrelated mRNAs and later tested on disease-related mRNAs. The *in vivo* selection procedure can be employed in the first step of this process,
screening a much larger number of ribozyme variants than could be tested individually.

The most efficient ribozyme repaired $3.8 \pm 0.5\%$ of the substrate mRNA in bacterial cells. This is 3.8-fold higher than the highest observed $trans$-splicing efficiency in bacterial cells ($1\%$ (3)). In contrast, significantly higher repair efficiencies can be achieved in mammalian cells (8, 9, 28), for two reasons. First, the mRNA life-time is much shorter in bacterial cells than in eukaryotic cells (29, 30). This gives the ribozyme less time to fold, anneal to the substrate mRNA, and catalyze $trans$-splicing, and thereby reduces the total amount of $trans$-spliced substrate. Second, most studies in mammalian systems used stronger promoters to express the ribozymes. The resulting high ribozyme concentrations are known to increase the fraction of repaired mRNA (8). Another factor that differs between an eventual application of ribozymes in therapy and our selection system is that the ribozyme and the substrate were encoded on the same plasmid molecule. This may result in partial co-localization of ribozyme and substrate, which would favor the ribozyme / substrate interactions. We hope that further studies of the selected ribozymes, and further selections will successively develop group I intron ribozymes that repair a high fraction of mRNAs at low ribozyme concentrations.
3.5 Materials and Methods

Selection of splice sites

The accessible splice sites on the chloramphenicol acetyltransferase gene were selected as described previously (7). The trans-tagging ribozyme with randomized IGS contained the sequence of the Tetrahymena group I intron ribozyme (Genbank accession number X54512) and a 77 nucleotide long 3'-exon from the alpha-mannosidase sequence from COS-7 cells (31), using PCR primers 1 and 2 (Tab. S3). The template for T7 RNA polymerase transcription of the substrate was PCR amplified from the plasmid pLysS (Novagen) in two PCRs, using primers 3 and 4 in the first PCR, then primers 5 and 6 in the second PCR. Ribozymes were transcribed from the PCR products by T7 RNA polymerase, then purified by denaturing 7M urea 5% PAGE.

The reactions for the splice site selection contained 100 nM substrate, 10 nM ribozyme, 0.2 mM GTP, 50 mM MOPS/KOH pH 7.0, 135 mM KCl, and 2 mM spermidine, and were incubated for 1 hour at 37°C. Magnesium concentrations of 5 mM, 2 mM and 1 mM gave consistent results so that the splice sites selected under all conditions were analyzed together. Reaction products were reverse transcribed with AMV reverse transcriptase (NEB) and primer 7 (Tab. S3). After RNA hydrolysis (10 min at 90°C in 200 mM NaOH), reverse transcription products were PCR amplified with PCR primers 8 and 9. PCR products were cloned into the BamHI and EcoRI sites of pUC19, and sequenced to determine the splice sites.
**Generation of the plasmid library**

The library plasmid is based on the plasmid pUC19 with the following modification: The BamHI and SacI sites were separated by digesting pUC19 with BamHI and HindIII then ligating with the 5'-phosphorylated and annealed DNA primers 10 and 11 (Tab. S3), creating plasmid pUC19b. The *cat* gene including its constitutive promoter and a hairpin terminator at its 3'-end was PCR amplified from the plasmid pLysS and ligated into pUC19b between the SacI and HindIII restriction sites. The resulting sequence of the substrate in the library plasmid is sequence 12. To create the frame shift that inactivates the *cat* gene, site directed mutagenesis with primers 13 and 14 was performed using the Stratagene quick-change protocol.

The ribozyme cassette of the library plasmid was obtained through two PCRs from a plasmid containing the sequence of the *Tetrahymena* group I intron. The first PCR introduced the promoter and silent mutations with primers 15 and 16. The second PCR used the primers 17 and 18. The resulting sequence encodes a ribozyme without an EGS. The promoter sequence in primers 15 and 17 was derived from plasmid pDSW204 (21). Its -30 box is mutated from TTGACA to TTTACA to reduce ribozyme expression. The 3'-exon of the ribozyme was generated through PCR using the *chloramphenicol acetyltransferase* gene in the plasmid pLysS (Novagen) as template and primers 19 and 20. After EcoRI digest it was joined by ligation with the EcoRI digested ribozyme cassette. This template for the ribozyme without EGS was used to generate the six sub-pools with partially randomized EGSs, in two PCRs. The
first PCR included the primer 21 and 22. In a second PCR the promoters and restriction sites were attached with primers 23 and 24. The PCR products were ligated in the BamHI and SacI sites of the pUC19b plasmid containing the substrate with inactivating frame shift to generate the library plasmid. The second series of PCRs was repeated for each of the six sub-pools.

**Generation of the *E. coli* library**

Electrocompetent *E. coli* cells Stbl4 (Invitrogen) were transformed with library plasmid (1 mm slit width, 1.8 kV, 25 mF, 200 Ω), plated on SOC plates containing 100 mg/mL ampicillin medium, and incubated at 30°C. We usually plated ~ 300,000 viable cells per 9 cm plate. After colonies were visible (16-20 hours) the plates were washed and cells were frozen with 50% glycerol (v/v) at -80°C. Replating efficiencies were determined by thawing the cell stocks and plating on SOCAMP plates. In our experience it was sufficient to calculate the replating efficiency from the OD$_{600}$ of the thawed cell suspension, with 1 OD$_{600}$ corresponding to 2 x 10$^8$ cells/mL. The number of cells growing on the SOCAMP plates was in some cases up to 2-fold lower than the value expected from the OD$_{600}$, presumably due to partial cell death during freeze/thawing.

To estimate the complexity of the *E. coli* library we considered two factors. First, 10 nucleotides were randomized (Fig. 3.2), therefore the maximal complexity was 4$^{10}$, about one million. Second, the number of colonies in the *E. coli* pool with a ribozyme gene was 1.5 x 10$^6$ cells for each sub-pool. This number was obtained by multiplying the number of viable cells with the fraction of
these cells that contained a ribozyme gene in the plasmid. The latter was found by colony PCR on at least 30 colonies. The fraction $f$ of the initial complexity that was represented in the *E. coli* library, was calculated by $f = 1-(1-(1/n))^m$, where $n$ is the initial complexity and $m$ is the number of colonies with a ribozyme gene. The resulting $f$ indicates that 76% of all possible pool sequences were represented in the initial pool.

We used the *E. coli* cell line Stbl4 (Invitrogen) in this study due to its reduced recombination frequency. However, we also performed a different selection with *E. coli* DH5a cells. The frequency of false positive colonies growing on chloramphenicol containing plates was about 1/2,000 for *E. coli* DH5a and about 1/500 for Stbl4, therefore we recommend the use of *E. coli* DH5a cells for the *in vivo* selection system.

**In vivo selection**

After thawing, the *E. coli* sub-pools were mixed to give equal viable cell numbers for each sub-pool, shaken one hour with 1mM IPTG at 30°C, then a total of 9 million clones were plated on 30 SOC plates with 8 µg/mL chloramphenicol and 1mM IPTG. After incubating at 30°C for about 20 hours colonies of different sizes were visible. The largest colonies (about 0.8 mm in diameter) corresponded to false positives as judged by their ability to grow equally well on plates with different chloramphenicol concentrations. The smaller colonies (between ~0.1 mm to ~0.5 mm) differed in size and number between different chloramphenicol concentrations. The cells were washed off the plates
and their plasmids isolated by standard alkaline lysis and column-based purification. The ribozyme genes were purified by digesting the plasmids with SacI / BamHI, performing electrophoresis on 1% agarose gels, excising the band with the correct size, and extracting from the gel. PCR amplification with primers 25 and 26 (Tab. S3) and digestion with SacI and BamHI allowed cloning into fresh library plasmid. The high fidelity DNA polymerase Phusion (NEB) was used for this PCR. Five rounds of selection were performed.

We estimated that each round of the selection enriched the functional ribozymes by a factor of 100 - 1,000 because the initial complexity was $9 \times 10^6$ and after 3 rounds of selection we found duplicate sequences, among 20 sequenced clones. At 5 rounds of selection few new sequences were found among 20 sequenced clones, and one sequence was found 19 times among the 60 clones of round 3, 4, and 5. The enrichment factor may differ for other splice sites on cat mRNA, other antibiotic resistance genes, other cis- and trans-splicing ribozymes, and different antibiotic concentrations on the selection plates but it illustrates the power of the in vivo selection technique.

**Determination of cell doubling times in liquid medium**

The doubling times of *E. coli* DH5a cells that express ribozymes were determined as follows. Five mL fresh overnight cultures (LB medium with 100 µg/mL ampicillin) were induced with 1mM IPTG and shaken at 37°C for 1 hour. The cultures were diluted to an OD$_{600}$ of 0.05 +/- 0.01 in LB medium with 2 mg/mL chloramphenicol and 1 mM IPTG. Cultures were grown until the OD$_{600}$
exceeded 1.0 or until the growth time reached 6 h. The OD$_{600}$ was measured after 60 minutes and every 30 minutes after. Growth rates were determined by nonlinear least squares fitting of exponential functions to the OD$_{600}$ as a function of growth time. Errors were calculated from at least 3 independent growth curves.

**Measurement of bacterial colony sizes**

Bacterial colonies of *E. coli* DH5a were grown on LB agar plates containing 8 µg/mL chloramphenicol and 1 mM IPTG by incubating for 20 hours at 37°C. Five representative colonies and a ruler were photographed under a dissection scope equipped with a digital camera. The images of the colonies were then superimposed on the ruler to measure the colony diameters. To control for clonal variation three cell stocks of *E. coli* DH5a clones with the same plasmid were measured. The average of 5 colonies gave the value for each clone, and the average from three clones gave the reported values with corresponding standard deviations.

**Structure prediction**

Secondary structures were predicted on a web-based server using the mFold algorithm (24). We used extended guide sequences, mRNA sequences, and 3′-exon sequences as shown in Fig. 3.1B. The parameters were 37°C and 1 M NaCl.
Quantitative RT-PCR

Total RNA was isolated from logarithmically growing *E. coli* cells after induction with 1 mM IPTG, using the RNeasy kit (Qiagen). The Qiagen RNeasy kit was used with RNAprotect Bacteria Reagent and on column DNase digestion. Three separate clones were used for each ribozyme. Two hundred ng of total RNA were reverse-transcribed using AMV-RT (NEB) with primer 33 for substrate, ribozyme, and product samples. The transcriptions were incubated for one hour at 42°C with AMV reverse transcriptase (NEB). One tenth of the reverse transcription reaction was used as template for quantitative PCR with the primers 35 and 36 for the substrate, and primers 37 and 38 for the product. The PCR was performed using the AB qPCR master mix on a Fast 7500 RT-PCR machine (Applied Biosystems). The method started with a 10 minute incubation at 95°C. Each PCR cycle included incubations of 30 seconds at 95°C, 30 seconds at 57°C, and 30 seconds at 72°C. The threshold cycle numbers from three 8-fold serially diluted samples confirmed the linearity of the assay. No significant amount of cross-amplification was detected. Data were analyzed by least squares fitting to an exponential function.

3.6 Acknowledgements

Joe Pogliano, Simpson Joseph, Akif Tezcan, Lorraine Pillus, and Dario Meluzzi are thanked for helpful discussions. Hector Viadiu and Thomas Hermann are thanked for critical reading of the manuscript. This work was supported by the National Institutes of Health [T32DK007233 to K.E.O; Hemoglobin and Blood
Protein Chemistry training grant to E. Komives]; and by the National Science Foundation [0743985 to U.F.M.].

Chapter 3, in full, is a reprint of the material as it appears in RNA, Olson, K.E. and U.F. Müller. RNA, 2012. 18(3): p. 581-9. The dissertation author was the first author on this paper.
3.7 Supporting Information

Table S3.1: Splice sites identified using the trans-tagging assay.

<table>
<thead>
<tr>
<th>splice site</th>
<th># sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3</td>
<td>2</td>
</tr>
<tr>
<td>U33</td>
<td>3</td>
</tr>
<tr>
<td>C43</td>
<td>1</td>
</tr>
<tr>
<td>A50</td>
<td>1</td>
</tr>
<tr>
<td>U54</td>
<td>1</td>
</tr>
<tr>
<td>U84</td>
<td>1</td>
</tr>
<tr>
<td>U87</td>
<td>1</td>
</tr>
<tr>
<td>U97</td>
<td>5</td>
</tr>
<tr>
<td>U99</td>
<td>1</td>
</tr>
<tr>
<td>U110</td>
<td>1</td>
</tr>
<tr>
<td>U177</td>
<td>1</td>
</tr>
<tr>
<td>U179</td>
<td>1</td>
</tr>
<tr>
<td>U448</td>
<td>1</td>
</tr>
</tbody>
</table>

Table S3.2: Number of clones identified from each sub-pool per round.

<table>
<thead>
<tr>
<th>sub-pool</th>
<th>-9</th>
<th>-6</th>
<th>-3</th>
<th>0</th>
<th>+3</th>
<th>+6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>14</td>
<td>7</td>
<td>10</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
Table S3.3: PCR primer sequences.
Primer 1: 5'-
GCGTAATACGACTCACTATAGNNNNNAAAAAGTTATCAGGCGATGCACC
Primer 2: 5'-GAAGCCTCTGAGCCGCAG
Primer 3: 5'-CAGGAGCTAAGGAAGCTAAATG
Primer 4: 5'-CGCCCCCCTGCCACTCATC
Primer 5: 5'-
GCGTAATACGACTCACTATAGCTAGGAGCTAAGGAAGCTAAATG
Primer 6: 5'-CGCCCCCCTGCCACTCATC
Primer 7: 5'-GAAGCCTCTGAGCCGCAG
Primer 8: 5'-CGAATTCCAGGAGCTAAGGAAGCTAAATG
Primer 9: 5'-GCATGCCACGAAGGTTCTCAACT...ATGAGTGCGAGGGCGGGCGCTA
GCATACCCCCTTTGGCTCTAATACGGGTCTTGAGGGGTTTTTGAAGCTT
Primer 10: 5'-GATCTCTTAGTGCAGCTGCGATGCTCGACTGACTGA
Primer 11: 5'-AGCTTCAGTAGCTAGCTGCTAGCTGCAGGCTAGCTCTAGAG
Sequence 12: 5'-
GCGGATCTGTGGTACAGCCTGCAGGAGATAGCAGTGAGAAG
GCAAAAGTTATCAGGCATGC
Primer 13: 5'-GCATGCCACGCATGGTGCAATGAAAG
Primer 14: 5'-ACGGCCAGTGAATTCGAGCTC
Primer 15: 5'-CGGGATCCTGTTTACAATTAATCATCCGGCTCGTATAATGTGTTGAGAAG
GCAAAAGTTATCAGGCATGC
Primer 16: 5'-
ACGGGATCTCCGGGTGTCGTTCAATAGCGTGCTAAGGATAGTGACTGAGTA
CTCCAAAAACTAACT (underlined positions denote silent mutations made in the
ribozyme 3'-exon)
Primer 17: 5'-GCATGCCACGCATGGTGCAATGAAAG
Primer 18: 5'-ACGGCCAGTGAATTCGAGCTC
Primer 19: 5'-CGGGATCCTGTTTACAATTAATCATCCGGCTCGTATAATGTGTTGAGAAG
GCAAAAGTTATCAGGCATGC
Primer 20: 5'-ACGGCCAGTGAATTCGAGCTC
Primer 21: 5'-
CGGCTCGTATAATGTGTTGAGAAXXXXXXXNNNNNNNNNNGGAAGGCAAAA
GTTATCAGG (the Xs represent the antisense nucleotides setting the bulge
register as denoted in Fig. 3 and the Ns represent the randomized nucleotides)
Primer 22: 5'-ACGGGATCTGAGCTC
Primer 23: 5'-
GCGGATCTCTTTACAATTAATCATCCGGCTCGTATAATGTGTTG
Primer 24: 5'-ACGGGATCTGAGCTC
Primer 25: 5'-
GCGGATCTCTTTACAATTAATCATCCGGCTCGTATAATGTGTTG
Primer 26: 5'-ACGGGATCTGAGCTC
Sequence 27: 5'-
GAGGATCCTAATACGACTCACTATTTANOGAAGGCAAAAAGTTATCAGG where
Nx represents the EGS region for the ribozyme being created
Primer 28: 5'-ACGGGATCTGAGCTC
**Table S3.3:** PCR primer sequences, continued.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>5'-GAGGATCCTAATACGACTCAC</td>
</tr>
<tr>
<td>30</td>
<td>5'-GGTATTCCTCACTCCAGAGCGATG</td>
</tr>
<tr>
<td>31</td>
<td>5'-GCATAATACGACTCCTATAGTCCTATAACCAGACGGTTC</td>
</tr>
<tr>
<td>32</td>
<td>5'-GGATGAGCATTCATCGGATGAGTGATAGTG</td>
</tr>
<tr>
<td>33</td>
<td>5'-AATTTAATACGACTCCTATAGCTAGCGGATGAGTGATAG</td>
</tr>
<tr>
<td>34</td>
<td>5'-GGTATTCCTCACTCCAGAGCGATG</td>
</tr>
<tr>
<td>35</td>
<td>5'-GCATAATACGACTCCTATAGTCCTATAACCAGACCGTTC</td>
</tr>
<tr>
<td>36</td>
<td>5'-GCCAGTCCACGCGGTAGGAGG</td>
</tr>
<tr>
<td>37</td>
<td>5'-GCCAGTCCACGCGGTAGGAGG</td>
</tr>
<tr>
<td>38</td>
<td>5'-GGTATTCCTCACTCCAGAGCGATG</td>
</tr>
<tr>
<td>39</td>
<td>5'-CACCCTCTTCATTGC</td>
</tr>
<tr>
<td>40</td>
<td>5'-GTATTCACGCGGCTGCTG</td>
</tr>
<tr>
<td>41</td>
<td>5'-GAGCTAGCGGATAGGATGTAG</td>
</tr>
<tr>
<td>42</td>
<td>5'-TTAAGCGTCAATAGATATGG</td>
</tr>
<tr>
<td>43</td>
<td>5'-GTACCTATAACCAGACCGTTC</td>
</tr>
<tr>
<td>44</td>
<td>5'-ATCAGGGCGGCAAGATGTG</td>
</tr>
<tr>
<td>45</td>
<td>5'-GTACCTATAACCAGACCGTTC</td>
</tr>
<tr>
<td>46</td>
<td>5'-TTAAGCGTCAATAGATATGG</td>
</tr>
<tr>
<td>47</td>
<td>5'-CTCCTACGGGAGGCCAGCAG</td>
</tr>
<tr>
<td>48</td>
<td>5'-GTATTCACGCGGCTGCTG</td>
</tr>
</tbody>
</table>
Figure S3.1: Predicted secondary structures between EGSs identified in an earlier selection, and the mRNA target site. The EGSs were selected from a single pool that contained 20 randomized nucleotides as the EGS, after two 5'-terminal As. Note that three of the four most efficient isolates (clone 141, clone 50, clone 24) show a C:A pair as the first pair of the P1 extension. The EGS of clone 77 was predicted to fold into two distinct secondary structures.
**Figure S3.2:** Library plasmid for the *in vivo* selection of ribozymes that can repair the mRNA of the *chloramphenicol acetyltransferase* (*cat*) gene. The plasmid contains the *cat* gene (red) under the control of the *cat* promoter (black; *cat*), inactivated with a frame shift (*X*), and terminated by a hairpin terminator (Hp; white). The ribozyme gene (white) is under control of a modified *trc* promoter (black; *trc1*), and is linked to a 5'-terminal, randomized EGS (green). The 3'-exon (blue) of the ribozyme is also terminated by a hairpin terminator (Hp; white). Restriction sites for cloning the *cat* cassette and the ribozyme cassette are indicated.

### 3.8 References


Chapter 4

Evolution of group I intron internal sequences for high trans-splicing efficiency in vivo

4.1 Introduction

The *Tetrahymena thermophila* group I intron is a catalytic RNA (ribozyme) that can catalyze its own removal from a pre-rRNA transcript without requiring the assistance of a protein (1). This *cis*-splicing reaction occurs through two consecutive transesterification reactions that ultimately ligate the exons and free the intron (2). In the first step of this *cis*-reaction, the intron folds to position the 5’ intron-exon junction in the catalytic core of the ribozyme. A transesterification reaction, initiated by an exogenous guanosine residue, cleaves the bond at the intron-exon junction and ligates the exogenous guanosine to the 5’-most nucleotide of the intron. Then, a conformational rearrangement takes place that aligns the 5’-exon with the 3’ intron-exon junction. The second transesterification, initiated by the 3’-hydroxyl of the 5’-exon, cleaves the 3’ intron-exon junction and ligates the two exons, releasing the intron (Fig. 4.1).
Figure 4.1: Steps of group I intron cis-splicing. An exogenous guanosine initiates the first transesterification, cleaving at the 5’ intron-exon junction. A conformational change then positions the 3’-exon in line with the 5’-exon. The 3’ hydroxyl of the 5’-exon initiates the second transesterification which cleaves the 3’ intron-exon junction and ligates the exons, releasing the intron.

Group I intron splicing naturally occurs via a cis-reaction, yet changes can be made to the sequence which allow the ribozyme to splice in trans (for review, see (3)). If the substrate in the trans-splicing reaction is an mRNA with a mutation 3’ of the splice site, the 3’-exon of the ribozyme, which can be altered as needed, will be transferred to replace and repair the mRNA sequence (Fig. 4.2). In trans-splicing the internal guide sequence (IGS) of the ribozyme binds to the substrate, thereby defining the splice site. Adding an antisense extended guide sequence
(EGS) 5’ of the IGS will increase both the specificity and efficiency of the trans-
reaction, while reducing splicing on non-target cellular RNAs (4). Although trans-
splicing efficiency for mRNA repair has been reported between 10-50% in vivo,
this efficiency has only been obtained under artificially high ribozyme
concentrations in the cell (5-7). The current trans-splicing ribozyme mediated
mRNA repair efficiency is not sufficient for therapeutic applications without these
increased ribozyme levels.

Figure 4.2: Conversion of the cis-splicing group I intron into a trans-splicing
intron. Panel A shows the cis-reaction with exons in red and the intron sequence
in black. Panel B shows the trans reaction with the substrate binding in trans in
red, the intron in black and the 3’-exon, which can be modified for repair, in blue.
Improving *trans*-splicing efficiency is essential to create a viable mRNA gene therapy and has been the aim of both rational design and selection (6, 8-11). Previous approaches to improve the ribozyme for *trans*-splicing efficiency have focused on optimizing the base pairing of the EGS and substrate or 3’-exon. Group I introns evolved for efficient *cis*-splicing in a specific sequence context. Converting these catalytic introns into *trans*-splicing ribozymes changes the demands on the ribozyme for efficient mRNA repair. Instead of being covalently attached to the 5’-exon, the intron must now bind to the specific splice site in *trans*. Locating this substrate inside a cell and pairing to the correct splice site may be more difficult than simply folding a covalently attached 5’-exon splice site into the catalytic core. Additionally, the structure of the ribozyme near the 5’-splice site has been optimized through evolution for the *cis*-reaction. However, the optimal *trans*-splicing structure is not known, may not be similar to the best *cis*-splicing structure, and could be different for each reaction depending on the specific target site sequence of the substrate. For efficient *trans*-splicing, ribozymes must also be stable before they bind to their substrates. If the ribozyme’s 3’-exon is cleaved before the substrate is found, repair cannot occur, a problem the *cis*-splicing ribozyme does not have. Lastly, the internal ribozyme sequence has not been evolved for a reaction in *trans*. It is likely that variants of the *Tetrahymena* intron sequence are better suited to improve the stability of the ribozyme’s 3’-exon and aid the finding and binding of the substrate mRNA.

Here we report an *in vivo* evolution of the internal sequence of the *Tetrahymena* group I ribozyme to improve *trans*-splicing. In this evolution, a
ribozyme with previously demonstrated efficient repair of a mutated chloramphenicol acetyltransferase (cat) mRNA in *E. coli* was the starting sequence (10). This sequence was evolved with the use of mutagenic PCR and recombination over 21 rounds of evolution with increasing selection pressure. Four mutations in the internal sequence of the ribozyme were identified that greatly increased cell survival in chloramphenicol medium. These may function by recruiting a cellular protein for improved activity.

**4.2 Results**

To improve group I introns for *trans* rather than *cis*-splicing with the eventual goal of *in vivo* mRNA repair, we evolved the ribozyme’s internal ribozyme sequence. We accomplished this through a previously developed selection procedure (10) introducing mutagenesis and recombination during the rounds of evolution. As a starting sequence we used one of the most active sequences from a previous study, termed R3C7 (10).

In this evolution, that took place over 21 rounds, a library plasmid was generated which included a chloramphenicol acetyltransferase (CAT) gene inactivated by a frame-shift mutation under constitutive expression. After mutagenesis by mutagenic PCR (mPCR) or recombination of the ribozyme sequence, the ribozyme was ligated into this library plasmid. The ligation utilized restriction sites that would properly orient the ribozyme into the plasmid and connect the intron 3’-exon junction (Fig. 4.3). After ligation, plasmids were transformed into high efficiency DH5α cells (NEB) and were plated on LB_Amp
plates. This plating step allowed for the assessment of pool sizes and ribozyme insert ratios. The colonies were then washed from each plate and pooled. An OD$_{600}$ measurement was obtained so the washed pool could be appropriately diluted to give approximately 300,000 colonies per plate. The pool clones were then plated on enough chloramphenicol selection plates to cover the pool size at least three fold, giving at least 95% coverage of all sequences in the pool. Plates were washed and plasmids were isolated for use as the template for the next round of evolution. Ribozyme sequences were PCR amplified and re-cloned in each round of evolution to both increase the mutations and to remove the possibility that the substrate cat gene lost its frame-shift mutation.

**Figure 4.3:** Plasmid map of the substrate and ribozyme used for each round of selection. The substrate with frame-shift mutation (red) is transcribed in one direction with a constitutive promoter and the ribozyme (green and white) with 3’-repair exon (blue) is transcribed in the opposite direction with a trc1 promoter. Both sequences have hairpin terminators to prevent read-through. The sequence between BamHI and BssHII was amplified and re-cloned in each round of the evolution.
In the first round of evolution, a high rate of mPCR was used (30 cycles following the protocol described in (12)) to generate a great diversity in the original pool. However, when additional mutagenesis was employed at the same rate, an unacceptably small number of clones survived on the selection plates. Therefore we did not further mutate the ribozyme sequence in round 2. Because the frequency of mutagenesis was found to be critical for establishing a sustainable line of evolution, the next rounds of evolution, rounds 3-8, were continued with low and medium rates of mutation (10 or 20 cycles of mPCR respectively) and produced stable lines of evolution.

The position and frequency of mutations was monitored with 10 clones sequenced per round of evolution. The increase in the frequency of mutations was identical between low and medium mutagenesis. For each round of mutagenesis about 1 mutation was gained per ribozyme with both low and medium levels of mutagenesis (Fig. 4.4). Additionally, the evolving population with low mutagenesis showed higher fitness as judged by the ability to survive on higher concentrations of chloramphenicol selection plates (Table 4.1). This suggested that low levels of mutagenesis were sufficient to evolve the ribozymes, whereas medium and high levels of mutagenesis generated mutations in the conserved core more frequently, deactivating a larger population of ribozymes. Therefore, in future rounds, the mPCR was kept at 10 cycles.
Figure 4.4: Average number of mutations per sequence over the 21 rounds of evolution. Rounds 0-2 are shown in purple. 20-cycle mutagenesis rounds 3-8 are shown in blue. 10-cycle mutagenesis rounds 3-8 are shown in green. Rounds of alternating mutagenesis and recombination are shown in orange. Final enrichment rounds only are shown in red. The standard error of the mean from 10 sequencing samples is used as the error.
Table 4.1: Concentration of chloramphenicol (μg/mL) used on the selection plates by round. The number of mutagenic PCR cycles is shown for rounds 1-8. In rounds 9-17, 10 cycles of mutagenesis was alternated with recombination. Round 2 and the final four rounds were selection only.

<table>
<thead>
<tr>
<th>Round</th>
<th>Cam (µg/mL)</th>
<th>Round</th>
<th>Cam (µg/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1-30</td>
<td>4</td>
<td>R9</td>
<td>4</td>
</tr>
<tr>
<td>R2</td>
<td>4</td>
<td>R10</td>
<td>4</td>
</tr>
<tr>
<td>R3-10</td>
<td>4</td>
<td>R11</td>
<td>6</td>
</tr>
<tr>
<td>R3-20</td>
<td>4</td>
<td>R12</td>
<td>6</td>
</tr>
<tr>
<td>R4-10</td>
<td>4</td>
<td>R13</td>
<td>8</td>
</tr>
<tr>
<td>R4-20</td>
<td>2</td>
<td>R14</td>
<td>8</td>
</tr>
<tr>
<td>R5-10</td>
<td>4</td>
<td>R15</td>
<td>10</td>
</tr>
<tr>
<td>R5-20</td>
<td>2</td>
<td>R16</td>
<td>12</td>
</tr>
<tr>
<td>R6-10</td>
<td>4</td>
<td>R17</td>
<td>16</td>
</tr>
<tr>
<td>R6-20</td>
<td>2</td>
<td>R18</td>
<td>20</td>
</tr>
<tr>
<td>R7-10</td>
<td>6</td>
<td>R19</td>
<td>28</td>
</tr>
<tr>
<td>R7-20</td>
<td>2</td>
<td>R20</td>
<td>50</td>
</tr>
<tr>
<td>R8-10</td>
<td>6</td>
<td>R21</td>
<td>70</td>
</tr>
<tr>
<td>R8-20</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To facilitate recombination of beneficial mutations onto the same ribozyme sequence, as well as to eliminate detrimental mutations from otherwise fit sequences, we introduced recombination into the evolution procedure starting with round 9. This was achieved using a method called the Staggered Extension Process (StEP). StEP is a variation on PCR that removes the hold at the extension temperature and permits short extensions during the cycling between heat denaturing and annealing temperatures (13). Recombination and mPCR were alternated during evolution rounds 9 through 17. During these rounds the increase in mutations was less than when only mutagenesis was used (Fig. 4.4). This can partially be explained because only every second round introduced
mutations. However, the increase in the mutations appeared to be less than half the rate observed during the previous rounds. This suggested that the recombination steps successfully removed mutations from the pool that reduced the activity.

After 17 rounds of evolution, the ribozymes with the highest activity were enriched in four rounds of selection without further mutation. Over these four rounds, the chloramphenicol concentration was increased from 20 µg/mL to a final concentration of 70 µg/mL in round 21 (Table 4.1). In these four rounds, the average number of mutations increased sharply (Fig. 4.4).

The selected and sequenced ribozymes from the last three rounds of selection, 30 clones total, were assumed to be the winners of the selection. The sequences were compared and 15 sequences were analyzed further that contained all mutations occurring at least twice among the 30 sequences. The repair efficiency for each ribozyme was compared by determining the doubling times of *E. coli* cultures with each selected plasmid in liquid medium containing 20 µg/mL chloramphenicol.

The most active clone, R21C24, contained 12 mutations relative to the starting ribozyme (R3C7). To determine which of these mutations were mediating the increased activity, clones knocking out individual or pairs of mutations were generated. Knocking out four specific mutations had a significant negative effect on the observed growth behavior. A clone was generated and tested via monitoring growth in liquid medium that contained only these four mutations. The
results showed these four mutations were necessary and sufficient to mediate the high activity \textit{in vivo} (Fig. 4.5).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.5.png}
\caption{Doubling times of the best selected ribozyme (R21C24) and knockout variants in \textit{E. coli}. Error bars are the standard deviation calculated from six measurements. The cultures were grown in LB medium with 1mM IPTG and 20 \textmu g/mL chloramphenicol.}
\end{figure}

These four mutations, (U236C, U238C, U239C, and U241A) were all clustered in the stem loop of P6b. In the 3D structure of the group I intron from \textit{Tetrahymena} this stem loop protrudes into the solvent away from the core of the ribozyme. The effect of the evolved mutations appeared to be increasing the size and decreasing the stability of the P6b loop. At least one of the four selected mutations probably removed a base pair of the P6b stem and increased the size
of the single-stranded portion of the P6b loop. To test this hypothesis, we modified the sequence at the P6b loop to determine if increasing the number of single-stranded bases in the loop would give an additional benefit to ribozyme fitness (Fig. 4.6a). The results showed that these mutants were as efficient as the selected sequence as determined by *E. coli* doubling time *in vivo*. To test whether the requirement was structure or sequence specific, we replaced the (probably single-stranded) C-rich loop with an A-rich loop which should also remain primarily single-stranded. By measuring the *in vivo* activity of these ribozymes in *E. coli* liquid culture, we found that even this A-rich loop improved *in vivo* activity, although the activity was not as high as with the C-rich loop. This strongly suggested that the effect of these mutations was not mediated by a sequence specific contact, for example, by recruiting the substrate of the trans-splicing reaction, instead it suggested that a non-sequence specific factor, for example a protein, may be recruited by the ribozyme to enhance the splicing activity. To test this hypothesis, we transferred the evolved sequence of the P6b loop onto the P8 loop, which is also protruding from the ribozyme at an adjacent position to the P6b stem loop. The measurement of the ribozyme activity in *E. coli* cells showed that these mutations increased the activity far above wild type activity, but not as high as with the selected 4 nucleotide mutant (Fig. 4.6b).
Figure 4.6: P6b and P8 stem loop mutations and their effects on the *in vivo* activity of the ribosome. A: Doubling times in liquid culture shown for P6b mutations versus R3C7, the parent clone for the selection. B: P8 loop sequence change and associated doubling times.

To decide whether a cellular factor may be interacting with the ribosome’s mutations, we performed *in vitro* splicing assays. The results showed no increased *in vitro* activity of the ribozymes with the four selected mutations. This supported the hypothesis that the mutations increased the ribozyme activity by recruiting a cellular factor such as a protein.

To identify any proteins in *E. coli* that were being recruited by the ribozymes, we performed pull down experiments with biotinylated ribozymes and *E. coli* cellular extracts (Fig. 4.7).
Figure 4.7: Schematic of protein pull-down by biotinylated RNA. The ribozyme, with incorporated biotin-UTP, (blue circles) is incubated with *E. coli* cell extracts. The RNA is pulled out by streptavidin (SA) coated magnetic beads. After two wash steps, proteins still interacting with the RNA are isolated.

Biotinylated ribozymes were generated through *in vitro* transcription using biotin-UTP. The cellular extract was prepared by treatment with lysozyme, freeze-thaw lysis and removal of genomic DNA and associated factors. After incubation of the extract with the biotinylated ribozymes and separation of the biotinylated ribozymes via streptavidin coated magnetic beads, the proteins binding to the ribozymes were analyzed by SDS polyacrylamide gel electrophoresis (PAGE) and silver staining (Fig. 4.8). In addition to the parent clone (R3C7) and minimized most active mutant (+4) the beads alone were incubated with cellular extract as a negative control.
Figure 4.8: Silver stained SDS PAGE of proteins separated through the pull down assay. Biotinylated ribozymes R3C7 and +4 were incubated with *E. coli* cell extract before capture on streptavidin beads. A negative control of beads only was also performed to distinguish the proteins specifically interacting with the RNA. The flow-through (FT) was diluted 1:100 for loading. The second wash (W2) and fraction pulled down by ribozymes or beads were not diluted. Apparent molecular weights of the proteins in the marker are listed on the left.

Several proteins appeared to be binding to the ribozymes. The strongest and most reproducible band that did not bind to the streptavidin beads alone had an apparent molecular weight of 66kD. To identify this protein, the band was cut out of the gel and subjected to a trypsin digest and mass spectrometry (MS) analysis. The results of the MS analysis from two independent experiments
unambiguously identified the 30S ribosomal protein S1 as the protein in question (Table 4.2).

**Table 4.2:** Results of mass spectrometry analysis. The most abundant protein bound to biotin labeled ribozymes (and not streptavidin beads) was identified via MS analysis and from two separate pull down assays. The trypsin identified is a result of trypsin digestion of cut out protein bands to enable MS analysis and the beta-D galactosidase is an internal standard of the MS measurements.

<table>
<thead>
<tr>
<th></th>
<th># of Peptides</th>
<th>Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1</td>
<td>17</td>
<td>30S S1</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Trypsin</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>beta-D gal</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>8</td>
<td>30S S1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>Trypsin</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>beta-D gal</td>
</tr>
</tbody>
</table>

**4.3 Discussion**

Through the evolution of the internal ribozyme sequence over 21 rounds we were able to identify a single clone with greatly increased growth behavior as compared to the parent clone. Responsible for this increased activity were four mutations in the P6b loop. These mutations may be recruiting a protein for increased *trans*-splicing activity. A protein that could fulfill this function was isolated in a pull down assay, and identified via mass spectrometry as the ribosomal protein S1. While it is currently unclear how this S1 protein may be helping the ribozyme, the *E. coli* S1 protein appears to have multiple functions. First, the protein binds upstream of the Shine-Dalgarno sequence and facilitates the recruitment of the ribosome in messenger RNAs that have a weak Shine-
Dalgarno sequence (14). Second, the S1 protein may recruit tmRNAs to the ribosome in order to release stalled ribosomes (15). However, this function is controversial (14). The S1 protein has at least two RNA binding domains, one of which appears to prefer a site on the ribosome, whereas the other may be general for single-stranded RNA (16). Our current hypothesis for the increased ribozyme activity mediated by the S1 protein is that the S1 protein binds upstream of the Shine-Dalgarno sequence of the cat mRNA and to the P6b loop of the ribozyme, helping to co-localize substrate and ribozyme (Fig. 4.9). Therefore, the four mutations evolved in the ribozyme may be responsible for better mimicking the protein binding site in the ribosome. Future work will test this hypothesis with in vitro rescue assays and binding assays to assess the interaction of the P6b loop with the S1 protein.

Figure 4.9: Proposed role of 30S ribosomal protein S1 in increasing the trans-splicing efficiency of the evolved R21C24 ribozyme. The S1 protein is known to have multiple RNA binding domains and may co-localize the substrate RNA and ribozyme, helping to increase the efficiency of splicing.

4.4 Acknowledgements

Chapter 4, in part, is currently being prepared for submission for publication, Olson, K.E. and U.F. Müller. The dissertation author was the first author on this material.
4.5 References


Chapter 5

Future Directions

Our *in vivo* selection method has allowed the optimization of the external guide sequence (EGS) and the evolution of the internal sequence of the *trans*-splicing *Tetrahymena thermophila* group I intron (Fig 5.1). Through these *in vivo* selections and evolutions, the activity of the *trans*-splicing ribozyme has increased. In the first selection, the best selected clone, with an improved 5’-EGS reduced the doubling time of the clone in *E. coli* liquid culture, in 2 µg/mL chloramphenicol, by half as compared to a clone without an EGS. In the evolution, the best selected clone had four mutations that are likely recruiting a protein for high activity. In liquid culture, with 20 µg/mL chloramphenicol, the doubling time of the evolved clone was reduced seven fold as compared to the parent clone in the evolution, the best selected clone from the EGS optimization.
Figure 5.1: Summary of the results from the previous in vivo selection and in vivo evolution. Selecting an efficient EGS increased the trans-splicing efficiency and reduced the doubling time of E. coli in 2 µg/mL chloramphenicol liquid culture by half. The subsequent evolution, which used the resulting clone from the EGS selection, produced an efficient clone with four new mutations. These mutations are believed to be recruiting a cellular protein for high activity, and reduced the doubling time of E. coli in 20 µg/mL chloramphenicol by seven fold.

So far in our studies the ribozyme targeted U177 in the chloramphenicol acetyltransferase (cat) mRNA to repair a frame shift mutation and allow growth of E. coli in medium containing chloramphenicol. However, several other splice sites or resistance genes could be utilized for similar selections. The internal sequence evolution appears to have generated four mutations in an exposed stem loop that has recruited a protein to facilitate high trans-splicing in vivo. While a protein performing this role has been preliminarily identified as the 30S
ribosomal protein S1, further experiments are necessary to test the requirement of the protein for increased ribozyme function.

Unfortunately, knockout of the S1 protein is lethal to \textit{E. coli} (1), therefore a comparison of growth \textit{in vivo} cannot be performed utilizing \textit{E. coli} cells that do and do not express the S1 protein. Although the species \textit{Bacillus subtilis}, a gram-positive bacterium, does not have an S1 protein (2), a growth comparison in this bacterium may not yield clear results as another protein with sequence homology is believed to perform the function of the S1 protein in these cells (3).

Therefore, experiments determining if the ribozyme is utilizing the S1 protein for increased efficiency must be made \textit{in vitro}. We have already demonstrated that the evolved best ribozyme sequence R21C24 has a \textit{trans}-splicing efficiency \textit{in vitro} similar to that of the parent clone R3C7. To test whether the S1 protein has an effect on the evolved ribozyme, the S1 protein can be expressed, purified and added into the splicing reaction \textit{in vitro}. If R21C24 then shows significantly higher splicing activity than R3C7, then the protein requirement is demonstrated.

It is possible, however, that after titrating the protein concentrations in the reaction, or altering the buffer to facilitate proper protein fold and ribozyme activity, the splicing activity of parent and evolved ribozyme remain similar. This result could indicate that the S1 protein has no role in the increased splicing activity or does not act alone. In this case, it may be possible to identify either a protein that acts with S1 or an alternate protein through less stringent washing of the ribozyme in pull down assays or through cross linking.
Once the protein has been identified and the interaction with the ribozyme substantiated, binding studies can be performed. If the protein in question is in fact S1, we already know that it interacts with both the parent clone R3C7 and the evolved clone containing four mutations in the P6b loop. To precisely measure the affinity for both ribozymes, filter binding assays can be performed while titrating the ribozyme, (or truncated P6b stem loop sequences) or the protein concentration (4). Sedimentation analysis can identify the stoichiometry between the protein and RNA interaction. A footprinting technique (5) can also be utilized to map the binding site on the ribozyme, and to test our hypothesis that the S1 protein may be helping to co-localize the substrate and ribozyme.

To directly determine the repair efficiency in vivo, qRT-PCR can be performed to analyze the percentage of the substrate that is repaired inside the cells. Lastly, chloramphenicol acetyltransferase activity assays can be performed to assess the activity level of the repaired CAT protein among different clones relative to the wild type gene. In this assay, active chloramphenicol acetyltransferase enzyme can be extracted from E. coli cells and used to modify a radiolabeled or fluorescently labeled chloramphenicol molecule. The fraction modified can then be separated on a TLC plate and and imaged to compare the amount of active enzyme in different samples (6).

Many variations of our selection procedure could be performed by varying the splice site, target mRNA, type of group I intron used and/or the type of cells the repair reaction is selected in. Each variation can be designed to address specific questions. Performing selections on a different target site or mRNA will
help determine if different EGS sequences and structures are favored at different splice sites. Engineering the evolved P6b loop (which may be recruiting a protein) onto another group I intron, or a group I intron targeting a different mRNA will help determine if this result is specific to a particular target gene or ribozyme, or if it is a universal way to increase *in vivo* activity. Additionally, performing internal sequence evolutions in yeast or mammalian cells may aid in finding proteins more relevant for gene therapy.

In contrast to the 5' trans-splicing we have used, at least one instance of a 3' trans-splicing reaction has been performed for repair on the mRNA level (also termed 5'-end replacement) (7). For splicing schematic, see figure 5.1a below. Selections and evolutions of 3' trans-repair reactions could be performed to better understand this reaction and optimize binding sites and efficiency of repair. This would again require a method for determining accessible splice sites and would help to determine the ribozyme requirements for efficient *trans*-splicing when repairing mRNA by transferring a 5'-exon. Once efficient ribozymes for 3' trans-repair are identified, a two ribozyme system can be developed to replace an internal portion of the mRNA transcript (Fig. 5.1b). This repair method will offer specific advantages. By replacing a small portion of the mRNA around the mutation, both the natural 5'-cap and the 3'-poly A tail will be retained, aiding in mRNA stability. This type of repair has potential significant advantages for long mRNA transcripts. Instead of replacing thousands of nucleotides by a single splice-site reaction, a fragment as small as 100 base pairs would be spliced in. Therefore using the two splice-site reaction could dramatically reduce the size of
the ribozyme-exon sequence that needs to be delivered to cells, which can be a concern depending on the capacity of the delivery system. Selections must be performed with this reaction as well, to ensure the two splice-site ribozyme is stable and efficient enough for *in vivo* repair.

Figure 5.2: Schematic of additional group I intron splicing reactions. A: In 3’-*trans*-splicing reactions, the substrate mRNA (red) binds in *trans* to the 3’-end of the ribozyme and the 5’-exon (blue) is transferred to the mRNA for repair. B: In the two splice site ribozyme, 5’ and 3’ *trans* ribozymes are connected by the exon fragment transferred for repair (blue). The two ribozymes bind to a substrate (red) in *trans* and through two *trans*-splicing reactions, replace an internal portion of the substrate with the exon sequence connecting the ribozymes.

Eventually, to create a viable gene therapy with mRNA repair, the principles of efficient ribozyme design must be better understood, or a selection procedure needs to be developed that can allow selection and evolution of ribozymes targeting disease genes in human cells. It may be possible to utilize a
fluorescent gene tag attached to the 3’ repair exon and to select ribozyme sequences that can splice efficiently via fluorescence activated cell sorting (FACS). FACS has previously been used to assess group I intron cis-splicing in vivo (8), and it should be possible to monitor trans-splicing as well. However, the 3’-tag may have an unintended or unknown effect on the trans-repair reaction, and splicing may not occur as efficiently when the tag is removed.

Alternatively, a cell surface enzyme linked immunosorbent assay (ELISA) could be utilized to identify translation products of repaired disease mRNAs inside human cells. Mutations that prevent membrane bound proteins from becoming incorporated in the membrane and thereby cause disease are optimal targets. For example, a known mutation in the vasopressin type 2 receptor (a type of G protein coupled receptor) generates a premature stop codon, preventing cells from sensing the hormone vasopressin, thereby causing nephrogenic diabetes insipidus (9). Another example target, is the trans-membrane ion channel cystic fibrosis transmembrane conductance regulator (CFTR). A known mutation in the cfr gene (ΔF508) is the most common cause of cystic fibrosis and predominantly produces a prematurely degraded protein that does not integrate into the membrane (10, 11). If plasmids encoding a library of trans-splicing repair mutants are transfected into a cell line developed for the study of cystic fibrosis (12), mRNA repair can be detected indirectly by a cell surface ELISA using primary antibodies that bind to the extracellular loops of CFTR (13).
In early rounds, transfected cells could be grown in 96 well plates containing several different ribozyme sequences. After sufficient incubation to afford the ribozyme time to repair the gene, an ELISA assay procedure could be performed to detect fluorescence in the wells where repair is occurring. Since it is likely a low proportion of ribozymes will repair the substrate, particularly in early selection rounds, selection is obtained by carrying forward only those wells that show measurable fluorescence. This would allow for a method of compartmentalizing the ribozyme sequences and assaying repair activity. As the evolution rounds progress, the pool complexity of ribozyme sequences will be lower and the concentration of the cell culture can be diluted more before aliquotting into the 96 well plates. This will reduce the number of different ribozymes in each well that have the chance to repair, thereby aiding in selecting the best sequences from the pool. The resulting selected ribozymes could then be utilized for studies in animal models and may one day lead to a disease cure.

References


Appendix: Materials and Methods

PCR Protocol

Polymerase Chain Reaction (PCR) allows for quick and efficient replication of DNA fragments in a single tube. Short DNA oligonucleotides (primers) that are complementary to the template define the sequence to be amplified and prime extension. A DNA polymerase such as Taq DNA polymerase, which comes from the organism *Thermus aquaticus*, will extend annealed primers, thereby replicating the DNA template, and allowing for exponential amplification in consecutive cycles. Each cycle has three temperature steps and begins with a heat denaturation to separate the template strands. The second step cools the reaction to an annealing temperature which allows the 5’ and 3’ primers to anneal to the template to create a new top and bottom strand respectively. The last step, extension, is carried out at the optimal temperature for the polymerase and allows the annealed primers to be extended to full length. In theory each cycle should be capable of doubling the amount of dsDNA from the previous cycle, but in practice the template amplifies approximately 1.6 fold from round to round. A timecourse, collecting a sample of the reaction toward the end of the extension step at regular cycle intervals (e.g. collecting 5µL every 3rd cycle of a 24 cycle timecourse), should be done to determine the appropriate number of cycles to amplify the template. After collecting samples during the timecourse, run them on an agarose gel to select the number of cycles that gives a clean, sharp PCR product at the length expected.

**Pipetting scheme for 100µL PCR / reaction**

<table>
<thead>
<tr>
<th>Volume</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 µL</td>
<td>Template (see notes below)</td>
</tr>
<tr>
<td>1 µL</td>
<td>5’ primer (100µM)</td>
</tr>
<tr>
<td>1 µL</td>
<td>3’ primer (100µM)</td>
</tr>
<tr>
<td>10 µL</td>
<td>10x PCR buffer</td>
</tr>
<tr>
<td>1 µL</td>
<td>dNTPs (25mM each)</td>
</tr>
<tr>
<td>85 µL</td>
<td>H2O</td>
</tr>
<tr>
<td>1 µL</td>
<td>Taq polymerase</td>
</tr>
</tbody>
</table>

Total 100 µL
Notes: Template can be varied as needed. In general, PCR from a plasmid for future agarose gel purification is done with 1µg, and PCR to make a T7 template is done with about 300ng.

**Example cycling profile**

![Example cycling profile diagram]

**PCR 10x Buffer for Taq**

- 600mM Tris HCl pH 8.3
- 170mM (NH₄)₂SO₄
- 15mM MgCl₂

**Troubleshooting**

**No PCR product:**

- Re-check that the correct primers are being used and that they were designed correctly (i.e. forgetting to order the reverse complement for the 3’-primer will not allow an amplification of product).
- Lower the annealing temperature and run another time course.
- Increase the number of cycles, but keep below 40 cycles in general.
- If a PCR product greater than 500bp is expected, increase the extension time estimating 1 kilo base extended by Taq per minute.

**Messy PCR product:**

- Check for alternate annealing sites, or self-complementarity of the primers used.
- Increase the annealing temperature to remove non-specific binding.
- Choose a lower PCR cycle for the final product.
- Gel isolate the product desired on a 1% agarose gel using a gel extraction kit. Note, a second PCR can be done directly from the punched out agarose gel purification, but not all types of agarose are compatible in PCR reactions.
• Try two PCR steps with nested primers. Begin with a set of primers that anneal outside of the fragment desired, then perform a second PCR with the first reaction as template using the primers that allow amplification of the sequence desired.

Tips
• It was previously found in our lab that the error rate of Taq is about 1/400 nt in 30 cycles of PCR. If greater fidelity is needed use a high fidelity enzyme such as Phusion from NEB, or Pfu from Stratagene. Changing the enzyme will change buffer composition, as well as dNTP and primer concentration requirements and cycling conditions.

Reference:
Designing Oligo DNAs

Design of oligos for PCR, site directed mutagenesis, reverse transcription, poisoned primer extension or for using as an adapter to modify a plasmid must be done with care. Ideally, the primers will be relatively short, less than 30 nucleotides, have a roughly even mix of bases spread over the primer length and avoid making self structures which can hinder the desired reactions. Primers that are longer than 30 nucleotides increase the chance for mis-annealing to the template or other DNA sequence in the reaction, as well as increasing the possibility of intramolecular base pairing which can lead to low product yield. It is important to ensure that the primer will have an annealing temperature (Tm, the temperature at which half of the primer sequence will be annealed to the complementary template) in a desired range for the intended application. The higher the annealing temperature, the greater the percentage of the primer that is annealed at a temperature below the Tm. PCR primers should ideally have a Tm between 55°C and 65°C, but the annealing step can be adjusted for primers with higher annealing temperatures. For site directed mutagenesis using Pfu Turbo, the primers should have a Tm that allows efficient annealing at the optimized annealing temperature of 55°C, but not require a Tm of above 68°C to avoid non-specific binding, as the extension of Pfu Turbo is optimized at 68°C. Poisoned primer extension and reverse transcription must be carried out at 42°C, therefore shorter primers are preferred to avoid non-specific binding.

For PCR Primers:

- Primers should be 20-25 nucleotides in length, although shorter (generally not shorter than 18nt) and longer (generally not longer than 60nt) primers can work if necessary.
- Primers should end with a G-C cap on the 3’-end of the oligo. This will increase binding affinity at the end of the primer which will be extended by the polymerase.
- If possible, choose primers with a similar Tm, generally between 55°C and 65°C
- Make sure the primer sequence is a mix of all four nucleotides, and does not have repeating patterns.
- For PCR and 5’, or forward and 3’ or reverse primer must be ordered. Ensure that the 3’-primer is ordered as the reverse complement of the template sequence.
- Check the primer for self-complementarity, or complementarity to the partner primer in the PCR.
• If the primer is adding nucleotides to the 5’ or 3’-end of the template, make sure the primer anneals to at least 20 nucleotides on the template, and calculate the Tm from these bases only.
  
  Simple Tm calculation:
  
  \[2^\circ C \times (A \text{ or } T) + 4^\circ C \times (G \text{ or } C) = T_m\]

  For Site Directed Mutagenesis:
  
  • Primers should be roughly 30-35 nucleotides long, end in a GC cap and have 10-15 nucleotides on either side of the mutation.
  
  • Primers should be designed so they are exactly complementary. Order the sequence you want the parent plasmid mutated to and its reverse complement.
  
  • If more than one mutation is being made immediately next to one another, increase the length of the primer. SDM has successfully been performed in our lab with primers up to 50 bases.
  
  • If more than one mutation is being made, but spaced a few nucleotides apart, consider using two consecutive SDM reactions.
  
  • The longest stretch of mutations in one SDM reaction was 10 base changes, more mutations, or mutations in a different sequence context may prove difficult.

  Reverse Transcription and Poisoned Primer Extension:
  
  • Primers should be short (15-18 nucleotides).
  
  • Only one primer is necessary and this primer must be complementary to the 3’-end of the RNA sequence.
  
  • Include a GC cap at the 3’-end of the primer if possible.
  
  • For poisoned primer extension, make sure that the primer binding site is designed in a region that will allow clear determination of reverse transcript termination at the chosen nucleotide. In general, two or three nucleotides difference between the positive and negative sequences is sufficient and resolution will be determined by the percentage and length of the PAGE gel run.

  For Adapter Oligos:
  
  • Deign two complementary oligos that leave appropriate 5’ and or 3’ overhangs, as determined by the restriction sites, to allow ligation into a digested plasmid.
• Due to the nature of the oligo synthesis process, they will arrive without a 5'-phosphate. Oligos will need to be kinased before ligation is possible into a dephosphorylated plasmid.
• Ensure that the oligos do not have self-complementarity and do not introduce an unintended restriction site.
• Design of adapter oligos can either preserve or destroy a restriction site. Double check the sequence before ordering to avoid problems downstream.
Colony PCR Protocol

Colony PCR is a method to pre-screen cloned plasmids before sequencing or to determine an insert ratio of a selection pool. After ligating a PCR amplified or restriction digest isolated insert into the vector plasmid and transforming into *E. coli*, colony PCR can be performed directly from the colony. The PCR conditions are similar to a standard PCR, using primers that either both anneal outside the multiple cloning site of the plasmid, or one that anneals to the plasmid and one that anneals inside the insert sequence. Using a pair of primers that anneal to the plasmid allows one to amplify a positive and negative control along with the clones being analyzed. Primer pairs that have one primer on the plasmid and the other on the insert allow for detection of both insert presence and directionality, especially in the cases where only one restriction site was used.

**Colony PCR premix for 36 samples**
- 15 µL 5’ primer (100µM)
- 15 µL 3’ primer (100µM)
- 150 µL 10x PCR buffer
- 15 µL dNTPs (25mM each)
- 1290 µL H2O
- 15 µL Taq polymerase

Total 1500 µL

**Notes:** The template for this PCR will be a sample of the *E. coli* colony. Begin by taking a sample of the colony with a pipette tip and touching the tip in the bottom of a PCR tube. Then streak the tip on a gridded agar master plate for later use. Add 30µL of the PCR premix to each tube and run 25 PCR cycles. Following the PCR, mix 5µL of each colony PCR reaction with 2µL 6x agarose loading buffer and load on a 2% agarose gel to determine which colonies have an insert.
Example cycling profile

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>94°C</td>
<td>2-5 min</td>
</tr>
<tr>
<td>94°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>72°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>72°C</td>
<td>2-5 min</td>
</tr>
<tr>
<td>57°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

PCR 10x Buffer for Taq
- 600 mM Tris HCl pH 8.3
- 170 mM (NH₄)₂SO₄
- 15 mM MgCl₂

Troubleshooting
No PCR product:
- Re-check primer identity, annealing site and concentration.
- Lower the annealing temperature and run another time course.
- Increase the number of cycles. Smaller colonies or low copy number plasmids will require more cycles, but stay below 40 cycles in general to avoid non-specific amplification.
- If a PCR product greater than 500bp is expected, increase the extension time estimating 1 kilo base extended by Taq per minute.

Messy PCR product:
- Ensure that the primer cannot anneal in more than one spot on the template.
- Increase the annealing temperature to remove non-specific binding.
- Choose a lower PCR cycle for the final product.

Tips
- When primer pairs are chosen that do not both anneal to the plasmid, no PCR product will be seen if there is no insert. Cloning directionality can be set by using two different restriction enzymes, and allowing the use of two primers that anneal only to the plasmid, outside of the cloning site.
- For efficient clonings, only a few colonies will need to be tested to find a few with the desired inserts. You can usually get a good idea of how
many colonies you will need to survey based on the number of clones on
the transformed plates and knowing the background of the particular
plasmid digestion.

Reference:
Microbiol Appendix 3*, Appendix 3D.
Mutagenic PCR Protocol

Mutagenic PCR (mPCR) uses the same principle as a standard PCR, but changes the annealing temperature, buffer conditions and dNTP concentrations to promote misincorporation of bases and increase the error rate of the enzyme. Our protocol is based off of the CadwellJoyce94 paper in which they describe a tested mPCR method that has been designed to counteract the natural A→T → G→C bias of Taq polymerase.

**Pipetting scheme for 100uL**

- 1 µL  Template (see notes below)
- 1 µL  5’ primer (100µM)
- 1 µL  3’ primer (100µM)
- 10 µL 10x mPCR buffer
- 10 µL 10x mPCR dNTP mix
- 65 µL H2O
- 10 µL 5mM MnCl₂
- 1 µL  Taq polymerase

Total 100 µL

**Notes:** Template can be varied as needed. In the paper, 20fmol of template is suggested for 30 cycles of mPCR. With our internal randomization selection, 100ng or 5ng of gel purified PCR template was used for 10 and 20 cycle mPCR in our internal randomization rounds, corresponding to approximately 335fmol and 16fmol respectively. Following the mutagenic PCR, perform a standard PCR to add restriction sites for cloning.

**Example cycling profile (method used for internal randomization):**

- 94°C 30 sec
- 50°C 30 sec
- 72°C 1 min
- 4°C ∞
**Troubleshooting**

Mutagenic PCR can often give messy products due to a low annealing temperature and buffer conditions that promote base mismatch incorporation. Because the protocol outlined by Cadwell and Joyce is optimized for a certain ratio of transitions (purine to purine or pyrimidine to pyrimidine changes) to transversions (purine to pyrimidine or vice versa) it is better to attempt to troubleshoot a difficult mPCR with template concentrations, primer design or annealing temperature changes. The standard mPCR protocol from Cadwell and Joyce was altered to increase the annealing temperature from 45°C to 50°C and to shorten the denaturation and annealing temperature to 30 seconds from 1 minute.

**Tips**

- Always add the first 6 ingredients, mix, add the manganese solution, mix and finally add the polymerase before mixing and starting the PCR. If the manganese is added to the buffer or before the water is added it may precipitate and affect the mutation rate.
- Mutagenic PCR is often cleaner from a PCR product rather than a plasmid. If the template is on a plasmid, it is recommended that a standard PCR be run, and the product gel purified from the plasmid to template the mPCR.
- Time courses should be run for each mPCR with a new set of primers, or if a certain number of cycles is desired, run a template concentration course to obtain a clean PCR.
- Sequencing of approximately 10 clones is recommended once the number of mPCR cycles is set to determine the number of mutations to expect per sequence and to ensure there is no sequence bias.

**References:**

StEP PCR Protocol

PCR amplification using the Staggered Extension Process (StEP) allows a pool of variants of a single sequence to be artificially recombined. The StEP protocol makes use of the fact that many polymerases have a lower level of extension activity at temperatures below the optimal extension temperature. Removing the extension time and cycling between denaturation and annealing temperatures limits the number of bases the polymerase can add before template and extended primer separate. In the next round, the partially extended primer can anneal to another template generating new combinations of the mutations in the pool. The following protocol is based off of the Zhao et al 1998 paper.

**Pipetting scheme for StEP PCR**

3 µL 50fmol/µL template
1 µL 5’ primer (30µM)
1 µL 3’ primer (30µM)
10 µL 10x PCR buffer
1 µL dNTPs 10mM ea.
83 µL H2O

1 µL Taq Polymerase

Total 100 µL

**Notes:** The StEP reaction should be run in thin walled tubes using a thermocycler that can ramp quickly, ideally at least 5°C per second. A timecourse should initially be run to 80 cycles or more to determine how many cycles it takes to obtain the size product desired without over PCR. If the product is not clean, gel purify the product to remove misannealing products. If the template was a plasmid and the PCR product is clean, DpnI digest the plasmid and follow with a standard PCR to add restriction sites for cloning.
Example cycling profile

\[
\begin{align*}
95^\circ C & \quad 5 \text{ min} \\
94^\circ C & \quad 30 \text{ sec} \\
57^\circ C & \quad 5 \text{ sec} \\
4^\circ C & \quad \infty
\end{align*}
\]

PCR 10x Buffer for Taq

- 600mM Tris HCl pH 8.3
- 170mM (NH₄)₂SO₄
- 15mM MgCl₂

Troubleshooting

No StEP product

- Adjust annealing temperature
- Increase number of cycles

Tips/Notes:

For the recombination rounds in the internal randomization selection of splice site 177, 40 cycles were sufficient to get a clean PCR product. We found it difficult to get more than one recombination per two sequences on average for a template of about 450 bases. The annealing time was reduced to 1 second, and although we found one sequence with 7 recombinations, the average was still approximately 1 in 2. A series of different thin walled tubes were also tested, but none gave a significantly different recombination frequency. It would be recommended that a new time course and sequencing be run on each new template and primer set to optimize this protocol.

Reference:

Site Directed Mutagenesis Protocol

Site Directed Mutagenesis (SDM) allows for mutations to be made on a plasmid without the need for cloning procedures. The mutation can be a single base change or several codon changes, and can span about 10 bases or more depending on the primer design. This protocol is often the easiest way to incorporate a mutation that is both in the middle of a sequence and distant from a natural and unique restriction site. Primers for this reaction should be designed that are exactly complementary and have the desired mutation(s) roughly in the middle. Each end of the primer should terminate in a GC clamp and be approximately 12-15 bases from the nearest mutation. In each cycle the goal is to extend the primers around the entire plasmid template, and after several rounds, digest the plasmid with DpnI to remove the template.

Pipetting scheme for SDM

| µL | 5 µL 10x rxn buffer |
|    | 1 µL 50ng/µL template |
|    | 1 µL 5’ primer (125ng/µL) |
|    | 1 µL 3’ primer (125ng/µL) |
|    | 1 µL dNTPs 10mM ea. |
|    | 40 µL H2O |
|    | 1 µL Pfu Turbo 2.5U/µL |
| Total | 50 µL |

Example cycling profile

95°C 95°C
30 sec 30 sec

55°C 68°C 68°C
1 min 5 min 7 min

4°C

Notes: The recommended number of cycles are; 12 for point mutations, 16 for single amino acid changes, and 18 for multiple codon changes, insertions or deletions. In general, use 1 minute per kilobase of plasmid template for the extension time and include a long final extension to allow all templates to
become fully elongated. The annealing temperature will likely work for most primer sets. It is recommended to run a gel before transforming to ensure you have sufficient product and did not design primers that anneal in multiple locations.

**DpnI digestion**
Add 1µL DpnI (NEB) directly to the SDM reaction and incubate for 1h at 37°C. If a different high fidelity enzyme is used, or if DpnI is used from a different supplier, the buffer may not be compatible with the enzyme. In this case, ethanol precipitate or purify on a kit to exchange the buffer before the DpnI digestion.

**Transformation and Sequencing**
After the DpnI digestion, transform the reaction into DH5α or other *E. coli* strain for cloning. If the DpnI digestion is directly used for a heat shock transformation, very few colonies should be observed. More colonies will be observed if the digestion is concentrated and transformed into electrocompetent cells. Mini prep a couple of colonies and send for sequencing to ensure the mutation was picked up.

**Troubleshooting**
No clones on transformed plate
- Make sure the plates had the correct antibiotic
- Run the SDM product on a gel to ensure the product is the size of a linearized plasmid
- Concentrate the reaction after DpnI digestion and/or use higher efficiency cells for the transformation.

No product observed on the gel
- First double check that the primers ordered were exact complements.
- Second, add 4-5 cycles and run again. Because this protocol is not a PCR, the template does not amplify and can at most double the initial template with each cycle.

Only template is sequenced
- Increase the DpnI digestion time
- Run the SDM product on a gel to ensure the product is the size of a linearized plasmid

**Additional resources:**
QuikChange Site Directed Mutagenesis Protocol by Stratagene
Quantitative Real Time PCR (qRT-PCR)

Quantitative Real Time PCR (qRT-PCR) is a method to quantify the amount of the DNA in a sample. This highly sensitive method involves running a PCR reaction in a 96 well PCR plate and monitoring the fluorescence of a DNA intercalator such as SYBR green over several rounds of PCR. SYBR green will bind preferentially to the double stranded DNA generated in the PCR and the DNA-dye complex will then absorb blue light and emit green. The fluorescence will increase with increasing DNA concentration generated in each round of PCR. The number of cycles required to cross a threshold level of fluorescence can be back calculated to the amount of DNA in the sample after taking into account the length of the PCR product. In general, two or three dilutions should be made of one sample to obtain an accurate reading.

To perform qRT-PCR from in vitro transcribed or total RNA extractions first follow the reverse transcription protocol for downstream qRT-PCR. After the initial 5x dilution of the reverse transcription reaction, the samples should be diluted between 100 and 1000 fold to ensure the DNA concentration is in a good range for qRT-PCR detection.

Primer and master mix premix:
1 part 2 µM each primer
5 parts Applied Biosystems qPCR master mix
- Mix primers and master for one extra reaction and mix thoroughly before pipetting onto the sample DNA. Make a separate premix for each primer pair needed.

Pipetting the samples:
- Pipette 4 µL of each sample into a separate well of a 96 well PCR plate.
- Add 6 µL of the primer/buffer premix and pipette up and down 2-3 times to mix. Make sure to fully empty the pipette tip.
- Write down the location of each sample used, especially if more than one primer pair is used on one plate.
- Seal the 96 well plate with adhesive film, making sure to seal around the top of each well.
- Spin the plate to ensure the samples are at the bottom of the wells.
Instructions for AppliedBiosystems qPCR machine:
1. Switch on the PCR machine (the computer should be on)
2. Open the drawer, place the plate inside making sure the A1 corner is in the top left and close the drawer.
3. On the computer, open the Fast System SDS software.
5. Select the appropriate area of wells that contain a sample.
6. Right click -> well inspector
7. Set task to unknown
8. Checkmark "use"
9. Detector style "GAPDH ..."
10. Reporter: SYBR
11. Reference: ROX
12. Click on the floppy disk icon -> save in data
13. Instrument: 10°/95°C, 30°/95°C, 30°/57°C, 30°/72°C. The samples should be run and monitored over 40 cycles.
14. Sample volume: 10 µL
15. Run Mode: Standard 7500
16. Data Collection: Step 2 (72°C)
17. Click on the floppy disk icon -> save
18. Start

Note: This run should take about 2-2.5 hours and will include a heat denaturation run at the end to help confirm the length of the PCR products.

Troubleshooting:
No PCR product is observed via fluorescence over the 40 rounds of PCR
  • Double check the template. If it came from the reverse transcription of an RNA, make sure the correct primer was used and run a sample on a gel to detect DNA production.
  • Double check the primer pair used includes both a 5’ and 3’ primer and both anneal to the template.
  • Check the length of the PCR product expected. The optimal PCR length should be between 100-300 bp, longer products may not amplify well in 30 seconds.
  • Run a few of the post PCR samples on a gel to check for product that may not have been observed by the machine. If clean product is seen here, check that the settings on the detector are correct.
  • If a lower dilution is available, run another qRT-PCR with these dilutions.
Threshold fluorescence is crossed at a low cycle number
- Check that the sample is a clean DNA product, or that the primers cannot anneal or mis-anneal to the sample or other DNA.
- Run the sample on a gel to make sure only one PCR product length is seen. This may include a large smear, but not distinctly different PCR products.
- Dilute the sample 10-100 fold and run the qPCR again.

Notes:
- If different DNA species are being compared with different PCR primers the PCR primers must be designed to give PCR products of the same length. The primer pairs must also be tested to ensure they give similar PCR amplification of a known template concentration.

Reference:
Reverse Transcription (cDNA synthesis)

Reverse transcriptase is an enzyme that catalyzes the generation of DNA from an RNA template also called reverse transcription. A short 3’ DNA primer is first annealed to the RNA template forming a binding site for the enzyme. The enzyme can then extend the 3’-end of the primer as templated by the RNA sequence. After reverse transcribing RNA into a DNA sequence, the DNA can then be used for PCR, cloning and sequence or qRT-PCR.

**Example Reverse Transcription for subsequent PCR:**
This protocol was used for reverse transcription of *in vitro* ribozyme reactions followed by PCR and cloning to identify splice sites.

**RT 5x Buffer**
250 mM Tris/HCl pH 8.3 @20°
375 mM KCl
15 mM MgCl₂

**Reverse Transcription (for 20µL)**
- 12 µL Template (directly from Ribozyme Reaction)
- 1 µL 3’ primer (100 uM)
- 4 µL 5 x Buffer
- 1 µL dNTP’s (25 mM each)
- 1 µL DTT (10mM)
- 1 µL Reverse Transcriptase

Total 20 µL

- Incubate at 47°C for 60 minutes for reverse transcription.
- Incubate on ice for 2 minutes to cool the reaction.
- Add 5 µL fresh 1M NaOH and incubate at 90°C for 10 minutes to degrade the RNA.
- Incubate on ice for 2 minutes to cool the reaction.
- Add 5 µL fresh 1M HCl to neutralize.
- Add 60 µL 10x PCR buffer to 30 µL of RT product for storage. This is done to ensure the product is buffered in a safe range for possible storage.
- Reaction products can be stored at -20°C overnight, or used immediately for the PCR time course.
PCR from RT (for 300µL)
45 µL Template with PCR buffer added
3 µL 5’ primer (100 µM)
3 µL 3’ primer (100 µM)
3 µL dNTP’s (25 mM each)
243 µL H₂O
3 µL Taq Polymerase
Total 300 µL

- Depending on the length of primers adjust PCR program annealing temperature, 45°C was used for 17bp primers.

Example Reverse Transcription for qRT-PCR template:
This protocol was used for reverse transcription of ribozyme, substrate and products RNAs from a total RNA extraction after in vivo splicing. This protocol works better for templates that come from total RNA extraction, which include more RNA than the desired template, because it includes a slow annealing step to promote annealing to the correct sequence.

4.5x Hybridization buffer
225 mM K-Hepes pH 7.0
450 mM KCl

10x Extension buffer
1.3 M Tris-HCl, pH 8.5
100 mM MgCl₂
100 mM DTT

Step 1: Hybridization of RNA and primer

Hybridization of total RNA template with primer:
4 µL 50 ng/µL RNA prep
2 µL 4.5x hybridization buffer
2 µL 1 µM primer
1 µL H₂O
9 µL
Hybridization of *in vitro* transcribed RNA and primer:

<table>
<thead>
<tr>
<th>Component</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>gel purified RNA template (1 ng/µL) or H2O for 16s RNA</td>
<td>2 µL</td>
</tr>
<tr>
<td>Background RNA (100 ng/µL)</td>
<td>2 µL</td>
</tr>
<tr>
<td>4.5x hybridization buffer</td>
<td>2 µL</td>
</tr>
<tr>
<td>1uM primer</td>
<td>1 µL</td>
</tr>
<tr>
<td>H2O</td>
<td>1 µL</td>
</tr>
</tbody>
</table>

9 µL

- Mix each hybridization reaction and heat to 90°C and cool at 7% ramp to 42°C to prepare for the extension.

**Step 2: Extension of hybridized primers**

Extension premix – depending on the number of samples, make a premix as follows.

<table>
<thead>
<tr>
<th>Extension premix</th>
<th>2 reactions</th>
<th>4 reactions</th>
<th>8 reactions</th>
<th>10 reactions</th>
<th>12 reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x extension buffer</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>dNTPs 25mM each</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>80</td>
<td>96</td>
</tr>
<tr>
<td>AMV-RT (15U/µL)</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Water</td>
<td>1.5</td>
<td>3.5</td>
<td>7</td>
<td>9</td>
<td>10.5</td>
</tr>
<tr>
<td>Total Volume</td>
<td>22</td>
<td>44</td>
<td>88</td>
<td>110</td>
<td>132</td>
</tr>
</tbody>
</table>

- Add 11 µL of the extension premix to each sample and mix.
- Incubate 1h at 42°C.
- Following the 1h incubation, dilute 5-fold by adding 80 µL H2O.
- Samples can be immediately used for qRT-PCR or stored at-20°C until needed.

Reference:
Poisoned primer extension

Poisoned primer extension is an assay that differentiates between RNAs with similar sequences. One nucleotide is chosen as the "poison" nucleotide and is included as a ddNTP which will terminate the reverse transcription. For the poisoned primer extension a 5'-end $^{32}$P labeled primer is first annealed to an RNA template. After annealing, the extension premix is added and reverse transcriptase can extend the primer until the first poison nucleotide is templated. The primers for this reaction must be carefully designed to anneal to a portion of the RNA sequence that is identical for each of the RNAs to be differentiated. The primer must also anneal just a few nucleotides from the point of difference in the sequence. This will allow for the primer to only be extended by a few nucleotides, and for the different species to be resolved on a PAGE gel and quantified. This protocol was successful in differentiating substrate, product, full length and cleaved ribozyme 3'-tails from a total RNA extraction.

Primer labeling

<table>
<thead>
<tr>
<th>Volume (µL)</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 µM Primer</td>
</tr>
<tr>
<td>1</td>
<td>10x PNK buffer</td>
</tr>
<tr>
<td>3</td>
<td>γ$^{[32P]}$ ATP (3000 ci/mmol)</td>
</tr>
<tr>
<td>4</td>
<td>H$_2$O</td>
</tr>
<tr>
<td>1</td>
<td>PNK</td>
</tr>
<tr>
<td>10</td>
<td>Total</td>
</tr>
</tbody>
</table>

Incubate 30min at 37°C and stopped with 2 volumes of formamide loading buffer. Page purify on an appropriate percentage PAGE gel (select based on the length and the percentage gel that will be used for running the extension products, make sure you get single nucleotide resolution). Notes:

- The amount of radioactive ATP can be adjusted depending on the volume and how radioactive the primer needs to be.
- Eluting in 300mM NaCl and 0.1% SDS may increase the amount eluted.

4.5x Hybridization buffer

425 mM K-Hepes pH 7.0
450 mM KCl
Hybridization

1 µL 4.5x Hybridization buffer
1 µL Primer (about 5,000 cpm/µL)
2.5 µL RNA template
4.5 µL Total

* Mix samples by vortexing or pipetting and spin down if necessary. Incubate the samples for 1 minute at 90°C then cool at 7% ramp to 42°C. At this point you can spin the tubes down shortly to get all the condensation to the bottom (although it is usually less than 0.5 µL that is on the sides of the tubes).

Note: Using about 5,000 cpm/µL has given a good signal when loading 4 µL of the 10 µL reaction (after precipitation and re-dissolving to 10 µL).

10x Extension buffer
1.3 M Tris-HCl, pH 8.5
100mM MgCl₂
100mM DTT

Extension – add 5.5 µL extension mix to the hybridization reaction after it has cooled to 42°C. Mix the reaction by pipetting and incubate at 42°C for 30 minutes.

<table>
<thead>
<tr>
<th>Extension premix</th>
<th>6 reactions</th>
<th>8 reactions</th>
<th>10 reactions</th>
<th>12 reactions</th>
<th>14 reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x extension buffer</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>0.25 mM d/ddNTPs</td>
<td>24</td>
<td>32</td>
<td>40</td>
<td>48</td>
<td>56</td>
</tr>
<tr>
<td>AMV-RT (15 U/µL)</td>
<td>0.5</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Water</td>
<td>2.5</td>
<td>3.5</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Total Volume</td>
<td>33</td>
<td>44</td>
<td>55</td>
<td>66</td>
<td>77</td>
</tr>
</tbody>
</table>

* After the reaction has finished, precipitate the extension products by adding 30 µL 3M NaOAc, pH 4.9 and 260 µL H₂O. Transfer to a 1.5 mL tube and fill with ethanol. After precipitating and drying the pellet, re-dissolve in 10 µL of formamide loading buffer and run on an appropriate percentage PAGE gel.
Reference:
In Vitro T7 Transcriptions

Through in vitro transcription a DNA template can efficiently be transcribed into RNA. This cell-free transcription system utilizes purified T7 RNA polymerase and a DNA template that includes a T7 promoter. In vitro transcriptions allow for the production of µg to mg amounts of RNA in under an hour. This RNA can then be used for research which would otherwise require extensive extraction and purification of RNAs produced inside cells. This type of transcription is particularly useful for RNAs that can re-fold efficiently through addition of buffer or via heat renaturation. Generating RNA ex vivo also allows for incorporation of non-natural NTPs. For example, RNAs can be made radioactive through the incorporation of α-³²P NTPs, tagged for pull down assays with biotin-UTP or labeled with a thio-modified NTP for retention on a mercury gel.

Cold transcription (for 100µL)                  Hot transcription (for 10µL)
10 µL 10x T7 Buffer                       1 µL 10x T7 Buffer
10 µL 10x Template (from PCR)             1 µL 10x Template (from PCR)
5 µL 20x DTT (100mM)                    0.5 µL 20x DTT (100mM)
10 µL 10x NTPs (25mM)                    1 µL 10x NTPs (25mM)
63 µL H₂O                              5 µL α[³²P] GTP (also for 20µL)
0.5 µL Pyrophosphatase                  0.75 µL H₂O
1.5 µL T7 Polymerase                    0.25 µL Pyrophosphatase
100 µL Total                            0.5 µL T7 Polymerase
                                          10 µL Total

Transcription of Biotin-UTP RNAs (for 100 µL)
10 µL 10x T7 Buffer
10 µL 10x Template (from PCR)
5 µL 100mM DTT
5 µL 100mM ATP
5 µL 100mM GTP
5 µL 100mM CTP
3 µL 100mM UTP
5 µL 100mM Bio-UTP
49.5 µL H₂O
1 µL Pyrophosphatase
1.5 µL T7 RNA Polymerase
Total 100 µL
Notes: The 10x PCR template is a 10x concentrated 100 µL reaction dissolved in 10 µL, 10mM Tris·HCl pH 8.0. Incubate the transcription for 1h at 37°C. For cold transcriptions, phenol, chloroform, chloroform extract the transcription and ethanol precipitate the product. After spinning down the ethanol precipitation, re-dissolve the transcription products in formamide loading buffer, heat two minutes at 70°C and purify on a PAGE gel. For hot transcriptions, add two volumes of formamide loading buffer, heat two minutes at 70°C and purify on a PAGE gel.

10x T7 Transcription Buffer
400 mM Tris pH 7.9 @20°
25 mM Spermidine
260 mM MgCl₂
0.1% Triton x-100

Troubleshooting
Low transcription product:
- If transcriptions are inefficient the reaction can be incubated longer than one hour.
- PCR template can be increased to increase RNA yield, but take care not to increase the volume of PCR product in 10mM Tris·HCl pH 8.0 to much more than a quarter of the reaction to avoid changing the buffer conditions.
- Ensure that the PCR product includes a T7 promoter or variant that will serve as T7 RNA polymerase binding site and transcription will start immediately downstream.

Ribozyme cleaves during transcription:
- If the RNA transcript is prone to self cleaving, transcriptions can be carried out at 30°C and/or shortened to 30 minutes.
- An alternative buffer can be used with lower magnesium concentration, see below for 10x mix. This buffer was found to be particularly useful for the 3'-trans ribozyme transcriptions, but less necessary for 5'-trans ribozymes.

10x low Magnesium transcription buffer
400 mM Tris/HCl pH 7.5
1.25 mM Spermidine
3.75 mM MgCl₂
Reference:
5’-Radiolabeling of RNAs

Radioactively labeling RNAs can be invaluable for monitoring the progress of reactions in vitro. In trans-splcing assays, labeling the substrate with $^{32}$P allows for monitoring the disappearance of the substrate, formation of the product and monitoring major side products or intermediates. While RNAs can be either 5’-end, 3’-end or internally labeled, 5’-labeling is preferred for 5’-trans splicing reactions in which the 3’-end of the substrate is replaced. By only labeling the 5’-end of the substrate, the gel runs of the reaction products are cleaner as the 3’-end cleaved off during the reaction does not show up on the scan of the phosphorimager screen. Additionally, quantification of the percentage of product formed during the reaction is more straightforward as there is only one label per 5’-end of the substrate, where as quantification of products with internally labeling substrates will depend on the number of radiolabeled nucleotides in the fragment.

To efficiently kinase a radioactive nucleotide to the 5’-end of an in vitro transcribed RNA, the RNA must first be dephosphorylated. Some PNKs offer exchange buffers which will exchange a cold γ-phosphate from the transcript with the γ-phosphate from a γ-$^{32}$P ATP, but this reaction is less efficient than the forward reaction to a dephosphorylated 5’-end.

**Dephosphorylation of transcribed RNAs (20µL)**

- x µL RNA to give 100pmol
- 2 µL 10x Antarctic Phosphatase buffer
- 16-x µL H$_2$O
- 2 µL Antarctic Phosphatase
- 20 µL Total

- Incubate the reaction 1h at 37°C to allow for efficient dephosphorylation. Phenol, chloroform, chloroform extract and ethanol precipitate the reaction products and re-dissolve in 20 µL water.
5'-radiolabeling dephosphorylated RNAs

- 4 µL Dephosphorylated RNA (20pmol)
- 2 µL 10x T4 PNK buffer
- 2 µL γ[^32P] ATP
- 11 µL H₂O
- 1 µL T4 Polynucleotide Kinase
- 20 µL Total

• Incubate the reaction 30 minutes at 37°C, then stop with 2 volumes FA loading buffer and purify the reaction product on the a PAGE gel.

Troubleshooting

Low yield of labeled RNA:

• RNAs that form stable structures at their 5'-ends do not efficiently dephosphorylate or kinase. If labeling is very inefficient, another substrate which is shorter could be attempted.

• The template could be annealed to a short oligo which is perfectly complementary to nucleotides near the 5'-end that should help unfold the RNA and aid both dephosphorylation and labeling.

• Double check the concentration of the RNA, buffer and age of the enzyme.

Reference:

**In Vitro Trans-Splicing Reactions**

Group I introns are capable of splicing *in vitro* without the need of proteins or other cellular factors. In an *in vitro* reaction, a radiolabeled substrate is incubated in splicing buffer with a *trans*-splicing ribozyme. Some ribozymes benefit from or require a pre-incubation to allow refolding into the active structure. After renaturation, the ribozyme and substrate are mixed and GTP is included to start the reaction. In the Tetrahymena thermophila ribozyme, the exogenous GTP will bind as a base triple in the active site, to the G264-C311 base pair in the P7 duplex using the Watson-Crick face. This GTP will initiate the first transesterification reaction, cleaving the substrate at the splice site. The last nucleotide of the intron, also termed the ωG, will then form a base triple in the same spot that the exogenous GTP bound to initiate the reaction. This will help to align the 5'-end of the substrate with the 3'-exon of the ribozyme. The 3'-hydroxyl of the 5'-portion of the substrate will then initiate the second transesterification reaction, ligating the substrate and 3'-exon and repairing the mRNA. The products of the *in vitro* reaction can then be run on a PAGE gel and unreacted substrate, product and intermediates can be quantified to follow the reaction kinetics. To ensure that substrate and product can be accurately quantified, the substrate either should be shortened, or the ribozyme should transfer a 3'-exon shorter than the full length required for mRNA repair.

**5x *in vitro* splicing buffer (optimized for 5'-trans T. th ribozyme)**

<table>
<thead>
<tr>
<th>Volume</th>
<th>Concentration</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>125 µL</td>
<td>(1M) MOPS</td>
<td>pH 7.0</td>
</tr>
<tr>
<td>2.5 µL</td>
<td>(1M MgCl&lt;sub&gt;2&lt;/sub&gt;)</td>
<td></td>
</tr>
<tr>
<td>90 µL</td>
<td>(3.75M) KCl</td>
<td></td>
</tr>
<tr>
<td>50 µL</td>
<td>(100mM) Spermidine</td>
<td></td>
</tr>
<tr>
<td>232.5 µL</td>
<td>H&lt;sub&gt;2&lt;/sub&gt;O</td>
<td></td>
</tr>
<tr>
<td>500 µL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** do not incorporate the GTP into the 5x buffer or the ribozyme may cleave the exon during the pre-incubation before the substrate is added. MOPS can be pH adjusted with NaOH or KOH. KOH may be preferred because it will not introduce a new ion into the reaction, but NaOH was used for the *in vitro* reactions in the internal randomization study.
Ribozyme pre-incubation
2 µL 5x buffer
4 µL 250nM gel purified ribozyme
4 µL H2O
10 µL

- Incubate the ribozyme mixed with buffer for 10 minutes at 37°C.

Substrate mix (for one reaction, scale up as needed)
2 µL 5x buffer
2 µL 200uM GTP
1 µL *Substrate (5,000 cpm/ µL)
5 µL H2O
10 µL

- Warm the premix to 37°C, then combine with the pre-folded ribozyme to start the reaction. Timepoints can be taken at any interval after initiating the reaction, by pipetting a 4 µL sample of the reaction with an equal volume formamide loading buffer. The time points can then be run on a PAGE gel, exposed overnight to a phosphorimager screen and scanned in the next morning for quantification.

Troubleshooting
Low product yield:
- Ensure the ribozyme concentration is correct.
- Sample at longer time points.
- Consider altering the EGS sequence or using a shortened substrate if possible.
- Optimize the buffer concentration, refolding or reaction conditions.

Reference:
Cloning Protocol

Cloning of DNA sequences into vector plasmids is a method to generate an easily replicatable stock of a desired sequence. These plasmids can then be used as templates for PCR, runoff transcription, or check the sequence of the DNA insert ligated into the plasmid. In general, to clone a DNA sequence, a DNA sample is digested with restriction enzymes to create unpaired 5' and 3'-end overhangs. These overhangs, generally of 4 bases each, can then anneal to a plasmid vector that has been treated with the same restriction enzymes, leaving a complementary overhang. The enzyme T4 DNA ligase can then connect the insert fragment to the plasmid vector and the re-circularized plasmid can be transformed and grown in *E. coli*. After growth, the plasmid can be purified from the cells and used for downstream applications.

**Digestion of the insert and plasmid vector**

15 µL DNA insert or plasmid (1 µg/µL)
5 µL 10x restriction digest buffer
25 µL H₂O
2.5 µL restriction enzyme 1
2.5 µL restriction enzyme 2
50 µL Total

- Incubate at 37°C between 8 and 24 hrs, most of the inserts will be cut after 1hr, but greater efficiency is desired as plasmids that do not get cut will transform efficiently and those that are cut only once can lead to multiple inserts or self re-ligation. The enzymes should retain about 90% efficiency up to about 16 hours. Note: Star activity (cutting/digesting at non-specific sites) will increase as the percentage of glycerol increases over 5%. The enzymes are stored in 50% glycerol so pay attention to reaction volumes and amount of both enzymes added.
- Ethanol precipitate the digested vector, then spin down and dry the pellet. Redissolve the pellet in 22 µL 10mM Tris·HCl pH 8.0 and use 2 µL to measure the A₂₆₀ with the nanodrop UV spectrophotometer.

**Dephosphorylation of the plasmid vector**

2.4 µL 10x Antarctic phosphatase buffer
20 µL dephosphorylated plasmid in 10mM Tris·HCl pH 8.0
2 µL Antarctic phosphatase
24 µL Total
• Mix the reaction and incubate at 37°C for at least 1 hour. Specific time requirements may differ depending on the type of overhangs (5' or 3') that are left by the restriction enzyme.
• Following the 1 hour dephosphorylation, heat inactivate the enzyme for 5 minutes at 65°C and cool to room temperature. Using the $A_{260}$ measured before the dephosphorylation, dilute the digested and dephosphorylated plasmid to 50 nM for use in the ligation reaction. Measuring DNA concentration in the dephosphorylation buffer is not reliable, if it needs to be re-measured, ethanol precipitate and dissolve in H$_2$O for measurement.
• Make sure to only dephosphorylate the plasmid and not the insert as this will prevent the ligation from happening.

Ligation with 15 nM insert and 5 nM vector

| X µL insert to make 15 nM |
| Y µL vector to make 5 nM |
| Z µL H$_2$O to total 20 µL |
| 2 µL 10x buffer |
| 1 µL T4 DNA ligase |
| 20 µL |

• Incubate ligation mixtures at 10°C overnight, ligations can be incubated longer if necessary before making the transformation.
• When ligation is complete, if electrocompetent cells are to be used, ethanol precipitate the reaction, wash with 70% ethanol to remove salts and re-dissolve in H$_2$O, or use a Zymogen Clean-N-Concentrate column to purify the reaction. The Zymogen kit will give higher yield and remove salts better than the ethanol precipitation. If heat shock competent cells are used the ligation can be used directly for the transformation.

Transformation
• Transform the ligation using either heat shock competent or electrocompetent cells. Recover and plate in two dilutions on LB or SOC plates with the appropriate antibiotic.
• Incubate at 37°C overnight.
• After 16-20 hours of growth the clones can be screened with colony PCR, washed from the plates to perform a selection or preserved in the cold room and will remain viable for several weeks.
Troubleshooting
Low insert ratio with low colony numbers:
- In general, this indicates the digestion, at least of the vector, was efficient, but ligation is lacking.
- Check the concentration of the plasmid and insert.
- Check that compatible restriction enzymes were used on the plasmid and insert.
- Double check the sequence of the plasmid and insert to make sure there is not more than one site the restriction enzymes can digest.
- Try a new tube of T4 DNA ligase as a method of testing the efficiency of the enzyme.

Low insert ration with high colony numbers:
- In general, this indicates the digestion and or the dephosphorylation of the vector was inefficient.
- Check that the digestion enzymes are working efficiently by digesting 1 µg of plasmid for 1 hour and running the product on a 1% agarose gel. If the enzymes are working efficiently, approximately 90% of the plasmid should be linearized, and run slower on the gel after the 1 hour digestion.
- Transform a portion of the digested vector and a portion of a vector ligation that does not include an insert. If the ligation of the vector gives many more colonies than the digestion only transformation this will indicate the dephosphorylation is not efficient enough.
- Check that the “no insert” clones have the expected colony PCR product length. If they have a different size insert cloned in, gel purification of the insert may be necessary.

Low colony yield:
- Use transformation competent cells with a higher efficiency.
- Concentrate the ligation or make multiple ligations and concentrate them before transformation.
- Ensure that the cells are recovered 1 hour after transformation in medium that does not have any antibiotic. If plating on kanamycin or chloramphenicol plates the cells seem to need more time after transformation to have a good survival rate on plates. If ampicillin resistance is carried on the plasmid, cells can be plated without the recovery.

Reference:
Buffers for Gels

5x TBE
54 g Tris Base
27.5 g Boric Acid
3.72 g Na₂EDTA (372 g/mol)
fill up to 1L with MilliQ-water, dissolve by stirring

50x TAE for Agarose gels
242 g Tris base
57.1 mL HOAc
50 mmol EDTA
fill up to 1L with MilliQ-water.

Sample buffer for urea PAGE
1 mL 10xTBE (Sequagel)
6.06 g urea
100uL 100x BPB/XC (each 30mg / 10mL)
fill up to 10mL with MilliQ-water, dissolve by inverting

6x Sample buffer for Agarose gels
30% (v/v) glycerol
0.25% (w/v) bromophenol blue
0.25% (w/v) Xylene Cyanol
2x TAE

7M urea 1xTBE 8% PAGE gel
size: 30cm x 40cm x 1.5mm: 220 mL gel solution made
22mL 10x TBE (Sequagel)
70.4mL Sequagel concentrate (25% AA/BAA 7M urea)
127.6mL Sequagel diluent (7M urea)
stir well
1100 uL 10% APS
220 uL TEMED
stir shortly, pour.
gel polymerized well, was used after 2 h
Phenol/Chloroform Extraction and Ethanol Precipitation

Phenol/chloroform extractions are a method for easy removal of enzymes from reactions such as PCR, T7 transcription or restriction digests. Phenol is an organic solvent, with higher density than water, in which proteins will preferentially partition. Nucleotides and nucleic acids will remain in the aqueous layer, allowing for the separation from proteins and for buffer exchange. Performing liquid-liquid extraction with an equal volume of phenol and aqueous reaction will remove the proteins. Removing the aqueous layer into chloroform will help to remove the phenol which would otherwise affect downstream reactions. Following the extractions, and ethanol precipitation is performed to further remove residual phenol and concentrate the nucleic acids in the aqueous layer. In the precipitation, the pH of the aqueous layer is lowered with sodium acetate and at least 3 volumes of ethanol are added to reduce the solubility of nucleic acids in the solution. The precipitation is then chilled to -20°C which will further reduce the solubility and cause nucleotides and nucleic acids to form a precipitate that can be pelleted.

**Phenol/Chloroform Extraction:**
- Mix an equal volume of phenol and aqueous reaction in a 1.5 mL tube and vortex for 30 seconds. If the aqueous reaction is less than 200 µL, add H2O. If the aqueous reaction is larger than 500 µL, separate into multiple tubes, or use tubes larger than 1.5 mL that can be centrifuged at high speeds.
- Centrifuge the 1.5 mL for 2 minutes at maximum speed.
- Remove the aqueous layer (top) into a new tube with an equal volume of chloroform.
- Mix by vortexing for 30 seconds and spin at maximum speed.
- Repeat the chloroform extraction, removing the second aqueous layer into an empty tube. If there is more than 300 µL, separate into multiple tubes.

**Ethanol Precipitation:**
- Add 1/9th volume of the aqueous reaction of 3M NaOAc, pH 4.9 into the 1.5 mL tube.
- Fill the 1.5 mL tube with 100% ethanol and vortex the sample to mix.
- Cool the precipitation for 20 minutes at -20°C to reduce solubility of the nucleic acids and nucleotides.
• Spin the tubes in a centrifuge cooled to 0°C at maximum speed for at least 20 minutes. The pellets will form on the top outside face of the tube, so make sure to orient them with the lid hinge at the top of the centrifuge.
• Pour off the supernatant, spin the tube shortly again and pipette off the remaining ethanol.
• If it is important to remove salts from the precipitation, wash the pellet with 100 μL ice cold 70% ethanol. Vortex, incubate on ice for 5 minutes, vortex again and spin the pellet for 10 minutes at maximum speed. Pipette off the 70% ethanol.
• Dry the pellet by laying the tube open for 5-20 minutes, or using the speed vacuum for about 1 minute.
• The pellet can now be re-dissolved in any desired buffer.

Notes: Phenol/chloroform extraction is not necessary before any ethanol precipitation, only when it is necessary to remove proteins from a solution. Nucleotides and primers will pellet with DNAs or RNAs in the ethanol precipitation. If these need to be removed, a PCR clean up kit or size exclusion buffer exchange column can be used instead.

Reference:
PCR and Gel Extraction Kit Protocol

PCR products can be isolated and separated from reaction components using a simple column purification procedure. DNA will adsorb to silica gel columns under high salt conditions. Once the DNA from the PCR reaction is bound, the column can be washed with high salt and ethanol buffers to remove primers, enzymes and unincorporated dNTPs. The PCR product can then be eluted in buffer and used for downstream reactions. If there may be more than one reaction product in the PCR (or when purifying DNA for subcloning), the PCR can first be gel purified on an agarose gel, cut out and then purified on a silica column.

**PCR Extraction with 5 Prime PCR extract**

1. Place the PCRExtract mini column in a 2 mL collection tube and add 500 µL buffer BL to equilibrate the column, and centrifuge the column for 1 minute.
2. Add 5 volumes of buffer PD to 1 volume of the PCR reaction (these columns are designed for 50 µL reactions, and will run out of buffer PD if processing more than 100 µL routinely.
3. Apply the sample to the PCRExtract mini column, incubate 2 minutes and centrifuge for 1 minute.
4. Discard flow-through and return the column to the collection tube.
5. Add 700 µL buffer PW to the column and centrifuge for 1 minute.
6. Discard the flow-through and return the column to the collection tube.
7. Add 500 µL buffer PW to the column and centrifuge for 1 minute.
8. Discard the flow-through, return the column to the collection tube and spin for an additional 2 minutes to remove residual wash buffer.
9. Remove the column and allow it to air dry with the cap open for several minutes to ensure any residual ethanol is removed.
10. To elute the DNA, place the column in a clean 1.5 mL tube, and 50 µL buffer PEB to the center of the membrane. Incubate 2 minutes and centrifuge for 1 minute. The eluate can be pipette back onto the column and spun again to increase the yield.
11. The eluted DNA can be used immediately or stored at -20°C.

**Mini Prep Procedure: Promega PureYield A1220, A1221 and A1222**

1. Purify the reaction on a 1% agarose gel and cut out the gel slice.
2. Place the gel slice into a pre-weighed, empty 1.5 mL tube, and obtain the weight of the gel slice.
3. Add 3 volumes buffer PS to 1 volume of gel (eg. 300 µL buffer PS to 100 mg of gel).

4. Heat the gel slice mixed with PS buffer to 50°C for about 10 minutes or until melted. Mix the tube during heating, and bring back to room temperature (not cooler) before loading onto the column.

5. Place the PCRExtract mini column in a 2 mL collection tube and add 500 µL buffer BL to equilibrate the column, and centrifuge the column for 1 minute.

6. Apply the sample to the PCRExtract mini column, incubate for 2 minutes and centrifuge for 1 minute.

7. Discard flow-through and return the column to the collection tube.

8. Add 700 µL buffer PW to the column and centrifuge for 1 minute.

9. Discard the flow-through and return the column to the collection tube.

10. Add 500 µL buffer PW to the column and centrifuge for 1 minute.

11. Discard the flow-through, return the column to the collection tube and spin for an additional 2 minutes to remove residual wash buffer.

12. Remove the column and allow it to air dry with the cap open for several minutes to ensure any residual ethanol is removed.

13. To elute the DNA, place the column in a clean 1.5 mL tube, and 50 µL buffer PEB to the center of the membrane. Incubate 2 minutes and centrifuge for 1 minute. The eluate can be pipette back onto the column and spun again to increase the yield.

14. The eluted DNA can be used immediately or stored at -20°C.

Reference:
Total RNA Extraction from *E. coli*

Total RNA extraction from *E. coli* in logarithmic growth allows for identifying and quantifying different populations of RNA species inside the cells. Using this procedure RNA can be extracted, free of proteins or DNA and reverse transcribed for sequencing or quantitative Real Time PCR (qRT-PCR). The cells are first grown, harvested and lysed in a buffer that will also remove proteins. The RNA is then bound to an anion exchange silica column and washed with buffers that remove cell debris and DNA contamination, allowing for pure RNA to be eluted. This extraction procedure is particularly useful to identify trans-splicing reaction products and determine the percentage of substrate spliced inside the cell.

Protocol for Qiagen RNeasy® RNA purification using RNAProtect® Bacteria Reagent and optional on-column DNase digest.

Day 1:

The evening before the RNA prep is to be processed, inoculate a 5mL overnight culture for each sample using LB medium and appropriate antibiotic, then incubate at 37°C with shaking.

Day 2:

1. If induction with IPTG is required to monitor expression levels or repair of a substrate by a ribozyme, add 5 µL 1M IPTG to each 5 mL overnight culture and continue shaking 1h at 37°C.
2. During the hour of induction or before processing the samples, prepare enough TE buffer for all samples (30 mM Tris.Cl, 1 mM EDTA, pH 8.0) containing 15 mg/mL lysozyme, each sample will require 200 µL. (Note: If samples will be stored at -20°C after adding RNAprotect Bacteria Reagent, do not make TE buffer until the RNA isolation will be performed).
3. Add 10 µL β-mercaptoethanol per 1mL Buffer RLT and mix. For each prep 700 µL RLT buffer is used. This buffer can be stored up to 1 month.
4. Dilute the overnight cultures 5 or 10 fold with LB and measure the OD<sub>600</sub>. Dilute each sample to approximately OD<sub>600</sub> = 0.100 in 20 mL LB medium with 1 mM IPTG (if monitoring expression or repair) and appropriate antibiotic. Monitor the growth of the culture and harvest when the OD<sub>600</sub> is between 0.400 and 0.600, ideally 0.500. (This should take about 1.5-3 hours depending on the individual growth rates).
5. Once the culture has reached OD<sub>600</sub> = 0.500, harvest 1.5 mL of the culture into a 2 mL tube and spin down 5 minutes at 5,000 x g. Remove 1 mL of
the supernatant, and resuspend the culture in the remaining 500 µL.
(Note: The cultures may not all reach the appropriate density at the same
time, process each culture at the appropriate density and hold after step 8
at -20°C).
6. Add 1 mL RNAprotect Bacteria Reagent to each tube of spun down and
resuspended cells and mix by vortexing for 5 seconds. Incubate at room
temperature 5 minutes further.
7. Centrifuge the samples 10 minutes at 5,000 x g.
8. Decant the supernatant. Remove the residual supernatant by gently
dabbing the inverted tube once onto a paper towel. Do not remove the
residual supernatant by pipetting, the pellet will be a clear film and may be
lost if pipetting is attempted. Do not leave more than 160 µL of
supernatant or the RNA prep will be compromised. (Pellets can be stored
at -20°C for up to two weeks, or at -70°C for up to 4 weeks. For
subsequent RNA purification, thaw pellets at room temp and proceed).
9. Add 10 µL of a 20 mg/mL proteinase K solution per 200 µL TE buffer with
lysozyme added, then add 210 µL of the mixture to each pellet. Carefully
resuspend the pellet by pipetting up and down several times.
10. Mix by vortexing for 10 seconds. Incubate at room temperature for 10
minutes on a rotator or shaker. If there is no rotator available, vortex for at
least 10 seconds every 2 minutes. Note: Since the RNA is stabilized, the
incubation time can be extended and may increase RNA yield.
11. Add 700 µL RLT (with β-mercaptoethanol added) and vortex vigorously. If
particulate material is visible, pellet it by centrifugation and use only the
supernatant in step 12. If centrifugation is necessary, spin 2 mL tubes 2
minutes at maximum speed in a microcentrifuge. Note: Flowthrough from
this and subsequent washing steps should be bleached before disposal.
12. Add 500 µL 100% ethanol to each tube and mix by pipetting. Note: a
precipitate may form after this step. This will not affect the RNeasy
procedure and does not need to be removed.
13. Transfer up to 700 µL lysate, including any precipitate, to an RNeasy Mini
spin column inside the 2mL collection tube supplied. Centrifuge the column
for 15 seconds between 8,000 and 10,000 x g. Discard the flow through
and apply and spin the remaining lysate on the column.
14. Add 350 µL Buffer RW1 to the RNeasy spin column and centrifuge for 15
seconds at a minimum of 8,000 x g.
15. Add 10 µL DNase I (1500 Kunitz diluted into 550 µL RNase-Free water
from Qiagen) to 70 µL Buffer RDD. Mix gently by inversion and centrifuge
to bring all liquid to the bottom of the tube.
16. Add 80 µL DNase I mix directly to the RNeasy spin column membrane and incubate at room temperature for 15 minutes.
17. Add 350 µL Buffer RW1 to the RNeasy spin column, wait 5 minutes, and then centrifuge for 15 seconds at a minimum of 8,000 x g.
18. Place the RNeasy spin column in a new 2 mL collection tube. Add 500 µL Buffer RPE to the column and centrifuge 15 seconds between 8,000 and 10,000 x g to wash the spin column membrane. Discard the flow through and re-use the collection tube in step 19.
19. Add another 500 µL Buffer RPE to the column and centrifuge for 2 minutes between 8,000 and 10,000 x g to wash the membrane and ensure there is no ethanol carryover. To further ensure no ethanol is carried over, remove the supernatant and spin again, or spin again in a new tube.
20. Place the spin column in a new 1.5 mL collection tube. Add 40 µL RNase-free water directly to the membrane (without touching it). Centrifuge for 1 minute between 8,000 and 10,000 x g.
21. If the expected RNA yield is >30 µg, repeat step 20 using another 40 µL RNase-free water, or eluate from step 20. Note: If using the eluate from step 20, the RNA yield will be 15-30% lower than that obtained using a second volume of water, but the final RNA concentration will be higher.
22. RNA can immediately be used, or stored at -20°C.
Mini and Midi Plasmid Prep Protocols

Transforming plasmids into *E. coli* allows for quick and efficient replication of the plasmid with high fidelity. Growing a single clone containing a transformed plasmid in liquid culture overnight will multiply the plasmid, and provide an easy medium to extract the plasmid. Mini, midi or maxi plasmid prep kits can be utilized to isolate the plasmid from the cells. The plasmid can then be used for sequencing, cloning or as a PCR template. Plasmid prep kits have three main steps. First the cells are lysed, breaking the bacterial cell wall and exposing the contents. After cell lysis the solution will be viscous and it is important not to vortex to avoid shearing the genomic DNA which will then purify with the plasmid. After lysis, the reaction is neutralized. In this step, the proteins and other cell debris will precipitate making for easy removal by centrifugation. Lastly, the plasmid containing solution is loaded onto a column. The DNA will bind to the column and can be further purified by washing with buffers high in alcohols such as ethanol and isopropanol. This will keep the plasmid insoluble in the solution and stuck on the column. After washing the plasmid clean plasmid can be eluted in H₂O. In general, mini preps give 10-20 µg of plasmid and midi preps can give 300 µg or more. The plasmid yield will depend on the density and volume of the overnight culture processed, copy number of the plasmid and the kit used.

Mini Prep Procedure: Promega Wizard A1270, A1330, and A1460

1. Grow a 5 mL culture of the colony of interest overnight in LB medium with the appropriate antibiotic.
2. The next morning, spin down 600 µL to 5 mL of the overnight culture at a minimum of 5,000 × g and up to maximum speed. The density of the culture and speed of the centrifuge will determine the volume and time of centrifugation. In general, a spin at maximum speed for 30 seconds will be the most efficient.
3. Decant and pipette off the supernatant leaving the pellet untouched.
4. Re-suspend the pellet in 250 µL resuspension buffer. Vortexing at this step is ok, but not after lysis buffer is added.
5. Add 250 µL of lysis buffer, mix by inversion. Incubate at room temperature until lysate clearing is observed for up to 5 minutes.
6. Add 350 µL neutralization solution, mix by snipping the tube several times until a cloudy precipitate forms.
7. Spin the solution at maximum speed (14,000 × g) for 10 minutes.
8. Pipette or decant the supernatant on to a minicolumn in collection tube provided in the kit and centrifuge for 1 minute at maximum speed.
9. Discard the flowthrough and replace the column into the collection tube.
10. Add 700 µL wash solution (with ethanol added) and centrifuge at maximum speed for 1 minutes.
11. Add 200 µL was solution and centrifuge another minute.
12. Air dry the tube for 10 minutes to ensure removal of the ethanol.
13. Place the mini column in a new, labeled 1.5mL tube. Add 50 µL H2O on to the center of the column. Incubate for 1 minute, then spin at maximum speed for at 1 minute.
14. The eluted DNA can be used directly for sequencing, transformations, or stored at -20°C until needed.

Mini Prep Procedure: Promega PureYield A1220, A1221 and A1222

1. Grow a 5 mL culture of the colony of interest overnight in LB medium with the appropriate antibiotic.
2. The next morning, spin down 600 µL to 3mL of the overnight culture at a minimum of 5,000 x g and up to maximum speed. The density of the culture and speed of the centrifuge will determine the volume and time of centrifugation. In general, a spin at maximum speed for 30 seconds will be the most efficient. (If using 600 µL, pipette the cell culture into a 1.5 mL tube and skip to step 5).
3. Decant and pipette off the supernatant leaving the pellet untouched.
4. Re-suspend the pellet in 600 µL water. Vortexing at this step is ok, but not after lysis buffer is added.
5. Add 100 µL of lysis buffer, mix by inversion. Incubate at room temperature for up to 2 minutes. The solution should appear mainly clear with a dark blue tint.
6. Add 350 µL ice cold neutralization solution, mix by inverting the tube several times until the solution changes completely from blue to yellow and a precipitate forms.
7. Spin the solution at maximum speed (14,000 x g) for 3 minutes.
8. Pipette or decant the supernatant on to a minicolumn in collection tube provided in the kit and centrifuge for 15 seconds at maximum speed.
9. Discard the flowthrough and replace the column into the collection tube.
10. Add 200 µL endotoxin removal wash, centrifuge 15 seconds at maximum speed.
11. Add 400 µL column wash solution (with ethanol added) and centrifuge at maximum speed for 30 seconds.
12. Place the mini column in a new, labeled 1.5mL tube. Add 30 µL H₂O on to the center of the column. Incubate for 1 minute, then spin at maximum speed for at least 15 seconds.
13. The eluted DNA can be used directly for sequencing, transformations, or stored at -20°C until needed.

Mini Prep procedure: 5 Prime FastPlasmid mini kit
1. Grow a 5 mL culture of the colony of interest overnight in LB medium with the appropriate antibiotic.
2. The next morning, spin down 1.5 - 3 mL of the overnight culture at maximum speed for 30 seconds.
3. Decant and pipette off the supernatant leaving the pellet untouched.
4. Add 400 µL of ice cold lysis buffer and vortex for 30 seconds or longer if the pellet is not completely resuspended.
5. Incubate 3 minutes for complete lysis.
6. Transfer the lysate to a column in a collection tube and spin the solution at maximum speed for 30-60 seconds.
7. Add 400 µL wash buffer, centrifuge 30-60 seconds at maximum speed.
8. Decant the filtrate from the collection tube and spin again 1 minute at maximum speed to remove all buffer from the column.
9. Place the column in a new tube and add 50 µL H₂O to the center of the column. Spin at maximum speed for 30-60 seconds.
10. The eluted DNA can be used directly for sequencing, transformations, or stored at -20°C until needed.
Pull-down of RNA-binding *E.coli* proteins

Proteins that interact with RNA inside a cell can be identified through the use of a pull-down assay. Biotin labeled RNAs are first transcribed with a small portion of one nucleotide included as a biotin labeled NTP. (As an alternative, the RNA can be transcribed with a streptavidin aptamer sequence tag instead of biotin labeling.) These RNAs are then incubated with *E. coli* cell extracts, giving the RNA and cellular proteins the opportunity to bind as they would *in vivo*. The RNA and cell extract are then mixed with magnetic beads coated with streptavidin. The tight binding interaction of the biotin and streptavidin will retain the RNA with the bead fraction. The beads are washed several times with a buffer, and isolated from the mix using a magnet. The wash steps will remove the majority of the non-interacting or weakly interacting proteins. If washing steps are done with the appropriate stringency, only strongly interacting proteins will remain bound to the RNA attached to the beads. Heating the bead fraction in SDS loading buffer will release the proteins, that can then be run on an SDS PAGE and stained. If a clear candidate of the interacting protein is identified, the band can be cut out and identified by mass spectrometry.

**Grow cells, wash and freeze aliquots:**
1 - Grow a 150 mL culture (or larger) of *E. coli* in LB medium until the OD$_{600}$ is approximately 0.500.
2 - Spin down the cells (4°C / 5,000 rpm / 5 min).
3 - Remove all supernatant (first pour off, then pipet off).
4 - Resuspend in 50 mL ice-cold 0.2x PBS, centrifuge as above
5 - Resuspend in a total of 1 mL 0.2x PBS (or as appropriate for larger cultures)
6 - Aliquot into 40 µL aliqots, freeze in liquid N2, store at -80°C

**Lyse the cells:**
1 - Suspend cells by adding 10 µL ice-cold solution of lysozyme 5 mg/mL (final 1 mg/mL) / 0.2x PBS / 1 mM EDTA / 0.1% TritonX100 / DNase (RNase free; 20-50 µg/mL) and pipetting. DNase is not always required and can be avoided if the cell debris and genomic DNA are removed in the spinning step after lysis.
2 - incubate 10min on ice
3 - Make 5 freeze-thaw cycles (never warm up)
4 - Add 100 µL ice-cold 0.2x PBS / 1mM EDTA, mix by pipetting
5 - Centrifuge 10 min / 14,000 rpm / 0°C
6 - Pipet supernatant in new tube on ice (try to remove the pellet in one piece before pipetting out the supernatant, it will be gooey). Use this supernatant for immediate capture.

**Incubate biotin-RNA with cell lysate**
1 - To a small volume of biotinylated RNA in 0.2xPBS, add cell lysis supernatant
2 - vortex shortly, and let stand on ice for 10 minutes

**Wash magnetic beads**
1 - Suspend beads in one tube by very careful pipetting with 1mL pipet. For all further steps, do not pipet: suspend by careful snipping
2 - wash magnetic beads, 50 µL for each experiment, (binding capacity ~50 pmol) 3x in 100 µL 0.5xPBS / 0.01% TritonX100.
3 - Remove the last supernatant only shortly before the RNA/cell lysate is added, to prevent drying of the beads (which leads to aggregation & inactivation).

**Capture, wash, and elute protein on magnetic beads**
After removing the third wash from the magnetic beads,
1 - take the tube out of the magnetic stand,
2 - add the RNA/cell lysate
3 - rotate 10min / 4°C
4 - capture for 30 sec on magnetic stand
5 - remove supernatant. Keep the supernatant from each wash step.
6 - take the tube out of the magnetic stand
7 - add 100uL 0.5x PBS / 0.1% TritonX100, mix by snipping
8 - go back to (4) and repeat the PBS/TritonX100 wash once
9 - after removing the wash supernatant, add 10 µL 0.2x PBS to the magnetic beads, add 2 µL 4xLDS and 1.3 µL 500 mM DTT and mix by pipetting (no need to be gentle).
10 - Transfer supernatant to a 0.5 mL tube, heat 2'/80°C, then capture beads with magnetic stand to remove supernatant for loading.
11 - For remaining samples mix 10 µL each wash with 2 µL 4xLDS and 1.3 µL 500 mM DTT.

**Run samples on 2 SDS PAGE gels**
Stain one with Coomassie
If necessary, stain the other with silver.

**PBS, 1X**
0.2g KCl
8.0g NaCl
0.2g KH2PO4
1.15g Na2HPO4
Add components one at a time to 900 mL of room temperature deionized water and stir until completely dissolved. Adjust the pH to 7.4 using 1N HCl or 1N NaOH, if necessary. Bring the final volume to 1 liter. If stored for long periods, filter the solution through a 0.45 μm filter and store in a tightly capped bottle.

References:

Growth and Preparation of Electrocompetent \textit{E. coli}

Electrocompetent \textit{E. coli} are useful for efficiently transforming larger plasmid libraries than can be transformed via heat shock. While heat shock cells are made competent chemically, electrocompetent \textit{E. coli} are made competent by removing the salts from the growth medium and storing in a 10\% glycerol solution. To prepare electrocompetent \textit{E. coli}, cells are grown from a cell stock and harvested in log phase at an OD$_{600}$ of approximately 0.600. This will ensure the culture is dense, but a majority of the cells have not transitioned to the stationary phase. The cells are then spun down, washed and stored at -80$^\circ$C until use. Because these cells generally do not have any antibiotic resistance, great care should be taken to use sterile equipment to avoid contamination of the cultures. During the electroporation, the cells are thawed, mixed with up to 1/10th volume of plasmid in water, and then 1.8 kilovolts are applied across an electroporation cuvette with 1 mm gap width, containing the cell and plasmid mix. This application of electricity will create holes in the \textit{E. coli} cell wall and allow the plasmids to enter the cells.

Protocol for preparation of electrocompetent \textit{E. coli}

- Inoculate 1-2L of sterile LB medium with 1-2mL thawed cell stock. If these cells already carry a plasmid with an antibiotic resistance gene, include that antibiotic in the culture.
- Monitor the OD$_{600}$ of the culture. The growth should take 3-5 hours depending on the density of the cell stock. Once the culture has reached an OD$_{600}$ near 0.600 (make sure not to get above 0.700) place the cells on ice for 20-30 minutes with shaking if possible. From this point on, keep the cells on ice at all times. Warming the cells up will result in reduced competency.
- Cool 1L autoclaved, sterile water and 500mL autoclaved, 10\% glycerol on ice for later use.
- Transfer the cells from the culture flasks into sterile (autoclaved or bleached and dried) centrifuge buckets. Centrifuge the cells at 5,000 x g for 10 minutes in a centrifuge cooled to 0-4$^\circ$C. Pour and then pipette off all the supernatant.
- Resuspend the cell pellet in 500mL ice cold water and centrifuge again 10 minutes at 5,000 x g. If two liters of culture were grown, combine both pellets before spinning.
- Pour and pipette off all of the supernatant, and repeat the wash and pelleting with water.
• Pour and pipette off all of the supernatant, and resuspend the cell pellet in almost all of the ice cold 10% glycerol.
• Pellet the cells 10 minutes at 5,000 x g and pour off but do not pipette off the supernatant.
• Resuspend the cell pellet in the remaining supernatant, and dilute about 100 fold in 10% glycerol to measure the OD$_{600}$. If the undiluted OD$_{600}$ is between 80-150, the cell density is in a good range. If they are more dense, dilute to this range with ice cold 10% glycerol, or if the cells are too dilute, spin and remove some of the supernatant before resuspending.
• After checking the OD$_{600}$ is in a good range, aliquot 40µL cells in 10% glycerol into 1.5mL tubes and freeze in a dry ice/ethanol bath and store at -80°C.

Quality control of newly made electrocompetent *E. coli*

Checking for contamination:
• Electrocompetent cell preps that do not already carry a plasmid giving antibiotic resistance are particularly susceptible to contamination by another bacterial species. Therefore, the newly made cells must be checked before extensive use to ensure they are not contaminated.
  o Plate the cells on LB plates with Amicillin, Chlorampenicol and if available Kanamycin (singly). If any cells are growing these cells are contaminated with another cell type and will cause problems with future transformations.

Calculation of the cell competency:
• It is useful to calculate the competency of the newly made cells to get a good estimate of the amount of plasmid which will need to be transformed if a certain number of colonies are desired, or if single colony resolution must be seen.
  o Take 1 ng of a purified plasmid (ideally 1 ng/µL) and transform into one aliquot of cells as below. Recover the cells 1h in LB medium, shaking at 37°C and plate the equivalent of 1, 10, 100 and 1000 µL of the recovered cells on an LB plate with the appropriate antibiotic.
  o Grow the plates overnight at 37°C and count the colony forming units (cfu) in the morning on a plate with between 200 and 1000 colonies. Multiply the number of colonies found on that plate to get the total number of colonies formed in the entire mL of the 1 ng transformation and then multiply by 1000 to get the number of colonies that would be transformed if 1 µg was used.
Electroporation of plasmids into competent cells.

- Thaw one aliquot of electrocompetent cells on ice per plasmid that needs to be transformed.
- Once thawed, mix 1-4 µL of plasmid DNA dissolved in water into the 40 µL cells by stirring with a pipette tip. It is important that the cells are not pipetted up and down at this point as it may cause shearing forces that break the cells open.
- Pre-cool the cuvettes on ice, then transfer the cell/DNA mix using a 1 mL pipette tip, to the 1 mm space between the cuvette’s electrodes.
- Dry the cuvette and zap the cells in the electroporator at 1.8 kV, 200 ohm, 25 µF. The time constant should be about 4.5 ms. If the time constant is much lower than 3, try drying the cuvette again and re-electroporating.
- Immediately after electroporation, rinse the cells out of the cuvette with 1mL SOC medium. This medium should not include antibiotics at this point.
- Incubate the cells in SOC for 1h at 37°C with shaking.
- Spread 100 µL of the cells on an LB plate with appropriate antibiotic.
- Spin the remaining 900 µL for 1 minute at 5,000 x g. Pour off the supernatant, resuspend the cell pellet in the residual supernatant and spread on a second plate. Note: depending on the expected number of colonies, dilutions can also be made to ensure there are single colonies.

Cleaning of electroporation cuvettes:

- Immediately after use, rinse and store cuvettes in an aqueous solution with a low percentage of SDS, 0.1% should be sufficient.
- When there are enough cuvettes to clean, rinse out the SDS solution. It is important to get rid of all of the SDS or a mess will be generated when neutralizing the H₂O₂ with KMnO₄.
- Fill the cuvettes with and incubate in 10% H₂O₂ for 5 minutes.
- Drain the H₂O₂ and rinse with plenty of water.
- Rinse each cuvette with Millipore purified water in the cuvette washer hooked up to the vacuum line.
- Pipette 1 mL 70% ethanol into each cuvette and rinse out with the cuvette washer.
- Lay cuvettes and lids out to dry.
• Neutralize the H$_2$O$_2$ with KMnO$_4$ in a fume hood. This will generate hydrogen and oxygen gas and take a few minutes. The remaining solution can go down the drain. If any residual SDS is in the H$_2$O$_2$ solution, a foam will form that can quickly increase in volume over 10 fold, if this happens, transfer into a larger container and wait.

References:
Growth Curve Protocol

Monitoring the optical density of bacterial cultures over time makes it possible to determine the doubling time of the culture. As the cells grow and divide, the medium will become turbid and absorb light at 600 nm. Healthy *E. coli* cultures, not challenged with antibiotics, will double approximately every 20 minutes. Including one antibiotic in the culture will slow the doubling to about 30 minutes and multiple antibiotics will slow the growth even further. For cultures that require mRNA repair by a ribozyme for antibiotic resistance, monitoring the growth behavior and calculating the doubling time will give an idea of the repair efficiency.

**Overnight culture:**
- Start a 5 mL overnight culture with LB medium and 100 µg/mL ampicillin (for pUC19 based plasmids).
- Inoculate each tube with a sample of the cell stock to be grown.
- Incubate the culture overnight at 37°C and shake at 250 RPM.

**Induction:**
- After approximately 16 hours of overnight growth, induce the culture with 5 µL 1M IPTG to promote ribozyme transcription.
- Incubate with further shaking for 1 hour.
- During this induction, prepare 20 mL LB cultures with 1 mM IPTG and desired concentration of the necessary antibiotic in sterilized 125 mL Erlenmeyer flasks.
- Pre-warm the cultures in the shaker for the remainder of the 1 hour incubation.

**Growth:**
- Dilute a sample of the induced cultures 10 fold and measure the OD$_{600}$.
- Inoculate the 20 mL cultures with the volume of overnight culture required to give an initial OD$_{600}$ of 0.050. Make sure to remove the same volume that is to be added before inoculation to keep the culture at 20 mL.
- Check that the OD$_{600}$ is between 0.040 and 0.060, if not add more culture or dilute with LB containing the same mix of IPTG and antibiotic.
- Monitor the OD$_{600}$ by taking 1 mL samples at 1 hour, and every 30 minutes thereafter through 5 hours or until the OD$_{600}$ is greater than 0.600.
- Plot the numbers and determine the doubling time.
His\textsubscript{6}-Tagged Protein Purification

Many proteins can be artificially expressed at high levels in \textit{E. coli}. Proteins are generally expressed in \textit{E. coli} cells of the strain BL21 (DE3) pLysS. This strain includes a chromosomal copy of T7 polymerase under a lac repressor. The lac repressor will bind to the lac operon, preventing the T7 RNA polymerase from binding to the promoter region preventing transcription. The lac repressor is removed after inducing the production of lactose, which will bind to the lac repressor causing an allosteric change releasing the operon and allowing mRNA transcription. If the protein is also tagged with six histidine residues, on the N- or C-terminus, it can be purified from the cell debris or other proteins using a nickel affinity column. The six histidine residues will bind strongly to Ni\textsuperscript{2+} affinity column. Washing the column with increasing concentrations of imidazole will remove the other proteins, and finally compete off the desired protein.

Expression:
- Grow a 100mL LB culture with appropriate antibiotic at 37°C, 250 RPM overnight.
- Inoculate a 2 liter LB culture with the appropriate antibiotic to OD\textsubscript{600}=0.1 and monitor the growth until OD\textsubscript{600}=0.6 (37°C / shaking / ~ 3 h).
- Induce protein production by adding IPTG to final 1mM
- Shake 3 more hours.
- Cool cultures on an ice bath.
- Pellet cells (5,000 g / 10' / 4°C).
- Either process the cell pellet immediately or store at -80°C.

Purification:
- Thaw and resuspend cell pellets from 2 liters in 40 mL lysis buffer with 5 mM imidazole, 1 mM PMSF and 1mg/mL lysozyme.
- Incubate 30 min on ice.
- Lyse cells by sonification on ice, using Branson sonifier and microtip at highest microtip setting, 4 times for 15 sec.
- Pellet cell debris and unlysed cells by centrifugation (10,000 g / 30' / 4°C)
- Wash 5 mL Ni-NTA agarose slurry (2.5mL agarose) with lysis buffer
- Add supernatant to 5mL Ni-NTA agarose (~35mL total; 60' / 4°C / gentle shaking).
- Fill washed Ni-NTA agarose into empty column
- Wash column with 10 mL lysis buffer /10 mM imidazole
- Wash column with 5 mL lysis buffer /20 mM imidazole
- Elute protein with 20 x 0.5 mL aliquots of lysis buffer /100 mM imidazole. To the fractions, add glycerol to 50% and freeze at -20°C.
• Check protein content by SDS PAGE analysis of fractions.

**Lysis buffer:**
50 mM Na$_4$(PO$_4$)$_y$ pH 8.0 (93.2% Na$_2$HPO$_4$; 6.8% NaH$_2$PO$_4$)
300 mM NaCl
5 mM β-Mercaptoethanol (14.3M)
5% glycerol (v/v)

To make 200mL lysis buffer:
12mL 5M NaCl
18.6mL 0.5M Na$_2$HPO$_4$
1.36mL 0.5M NaH$_2$PO$_4$
70uL bME
10mL glycerol
158mL H$_2$O

**Imidazole stock:**
prepare a 1M stock solution, titrate to pH 8.0 with H$_3$PO$_4$ (takes about 1.5 mL of 86% H$_3$PO$_4$/100mL 1M imidazole).

Reference: