UNIVERSITY OF CALIFORNIA, SAN DIEGO

Design and Evolution of *Trans*-Splicing Group I Intron Ribozymes

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

in

Chemistry

by

Zhaleh N. Amini

Committee In Charge:

Professor Ulrich Müller, Chair
Professor Daniel Donoghue
Professor Jens Lykke-Andersen
Professor Charles Perrin
Professor Navtej Toor

2015

The dissertation of Zhaleh N. Amini is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2015

Table of Contents

Signature Page iii

Table of Contents iv

List of Figures vii

Acknowledgements x

Vita xii

Abstract of the Dissertation xiii

Chapter 1: Introduction 1

1.1 Discovery and *Cis*-Splicing of the Group I Intron Ribozyme 1

1.2 Structure of the Group I Intron 2

1.3 Phylogenetic distribution and Classes of Group I Introns 4

1.4 Evolution of the Group I Intron 6

1.5 Group I Intron *Trans*-Splicing 9

1.6 *Trans*-excision 13

1.7 Goals of the Dissertation 14

1.8 References 16

Chapter 2: Low Selection Pressure Aids the Evolution of Cooperative Ribozyme Mutations in Cells 23

2.1 Abstract 23

2.2 Introduction 24

2.3 Experimental procedures 28

2.4 Results 31
2.5 Discussion 48
2.6 Acknowledgements 55
2.7 Supporting Information 56
2.8 References 59

Chapter 3: Spliceozymes: Ribozymes that Remove Introns from Pre-mRNAs in *Trans* 66

3.1 Abstract 66
3.2 Introduction 66
3.3 Results 69
3.4 Discussion 86
3.5 Materials and methods 88
3.6 Acknowledgements 96
3.7 Supporting Information 96
3.8 References 97

Chapter 4: Decreased Side Product Formation in Evolved Group I Intron Spliceozymes 103

4.1 Abstract 103
4.2 Introduction 103
4.3 Results 107
4.4 Discussion 122
4.5 Materials and methods 125

4.6 Acknowledgements 132
4.7 Supporting Information 133
4.8 References 135
Chapter 5: Future Directions 139
5.1 Introduction 139
5.2 Future Spliceozyme Applications 140
5.3 Future Applications for the Analysis of Evolutionary Parameters 146
5.4 References 148

List of Figures

Figure 1.1 *Cis*-splicing reaction of the *Tetrahymena thermophila* Tth.L 1925 IC1 group I intron ribozyme 2

Figure 1.2 3D model of the *cis*-splicing Tth.L 1925 IC1 group I intron from *Tetrahymena thermophila* (13) 4

Figure 1.3 Cyclical gain and loss model of group I intron evolution proposed by Goddard and Burt (35). 7

Figure 1.4 *Cis*- and *trans*-splicing reaction of the *Tetrahymena thermophila* Tth.L 1925 IC1 group I intron 10

Figure 2.1 *Trans*-splicing ribozyme variant of the *Tetrahymena* group I intron that was used as parent construct for the evolutions 26

Figure 2.2 Evolution of the *trans*-splicing ribozyme under four different conditions..... 35

Figure 2.3 Secondary structure representations of the mutations identified after 14 rounds of evolution in each of the four lines 38

Figure 2.4 Analysis of mutations generated by PCR mutagenesis..... 40

Figure 2.5 Identification of mutations that increased ribozyme activity, among the 11 mutations in clone IV-12-10 42

Figure 2.6 Accumulation of evolutionary intermediates of the M5 ribozyme in rounds 10-12 of the evolution 44

Figure 2.7 Fitness profile of all evolutionary intermediates from the parent ribozyme (M0) to the M5 ribozyme (M5), at three selection pressures 46

Figure 2.8 Accumulation of evolutionary intermediates for the five mutations of the M5 ribozyme 57

Figure 2.9 Values for the growth of *E. coli* cells on LB medium containing chloramphenicol, mediated by all evolutionary intermediates between the parent ribozyme and the M5 ribozyme (M9/236/238/239/241) 59

Figure 3.1 Secondary structure of the spliceozyme, based on the group I intron ribozyme from *Tetrahymena* 69

Figure 3.2 Influence of the spliceozyme 5'-terminus design on the reaction *in vitro* 71

Figure 3.3 Effect of substrate recognition sequences on product formation 74

Figure 3.4 Spliceozyme activity in *E. coli* cells 77

Figure 3.5 Effect of the internal intron sequence on spliceozyme activity in *E. coli* cells 79

Figure 3.6 Effect of the 3'-terminal intron sequence on spliceozyme activity in *E. coli* cells 81

Figure 3.7 Quantitation of CAT enzyme activity, *CAT* RNA level, and plasmid level in *E. coli* cells 83

Figure 3.8 Fraction of 5′-exon side products during the *in vitro* reaction of the spliceozyme 97

Figure 4.1 Secondary structure of spliceozyme used as parent in the evolution 105

Figure 4.2 Evolution of spliceozymes in *E. coli* cells 108

Figure 4.3 Identification of mutations necessary for full activity in the evolved spliceozyme W11 111

Figure 4.4 Activities of evolved spliceozyme mutations in *E. coli* cells..... 112

Figure 4.5 Effect of evolved spliceozyme mutations on the product pattern of *in vitro* splicing reactions 114

Figure 4.6 Effect of evolved mutations in the spliceozyme on the product pattern of *in vitro* splicing reactions 116

Figure 4.7 Correlation of product formation *in vitro* with CAT activity in *E. coli* cells..... 118

Figure 4.8 Identification of side products from the *in vitro* splicing reaction 120

Figure 4.9 Crystallographic structures of group I introns that suggest a model for the function of mutation U271C 124

Figure 4.10 Enrichment of key mutations during evolution..... 133

Figure 4.11 Growth inhibition of *E. coli* via the expression of spliceozyme variants and CAT pre- mRNA..... 134

Figure 4.12 Off-target effects of the spliceozyme on three essential mRNAs in *E. coli* 135

Acknowledgements

I am grateful for several individuals who have helped me grow throughout my graduate career through their encouragement, support and valuable discussions. To my advisor, Professor Ulrich Müller, thank you for your support and guidance throughout my graduate education. I greatly appreciate your incredible mentorship and genuine passion for research.

To each of the lab members with whom I have worked, Greg Dolan, Karen Olson, Janina Moretti, Chengguo Yao, Logan Norrell and Patricia Dewi, as well as members of the Toor, Joseph and Yang labs, thank you for the knowledge and support you have shared. The friendships that have formed are ones that have helped me endure the ups and downs of this journey, and I hope are long lasting.

Finally, to my mother, Nasrin Arzani, father, Behrouz Amini, along with all of my family members and friends, thank you immensely for your incredible care, love and encouraging words throughout the years. I am truly grateful to have every one of you in my life.

Chapter 2, in full, is a reprint of the material as it appears in the Journal of Biological Chemistry, **Amini Z.N.** and Muller U.F. (2013) Low selection pressure aids the evolution of cooperative ribozyme mutations in cells. J Biol Chem. 288(46): 33096-33106. The dissertation author is the first author on this paper.

Chapter 3, in full, is a reprint of the material as it appears in PLoS One, **Amini Z.N.,** Olson K.E. and Muller U.F. (2014) Spliceozymes: ribozymes that remove introns from pre-mRNAs in *trans*. PLoS One. 9(7): e101932. The dissertation author is the first author on this paper.

Chapter 4, in full, is currently being prepared for submission for publication, **Amini Z.N.** and Muller U.F. Decreased side product formation in evolved group I intron spliceozymes. The dissertation author is the first author on this paper.

This work was supported, in part, by the National Institute of Health Molecular Biophysics Training Grant and the Hellman Foundation

Vita

Cornell University

2009 Bachelor of Arts, Chemistry

University of California, San Diego

2011 Master of Science, Chemistry

2015 Doctor of Philosophy, Chemistry

Publications

Amini Z.N. and Muller U.F. (2013) Low selection pressure aids the evolution of cooperative ribozyme mutations in cells. J Biol Chem. 288(46):33096-33106.

Amini Z.N., Olson K.E. and Muller U.F. (2014) Spliceozymes: ribozymes that remove introns from pre-mRNAs in *trans*. PLoS One. 9(7):e101932.

Amini Z.N. and Muller U.F. Decreased side product formation in evolved group I intron spliceozymes. (Manuscript In Preparation).

ABSTRACT OF THE DISSERTATION

Design and Evolution of *Trans*-Splicing Group I Intron Ribozymes

by

Zhaleh N. Amini

Doctor of Philosophy in Chemistry

University of California San Diego, 2015

Professor Ulrich Müller, Chair

Group I introns are catalytic RNAs (ribozymes) capable of catalyzing their self-excision from precursor RNAs through two consecutive transesterification reactions. Although the ribozyme has evolved to perform this *cis*-splicing reaction in nature, man-made modifications to the 5' end of the ribozyme have allowed it to catalyze a *trans*-splicing reaction, in which it is able to replace the 3' portion of a substrate RNA with it's own 3' tail. The *trans*-splicing group I introns used in this thesis were variants of the Tth.L 1925 IC1 group I intron ribozyme found in *Tetrahymena thermophila*. The work contained in this dissertation aims to both utilize *trans*-splicing group I introns to further understand principles of

RNA evolution, as well as develop and optimize a new *trans*-splicing variant of the ribozyme for future use in therapy.

In the first study of this dissertation, a *trans*-splicing group I intron was used as a model system to examine the effect of selection pressure and recombination on evolving populations of RNA in a cellular environment. Four parallel evolutions were completed, two employing a low selection pressure, and two employing a high selection pressure. Ribozyme populations with higher efficiency, measured by cellular growth conferred by the ribozyme, resulted from evolutions performed at a low selection pressure. It was found that this increase in fitness was the result of a set of four mutations acting cooperatively. Fitness profiles of evolutionary intermediates revealed that a low selection pressure can increase the accessibility of evolutionary paths leading to the evolution of cooperative mutations. This finding not only adds to the understanding of natural RNA evolution, but also aids in the design of more efficient evolutions of RNA species.

In the second study of this dissertation, a new *trans*-splicing variant of the group I intron was developed, capable of catalyzing the removal of internal sequences from pre-mRNA and joining the two flanking sequences, thereby generating a functional RNA. This group I intron has been termed the 'spliceozyme' because its action is analogous to that of the spliceosome. The action performed by the spliceozyme give this system the ability to repair certain types of diseases caused by mis-splicing and therefore, the potential to be used therapeutically. To increase the efficiency and therapeutic potential of this

system, the spliceozyme was evolved in *E. coli* cells, challenging it to more efficiently catalyze the removal of internal sequences. The most efficient variant contained a set of mutations resulting in increased product formation and decreased side product formation. This observed effect was seen *in vitro*, suggesting that this effect may increase spliceozyme efficiency in a range cell types. Future work will move this system into mammalian cells and optimize the spliceozyme for use in a mammalian system, thus developing it as a therapeutic tool.

Chapter 1: Introduction

1.1 Discovery and *Cis*-Splicing of the Group I Intron Ribozyme

In 1968 Carl Woese, Leslie Orgel and Francis Crick proposed that RNA could act to both store genetic information as well as catalyze chemical reactions (1–3). It wasn't until 1982 that this hypothesis came to fruition when researchers from the laboratory of Thomas Cech discovered the ability of the group I intron from *Tetrahymena thermophila* to catalyze it's own self-excision from precursor rRNA, thus demonstrating an RNA acting as a catalyst (4). It was discovered than the intron was able to be spliced out of the precursor rRNA *in vitro* with only guanosine as a cofactor (4,5). Around the same time, the laboratory of Sidney Altman made the discovery the RNA portion of *E. coli* RNAse P functioned as a catalyst (6). These findings, showing that the RNA portion of could act as a catalyst, constituted a major breakthrough that paved the way for RNA catalysis as a new field of study.

It is now known that group I introns are able to catalyze their self-excision from precursor RNA and the joining of the two flanking exons by two sequential transesterification reactions (Figure 1.1) (4,7,8). During the first step of this *cis*-splicing reaction, an exogenous guanosine (or GMP, GDP or GTP), termed 'exoG,' binds in the guanosine binding site in the catalytic core of the ribozyme, acting as a nucleophile and attacking the 5' splice site. The guanosine is then bound to the 5' end of the group I intron, leaving behind a 3' OH at the 5' splice

site. After a conformational rearrangement, the exoG exits the active site, allowing the last nucleotide of the intron, termed 'ωG,' to take its place. In this conformation, the 3'OH of the 5' exon is able to act as a nucleophile, attacking the 3' splice site. This results in ligation of the two flanking exons, and dissociation of the group I intron from the mRNA (4,7–9).

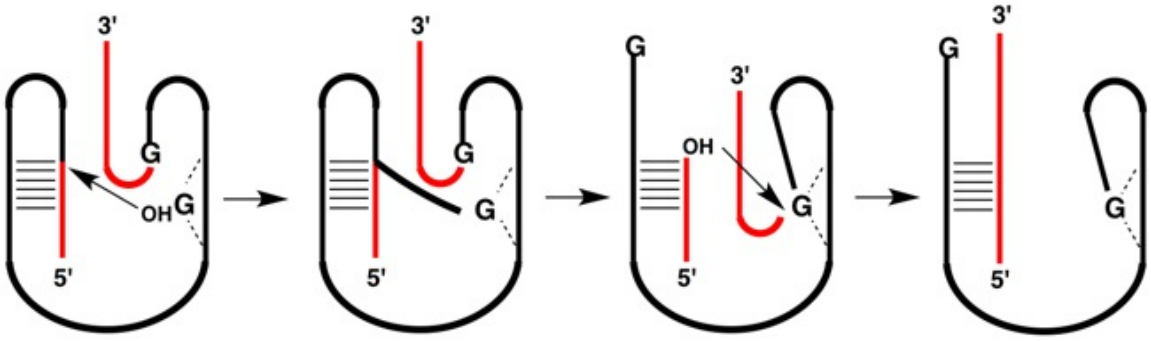


Figure 1.1 *Cis*-splicing reaction of the *Tetrahymena thermophila* Tth.L 1925 IC1 group I intron ribozyme. During *cis*-splicing, the ribozyme (black) catalyzes its own self-excision and the joining of the two flanking exons (red) through two successive transesterification reactions.

1.2 Structure of the Group I Intron

Early structural information about the group I intron was gained through phylogenic comparison, energy minimization models and structural probing experiments (10–13). Data collected in these experiments led to a 3D model prediction (13), most of which was later confirmed by X-ray crystallography (Figure 1.3). Crystal structures of Tetrahymena IC1, Azoarcus IC3 and Twort IA2 group I introns have now been solved at different stages along the splicing pathway (14–17). It is now understood that the group I intron is comprised of coaxially stacked P3-P9 and P4-P6 helical domains that make up the conserved core and contain the active site. The conserved core of different group I introns

show a high degree of structural similarity. Peripheral helices, which vary in number depending on the class of group I intron, wrap around the ribozyme core (13,18–20). Although many of the peripheral domains can be deleted while still maintaining catalytic function, it is believed that these domains assist in stabilizing the structure of the ribozyme (20).

Important insight regarding the architecture of the active site of the group I intron was gained, in part, from the crystal structures of group I introns. These structures reveal that guanosine is able to bind in the guanosine-binding site through a base triple interaction with the G264-C11 base pair, as well as stacking between two additional base triple interactions (14,16,21). To understand how the group I intron is able to act as catalyst while not containing functional groups with pKa values near neutrality, many focused on possible interaction with metal ions. Metal ion rescue experiments, paired with thermodynamic fingerprint analysis have shown interaction with Mg^{+2} ions that have the effect of stabilizing the attacking group by lowering its pKa, stabilizing the developing negative charge leaving group and positioning the structure for an in-line attack (14,22–26).

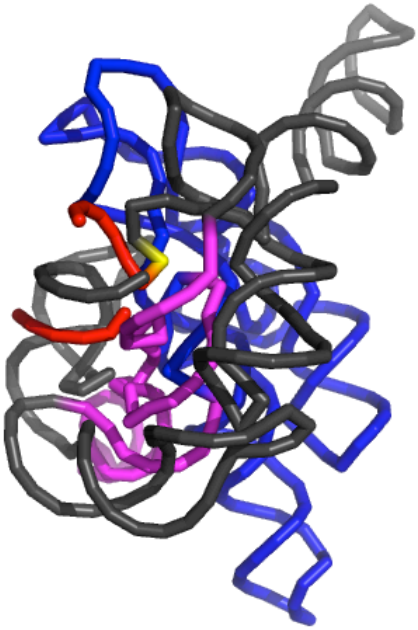


Figure 1.2 3D model of the *cis*-splicing Tth.L 1925 IC1 group I intron from *Tetrahymena thermophila* (13). The ribozyme conserved core, consisting of the P4-P6 domain (blue), P3-P8 domain (purple) and ωG (yellow), and peripheral domains (grey) are shown, along with the last 6 nucleotides of the 5' exon and first 8 nucleotides of the 3' exon (red).

1.3 Phylogenetic distribution and Classes of Group I Introns

The distribution of group I introns are phylogenetically widespread with approximately 1500 group I introns having been identified in nature, ranging in length from 250 to 500 nucleotides (27). Of these, 1400 have been found in eukaryotic genomes; 800 of which are located in DNA contained within the nucleus, 220 in mitochondrial genes and 370 in plastid DNA. Within the nucleus of eukaryotes, the group I intron has been found at 47 sites in the small subunit rRNA and 44 sites in the large subunit rRNA. With the exception of sea anemone and coral, the group I intron has not been identified in animals. In contrast to the dispersion of group I introns in eukaryotes, only 50 group I introns

have been identified in bacteria, at 9 insertion sites within tRNA, rRNA and recA genes. Even more infrequent is the distribution of group I introns in virus and phage DNA. In Archaea group I introns have not yet been identified (27). Interestingly, the evolution of group I introns shows evidence of both horizontal and vertical transmission, with group I introns from distant organisms being closely related. This occurrence makes it difficult to study the natural evolutionary history of group I introns.

Structural and sequence distinctions have allowed group I introns to be separated into five main groups, termed IA-E, and further classified into several distinct subgroups (28–30). Group I introns found in the nucleus are classified into the IB, IC1 or IE groups, with the group I intron from *Tetrahymena thermophila* being a member of the IC1 group (27,31,32). Conserved in all group I introns are two sets of helices, termed the ‘P3-P8’ and ‘P4-P6’ domains, that comprise the ribozyme core. The presence of other peripheral helices varies among group I intron subgroups (18–20). Although group I introns contain conserved secondary and tertiary structural features, very little sequence conservation is observed in primary sequences comparisons (20,30,31,33,34). The diversity of group I intron primary sequences has provided information regarding of the positions at which nucleotide co-variation is tolerated or conservation is necessary. In addition, this suggests that a large sequence space can be explored during artificial evolutions of group I introns.

1.4 Evolution of the Group I Intron

Sporadic distribution of group I introns, paired with the observation that introns from homologous sites appear to be closely related, even when found on distantly related organisms, supports the cyclical gain and loss model of group I intron evolution (Figure 1.3) (27,35). Group I introns often encode a homing endonuclease gene (HEG), capable of lateral transfer of intervening sequences (introns or inteins) into homologous alleles that lack the sequence. During this process, the homing endonuclease catalyzes a double strand break after recognizing a short, 14-40 nucleotide, target sequence, resulting in duplication of the intervening sequence into the cleaved allele by double strand break repair (36). Being encoded in self-splicing sequences ensures the propagation of the HEG without harm to the host. The cyclical gain and loss model of group I intron evolution proposes that introns are horizontally transferred into an intronless site via homing, resulting in an intron with a functional HEG becoming fixed in a population. Once fixed in a population there is no selective pressure for the HEG to remain functional; it is therefore subsequently degraded and lost. This is followed by loss of the intron in the population, allowing the cycle to repeat (35).

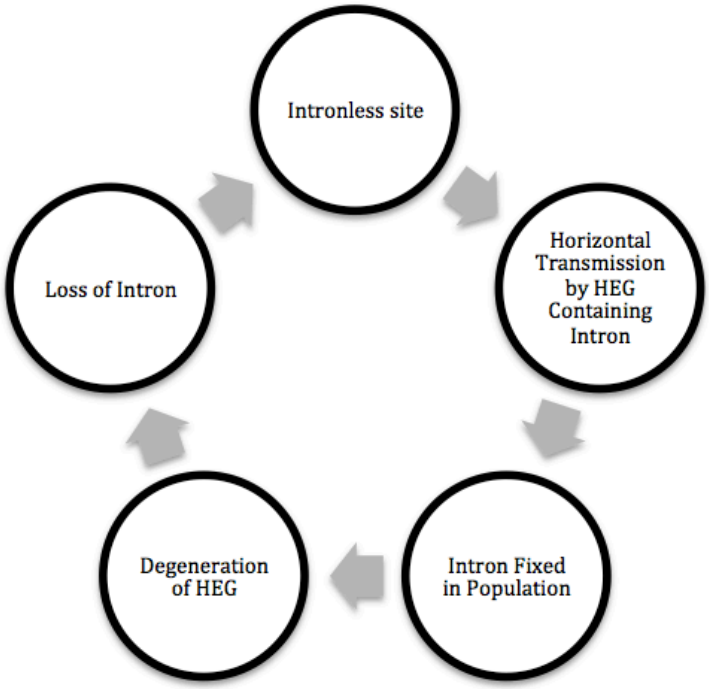


Figure 1.3 Cyclical gain and loss model of group I intron evolution proposed by Goddard and Burt (35).

Select group I introns, like other cellular RNAs, have evolved to utilize cellular proteins for stability or as chaperones (37–39). The few cases in which a group I intron appears to have been strictly vertically inherited for long periods of time allows for the study of the loss of secondary structure elements, and concurrent loss of independent splicing ability (40). The paucity of group I introns that have undergone long term strictly vertical inheritance, leaves a lack of knowledge regarding evolutionary intermediates, as well as evolutionary conditions that facilitate the recruitment of host proteins by the group I intron. Knowledge regarding the evolution of not only group I introns, but also RNA more broadly, is important for our understanding of RNA interactions with cellular factors including proteins, and for future directed evolutions of RNA.

Artificial evolutions, conducted under controlled conditions, are able to provide more insight into principles guiding the evolution of RNA. Notable previous RNA evolution studies have been conducted *in vitro* and have demonstrated that evolutionary adaptation of an RNA enzyme is enhanced by higher genetic diversity in the starting population (41), the co-evolution of distinct ribozymes can lead each to occupy separate niches (42) and that high mutational loads can lead to extinction of small populations of RNA enzymes, but recombination can act to reduce this mutational load (43). Although *in vitro* studies do provide insight into important principles in RNA evolution, there is still a lack of studies investigating RNA evolution in a cellular environment, in which RNA species can interact with a host of cellular factors. It is still unclear how proteins are recruited by RNA species and the evolutionary conditions that encourage or discourage the formation of RNA-protein complexes. One recent study evolving a group I intron in *E. coli* cells demonstrated the evolution a new RNA-protein interaction in which the evolution of four mutations resulted in recruitment of the transcription termination factor Rho by the group I intron. Interestingly, the evolved group I intron did not show evidence of improved *trans*-splicing efficiency, but instead resulted in more efficient translation of the splicing product (44). The group I intron provides an excellent model system to further study RNA evolution because it acts as a reporter system on the evolution of many structural and functional properties. Group I introns are highly structured RNAs that undergo a conformational change, catalyze two reactions and the

trans-splicing variants require binding two species (an RNA substrate and a nucleotide) (9,14,20,34,45,46).

1.5 Group I Intron *Trans*-Splicing

Although the group I intron evolved in nature to catalyze a *cis*-splicing reaction, alterations to the ribozyme have allowed it to catalyze several non-native reactions. Sequence alterations to the 5' end of the group I intron have allowed the ribozyme to catalyze the splicing reaction in *trans*, on an exogenous substrate RNA (Figure 1.4). By removing the 5' exon and modifying the sequence at the 5' end of the ribozyme, this region is able to serve as a recognition sequence, termed the internal guide sequence (IGS) allowing the ribozyme to hybridize to a target substrate. Once bound, the ribozyme is able to replace the 3' portion of the substrate RNA with the its own 3' exon (47,48). This reaction can be widely applied to a range of RNA substrates, as the only sequence requirement for *trans*-splicing is the presence of a U to serve as the 5' splice site. Similarly, the group I intron has also been modified to allow it to catalyze the replacement of the 5' portion of a substrate RNA in *trans*. During this reaction, the ribozyme 3' sequence is modified to serve as a recognition sequence for an exogenous substrate. Once bound, the ribozyme is able to replace the 5' end of the substrate with it's own 5' exon (49).

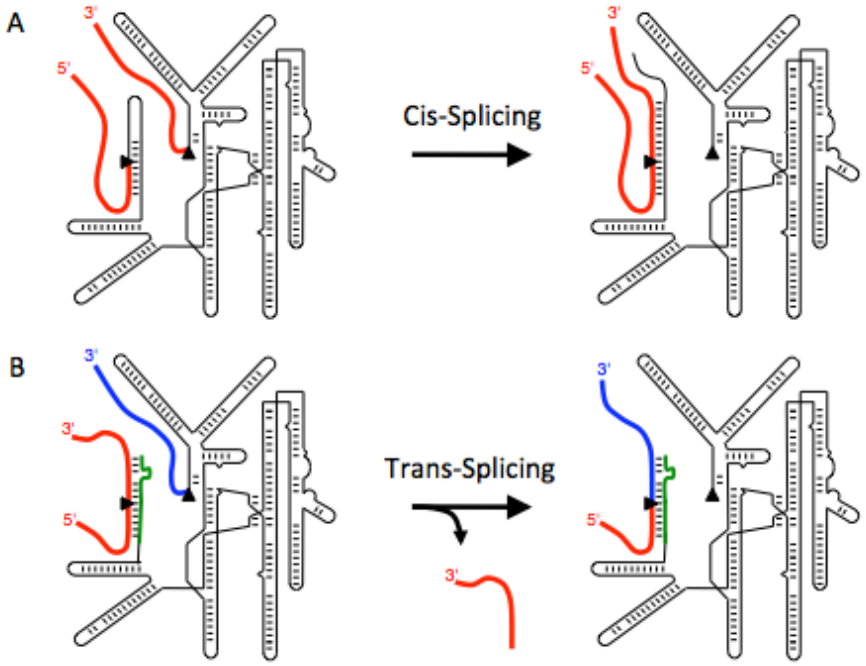


Figure 1.4 *Cis*- and *trans*-splicing reaction of the *Tetrahymena thermophila* Tth.L 1925 IC1 group I intron shown using secondary structure representations of the ribozyme. (A) *Cis*-splicing reaction in which the ribozyme (black) is able to catalyze its own self-excision and the joining of the two flanking exons (red). (B) The group I intron (black) can be converted into a *trans*-splicing format through sequence modifications to its 5' EGS (green). In this format, the group I intron is able to bind to an exogenous RNA substrate (red) and subsequently replace the 3' end of the substrate with it's own 3' tail (blue) in a reaction analogous to the *cis*-splicing reaction.

The ability of the group I intron to replace a portion of substrate RNA has the potential for a wide range applications in both therapy and research (50). Replacing a portion of a mutated cellular mRNA with a wild type sequence could be used to treat genetic disorders. Studies have shown repair of sickle beta-globin mRNAs and trinucleotide repeat expansions, among others (47,51,52). This method of treatment has the benefit of maintaining endogenous patterns of gene expression while replacing a mutated protein with a functional one. In addition to repair of mutated mRNA, *trans*-splicing ribozymes can selectively

induce apoptosis in diseased or virally infected cells by splicing a sequence encoding a toxic peptide into virus or cancer specific mRNAs (53–56). In addition to the therapeutic potential of the *trans*-splicing group I intron, this system can also be used to splice a tag onto target RNA species *in vivo*. This allows for detection of the target RNAs through various methods including molecular imaging (57,58).

Despite the potential to be used in both therapy and research, the use of *trans*-splicing group I introns have been limited, in part, by low efficiency. Splicing efficiencies of 50% have been observed when using ribozyme concentrations too high for clinical purposes. When using ribozymes at clinically relevant concentrations, splicing efficiencies of 10% or less have been observed (47,50,59–62). Low efficiency may be caused by a number of factors, one of which being ribozymes folding into non-native states. The distribution of ribozymes folding into native and non-native states is dependent on a number of factors including flanking exons, promoters (63,64) and the presence or lack of specific protein interactions that serve to stabilize or destabilize the ribozyme (65–67). Depending on the nature of mis-folding, ribozymes must undergo local or global rearrangements before assuming the native conformation allowing for splicing to occur(68).

In addition to mis-folding, weak interaction between ribozyme and substrate can also lead to low efficiency. This phenomenon can be attributed to lack of stability between the ribozyme IGS and its target substrate or inaccessibility of splice sites due to the nature of secondary structures formed by

the substrate. It has also been observed that the IGS often interacts with cellular RNA species other than its target substrate, leading to non-specific splicing (69). These issues have been partially overcome by adding an extended recognition sequence on the 5' end of the ribozyme, termed the extended guide sequence (EGS), thus providing greater complementarity between ribozyme and substrate. Increased splicing efficiency has been observed from the addition of this EGS region, and in some circumstances, adjustments to the recognition sequence to allow for formation of the P10 helix (61,70). In addition, a useful *trans*-tagging assay has been developed to identify accessible splice sites in the substrate mRNA, (71). Although this method has proven to be valuable, a far less time consuming method, involving the computational prediction of binging free energies between ribozyme and substrate RNA, has aided in the selection of efficient splice sites with a higher accuracy than the *trans*-tagging assay (72).

Selection and evolution are two techniques that can also be employed to improve ribozyme efficiencies in cells. A method was developed for *in vivo* selection of group I intron ribozymes with randomized sequences that can be used for identification of ribozymes with greater *trans*-splicing efficiencies (61). An additional method, relying on many of the same principles, was also developed that allows for the *in vivo* evolution of group I intron ribozymes through successive rounds of mutagenesis or recombination, followed by selection. This method allows for the optimization of ribozymes in cells (44).

In addition to the need for improved *trans*-splicing efficiency, lack of an effective clinical delivery method is another obstacle faced by the group I intron

as well as other therapeutic RNA. Common current approaches for RNA delivery include the use of liposomes, nanocapsules and viral vector delivery systems, as well as chemical modifications that improve RNA stability *in vivo* (73–77). Work is still needed, however, to develop a safe and effective method of delivery or RNA therapeutics.

1.6 *Trans*-excision

The *Tetrahymena thermophila* Tth.L 1925 IC1 group I intron has been modified previously to perform an additional non-native reaction, termed a *trans*-excision reaction (78). In this reaction, the group I intron is able to excise internal fragments of substrates, resulting in joining of the two flanking sequences and the excised fragment attached to the 3' end of the ribozyme (79). In a cellular environment, this system has been able to excise single nucleotide fragments, with an estimated splicing efficiency of 0.6-12% (80). *In vitro*, fragments of up to 28 nucleotides in length have been removed (78). Like the *trans*-excision reaction, the group I intron has also been shown to be able to catalyze a *trans*-insertion reaction in which fragments of up to six nucleotides in length were able to be inserted into a substrate RNA (81).

Similar to the *trans*-splicing reaction, the *trans*-excision and *trans*-insertion reactions could be developed for use in the therapy of genetic diseases. Changes in sequences recognized by the spliceosome and cellular splicing factors often leads to mis-splicing of pre-mRNA transcripts. These events can result in complete or partial exon skipping, pseudoexon inclusion, intron skipping and a change in alternative splicing patterns (82). It is estimated that 10-60% of

all disease causing mutations result in aberrant splicing events (82,83). Techniques that are able to repair or reduce mis-splicing events could be developed to treat a wide range of diseases.

In order to be used clinically, the reported reaction efficiencies of the *trans*-excision reaction of 0.6-12% in a cellular environment would likely need to be improved. The ability of excising fragments greater than one nucleotide in length would have a wider range of therapeutic applications. In addition, due to the nature by which this reaction proceeds, multiple turnover is not possible with the *trans*-excision reaction (79). Like the *trans*-splicing group I intron, an effective delivery method would need to be developed and optimized before use in a clinical setting.

1.7 Goals of the Dissertation

Despite the many potential therapeutic applications of *trans*-splicing group I introns, ribozymes with improved efficiency and high splicing specificity are still needed before the system can be used for clinical applications. Evolution is a promising and powerful tool that may increase the *trans*-splicing efficiency and specificity of these group I introns. It is currently unclear, however, what evolutionary parameters are best to optimize these ribozymes. The following chapters detail the study of principles governing RNA evolution in a cellular environment and the application of these principles in the evolution and optimization of a novel variant of a *trans*-splicing group I intron with therapeutic applications.

Chapter 2: Low Selection Pressure Aids the Evolution of Cooperative Ribozyme Mutations In Cells

Understanding the evolution of RNA species in cells provides insight into the natural evolution of RNA, and helps the design of more efficient directed RNA evolution experiments. In this study, the effects of two evolutionary parameters, selection pressure and recombination, were studied in an evolving population of *trans*-splicing group I intron ribozymes in cells. It was found that the emergence of a set of cooperative mutations was facilitated more efficiently when a low, rather than high, selection pressure was applied. This finding not only enhances the understanding of RNA evolution in biology, but is also applicable to the development of more efficient *trans*-splicing group I intron for use in therapy as well as the design of more efficient evolutions of other RNA species.

Chapter 3: Spliceozymes: Ribozymes that Remove Introns from Pre-mRNAs in *Trans*

In this study, a new variant of the *trans*-splicing group I intron ribozyme was developed, and is termed the 'spliceozyme'. The spliceozyme is able to catalyze the removal internal sequences from pre-mRNA substrates in *trans*, and the joining of the two flanking exons, thereby generating a functional mRNA. Although a similar '*trans*-excision ribozyme' has previously been developed, this system was only shown to be able to catalyze the removal of fragments one nucleotide in length in cells, and design principles prevented this system from the potential to perform multiple turnover reactions. The spliceozyme system, employing different design principles, has been shown to be active both *in vitro*

and in cells, in which it is able to remove a wide variety of 100 nucleotide long fragments. In addition, the spliceozyme design allows for multiple turnover reactions and can be designed to target a variety of splice sites, thus having many potential applications in both research and therapy.

Chapter 4: Decreased Side Product Formation in Evolved Group I Intron Spliceozymes

In order to increase the efficiency of the spliceozyme, and subsequently the potential applications of the system, an evolution was performed in cells. In this evolution, the spliceozyme was challenged to more efficiently remove internal sequences from substrate pre-mRNA. Characterization of one of the fittest clones resulting from this evolution identified necessary mutations in both the conserved core and peripheral domains of the ribozyme. This set of mutations results in greater product formation, as well as decreased side product formation, qualities that may increase the therapeutic potential of the spliceozyme.

1.8 References

1. Orgel L (1968) Evolution of the Geneitc Apparatus. *J Mol Biol* 38:381–393.

2. Crick F (1968) The Origin of the Genetic Code. *J Mol Biol* 38:367–379.

3. Woese C (1967) *The Genetic Code: the Molecular Basis for Genetic Expression* (Harper & Row, New York).

4. Kruger K, Grabowski P, Zaug A, Sands J, Gottschling D, Cech T (1982) Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* 31:147–157.

5. Cech T (1992) Nobel Lecture, December 8, 1989. *Nobel Lectures: Chemistry 1981-1990* (World Scientific Publishing, Singapore).

6. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849–857.

7. Zaug A, Grabowski P, Cech T (1993) Autocatalytic cyclization of an excised intervening sequence RNA is a cleavage-ligation reaction. *Nature* 301(5901):578–83.

8. Price J (1987) Origin of the phosphate at the ligation junction produced by self-splicing of Tetrahymena thermophila pre-ribosomal RNA. *J Mol Biol* 196(1):217–21.

9. Guo F, Gooding A, Cech T (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16(3):351–62.

10. Michel F, Dujon B (1983) Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO J* 2:33–38.

11. Cech T, Tanner N, Tinoco I, Weir B, Zuker M, Perlman P (1983) Secondary structure of the Tetrahymena ribosomal RNA intervening sequence: Structural homology with fungal mitochondrial intervening sequence. *Proc Natl Acad Sci* 80:3903–3907.

12. Kim S, Cech T (1987) Three-dimensional model of the active site of the splicing rRNA precursor of Tetrahymena. *Proc Natl Acad Sci* 84:8788–92.

13. Lehnert V, Jaeger L, Michel F, Westhof E (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the Tetrahymena Thermophila ribozyme. *Chem Biol* 3:993–1009.

14. Adams P, Stahley M, Kosek A, Wang J, Strobel S (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430(6995):45–50.

15. Adams P, Stahley M, Gill M, Kosek A, Wang J, Strobel S (2004) Crystal structure of a group I intron splicing intermediate. *RNA* 10(12):1867–87.

16. Golden B, Cech T (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12:82–89.

17. Guo F, Gooding A, Cech T (2004) Structure of the Tetrahymena ribozyme: Base triple sandwich and metal ion at the active-site. *Mol Cell* 16:351–62.

18. Burke J (1988) Molecular genetics of group I introns: RNA structures and protein factors required for splicing - A review. *Gene* 73:273–294.

19. Cech T (1988) Conserved sequences and structues of group I introns: Building an active site for RNA catalysis - A review. *Gene* 73:259–271.

20. Cech T (1990) Self-splicing of group I introns. *Annu Rev Biochem* 59:543–568.

21. Guo F, Gooding A, Cech T (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16(3):351–62.

22. Shan S, Yoshida A, Sun S, Piccirilli J, Herschlag D (1999) Three metal ions at the active site of the Tetrahymena ribozyme reaction. *Proc Natl Acad Sci* 96:12299–12304.

23. Shan S, Kravchuk A, Piccirilli J, Herschlag D (2001) Defining the catalytic metal ion interactions in the Tetrahymena ribozyme reaction. *Biochemistry* 40:5161–5171.

24. Shan S, Herschlag D (2002) Dissection of a metal-ion mediated conformational change in Tetrahymena ribozyme catalysis. *RNA* 8:861–872.

25. Shan S, Herschlag D (1999) Probing the role of metal ions in RNA catalysis. Kinetic and thermodynamic characterization of a metal ion interaction with the 2' moiety of the guanosine nucleophile in the Tetrahymena group I ribozyme. *Biochemistry* 38:10958–75.

26. Yoshida A, Sun S, Piccirilli J (1999) A new metal ion interaction in the Tetrahymea ribozyme reaction revealed by double sulfur substitution. *Nat Struct Biol* 6:318–321.

27. Haugen P, Simon D, Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* 21(2):111–119.

28. Suh S, Jones K, Blackwell M (1999) A group I intron in the nuclear small subunit rRNA gene of Cryptendoxyla hypophloia, an ascomycetous fungus: evidence for a new major class of group I introns. *J Mol Evol* 48(5):493–400.

29. Jaeger L, Michel F, Westhof E (1997) The structure of group I ribozymes. *Nucleic Acids Mol Biol* 10:33–51.

30. Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610.

31. Cannone J, Subramanian S, Schnare M, Collett J, D'Souza L, Du Y, Feng B, Lin N, Madabusi L, Muller K, Pande N, Shang Z, Yu N, Gutelle R. (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3(1):15.

32. Jackson S, Cannone J, Lee J, Gutell R, Woodson S (2002) Distribution of rRNA introns in the three-dimensional structure of the ribosome. *J Mol Biol* 323(1):35–52.

33. Vicens Q, Cech T (2006) Atomic level architecture of group I introns revealed. *Trends Biochem Sci* 31(1):41–51.

34. Woodson S (2005) Structure and assembly of group I introns. *Curr Opin Struct Biol* 15:324–330.

35. Goddard M, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci USA* 96:13880–13885.

36. Chevalier B, Stoddard B (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res* 29:3757–3774.

37. Lewin A, Thomas J, Tirupati H (1995) Cotranscriptional splicing of a group I intron is facilitated by the Cbp2 protein. *Mol Cell Biol* 15(12):6971–8.

38. Majumder A, Akins R, WIlkinson J, Keley R, Snook A, Lambowitz A (1989) Involvement of tyrosyl-tRNA synthetase in splicing of group I introns in Neurospora crassa mitochondria: biochemical and immunochemical analyses of splicing activity. *Mol Cell Biol* 9(5):2089–2104.

39. Waldsich C, Grossberger R, Schroeder R (2002) RNA chaperone StpA loosens interactions of the tertiary structure in the td group I intron in vivo. *Genes Dev* 16(17):2300–12.

40. Haugen P, Runge H, Bhattacharya D (2004) Long-term evolution of the S788 fungal nuclear small subunit rRNA group I introns. *RNA* 10(7):1084–96.

41. Hayden E, Ferrada E, Wagner A (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474:92–95.

42. Voytek S, Joyce G (2009) Niche partitioning in the coevolution of two distinct RNA enzymes. *Proc Natl Acad Sci USA* 106:7780–7785.

43. Soll S, Diaz Arenas C, Lehman N (2007) Accumulation of deleterious mutations in small abiotic populations of RNA. *Genetics* 175:267–275.

44. Olson K, Dolan G, Muller U (2014) In Vivo Evolution of a Catalytic RNA Couples Trans-Splicing to Translation. *PLoS One* 9(1):e86473.

45. Golden B, Gooding A, Podell E, Cech T (1998) A preorganized active site in the crystal structure of the Tetrahymena ribozyme. *Science (80-)* 282(5387):259–64.

46. Emerick V, Pan J, Woodson S (1996) Analysis of rate determining conformational changes during self splicing of the Tetrahymena intron. *Biochemistry* 35(41):13469–77.

47. Sullenger B, Cech T (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* 371(6498):619–622.

48. Inoue T, Sullivan F, Cech T (1985) Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. *Cell* 43(2Pt1):431–437.

49. Alexander R, Baum D, Testa S (2005) 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme. *Biochemistry* 44(21):7796–804.

50. Fiskaa T, Birgisdottir A (2010) RNA reprogramming and repair based on trans-splicing group I ribozymes. *N Biotechnol* 27(3):194–203.

51. Lan N, Howrey R, Lee S, Smith C, Sullenger B (1998) Ribozyme-mediated repair of sickle β-globin mRNAs in erythrocyte precursors. *Science (80-)* 280(5369):1593–6.

52. Phylactou L, Darrah C, Wood M (1988) Ribozyme-mediated trans-splicing of a trinucleotide repeat. *Nat Genet* 18(4):378–381.

53. Shin K, Sullenger B, Lee S (2004) Ribozyme-mediated induction of apoptosis in human cancer cells by targeted repair of mutant p53 RNA. *Mol Ther* 10(2):365–72.

54. Ban G, Song M, Lee S (2009) Cancer cell targeting with mouse TERT-specific group I intron of Tetrahymena thermophila. *J Microbiol Biotechnol* 19(9):1070–1076.

55. Ryu K, Kim J, Lee S (2003) Ribozyme-mediated selective induction of new gene activity in hepatitis C virus internal ribosome entry site-expressing cells by targeted trans-splicing. *Mol Ther* 7(3):386–95.

56. Ayre B, Kohler U, Goodman H, Haseloff J (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc Natl Acad Sci USA* 96(7):3507–12.

57. So M, Gowrishankar G, Hasegawa S, Chung J, Rao J (2008) Imaging target mRNA and siRNA-mediated gene silencing in vivo with ribozyme-based reporters. *Chembiochem* 9(16):2682–91.

58. Hong S, Jeong J, Lee Y, Jung H, Kim K, Kim Y, Lee Y, Lee S, Bae C, Park J, Kim I (2007) Molecular imaging of endogenous mRNA expression in a mouse tumor model by adenovirus harboring trans-splicing ribozyme. *FEBS Lett* 581(28):5396–400.

59. Byun J, Lan N, Long M, Sullenger B (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes. *RNA* 9(10):1254–63.

60. Jones J, Sullenger B (1997) Evaluating and enhancing ribozyme reaction efficiency in mammalian cells. *Nat Biotechnol* 15(9):902–5.

61. Olson K, Muller U (2012) An in vivo selection method to optimize trans-splicing ribozymes. *RNA* 18:581–89.

62. Rogers C, Vanoye C, Sullenger B, George A (2002) Functional repair of a mutant chloride channel using a trans-splicing ribozyme. *J Clin Invest* 110:1783–89.

63. Koduvayur S, Woodson S (2004) Intracellular folding of the Tetrahymena group I intron depends on exon sequence and promoter choice. *RNA* 10(10):1526–32.

64. Nikolcheva T, Woodson S (1999) Facilitation of Group I Splicing in Vivo: Misfolding of the Tetrahymena IVS and the Role of Ribosomal RNA Exons. *J Mol Biol* 292:557–67.

65. Chen X, Mohr G, Lambowitz A (2004) The Neurospora crassa CYT-18 protein C-terminal RNA-binding domain helps stabilize interdomain tertiary interactions in group I introns. *RNA* 10(4):634–44.

66. Weeks K, Cech T (1995) Protein facilitation of group I intron splicing by assembly of the catalytic core and the 5’ splice site domain. *Cell* 82:221–230.

67. Herschlag D (1995) RNA chaperones and the RNA folding problem. *J Biol Chem* 270:20871–74.

68. Wan Y, Suh H, Russell R, Herschlag D (2010) Multiple unfolding events during native folding of the Tetrahymena group I ribozyme. *J Mol Biol* 400(5):1067–77.

69. Jones J, Seong-Wook L, Sullenger B (1996) Tagging ribozyme reaction sites to follow trans-splicing in mammalian cells. *Nat Med* 2(6):643–8.

70. Kohler U, Ayre B, Goodman H, Haseloff J (1999) Trans-splicing ribozymes for targeted gene delivery. *J Mol Biol* 285(5):1935–50.

71. Jones J, Lee S, Sullenger B (1996) Tagging ribozyme reaction sites to follow trans-splicing in mammalian cells. *Nat Med* 2:643–648.

72. Meluzzi D, Olson K, Dolan G, Arya G, Muller U (2012) Computational prediction of efficient splice sites for trans-splicing ribozymes. *RNA* 18:590–602.

73. Guo P, Coban O, Snead N, Trebley J, Hoeprich S, Guo S, Shu Y (2010) Engineering RNA for targeted siRNA delivery and medical application. *Adv Drug Deliv Rev* 62:650–66.

74. Gao Y, Liu X, Li X (2011) Research progress on siRNA delivery with nonviral carriers. *Int J Nanomedicine* 6:1017–25.

75. Higuchi Y, Kawakami S, Hashida M (2010) Strategies for in vivo delivery of siRNAs: recent progress. *BioDrugs* 24(3):195–205.

76. Kotterman M, Schaffer D (2014) Engineering adeno-associated viruses for clinical gene therapy. *Nat Rev Genet* 15:445–51.

77. Pereira D, Rodrigues P, Borralho P, Rodrigues C (2013) Delivering the promise of miRNA cancer therapeutics. *Drug Discov Today* 18(5-6):282–89.

78. Bell M, Johnson A, Testa S (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs. *Biochemistry* 41(51):15327–33.

79. Dotson P, Sinha J, Testa S (2008) A Pneumocystis carinii Group I Intron-Derived Ribozyme Utilizes an Endogenous Guanosine as the First Reaction Step Nucleophile in the Trans Excision-Splicing Reaction. *Biochemistry* 47(16):4780–7.

80. Baum D, Testa S (2005) In vivo excision of a single targeted nucleotide from an mRNA by a trans excision-splicing ribozyme. *RNA* 11(6):897–905.

81. Johnson A, Sinha J, Testa S (2005) Trans insertion-splicing: ribozyme-catalyzed insertion of targeted sequences into RNAs. *Biochemistry* 44(31):10702–10.

82. Hammond S, Wood M (2011) Genetic therapies for RNA mis-splicing diseases. *Trends Genet* 27(5):196–205.

83. Wang G, Cooper T (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8(10):749–61.

Chapter 2: Low Selection Pressure Aids the Evolution of Cooperative Ribozyme Mutations in Cells

2.1 Abstract

Understanding the evolution of functional RNA molecules is important for our molecular understanding of biology. Here we tested experimentally how two evolutionary parameters - selection pressure and recombination - influenced the evolution of an evolving RNA population. This was done using four parallel evolution experiments that employed low or gradually increasing selection pressure, and recombination events either at the end or dispersed throughout the evolution. As model system, a *trans*-splicing group I intron ribozyme was evolved in *E. coli* cells over 12 rounds of selection and amplification, including mutagenesis and recombination. The low selection pressure resulted in higher efficiency of the evolved ribozyme populations, while differences in recombination did not have a strong effect. Five mutations were responsible for the highest efficiency. The first mutation swept quickly through all four evolving populations, while the remaining four mutations accumulated later and more efficiently under low selection pressure. To determine why low selection pressure aided this evolution, all evolutionary intermediates between the wild type and the 5-mutation variant were constructed, and their activities at three different selection pressures were determined. The resulting fitness profiles showed a high cooperativity between the four late mutations, which can explain why high

selection pressure led to inefficient evolution. These results show experimentally how low selection pressure can benefit the evolution of cooperative mutations in functional RNAs.

2.2 Introduction

Catalytic RNAs (ribozymes) are essential in all life forms, playing important roles in gene expression and regulation. Natural ribozymes include the ribosome (1,2), RNase P (3), the spliceosome (4,5), group I (6) and group II introns (7), and six small, self-cleaving ribozymes (8-13), most of which are variants of ancestral ribozymes that originated more than a billion years ago. The spliceosome, for example, appears to share a common ancestor with self-splicing group II introns (14,15), RNase P evolved into RNA-protein particles with different sets of proteins in bacteria, archaea and eukarya (16), and group I introns appear to follow an evolutionary round of invasion, degeneration, and loss (17). The correlation of sequences for group I intron ribozymes that were vertically inherited over hundreds of millions of years has provided insight into their loss of secondary structure elements and the concomitant loss of splicing activity (18). The comparison of biological sequences, however, cannot recapitulate their evolution because the evolutionary parameters, and evolutionary intermediates, are lost to history. In contrast, controlled evolution experiments in the lab are a powerful tool to understand the parameters that shape RNA evolution.

Previous experimental studies have elucidated several central features of RNA evolution *in vitro*. These studies showed that genetic diversity of a starting

population increases the rate of adaptive evolution (19) that recombination can benefit an evolving population by reducing mutational load (20), and that two distinct, coevolving ribozymes can diversify such that each ribozyme dominates a different niche (21). In contrast to these RNA evolution studies *in vitro* and RNA *selection* experiments in cells, according to our knowledge, no experimental studies have addressed RNA *evolution* in cells. Evolution differs from selection by the repeated application of mutagenesis between multiple selection steps, an important difference that facilitates the successive optimization of sequences. Because nature optimizes RNA sequences by evolution it is desirable to understand the parameters affecting RNA evolution in cells. The dynamics of RNA evolution in cells are likely affected by the cellular environment, in which RNAs may be able to recruit and utilize cellular factors. This study utilizes an *in vivo* evolution system that allows studying the evolution of functional RNAs in *E. coli* cells (70).

As a model system for RNA evolution in cells we utilize group I intron ribozymes, which are able to excise themselves from pre-mRNAs *in vitro* and *in vivo* (6). To do this, the introns fold into three-dimensional structures and catalyze two transesterification reactions. This results in the excision of the intron sequence and the joining of the flanking exons. These group I intron ribozymes can be converted to a *trans*-splicing format by removing the 5'-exon and adding a short substrate recognition sequence to the new ribozyme 5'-terminus (22,23) (Fig. 2.1A). In this new format the ribozyme can specifically recognize a target site on a substrate RNA and replace the substrate 3'-portion with its own 3'-

terminus. Group I intron ribozymes show *trans*-splicing activity *in vitro* (22,23), in bacterial cells (23,24) and in mammalian cells (25,26).

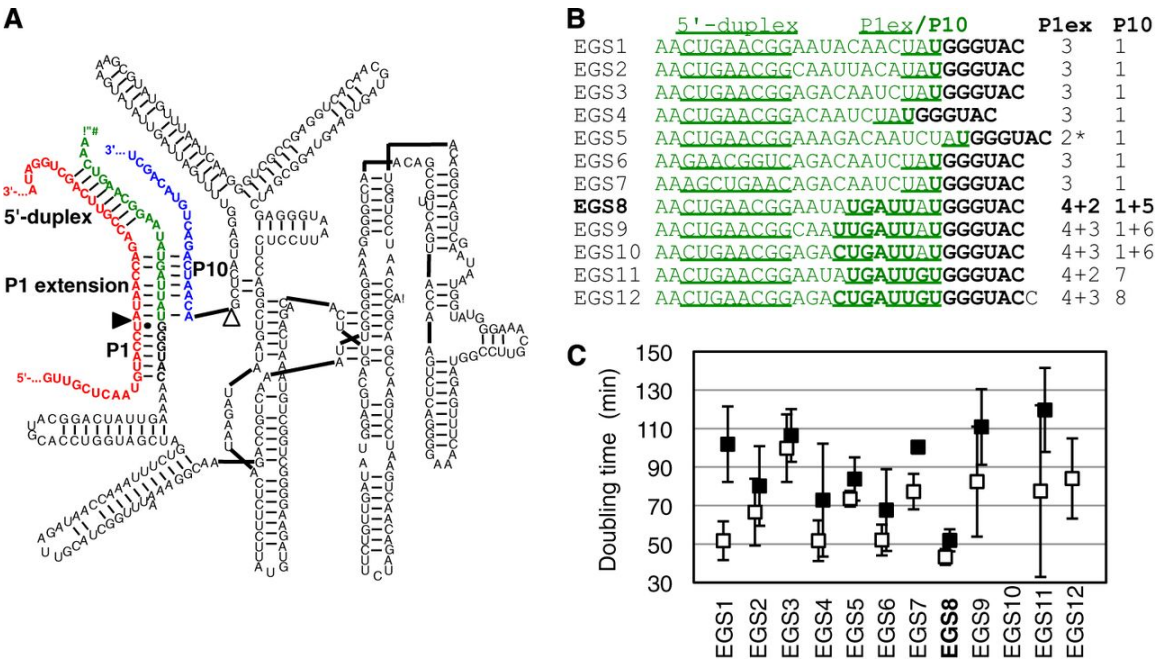


Figure 2.1 *Trans*-splicing ribozyme variant of the *Tetrahymena* group I intron that was used as parent construct for the evolutions. (A) Secondary structure of the ribozyme (black) with its 5'-terminal extended guide sequence (EGS; green) base paired to the target site on the substrate (red). The ribozyme 3'-exon is in blue. The 5'-splice site is marked by a black triangle, the 3'-splice site with an open triangle. The 5'-duplex, P1 helix, the P1 extension helix, and the P10 helix are labeled. The EGS used in this secondary structure is sequence 8 from sub-figure (B). (B) Sequences of 12 designed EGSs that were tested on the ribozyme 5'-terminus. For each sequence, the predicted number of base pairs in the P1 helix and the P10 helix are given on the right. The underlined portion corresponds to the predicted 5'-duplex and the P1 extension helix, the bold portion to the P10 helix. The asterisk denotes a sequence where three base pairs could be formed in the P1 extension helix but are not predicted to form due to a self-structure of the EGS. (C) Ability of the 12 different EGSs to increase E. coli growth rates in the presence of chloramphenicol due to the repair of CAT pre-mRNA in the cells. The doubling time of E. coli cells is given for all 12 constructs, in LB medium containing 2 µg/mL chloramphenicol (open squares) and 6 µg/mL chloramphenicol (filled squares). Note that the symbols are slightly offset to clarify the error bars. EGS variant 10 did not mediate measurable growth. Note that low doubling times correspond to high ribozyme efficiencies. EGS variant 8 was chosen for the parent ribozyme of the evolution.

Trans-splicing ribozymes could be useful for therapeutic applications (for review see (27)) to correct genetic mutations on the mRNA level (25,26,28,29), and specifically kill virus-infected cells (30,31) or cancerous cells (32). Therapeutic applications are currently limited by inefficient delivery of the ribozymes into cells, and by the low *trans*-splicing efficiency of the ribozymes in cells. The *trans*-splicing efficiency in cells has exceeded 10% only in exceptional cases (26). Efficiency is dependent upon on the choice of the splice site (33), the design of a 5'-terminal extension of the ribozyme sequence, the extended guide sequence (EGS; (24,26,34), and, importantly, on the sequence of the ribozyme itself. While the ribozyme sequence evolved in nature for *cis*-splicing as opposed to *trans*-splicing, experimental evolution could be used to adapt it to *trans*-splicing. Therefore, the evolution of *trans*-splicing ribozymes in cells serves two purposes. First, the evolution itself allows determining and quantifying the factors that guide RNA evolution in cells, through controlled experiments in the lab. Second, the products of the evolution experiments - more efficient *trans*-splicing ribozymes - could be useful tools in research and therapy.

This study analyzed the effects of selection pressure and recombination on the evolution of *trans*-splicing group I intron ribozymes in *E. coli* cells. Four lines of evolution were conducted that differed in the application of selection pressure and recombination. The evolution of the fittest phenotype, which relied on five mutations, was most efficient under low selection pressure. Analysis of the sequences during evolution, and of the evolutionary intermediates between parent ribozyme and the most efficient ribozyme, found that four highly

cooperative mutations resulted in two disadvantages for evolution under high selection pressure.

2.3 Experimental procedures

Library plasmid

The library plasmid expressed a *trans*-splicing variant of the group I intron ribozyme from *Tetrahymena,* and the chloramphenicol acetyl transferase *(CAT)* pre-mRNA as described previously (24). The expression of the ribozyme was driven by a downregulated version of the IPTG-inducible *trc* promoter (35), and the ribozyme carried a 3'-terminal hairpin transcription terminator. The *CAT* pre-mRNA was encoded in the opposite direction, and its expression was driven by the constitutive promoter derived from its parent plasmid pLysS (Novagen). It carried the frameshift mutation ΔG322 (counted with the A of the start codon as position 1), which abolished CAT activity.

In vivo evolution

The *in vivo* evolution was done essentially as described[1] but with careful control of the population sizes during the selective step, the rate of mutagenesis, and the selection pressure. Briefly, electrocompetent *E. coli* DH5a cells (NEB) were transformed with library plasmids containing pools of ribozyme sequences, plated on LB plates containing 100 µg/mL ampicillin, and incubated at 37 °C for ~17 hours. The complexity of viable cells in each pool was estimated using dilution series plated on medium containing 100 µg/mL ampicillin. These *E. coli* libraries were then washed from the plates with LB medium, the OD_{600} of the resulting cell suspension was measured, and the suspension was diluted to an

OD_{600} of 0.015. Each cell suspension was induced with a final concentration of 1mM IPTG and shaken for 1 hr at 37°C. One million cells that contained plasmids with ribozyme inserts (as determined by colony PCR) were then plated on medium containing the appropriate chloramphenicol concentration. Chloramphenicol concentrations for evolutionary lines I, and III were determined by plating each selection on plates containing 3 different concentrations of chloramphenicol. The concentration corresponding to ~10% of plated clones forming visible colonies, was then used for selection. After incubation at 37 °C for ~17 hours, plasmids were isolated from the grown colonies, then the ribozyme genes were subjected to mutagenic PCR (36) or recombination using the staggered extension process (StEP) (37), and re-cloned into new library plasmids. The rate of mutagenesis was ~2.4 mutations per ribozyme corresponding to 10 rounds of mutagenic PCR (36). Five to ten plasmids were randomly selected from each population after the selective step of evolution, to determine ribozyme sequences and monitor the progress of the evolution.

Construction of ribozyme variants

The design of variants in EGS at the ribozyme 5'-termini (see figure 2.1) used the mfold algorithm (38) to predict the structures formed by the EGS with the substrate and the ribozyme 3'-terminus. Ribozyme variants with modifications in the 5'-terminal EGS were generated by PCR with 5'-PCR primers that inserted the sequence mutations using PrimeSTAR DNA polymerase (TaKaRa). Ribozyme variants with internal mutations were constructed by site directed mutagenesis (39). Briefly, each 50 µL reaction contained 1.25 pmol forward

primer, 1.25 pmol reverse primer, 2.5 nmol of each dNTP, 10µL 5x PrimeSTAR GXL Buffer, 2.5 units of PrimeSTAR GXL Polymerase, and 2 ng of template plasmid. After PCR (5 min/95°C, then 18 rounds of 50 sec/95°C, 50 sec/60°C, 4 min 45 sec/68°C and a final 7 min/68°C), 20 units of DpnI were added and the mixture was incubated at 37°C for 1 hour. Reactions were purified using the DNA Clean and Concentrate kit (Zymogen) and transformed into *E. coli* DH5α cells by electroporation.

Measurement of cell growth in liquid medium

The measurement of doubling times in liquid culture was done essentially as described (24). Briefly, fresh overnight cultures of *E. coli* DH5α cells with the library plasmid and the respective ribozyme variant in LB medium containing 100 µg/mL ampicillin were induced with 1 mM IPTG and shaken for one hour at 37°C. Each culture was diluted to an OD_{600} of 0.05 with LB medium containing 1 mM IPTG and the indicated concentration of chloramphenicol, and shaken at 37°C. The OD600 of each culture was measured every 30 minutes until it had exceeded 1.0. Growth rates were determined by least squares fitting.

Measurement of cell growth on LB-agar plates

Fresh overnight cultures in LB medium containing 100 µg/mL ampicillin were diluted to an OD_{600} of 0.0025, induced with a final concentration of 1 mM IPTG, and shaken at 37°C for 1 hour. Of each culture, 0.1 mL (~50,000 viable cells) was plated on one LB agar plate containing 100µg/mL ampiclliin and one LB agar plate containing the noted concentration of chloramphenicol and IPTG at a concentration of 1 mM. Cells were grown at 37°C for 16 hours and then each

plate was washed with 1.6 mL PBS. The OD_{600} of each suspension from plates containing chloramphenicol was measured and normalized by the OD_{600} from plates containing ampicillin. For most experiments the growth was measured on plates because the selection was done on plates, and the aim of measuring the growth of individual clones was to draw conclusions about their behaviour during the evolution.

2.4 Results

To analyze the effect of evolutionary parameters on ribozyme evolution in cells we used an *in vivo* selection system developed for the evolution of *trans*-splicing group I intron ribozymes (24). The core of this system is a plasmid encoding a gene for a *trans*-splicing variant of the *Tetrahymena* group I intron ribozyme, and an inactivated gene of chloramphenicol acetyl transferase (CAT). When both the inactivated *CAT* mRNA and the ribozyme are expressed, the ribozyme is able to repair the *CAT* mRNA using *trans*-splicing, allowing the translation of functional CAT enzyme. The CAT enzyme uses acetyl-CoA to acetylate the antibiotic chloramphenicol (40), which abolishes binding of chloramphenicol to the ribosome (41) and mediates resistance to chloramphenicol. Repair of the inactivated *CAT* mRNA was facilitated by the *trans*-splicing ribozyme, containing a 3'-exon designed to repair the 3'-terminus of the mutated *CAT* mRNA (Fig. 2.1A). Repair of *CAT* mRNA by the *trans*-splicing group I intron ribozyme facilitated growth of *E. coli* cells on medium containing chloramphenicol, allowing the selection of efficient *trans*-splicing ribozymes in *E. coli* cells. The number of cells from the *E. coli* library that was plated on selection

medium in each round of the evolution was maintained at 10^6. Increasing the concentration of chloramphenicol in the selection medium raised the selection pressure, to select for increased ribozyme efficiency. Mutations were introduced into the ribozyme genes by mutagenic PCR (42), and recombination events were generated by the PCR-based StEP technique (37). Repeated rounds of mutagenesis (or recombination) and selection allowed evolving successively more active ribozyme variants [1]. This *in vivo* evolution system was used to study the influence of selection pressure and recombination on the efficiency of the resulting ribozyme population, while keeping other evolutionary parameters constant including the rate of mutagenesis, the rate of recombination, and the population size.

As starting point for the evolution, a *trans*-splicing ribozyme was designed that recognizes splice site 97 on the *CAT* mRNA, differing from a previous evolution on splice site 177 [1]. Because each splice site requires its own optimized ribozyme 5'-terminus (the extended guide sequence, or EGS (24,26,30,34)) 12 different EGS sequences were designed and tested, to identify a 5'-terminus that could mediate efficient growth of *E. coli* cells in medium containing chloramphenicol (Fig. 2.1B, C). The constructs differed in several aspects of the EGS design: the absence and presence of a P10 helix, the length of the P1 extension (3 or 4 bp), three different registers of the internal loop, and the specific sequence of the internal loop. The sequences of these constructs were designed such that the predicted back-folding energy of the EGS was weaker than -0.9 kcal/mol because self-structure formation of the EGS would

have prevented the EGS from base pairing with the target site on the *CAT* pre-mRNA. The 12 constructs were measured for their ability to mediate chloramphenicol resistance due to the repair of *CAT* mRNA (Fig. 2.1C). Construct 8, which was predicted to form a 4-base pair P1 extension helix and a strong P10 helix, showed the best growth at two different chloramphenicol concentrations, and was chosen as the starting point for the evolutionary study.

To investigate the influence of selection pressure and recombination on the evolving ribozymes, four lines of evolution were performed in parallel, systematically varying the application of selection pressure and recombination (Fig. 2.2A). The first three rounds of evolution were performed as one single population, after which the population was divided into four lines, I-IV. In two of the four lines, I and III, high selection pressure was applied by successively raising the concentration of chloramphenicol from a level of 10 μg/mL in round 3, to 240 μg/mL in round 12 of the evolution (Fig. 2.2B). The chloramphenicol concentration was chosen in each cycle such that ~10% of the ribozyme population formed visible colonies. In the two other lines, II and IV, a low selection pressure was applied, by maintaining the chloramphenicol concentration at 10 μg/mL. Under this low selection pressure, approximately 30% of the population formed visible colonies (average for evolution rounds 8 to 12). Recombination events were applied in three of the first 12 rounds of evolution, for all four lines (Fig. 2.2A). In lines I and II, the recombination events were clustered in the last three rounds, whereas in lines III and IV they were evenly distributed throughout the 12 rounds. The progress of the evolution was followed by

sequencing 5-10 individual clones of the library for each line of evolution, during each round. The average number of mutations per ribozyme increased steadily, by about 1 mutation per ribozyme during each round of evolution (Fig. 2.2C).

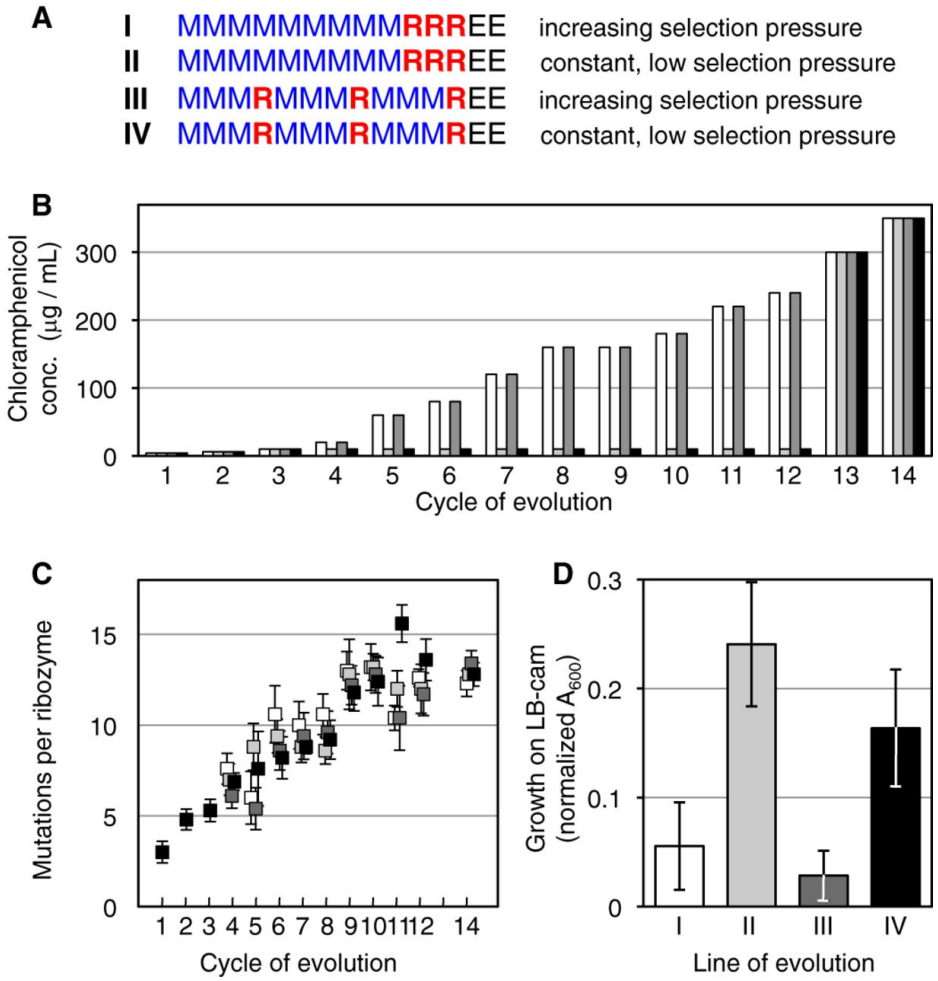


Figure 2.2 Evolution of the *trans*-splicing ribozyme under four different conditions (I - IV), which differed in selection pressure and recombination. (A) Schematic for the 12 rounds of evolution for the four separate lines. Nine rounds with mutagenesis (M, blue) and three rounds with recombination (R, red) were followed by two rounds without mutagenesis or recombination (E, black). (B) Selection pressure over the course of the evolution, given as the concentration of chloramphenicol in selection medium. The selection pressure for the lines I (white) and III (dark grey) increased with the activity of the pool, while the selection pressure for the lines II (light grey) and IV (black) was kept low, never exceeding 10 µg/mL chloramphenicol concentration. The selection pressures in rounds 13 and 14 were at the same high levels for all four lines, to select the most active ribozymes from each population. (C) The average number of mutations per ribozyme is plotted as a function of the evolution rounds, for line I (white), II (light grey), III (dark grey), and IV (black). Each value is the average from 5-10 sequences, with error bars denoting the standard error of the mean. Note that the symbols are slightly offset to clarify the error bars. (D) Activities of ribozyme pools after evolution round 14, measured as cell growth on plates containing 350 µg/mL chloramphenicol. The cell growth was normalized for growth on medium with ampicillin. Error bars are standard deviations of three experiments.

To identify the lines of evolution that generated the most efficient ribozymes, all four lines of evolution were subjected to two rounds of evolution under high selection pressure (round 13 and 14), at chloramphenicol concentrations of 300 and 350 µg/mL (Fig. 2.2A, B). These two rounds did not contain mutagenesis or recombination events so that only sequences were enriched that evolved in rounds 1-12. After these two rounds of enrichment the activity of the four populations was quantified by determining the fraction of cells that grew on plates with 350 µg/mL of chloramphenicol (Fig. 2.2D). Surprisingly, the two populations that had evolved under the lowest selection pressure showed the highest activities. The differences in recombination did not cause a significant difference between the four lines of evolution.

The mutations that generated the advantage for the two populations evolved under low selection pressure were identified by comparing sequences from lines II and IV (low selection pressure) with lines I and III (high selection pressure) (Fig. 2.3). Ten ribozymes were sequenced from each of the four populations, after evolution round 14. Mutations near the ribozyme 5'-terminus were common, with the mutation G9U dominating all four lines. Similarly, all four lines showed several mutations in the P9.2 helix. In contrast, four mutations in the P6b stem-loop clearly differentiated lines I and III from lines II and IV: U236C, U238C, U239C, and U241A. All four mutations were present in 18/20 sequences from populations II and IV, but in 0/20 in lines I and III. Three of the four mutations were in 20/20 clones from lines II and IV, and in 3/20 clones from lines I and III. Conversely, 17/20 clones from lines I and III had none or one of the four

mutations. Therefore, the increased fitness in evolutionary lines II and IV was correlated with the appearance of the four mutations in the P6b loop.

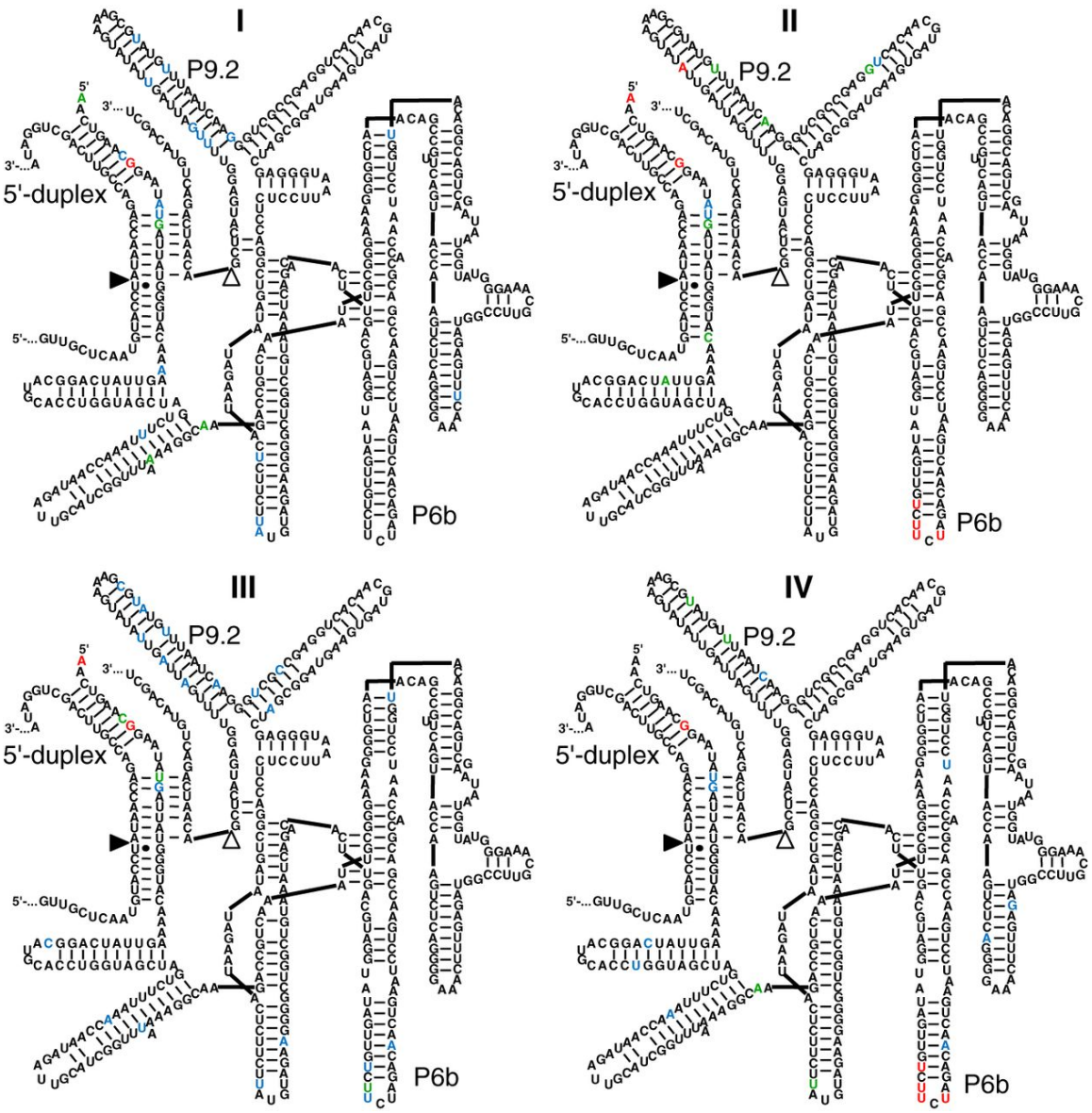


Figure 2.3 Secondary structure representations of the mutations identified after 14 rounds of evolution in each of the four lines. Note that these mutations reflect the most active ribozymes after two rounds of enrichment at high selection pressure (rounds 13 and 14). The line (I - IV) is given for each secondary structure. For each structure, 10 sequences were analyzed. The color of the nucleotide corresponds to the frequency with which the nucleotide was found mutated. Red: 10-8 mutations; Green: 7-5 mutations; Blue: 4-2 mutations; Black: 1-0 mutation. See figure 2.1 for explanations on the secondary structure. The positions of the P6b stem-loop, P9.2 stem-loop, and the 5'-duplex are indicated. Note that the mutations in the P6b loop are highly enriched in lines II and IV but not in lines I and III.

To test whether the emergence of the four clustered mutations in the P6b loop might have been helped by an uneven coverage of mutations over the length of the ribozyme we analyzed 128 mutations in 30 sequences generated by the mutagenic PCR protocol used during the evolution (Fig. 2.4). No clustering of mutations and no significant deviation from random distributions was detected. As expected, the types of mutations showed a significant bias among the six possible mutations (figure 2.4C). The two types of mutations that were necessary to generate the clustered P6b mutations (U to C and U to A) were the two most frequent types of mutations obsered. Therefore, the emergence of the P6b loop mutations was likely not aided by an uneven distribution of mutations over the ribozyme but benefitted from a mutational bias in the mutagenic PCR method.

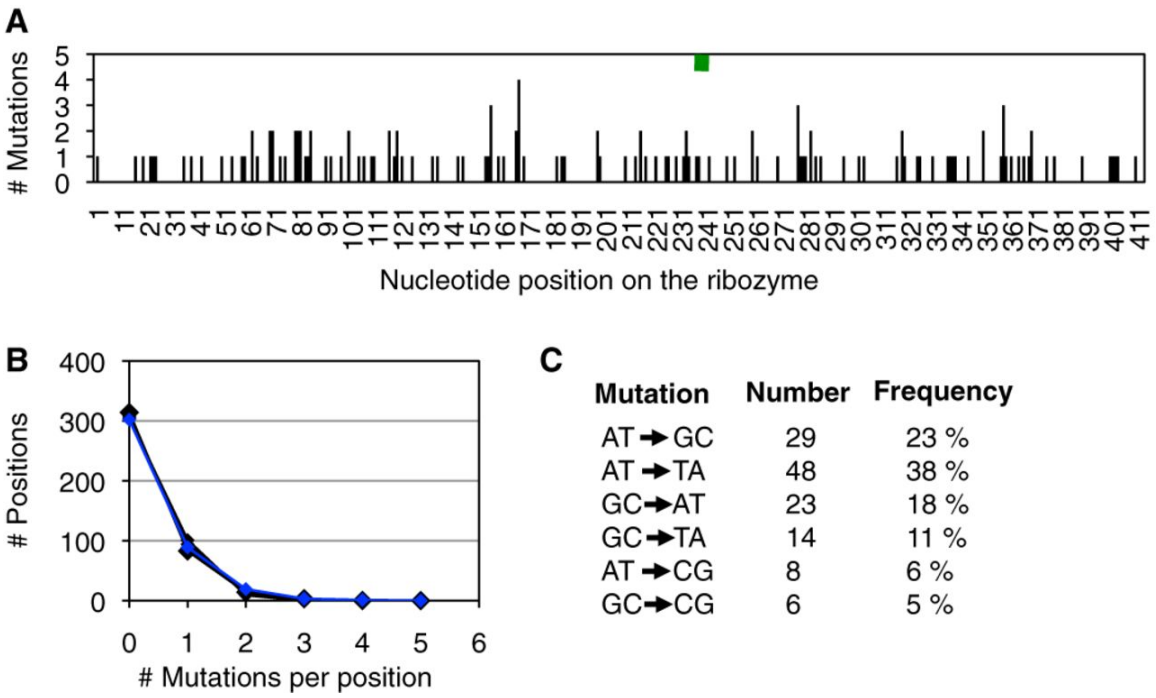


Figure 2.4 Analysis of mutations generated by PCR mutagenesis. (A) Distribution of 128 mutations generated by mutagenic PCR (blue) over the 414 nucleotide positions of the Tetrahymena ribozyme. The green rectangle shows the position of the P6b stem-loop mutations (positions 236-241). (B) Frequency of multiple mutation occurrences at the same position. Of the 414 positions, the number of positions is shown with 0, 1, 2, 3, and 4 mutations (blue). Five data sets with a random distribution are shown as comparison (black). (C) Mutational bias in the type of mutations. For each of the six types of mutations, their number of occurrences (total = 128) and their relative frequencies are shown. One insertion and 12 deletions were also detected (not included).

As a more rigorous way to identify the individual mutations that gave rise to high activity, the 40 clones isolated after round 14 of the evolution were screened for chloramphenicol resistance. In a semi-quantitative assay for growth on medium containing 100 µg/mL chloramphenicol, four ribozyme clones were found to mediate the most efficient growth, all of which contained the four mutations in the P6b stem-loop (data not shown). One of these clones, clone IV-12-10, was chosen for further analysis. Using site-directed mutagenesis, each of

the 11 point mutations present in ribozyme IV-12-10, was individually reverted to the wild-type sequence, and the resulting ribozymes were tested in growth assays (Fig. 2.5A). Four mutations were necessary for maximum growth of clone IV-12-10 (G9U, U236C, U238C, and U239C). To test whether these four mutations alone were sufficient to mediate the same chloramphenicol resistance as clone IV-12-10, they were inserted into the sequence of the parent ribozyme, generating clone M4. While the four mutations in clone M4 alone were not sufficient, a fifth mutation (U241A) completed the motif, generating clone M5, which was able to mediate growth at the level of IV-12-10 (Fig. 2.5B). These five mutations were identical to the mutations identified from the comparison of 40 sequences between evolutionary lines I-IV (see above), suggesting that the emergence of these 5 mutations in lines II and IV gave the fitness advantage to these lines.

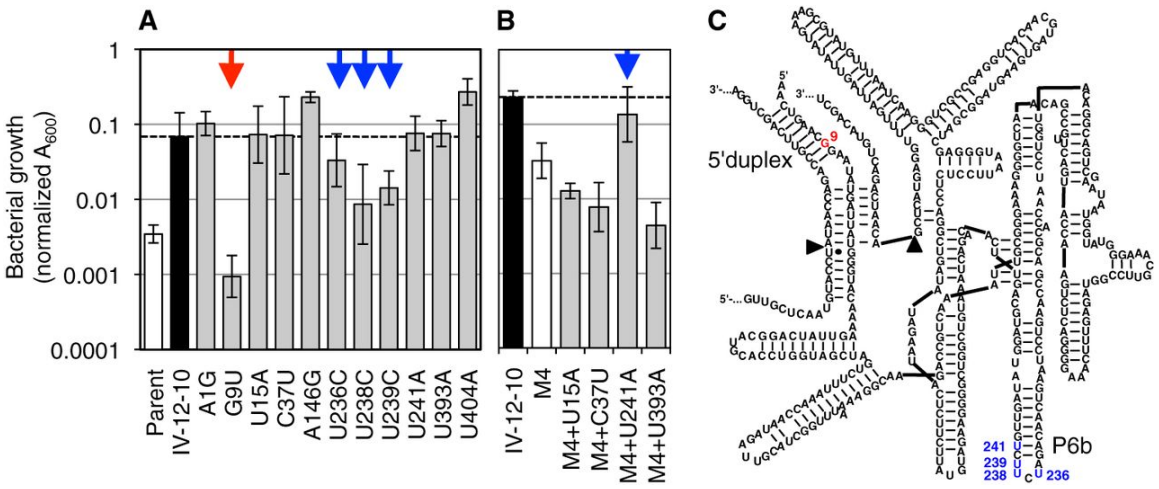


Figure 2.5 Identification of mutations that increased ribozyme activity, among the 11 mutations in clone IV-12-10. (A) Activity of clone IV-12-10 variants that carried single reversion mutations towards the parent ribozyme. Activity was measured as growth on LB-agar containing 200 µg/mL chloramphenicol. Note the logarithmic scale that shows the growth as measured as OD600. Growth of the parent ribozyme is shown as comparison, and a horizontal dashed line is shown for comparison with clone IV-12-10. The red arrow indicates a mutation reversion with >10-fold effect (G9U); the blue arrows indicate mutation reversions with < 10-fold effect. (B) Five mutations were necessary to mediate full activity. A ribozyme containing only the four mutations identified in sub-figure A (ribozyme variant M4) did not mediate full activity compared to the IV-12-10 variant. Three additional candidate mutations were added to ribozyme M4 and measured for growth in the presence of 200 µg/mL chloramphenicol. Mutation U241A, in combination with the four mutations identified in sub-figure A, was found to be necessary and sufficient for full activity observed in the IV-12-10 variant. Secondary structure of the ribozyme with the positions of beneficial mutations indicated. Color-coding is as in (A). The helices containing these mutations are labeled as 5'-duplex, and P6b.

If the five M5 mutations were responsible for the difference in fitness between the four evolutionary lines, it might be possible to detect their evolutionary precursors in lines II and IV, *before* the enrichment for the most efficient ribozymes, i.e. before round 13. To do this, the co-occurrence of M5 mutations in 80 sequences of evolution rounds 10-12 was used to identify specific evolutionary intermediates (Fig. 2.6). The majority of sequences in all four lines contained the G9U mutation, consistent with its dominance after round 14. The difference between the lines of evolution with regard to the M5 mutations

was again correlated with high and low selection pressure: In lines I and III (high selection pressure), only single mutations were detected in addition to G9U (M2). In contrast, in lines II and IV (low selection pressure) ribozyme sequences also with two and three additional mutations (M3 and M4) were found. At this point in the evolution, M5 mutants did not dominate their populations likely because the evolutionary lines had not experienced the strong selection pressure to enrich for the most efficient ribozymes, which was applied in evolution rounds 13 and 14. The increased frequency of M3 and M4 evolutionary intermediates during rounds 10-12 suggested that low selection pressure (lines II and IV) allowed a more efficient exploration of M5 mutations. A more detailed analysis of the accumulation of M5 mutations during evolution is given in Fig. 2.8.

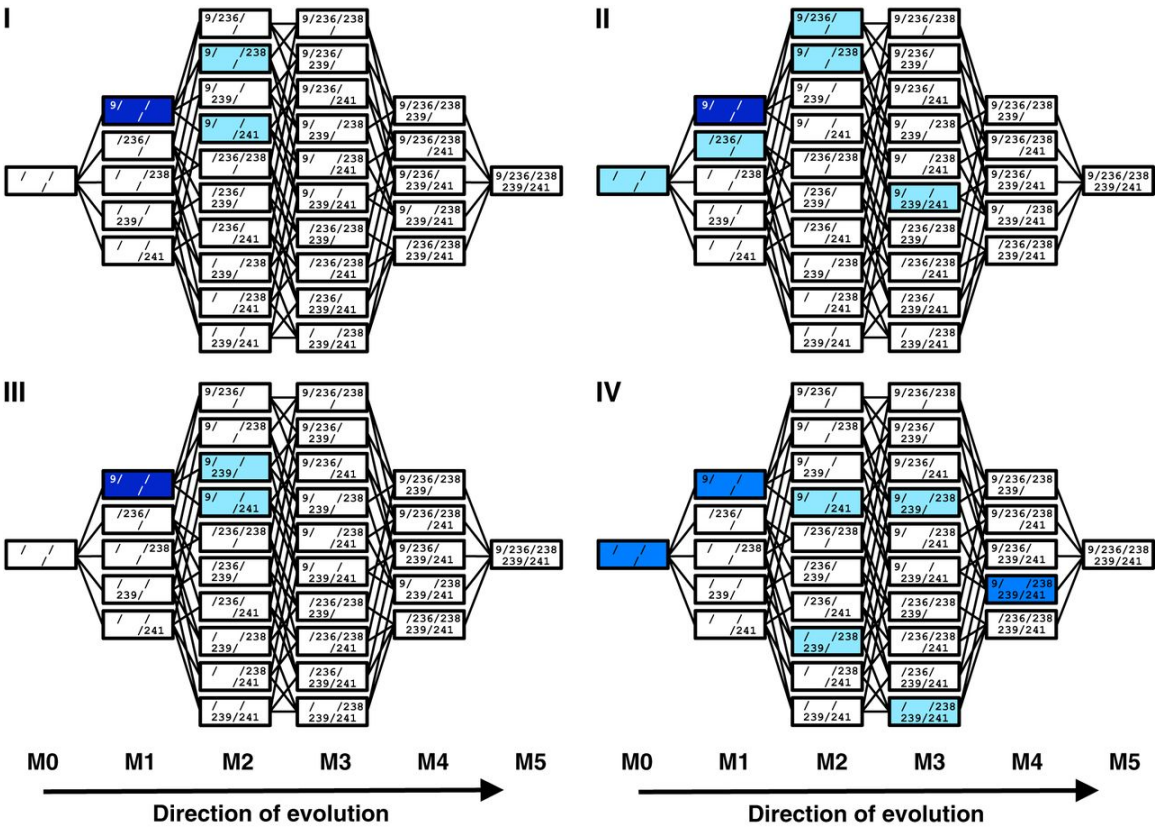


Figure 2.6 Accumulation of evolutionary intermediates of the M5 ribozyme in rounds 10-12 of the evolution. For each of the four lines of evolution (I, II, III, IV), all 32 evolutionary intermediates are shown, with each intermediate represented as one box. The individual mutations are listed inside. Parent ribozymes (M0) are shown on the left, and the 5-mutant ribozymes (M5) on the right. Colors illustrate the frequency with which the specific evolutionary intermediates were identified among 20 clones for each line of evolution. Light blue: 1-2 clones; blue: 3-10 clones; dark blue: 11-20 clones.

To understand how low selection pressure could have aided the evolution from the parent ribozyme to the M5 ribozyme we constructed all 30 evolutionary intermediates between the wild type ribozyme and the M5 ribozyme (see Fig. 2.6). The activities of these evolutionary intermediates were quantified using the same conditions as were used during the evolution (Fig. 2.7). Bacterial growth was measured on plates containing three different concentrations of chloramphenicol, corresponding to low selection pressure (10 μg/mL; Fig. 2.7A),

medium selection pressure (100 μg/mL; Fig. 2.7B), and high selection pressure (200 μg/mL; Fig. 2.7C). The resulting fitness profiles confirmed that the first step of evolution, from parent ribozyme (M0) to the G9U mutant, was of large fitness benefit at low and medium selection pressure. This was consistent with the observation that mutation G9U quickly swept through all four evolving populations (Fig. 2.8).

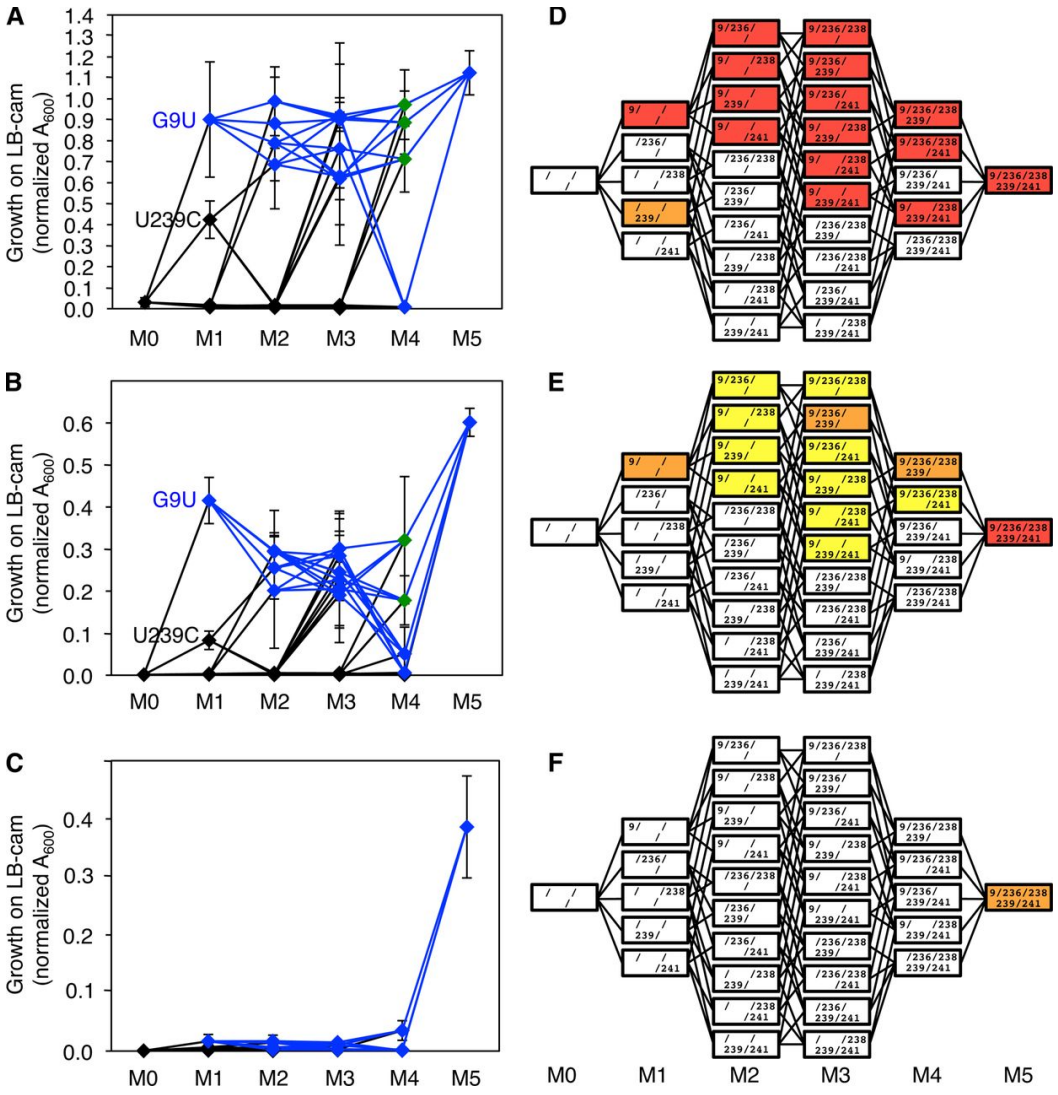


Figure 2.7 Fitness profile of all evolutionary intermediates from the parent ribozyme (M0) to the M5 ribozyme (M5), at three selection pressures. Intermediates are labeled according the their number of M5 mutations (M1, M2, M3, M4). The selection pressures correspond to chloramphenicol concentrations of (A) 10 μg/mL, (B) 100 μg/mL, and (C) 200 μg/mL. The fitness was measured as growth on medium containing the respective chloramphenicol concentration. The two M1 ribozymes that display significant growth are labeled (G9U and U239C). All evolutionary intermediates with the G9U mutation are shown in blue or green, and connected by blue lines. Green symbols highlight the 'gatekeeper' intermediates (see text). The values for all intermediates at all selection pressures are given in figure 2.9. Error bars are standard deviations from three biological experiments. The corresponding heat maps are displayed in (D) 10 μg/mL, (E) 100 μg/mL, and (F) 200 μg/mL. The arrangement of evolutionary intermediates is the same as in figure 2.6. Colors denote fitness values larger than 0.6 (red), 0.6-0.3 (orange), 0.3-0.1 (yellow), and smaller than 0.1 (white).

In contrast, the remaining four mutations in the P6b stem-loop acted highly cooperatively because when they were added to the G9U mutation (blue lines in figure 2.7) they did not mediate a fitness increase for any one, two, or three additional mutations (M2, M3, and M4, respectively). Only when all four mutations were combined in the M5 ribozyme was there a strong increase in activity. The M2, M3, and M4 intermediates even appeared to show a slight reduction in fitness relative to the G9U single mutant (Fig. 2.7B). In addition, several of the M4 evolutionary intermediates mediated only weak growth or no growth, such that at low and medium selection pressure, the remaining three and two evolutionary intermediates acted as 'gatekeepers' on the path from the M3 intermediates to the M5 ribozyme (green symbols in figure 2.7 A, B). Lastly, the high cooperativity of the mutations resulted in very low activity of all evolutionary intermediates at high selection pressure (Fig. 2.7C). This means that the M1-M4 evolutionary intermediates would not allow access to the M5 ribozyme at high selection pressure, which corresponded to the final 2-5 rounds of evolution for the evolutionary lines I and III (200 µg/mL; Fig. 2.2B).

The fitness of all evolutionary intermediates (Fig. 2.7 D,E,F) outlines the evolutionary paths that were accessible at different selection pressures. They can be used to explain why specific evolutionary intermediates were enriched during the evolution of the four strains (figure 2.6). Only two evolutionary intermediates in line IV did not appear to follow this pattern: 239/239 and 238/239/241. Both could be explained by recombination events: Intermediate 238/239 could have been generated by recombinational loss of mutation at

position 9 from the detected intermediate 9/238/239. Intermediate 238/239/241 could have been generated similarly from intermediate 9/238/239/241. In both cases, the large distance to position 9 made these transitions likely. In summary, the high cooperativity of the four mutations in the P6b stem-loop generated several constraints on the emergence on this motif that were exacerbated under high selection pressure, giving an advantage to their evolutionary emergence under low selection pressure.

2.5 Discussion

To study the effect of evolutionary parameters on RNA evolution *in vivo* we evolved four ribozyme populations, systematically varying the selection pressure and the pattern of recombination events. Lower selection pressure led to ribozyme populations with higher fitness. To determine the cause for this behavior, the ribozyme sequences were compared between the evolutionary lines. Five mutations (M5) mediating high activity were identified, and all 30 evolutionary intermediates between the parent ribozyme and the M5 ribozyme were constructed and tested for activity. Four mutations in the P6b loop of the ribozyme were highly cooperative, suggesting an explanation for the success of the evolutionary lines under low selection pressure.

Our study identified two factors that appear to have contributed to low selection pressure aiding the evolution of the four highly cooperative P6b stem-loop mutations. First and most importantly, increased selection pressure strongly reduced the survival of all evolutionary intermediates (Fig. 2.7). High selection pressure (200 μg/mL chloramphenicol) dramatically reduced the growth of all

evolutionary intermediates, with exception of the M5 ribozyme (Fig. 2.7C,F). This made the M5 ribozyme evolutionarily inaccessible at high selection pressure. At intermediate selection pressure (100 μg/mL chloramphenicol), the fitness profile generated a slight valley in the fitness landscape for the evolutionary intermediates M2, M3, and M4 relative to the M1 intermediate G9U (Fig. 2.7B,E). Such a valley generates a disincentive for the population to enter these pathways, and larger drops in activity can even prevent populations from crossing the fitness valleys (43). In other words, if the evolving ribozyme populations I, and III had not reached the M5 ribozyme in evolution round 7, it would have become unlikely to traverse to the M5 ribozyme, and at the evolutionary round 11 the evolution of the M5 ribozyme would have become nearly impossible (Fig. 2.2B).

The second, but probably less important advantage of low selection pressure was in the larger population sizes at low selection pressure. Under high selection pressure, only ~10% of the cells formed visible colonies, whereas this fraction was ~30% under low selection pressure. This generated ~3-fold larger effective population sizes at low selection pressure. Large population sizes were important to evolve the four clustered mutations due to their high cooperativity, which required sampling the four-mutation sequence space without fitness benefit for any three of the four mutations. Because the ribozyme had a length of 414 nucleotides, this sequence space ($414 \cdot 413 \cdot 412 \cdot 411 \sim 10^{10}$) was significantly larger than the population size of 10^6 cells that were plated in each selection step. The mutations, therefore, had to successively accumulate over

multiple rounds of evolution, and the effective population sizes became limiting factors in the accumulation of the mutations. This constraint was strengthened by the low activity of several M4 intermediates under low and intermediate selection pressure, which reduced the number of possible evolutionary pathways to only two or three M4 'gatekeeper' intermediates that could lead from M3 intermediates to the M5 ribozyme (green symbols, Fig. 2.7). Therefore, the characteristics of the fitness profiles under different selection pressures generated a requirement for large population sizes, which was satisfied better at low than at high selection pressure.

Why did the four P6b stem-loop mutations act cooperatively? The same four mutations were identified in a previous study to increase the ribozyme efficiency in *E. coli* cells [1]. These four mutations did not increase the *in vitro* *trans*-splicing efficiency of the ribozymes. Instead, they specifically bound the transcription termination factor Rho (in *E. coli* cell lysate), and increased the assembly of polysomes and the translation of the *trans*-spliced mRNA. Because Rho regulates the expression of many RNAs (44) these data suggested that the four P6b stem-loop mutations evolved to modulate the expression of its splicing product, the *CAT* mRNA. The mutations in the P6b stem-loop could have recruited Rho because Rho binds $(C)_7$ and $(C)_8$ sequences with micromolar affinity (45), poly(C) acts inhibitory to Rho function (46), and three of the four P6b mutations in the M5 ribozyme (U236C, U238C, and U238C) generated a $(C)_5$ oligomer (Fig. 2.5). The fourth mutation (U241A) may make the $(C)_5$ sequence more accessible for interaction with Rho because it is predicted to increase the

size of the P6b loop (70). This model is consistent with a cooperative behavior of the four P6b stem-loop mutations because the lack of any of these mutations could reduce the accessibility of the loop or reduce the length of the oligo(C) sequence, and thereby drastically reduce the affinity to Rho due to the length dependence of Rho binding to oligo(C) sequences (45).

Recombination did not show a strong effect on the outcome of the evolution. After enriching for the most active ribozyme variants in rounds 13 and 14 of the evolution, the activities of lines I and III were not significantly different, as were the activities between lines II and IV (Fig. 2.2D). Similarly, the mutations of the M5 motif that were explored in line I were almost identical to those of line III (Figs. 2.3 and 2.6). The lack of an effect by recombination on the ribozymes with the M5 mutations can be explained by the omnipresence of the G9U mutation (if all sequences contain the same mutation then recombination cannot generate a difference), and by the clustering of the four mutations in the P6b loop within six nucleotides (recombination is unlikely between closely spaced mutations).

The 12 designed 5'-terminal extended guide sequences (EGSs) showed very different activities in cells, highlighting that the design principles for an EGS of *trans*-splicing ribozymes are not yet fully understood (24,26,34). EGS 8, which showed the highest activity and was chosen as the starting point for the evolution, was predicted to form a P10 helix with 6 base pairs, which was consistent with earlier studies (26,34) and supports the interpretation that the benefit of a P10 helix is dependent on the splice site (24). Interestingly, the

evolved, strongly beneficial mutation G9U truncates the 5'-duplex from 8 base pairs to 6 or 7 base pairs and may increase the size of the adjacent internal loop. These results suggested that currently the most reliable way to identify the optimal EGS for a given splice site is by a combined approach between design and an *in vivo* selection procedure (24).

Previous *in vitro* evolution studies used high selection pressure to generate more efficient catalytic RNAs, which stands in contrast to the central finding of our study. High selection pressure is expected to help enriching the most efficient phenotype because it efficiently removes less active phenotypes from the population. This is especially clear for *selection* experiments, where all sequence diversity is contained in the starting population, and the task is to identify the most efficient sequence of that population, with selection steps as stringent and as few as possible (47-49). In contrast, *evolution* experiments usually do not contain the fittest phenotype in the starting population, and it is necessary to accumulate multiple mutations over successive cycles of mutagenesis - selection - amplification in order to access the fittest individual (50-53). Therefore, in evolution studies the fitness of evolutionary intermediates and the roughness of the fitness landscape become important. In a smooth fitness landscape, when each successive mutation increases fitness until the fitness peak is reached, high selection stringency helps a fast climb to the peak. However, at least some RNA fitness landscapes are rough (54), where high selection pressure would doom a population with low diversity to extinction (55). A larger genetic diversity speeds up the evolution of ribozymes (19) but the

mutational load that is used to generate this diversity can lead to the extinction of populations, especially at small population sizes (20). This illustrates that a combination of several evolutionary parameters determines the benefit of low or high selection pressure. Our evolution in *E. coli* cells combined several factors that could present an obstacle for evolution at high selection pressures: The effective population sizes (~100,000 - 300,000) were several orders of magnitude lower than those of most *in vitro* experiments, the mutagenic rate was only 3-fold below a level that previously led to extinction (70), and the cooperativity of the four P6b stem-loop mutations required the exploration of a four-mutation sequence space without gain in activity for the ribozymes with 1-4 mutations. Future experiments in different evolution model systems and under varying conditions are necessary to determine more generally when low selection pressure benefits the evolution of fitter phenotypes.

Previous studies selected RNAs in cells (56-60) but did not include the multiple rounds of mutagenesis and selection. In contrast to *selection* experiments, *evolution* experiments require multiple rounds of mutagenesis, selection, and amplification, such as in the current study. The experimental system employed in this study evolved a single RNA molecule inside cells, in a constant genetic background. This setup allowed a stringent control of the evolutionary parameters, the application of high mutation rates, and a relatively simple analysis and interpretation of the results. In contrast, the sequencing of complete genomes made it possible to analyze mutations in *E. coli* populations that evolved under experimental conditions for 2,000 generations in the lab

(61,62). Here, the identification of the mutations that cause the improved phenotype is quite laborious (63), creating an obstacle for experiments that evolve the complete genome. The simpler approach described in this study allows answering specific questions that can be addressed by following the evolution of specific macromolecules in cells.

Trans-splicing group I intron ribozymes were originally developed for the possible use in therapeutic applications (23). These ribozymes could be employed for the treatment of genetic disorders by repairing the mutations on the level of mRNAs (25,26,28,29), and for the selective killing of cancerous or virally infected cells by splicing toxin-eoncoding RNA sequences into hTert mRNA or viral RNAs, respectively (30-32). The evolution experiments described here did not directly generate ribozymes that could be used in therapy because the most efficient ribozymes appear to rely on interactions with the bacterial protein Rho [1], which does not exist in human cells. However, similar evolution experiments can now be carried out in human cell lines. The results of the current study suggested that this would be done most efficiently by evolution experiments under low selection pressure, to allow for the enrichment of highly cooperative mutations.

Studies of protein evolution have resulted in several findings that can be compared to our study of RNA evolution. One study focused on the evolution of bacterial ß-lactamase, in which a 5-mutation variant mediated resistance to the antibiotic cefatoxin (64). By testing all 32 evolutionary intermediates between the wild type and the 5-mutation variant, it was found that only a fraction of the evolutionary paths were accessible under the used conditions, mirroring the

results in our study (Fig. 2.7). However, the epigenetic interactions that generated these paths were pleiotropic effects such as increased protein aggregation and reduced thermodynamic stability, in contrast to the cooperativity of RNA mutations in our study, which was probably caused by a single factor, binding to the Rho protein (70). Other studies found that catalytic promiscuity can help in the evolution of a new function both in proteins (65,66) and in RNAs (67). Currently the most powerful experimental system to study macromolecular evolution appears to be phage-assisted continuous evolution (PACE) (68). Here the evolving molecule, usually T7 RNA polymerase, is repeatedly selected for high activity inside *E. coli* cells, allowing for the completion of hundreds of evolution rounds within a few days. With this technique, populations that were evolved under low selection stringency followed by high selection stringency reproducibly arrived at different sets of mutations than when evolution was done at high stringency alone, similar to the results of our study (69). Additionally, low selection stringency generated larger genetic diversity, which appeared to be the case in our study as well. Future studies will show to what extent the different chemistries of RNAs and proteins cause different evolutionary behavior.

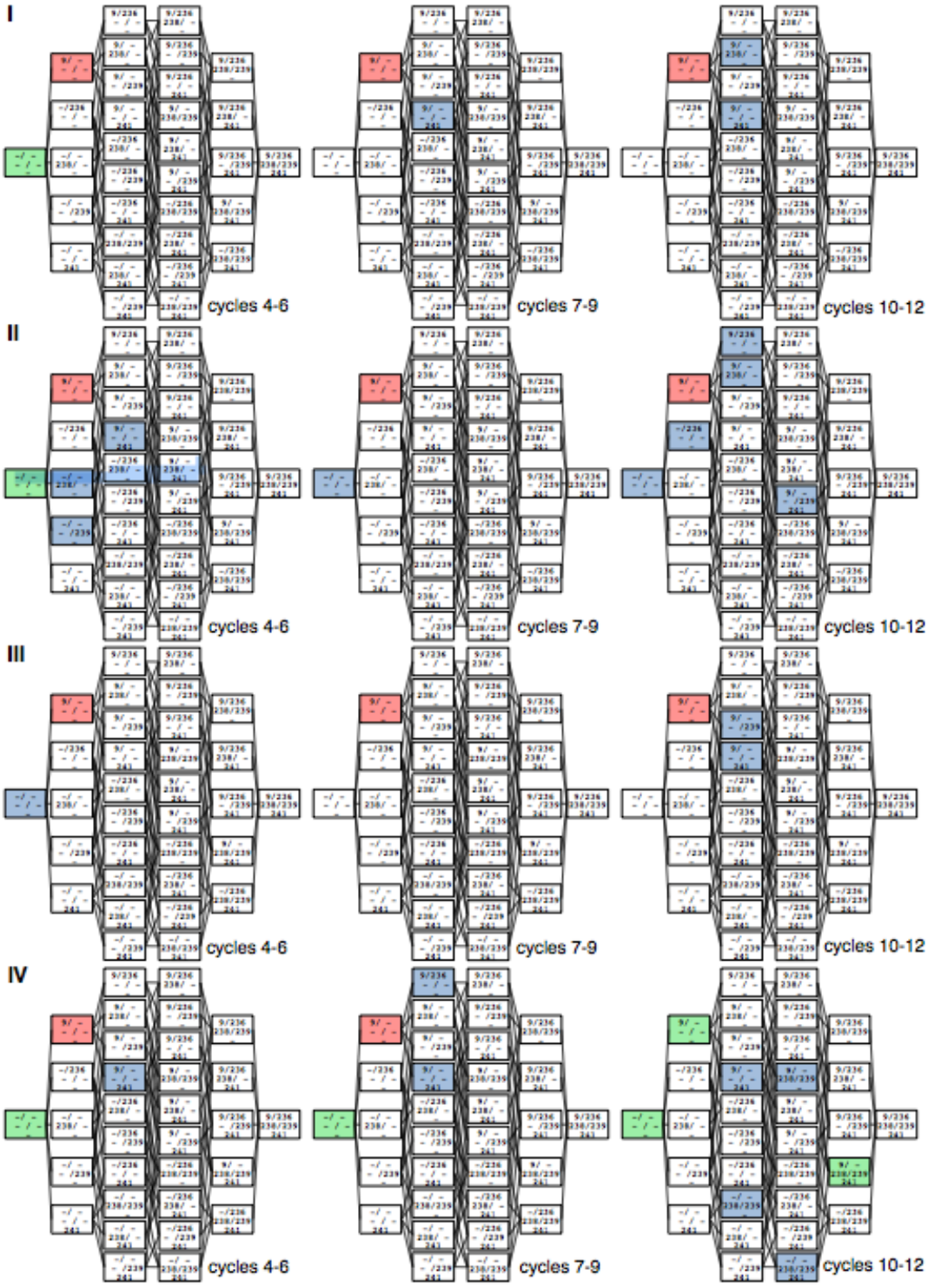
2.6 Acknowledgements

We thank Karen Olson and Gregory Dolan for helpful discussions. William Sinko is thanked for computational help for the identification of mutations. Funding was provided by the NIH molecular biophycis training grant to Susan Taylor (T32 GM008326) and two grants by the Hellman Family foundation to U.F.M. (2011/12 and 2012/13).

Chapter 2, in full, is a reprint of the material as it appears in the Journal of Biological Chemistry, **Amini Z.N.** and Muller U.F. (2013) Low selection pressure aids the evolution of cooperative ribozyme mutations in cells. J Biol Chem. 288(46): 33096-33106. The dissertation author is the first author on this paper.

2.7 Supporting Information

Figure 2.8 Accumulation of evolutionary intermediates for the five mutations of the M5 ribozyme. The figure separates the accumulation in lines I- IV. All 32 evolutionary intermediates are shown for three intervals of the evolution, representing cycles 3-5 (left), cycles 7-9 (middle), and cycles 10-12 (right). For each 32 evolutionary intermediates, the parent ribozyme is shown on the left and the 5-mutant ribozyme on the right. Each box represents one evolutionary intermediate, with the individual mutations listed inside. Colors represent the frequency with which the evolutionary intermediates were observed. Red: 51-100%; green: 11-50%; Blue: 1-10%; No color: <1%.


I
cycles 4-6
cycles 7-9
cycles 10-12
II
cycles 4-6
cycles 7-9
cycles 10-12
III
cycles 4-6
cycles 7-9
cycles 10-12
IV
cycles 4-6
cycles 7-9
cycles 10-12

	10ug/mL cam		100ug/mL cam		200ug/mL cam	
Mutation	**Average**	**StDev**	**Average**	**StDev**	**Average**	**StDev**
Parent	0.0303	0.0217	0.0012	0.0005	0.0005	0.0004
M9	0.9016	0.2750	0.4154	0.0549	0.0169	0.0116
M236	0.0172	0.0069	0.0017	0.0005	0.0012	0.0004
M238	0.0136	0.0031	0.0031	0.0019	0.0025	0.0016
M239	0.4239	0.0892	0.0830	0.0218	0.0065	0.0041
M241	0.0071	0.0006	0.0018	0.0011	0.0012	0.0005
M9/236	0.7896	0.3140	0.2014	0.1373	0.0065	0.0039
M9/238	0.8830	0.0957	0.2959	0.0959	0.0171	0.0103
M9/239	0.6872	0.0778	0.2553	0.0742	0.0145	0.0041
M9/241	0.9888	0.1640	0.2938	0.0384	0.0024	0.0017
M236/238	0.0112	0.0028	0.0014	0.0008	0.0011	0.0002
M236/239	0.0104	0.0009	0.0015	0.0015	0.0005	0.0002
M236/241	0.0054	0.0032	0.0009	0.0002	0.0009	0.0002
M238/239	0.0163	0.0058	0.0051	0.0049	0.0047	0.0011
M238/241	0.0071	0.0023	0.0013	0.0004	0.0016	0.0008
M239/241	0.0151	0.0048	0.0026	0.0027	0.0014	0.0007
M9/236/238	0.6198	0.3173	0.2050	0.1275	0.0099	0.0020
M9/236/239	0.7624	0.2431	0.3012	0.0834	0.0101	0.0014
M9/236/241	0.9211	0.3458	0.2273	0.1152	0.0034	0.0011
M9/238/239	0.6309	0.2315	0.1906	0.0782	0.0149	0.0018
M9/238/241	0.9022	0.2634	0.2450	0.1271	0.0033	0.0005
M9/239/241	0.9143	0.0687	0.2840	0.1063	0.0021	0.0005
M236/238/239	0.0162	0.0057	0.0048	0.0026	0.0029	0.0010
M236/238/241	0.0058	0.0018	0.0024	0.0021	0.0015	0.0002
M236/239/241	0.0016	0.0004	0.0028	0.0014	0.0021	0.0006
M238/239/241	0.0046	0.0001	0.0011	0.0004	0.0015	0.0007
M9/236/238/239	0.7129	0.1573	0.3212	0.1512	0.0354	0.0169
M9/236/238/241	0.9725	0.1658	0.1780	0.0585	0.0020	0.0009
M9/236/239/241	0.0087	0.0012	0.0058	0.0005	0.0015	0.0006
M9/238/239/241	0.8869	0.1506	0.0509	0.0634	0.0010	0.0000
M236/238/239/241	0.0079	0.0045	0.0011	0.0008	0.0016	0.0007
M9/236/238/239/241	1.1242	0.1053	0.6015	0.0332	0.3860	0.0878

Figure 2.9 Values for the growth of *E. coli* cells on LB medium containing chloramphenicol, mediated by all evolutionary intermediates between the parent ribozyme and the M5 ribozyme (M9/236/238/239/241). Values were obtained by growing cells on LB-agar plates, washing the plates, and measuring the OD600. Each value was normalized for the respective clone's growth on LB medium containing 100 µg/mL ampicillin. Details are given in the legend of figure 6 and the materials and methods. Standard deviations are from three biological replicates.

2.8 References

1. Noller, H. F., Hoffarth, V., and Zimniak, L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* 256, 1416-1419

2. Nissen, P., Hansen, J., Ban, N., Moore, P. B., and Steitz, T. A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* 289, 920-930

3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849-857

4. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) Splicing of messenger RNA precursors. *Annu Rev Biochem* 55, 1119-1150

5. Valadkhan, S. (2005) snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol* 9, 603-608

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

7. van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell* 44, 225-234

8. Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. (1986) Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* 14, 3627-3640

9. Buzayan, J. M., Gerlach, W. L., and Bruening, G. (1986) Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc Natl Acad Sci U S A* 83, 8859-8862

10. Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol* 62, 2674-2679

11. Saville, B. J., and Collins, R. A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* 61, 685-696

12. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* 428, 281-286

13. Roth, A., Weinberg, Z., Chen, A. G. Y., Kim, P. B., Amer, T. D. A., and Breaker, R. R. (2013) A novel class of self-cleaving ribozymes is prevalent in many species of bacteria and eukarya. *Nature Chemical Biology* accepted

14. Dayie, K. T., and Padgett, R. A. (2008) A glimpse into the active site of a group II intron and maybe the spliceosome, too. *RNA* 14, 1697-1703

15. Koonin, E. V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct* 1, 22

16. Walker, S. C., and Engelke, D. R. (2006) Ribonuclease P: the evolution of an ancient RNA enzyme. *Crit Rev Biochem Mol Biol* 41, 77-102

17. Goddard, M. R., and Burt, A. (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* 96, 13880-13885

18. Haugen, P., Runge, H. J., and Bhattacharya, D. (2004) Long-term evolution of the S788 fungal nuclear small subunit rRNA group I introns. *RNA* 10, 1084-1096

19. Hayden, E. J., Ferrada, E., and Wagner, A. (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474, 92-95

20. Soll, S. J., Diaz Arenas, C., and Lehman, N. (2007) Accumulation of deleterious mutations in small abiotic populations of RNA. *Genetics* 175, 267-275

21. Voytek, S. B., and Joyce, G. F. (2009) Niche partitioning in the coevolution of 2 distinct RNA enzymes. *Proc Natl Acad Sci U S A* 106, 7780-7785

22. Inoue, T., and Cech, T. R. (1985) Secondary structure of the circular form of the Tetrahymena rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci U S A* 82, 648-652

23. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* 371, 619-622

24. Olson, K. E., and Muller, U. F. (2012) An in vivo selection method to optimize trans-splicing ribozymes. *RNA* 18, 581-589

25. Rogers, C. S., Vanoye, C. G., Sullenger, B. A., and George, A. L., Jr. (2002) Functional repair of a mutant chloride channel using a trans-splicing ribozyme. *J Clin Invest* 110, 1783-1789

26. Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes. *RNA* 9, 1254-1263

27. Fiskaa, T., and Birgisdottir, A. B. (2010) RNA reprogramming and repair based on trans-splicing group I ribozymes. *N Biotechnol* 27, 194-203

28. Watanabe, T., and Sullenger, B. A. (2000) Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc Natl Acad Sci U S A* 97, 8490-8494

29. Kastanos, E., Hjiantoniou, E., and Phylactou, L. A. (2004) Restoration of protein synthesis in pancreatic cancer cells by trans-splicing ribozymes. *Biochem Biophys Res Commun* 322, 930-934

30. Ayre, B. G., Kohler, U., Goodman, H. M., and Haseloff, J. (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc Natl Acad Sci U S A* 96, 3507-3512

31. Ryu, K. J., Kim, J. H., and Lee, S. W. (2003) Ribozyme-mediated selective induction of new gene activity in hepatitis C virus internal ribosome entry site-expressing cells by targeted trans-splicing. *Mol Ther* 7, 386-395

32. Song, M. S., and Lee, S. W. (2006) Cancer-selective induction of cytotoxicity by tissue-specific expression of targeted trans-splicing ribozyme. *FEBS Lett* 580, 5033-5043

33. Meluzzi, D., Olson, K. E., Dolan, G. F., Arya, G., and Muller, U. F. (2012) Computational prediction of efficient splice sites for trans-splicing ribozymes. *RNA* 18, 590-602

34. Kohler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery. *J Mol Biol* 285, 1935-1950

35. Weiss, D. S., Chen, J. C., Ghigo, J. M., Boyd, D., and Beckwith, J. (1999) Localization of FtsI (PBP3) to the septal ring requires its membrane anchor, the Z ring, FtsA, FtsQ, and FtsL. *J Bacteriol* 181, 508-520

36. Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 2, 28-33

37. Zhao, H., Giver, L., Shao, Z., Affholter, J. A., and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol* 16, 258-261

38. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415

39. Weiner, M. P., and Costa, G. L. (1994) Rapid PCR site-directed mutagenesis. *PCR Methods Appl* 4, S131-136

40. Shaw, W. V. (1967) The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant Escherichia coli. *J Biol Chem* 242, 687-693

41. Tritton, T. R. (1979) Ribosome-chloramphenicol interactions: a nuclear magnetic resonance study. *Arch Biochem Biophys* 197, 10-17

42. Cadwell, R. C., and Joyce, G. F. (1994) Mutagenic PCR. *PCR Methods Appl* 3, S136-140

43. Weissman, D. B., Desai, M. M., Fisher, D. S., and Feldman, M. W. (2009) The rate at which asexual populations cross fitness valleys. *Theor Popul Biol* 75, 286-300

44. Cardinale, C. J., Washburn, R. S., Tadigotla, V. R., Brown, L. M., Gottesman, M. E., and Nudler, E. (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science* 320, 935-938

45. Wang, Y., and von Hippel, P. H. (1993) Escherichia coli transcription termination factor rho. II. Binding of oligonucleotide cofactors. *J Biol Chem* 268, 13947-13955

46. Walstrom, K. M., Dozono, J. M., and von Hippel, P. H. (1997) Kinetics of the RNA-DNA helicase activity of Escherichia coli transcription termination factor rho. 2. Processivity, ATP consumption, and RNA binding. *Biochemistry* 36, 7993-8004

47. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505-510

48. Ellington, A. D., and Szostak, J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346, 818-822

49. Bartel, D. P., and Szostak, J. W. (1993) Isolation of new ribozymes from a large pool of random sequences. *Science* 261, 1411-1418

50. Lehman, N., and Joyce, G. F. (1993) Evolution in vitro of an RNA enzyme with altered metal dependence. *Nature* 361, 182-185

51. Wright, M. C., and Joyce, G. F. (1997) Continuous in vitro evolution of catalytic function. *Science* 276, 614-617

52. Ordoukhanian, P., and Joyce, G. F. (1999) A molecular description of the evolution of resistance. *Chem Biol* 6, 881-889

53. Schmitt, T., and Lehman, N. (1999) Non-unity molecular heritability demonstrated by continuous evolution in vitro. *Chem Biol* 6, 857-869

54. Lehman, N., Donne, M. D., West, M., and Dewey, T. G. (2000) The genotypic landscape during in vitro evolution of a catalytic RNA: implications for phenotypic buffering. *J Mol Evol* 50, 481-490

55. Wright, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the sixth international congress of genetics* 1, 356-366

56. Guo, F., and Cech, T. R. (2002) In vivo selection of better self-splicing introns in Escherichia coli: the role of the P1 extension helix of the Tetrahymena intron. *RNA* 8, 647-658

57. Ayre, B. G., Kohler, U., Turgeon, R., and Haseloff, J. (2002) Optimization of trans-splicing ribozyme efficiency and specificity by in vivo genetic selection. *Nucleic Acids Res* 30, e141

58. Unwalla, H. J., Li, H., Li, S. Y., Abad, D., and Rossi, J. J. (2008) Use of a U16 snoRNA-containing ribozyme library to identify ribozyme targets in HIV-1. *Mol Ther* 16, 1113-1119

59. Wieland, M., and Hartig, J. S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl* 47, 2604-2607

60. Chen, X., Denison, L., Levy, M., and Ellington, A. D. (2009) Direct selection for ribozyme cleavage activity in cells. *RNA* 15, 2035-2045

61. Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., and Cooper, T. F. (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332, 1193-1196

62. Cooper, T. F., and Lenski, R. E. (2010) Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol Biol* 10, 11

63. Chou, H. H., Chiu, H. C., Delaney, N. F., Segre, D., and Marx, C. J. (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332, 1190-1192

64. Weinreich, D. M., Delaney, N. F., Depristo, M. A., and Hartl, D. L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312, 111-114

65. O'Brien, P. J., and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol* 6, R91-R105

66. Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C., and Tawfik, D. S. (2005) The 'evolvability' of promiscuous protein functions. *Nat Genet* 37, 73-76

67. Lau, M. W., and Unrau, P. J. (2009) A promiscuous ribozyme promotes nucleotide synthesis in addition to ribose chemistry. *Chem Biol* 16, 815-825

68. Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011) A system for the continuous directed evolution of biomolecules. *Nature* 472, 499-503

69. Leconte, A. M., Dickinson, B. C., Yang, D. D., Chen, I. A., Allen, B., and Liu, D. R. (2013) A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* 52, 1490-1499

70. Olson, K. E., Dolan, G. F., and Müller, U. F. (2014) In vivo evolution of a catalytic RNA couples trans-splicing to translation. *PLoS One* 9, e86473

Chapter 3: Spliceozymes: Ribozymes that Remove Introns from Pre-mRNAs in Trans

3.1 Abstract

Group I introns are pre-mRNA introns that do not require the spliceosome for their removal. Instead, they fold into complex three-dimensional structures and catalyze two transesterification reactions, thereby excising themselves and joining the flanking exons. These catalytic RNAs (ribozymes) have been modified previously to work in *trans*, whereby the ribozymes can recognize a splice site on a substrate RNA and replace the 5'- or 3'-portion of the substrate. Here we describe a new variant of the group I intron ribozyme from *Tetrahymena* that recognizes two splice sites on a substrate RNA, removes the intron sequences between the splice sites, and joins the flanking exons, analogous to the action of the spliceosome. This 'group I spliceozyme' functions *in vitro* and *in vivo*, and it is able to mediate a growth phenotype in *E. coli* cells. The intron sequences of the target pre-mRNAs are constrained near the splice sites but can carry a wide range of sequences in their interior. Because the splice site recognition sequences can be adjusted to different splice sites, the spliceozyme may have the potential for wide applications as tool in research and therapy.

3.2 Introduction

Group I introns are intervening sequences in pre-mRNAs that do not require the spliceosome for their removal (1). Instead, they fold into three-dimensional structures and catalyze two transphosphorylation reactions resulting

in their excision from the primary transcript and the joining of their flanking exons (1,2,3). The biochemistry of group I intron ribozymes has been studied extensively (for review, see (4,5)) and crystal structures have been reported for group I introns from three different species (6,7,8).

Group I intron ribozymes have been engineered to catalyze several non-native reactions. For example, they can be converted from the natural *cis*-splicing format into a *trans*-splicing format, by removing the ribozyme 5'-exon and replacing it with a short substrate recognition sequence (9,10). This substrate recognition sequence base pairs to a complementary target site on the substrate RNA, thereby specifying the splice site (11,12). In the subsequent reaction, the ribozyme replaces the 3'-portion of the substrate RNA with its own 3'-exon; this has been demonstrated *in vitro* and in bacterial and mammalian cells (10,13,14). In addition to replacing the substrate 3'-portion by the ribozyme 3'-exon, group I intron ribozymes are also able to replace the substrate 5'-portion by the ribozyme 5'-exon. In this case, the recognition of the splice site is more complicated, using the concerted action of the ribozyme 5'-terminus and 3'-terminus (15,16).

When the two types of reactions - replacement of the substrate 3'- and 5'-portion - are combined, group I intron ribozymes can facilitate *trans*-excision reactions - the removal of an internal RNA segment and the joining of the flanking exons (17). This *trans*-excision reaction relies upon three recognition sequences, consisting of the formation of the P1, P9.0, and P10 helices between substrate and ribozyme (16,18). These *trans*-excision ribozymes were shown to excise fragments of up to 28 nucleotides from short substrates *in vitro* (17), and

remove single nucleotides from full-length mRNA in *E. coli* cells (19). No evidence has indicated that longer fragments can be excised in cells. To obtain ribozymes that can efficiently excise long internal substrate fragments from full length mRNAs under physiological conditions, a different splice site recognition principle was used here to recognize the 5'- splice site and the 3'- splice site with two separate structural elements, the P1 and P9.2 helix, respectively.

Here we demonstrate that a new variant of the group I intron from *Tetrahymena* can remove introns from a full-length mRNA in *trans*. The ribozyme differs from the described *trans*-excision ribozymes by using its 3'-terminal P9.2 helix for recognition of the 3'-splice site (figure 3.1), which leads to the efficient excision of introns with a length of 100 nucleotides. The excision is accurate *in vitro* and in *E. coli* cells, and tolerates many sequences within the intron. We termed this RNA 'spliceozyme' because it is a ribozyme that functions like a spliceosome. After further developments, the spliceozyme may serve as a versatile tool to remove internal RNA sequences for applications in research and in therapy.

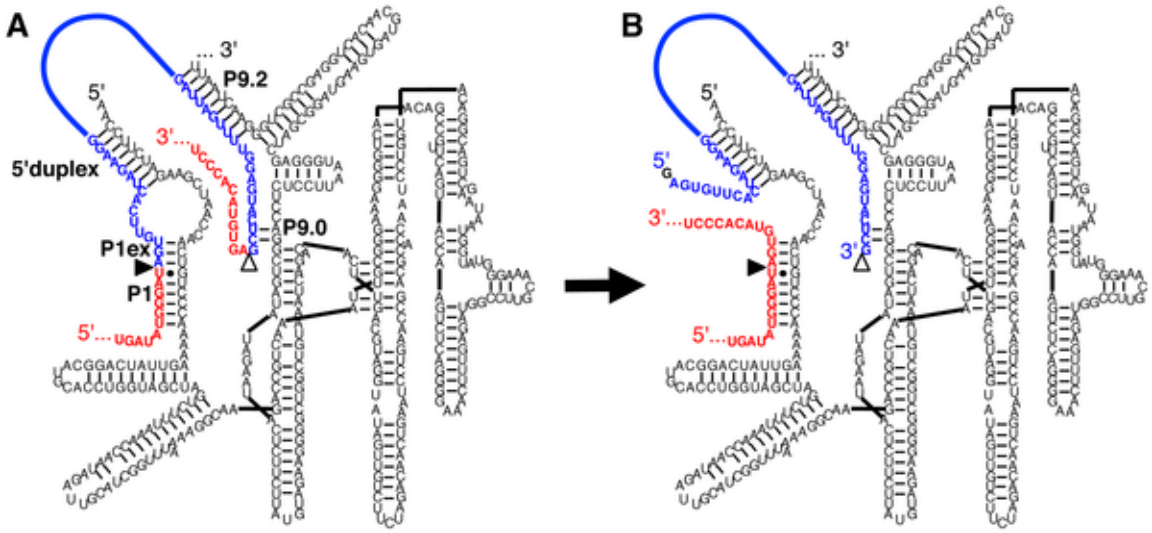


Figure 3.1 Secondary structure of the spliceozyme, based on the group I intron ribozyme from *Tetrahymena*. (A) Before the reaction, the substrate consists of two exons (red) and an intervening intron (blue). The spliceozyme (black) uses its 5'-terminus to form the P1 duplex, the P1 extension (P1ex), and the 5'-duplex to position the substrate 5'-splice site (filled triangle). The ribozyme 3'-terminus forms the P9.2 duplex and the P9.0 duplex to position the substrate 3'-splice site (empty triangle). (B) After two transphosphorylation reactions the 5'-exon and the 3'-exon of the substrate are joined. The 5'-terminus of the intron is capped by an exogenous G (black) and the 3'-terminus of the intron is liberated. The sequences correspond to the removal of an intron from splice site 258 of the CAT pre-mRNA used in this study. The internal sequence of the intron is drawn as a bold blue line. Note that a hairpin terminator is added to the 3'-terminus for all reactions in cells. The secondary structure of the ribozyme is based on (58) with the alteration that the P4-P6 domain was positioned on the right side for clarity.

3.3 Results

Design of the Spliceozyme

The spliceozyme described here is based on the group I intron ribozyme from *Tetrahymena* (figure 3.1). For recognition of the 5'-splice site, the substrate sequence immediately adjacent to the splice site was recognized by forming the P1 helix and the P1 extension helix, followed by an internal loop and a duplex at the ribozyme 5'-terminus (12,13,20,21). The ribozyme design with only a P1 duplex was labeled P1; the design with a P1 duplex and a P1 extension was labeled P1ex, and the design that included all elements of the extended guide

sequence (P1 helix, P1 extension, internal loop, and 5'-duplex) was labeled EGS. For recognition of the substrate 3'-splice site, the P9.2 helix was designed to form in *trans*, and was truncated to 9 base pairs. The substrate for the *trans*-splicing reactions described here was the 678-nucleotide long mRNA encoding the enzyme chloramphenicol acetyl transferase (CAT), with an intron of 64 or 100 nucleotides in length inserted at position 258.

Spliceozyme activity in vitro

To test whether the spliceozyme could function *in vitro*, the *CAT* pre-mRNA was radiolabeled internally and incubated with spliceozymes targeting the two exon / intron junctions of a 100-nucleotide long intron. Reaction conditions used near-physiological conditions, including 5 mM $MgCl_2$ at 37°C, 100 nM *CAT* pre-mRNA, and spliceozyme concentrations of 100 nM and 1 μM. Splicing products of the correct size were detectable after only a few minutes of reaction time (figure 3.2, upper panels). Quantitation of the signals showed that within five minutes, 3-7 % and 23-36 % of the pre-mRNA were converted to a splicing product of the correct size by 100 nM and 1 μM spliceozyme, respectively (figure 3.2, lower panels). The ribozyme with a P1 5'-terminus reacted approximately 2.5-fold faster than the other ribozymes (0.25 min^{-1} versus 0.09-0.12 min^{-1}, respectively), each with a P9.2 duplex length of 9 base pairs (compare figures 3.2A, B, and C). This suggests a possible inhibitory role of extensions at the 5'-terminus in the P1ex and EGS constructs, and is consistent a previous suggestion that the ribozyme 5'-structure could indeed be partially inhibitory for *trans*-splicing (22).

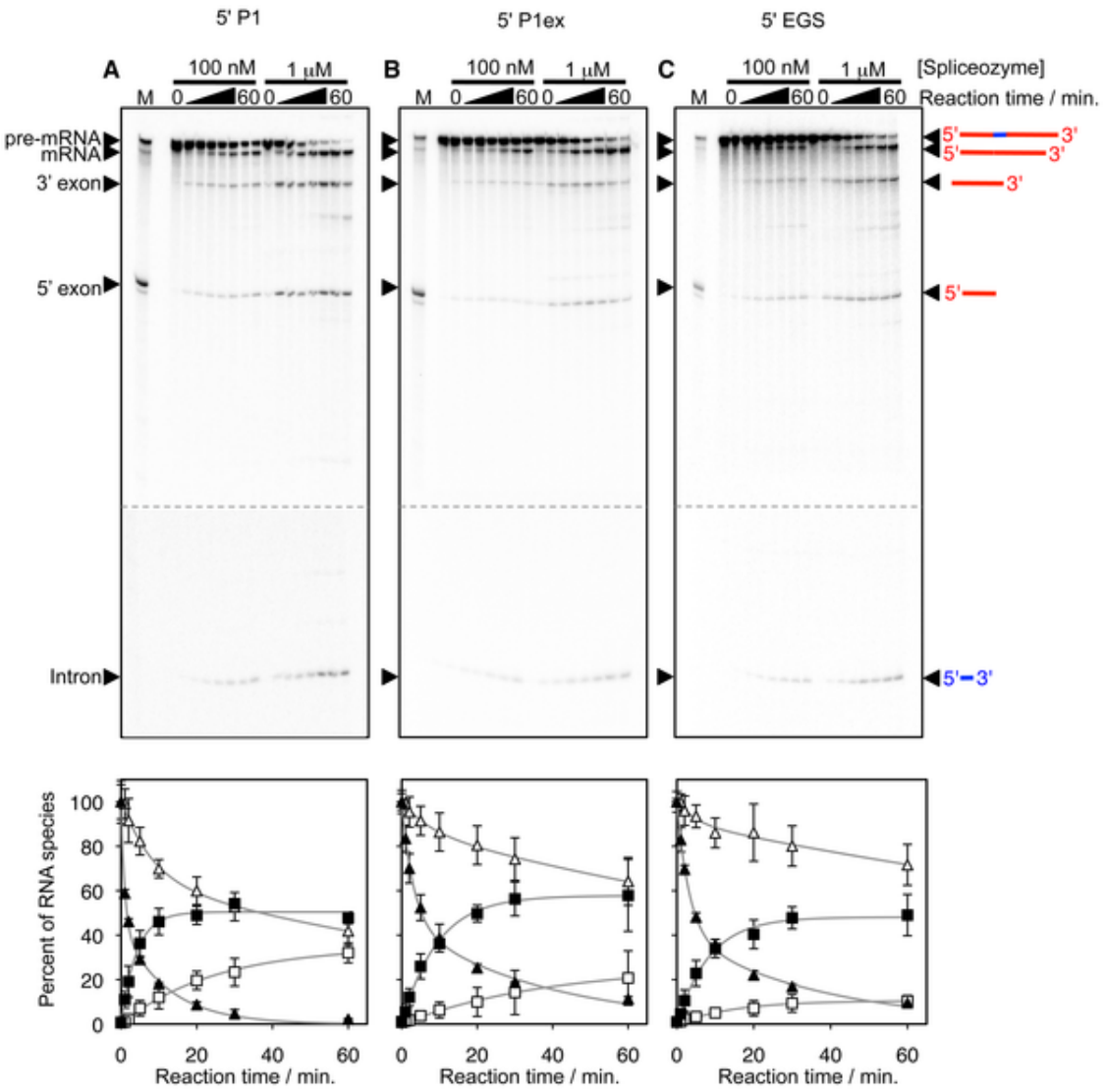


Figure 3.2 Influence of the spliceozyme 5'-terminus design on the reaction *in vitro*. The ribozyme 5'-terminus ends in (A) a P1 duplex, (B) a P1 extension, or (C) a 5'-duplex (see figure 3.1). The corresponding constructs are labeled as 5' P1, 5' P1ex, and 5' EGS, respectively. The top panels show autoradiograms of internally radiolabeled splicing products after separation by denaturing polyacrylamide gel electrophoresis. The marker (M) shows the position of pre-mRNA (778 nt), mRNA (678 nt), and 5'-exon (278 nt). The 3'-exon and the intron had a length of 400 nt and 100 nt, respectively. For each spliceozyme construct, two splicing reactions are analyzed with spliceozyme concentrations of 100 nM and 1 μM. Samples were taken at reaction times between 0 and 60 minutes. A schematic of the substrates and reaction products is shown to the right of the image, with exons in red and the intron in blue. Bottom panels show the quantitation of the disappearance of the CAT pre-mRNA (triangles) and the appearance of the CAT mRNA (squares). Empty symbols correspond to 100 nM spliceozyme, while filled symbols correspond to 1 μM spliceozyme concentration. Grey lines show single-exponential curve fits to the products and double-exponential curve fits to the reaction products. Error bars are standard deviations from three experiments.

To determine if the products contained the correct splicing junction, the reaction products were subjected to RT-PCR, specifically amplifying sequences that contained both the 5'-exon and the 3'-exon of the *CAT* mRNA. Cloning and sequencing showed the correct exon junction sequence in ten out of ten *trans*-splicing products (data not shown). These results demonstrated that a spliceozyme using its 5'-terminus for recognition of the 5'-splice site and its 3'-terminus for recognition of the 3'-splice site via formation of the P9.2 helix was able to efficiently and accurately remove the intron from a pre-mRNA *in vitro*.

Between 8% and 46% of the substrate molecules lost their 5'-exon and were thereby converted to side products within one hour (figure 3.8A). The differences were correlated with two factors: spliceozyme concentration and design of the spliceozyme 5'-terminus. First, a spliceozyme concentration of 1 mM led to approximately 4-fold greater release of the 5'-exon compared to the same constructs at a ribozyme concentration of 100 nM, likely a result of more reaction events generating more side reaction events. Second, in constructs differing in the ribozyme 5'-terminus (all with a P9.2 duplex length of 9 base pairs) approximately two-fold higher 5'-exon release was observed in the absence of a P1 extension duplex (construct P1) compared to the other two constructs that contained a P1 extension duplex (P1ex, EGS). One possible explanation for this behavior is that the ribozyme 5'-terminus forming the P1 extension helps position the 3'-splice site at the catalytic site by forming the P10 helix (23); this helix formation may therefore shorten the time between the first and the second catalytic step of splicing, during which the 5'-exon can escape

from the spliceozyme. At 1 mM spliceozyme concentration additional side products are detectable in figure 3.2; the identity of the corresponding side products is currently unclear.

Influence of the P9.2 duplex length on spliceozyme activity in vitro

Truncation of the P9.2 helix from 9 base pairs to 8, 7, or 6 base pairs did not significantly affect spliceozyme efficiency (figure 3.3A). Only when the length of the P9.2 helix was reduced to 5 base pairs was a lower fraction of substrate converted to product than ribozymes with a P9.2 helix of six to nine base pairs (31% versus 45%-58%, respectively). This is consistent with the finding that duplexes shorter than 7 base pairs are below the optimal length for the specific recognition of target sites (24). The length of the P9.2 duplex did not strongly influence the amount of 5'-exon release (figure 3.8B).

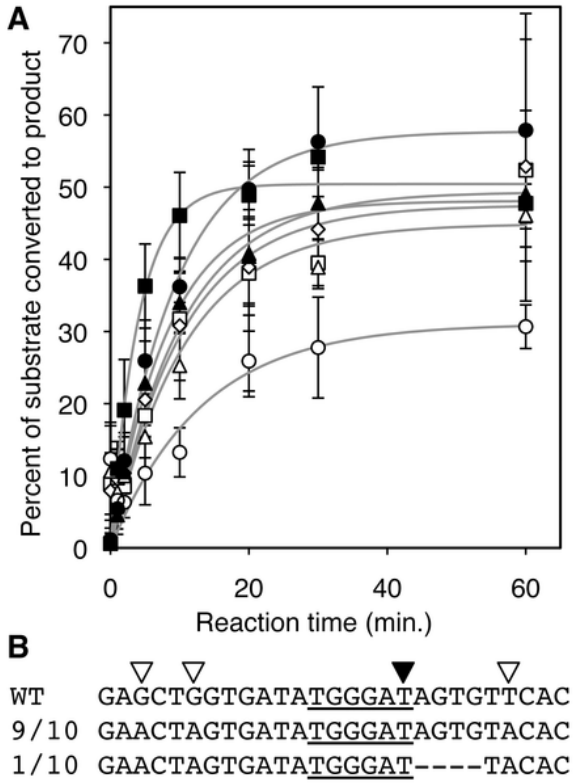


Figure 3.3 Effect of substrate recognition sequences on product formation. (A) Kinetics of spliceozyme splicing *in vitro*, measured as the molar percentage of the initital substrate concentration (CAT pre-mRNA) converted to product (CAT mRNA). The reactions were analyzed as in figure 3.2. Data from spliceozymes with a 3'-terminal P9.2 helix of 9 base pairs are shown with filled symbols, including spliceozyme 5'-termini of a 5'-duplex (triangles), P1 extension helix (circles), and P1 helix (squares). Empty symbols denote results from spliceozymes terminating in a P1 helix at the 5'-terminus, and the 3'-terminal P9.2 helix truncated to 8 base pairs (diamonds), 7 base pairs (squares), 6 base pairs (triangles), and 5 base pairs (circles). Error bars correspond to standard deviations from three experiments. (B) Sequences of splicing junctions resulting from RT-PCR, cloning and sequencing of *trans*-splicing products from the reaction with a P9.2 helix of 5 base pairs and a P1 helix at the 5'-terminus. Note that the wild type sequence (WT) differs from the splicing products in three silent mutations (empty triangles), which confirmed that the sequences were splicing products and not a contamination by the wild type gene. Nine out of ten sequences showed the sequence expected from correct splicing at the 3'-terminal G, whereas one sequence indicated that the guanosine four nucleotides downstream of the intended 3'-splice site was used instead. The sequence participating in the P1 helix is underlined, containing the 5'-splice site (filled triangle).

To estimate the effect of truncating the P9.2 helix on splicing accuracy, ten splice site junctions were determined by reverse transcription, cloning and sequencing of the splicing products resulting from a P1 helix of 6 base pairs and

a P9.2 helix with 5 base pairs (figure 3.3B). Nine out of ten splice site junctions had the correct sequence. One sequence gave evidence for an event of aberrant splicing, in which the guanosine four nucleotides downstream of the correct guanosine was chosen as 3'-splice site. In contrast, ten out of ten sequences showed the correct product when the 3'-splice site was recognized by a 9 base pair helix (data not shown). This suggested that a P9.2 duplex length of 5 base pairs may be too short for reliable recognition of the correct 3'-splice site, consistent with earlier considerations on substrate recognition helices (24). Together, the truncation experiments at the ribozyme 5'- and 3'-terminus suggested that a P1 extension at the 5'-splice site reduced the loss of 5'-exons, and that the 3'-splice site required a P9.2 duplex with at least 6 base pairs.

Spliceozyme activity in cells

To assess the activity of the spliceozyme in cells, a plasmid encoding the spliceozyme and a *CAT* pre-mRNA with a 64-nucleotide long intron was generated, and transformed into *E.coli* cells (figure 3.4A). In order for bacterial resistance to chloramphenicol to be conferred, it is necessary for the spliceozyme in the cells to process *CAT* pre-mRNA to *CAT* mRNA, which is then translated to the CAT enzyme. The CAT enzyme acetylates chloramphenicol with acetyl-coA as acetyl donor (25), rendering chloramphenicol unable to inhibit the ribosome (26), thereby allowing *E. coli* cells to grow in the presence of chloramphenicol. Indeed, *E.coli* cells transformed with the spliceozyme construct were able to grow on medium containing 8 µg/mL chloramphenicol (figure 3.4B; grey columns), at levels similar to the growth mediated by *CAT* mRNA without an

intron (figure 3.4B, black column on the left). To test whether this growth was a result of spliceozyme-mediated catalysis, the catalytic center of the spliceozyme was inactivated by mutations. A mutation at the guanosine binding site (G264A) was found previously to strongly reduce ribozyme activity while maintaining the overall ribozyme fold (27,28,29). Because the remaining weak activity appeared to generate some bacterial growth (data not shown) the ribozyme was further inactivated with a total of six mutations (A261U, C262A, A263G, G264A, A265U, and C266A). The modified construct did not mediate detectable bacterial growth (figure 3.4B, white columns) above the background level of the assay (figure 3.4B, black column on the right), approximately 1/1,000 of the signal from the functional spliceozyme. Two different introns in the pre-mRNA resulted in similar levels of growth with active spliceozyme and no growth with inactive spliceozyme (constructs 64-SiC2 and 64-SiC3; figure 3.4B). Note that these two intron sequences were not arbitrarily chosen but were the result of a selection (see below). To estimate the splicing accuracy of the spliceozyme in cells total RNA was isolated and used to determine the sequence of splice junctions in the *CAT* mRNA as described for the *in vitro* reaction. Ten out of ten *trans*-splicing products had the correct sequence (data not shown). These results demonstrated that catalysis by the spliceozyme removed the intron from a pre-mRNA in *E. coli* cells, and did so accurately and efficiently enough to generate an antibiotic resistance phenotype.

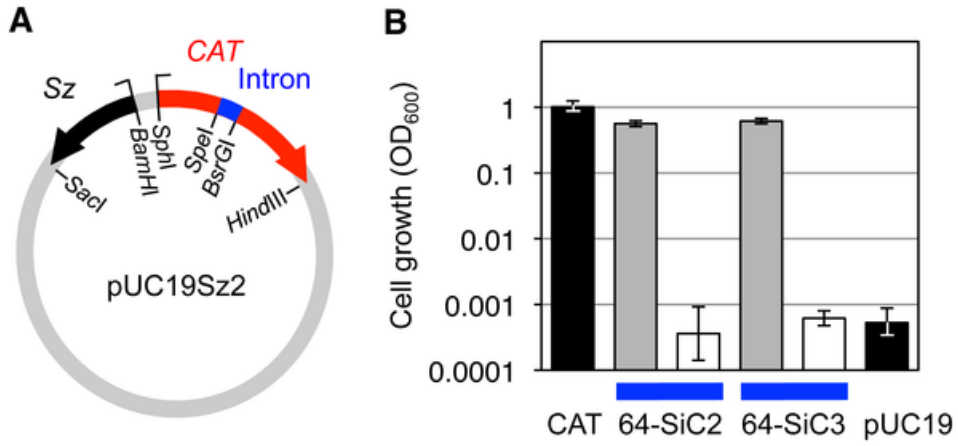
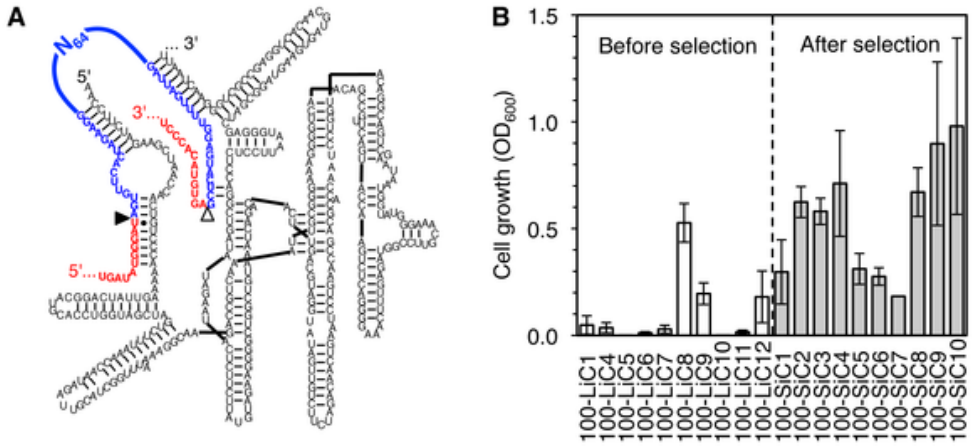


Figure 3.4 Spliceozyme activity in *E. coli* cells. (A) Schematic of the plasmid used for the expression of spliceozyme (Sz; black) and CAT pre-mRNA (CAT; red) containing an intron (blue) in E. coli cells. Restriction sites used in cloning of the plasmid are indicated. (B) Quantitation of E. coli cell growth on LB-agar plates containing 8 µg/mL chloramphenicol, by measuring the A600 of cell suspensions from washing the plates after incubation. The black column on the left shows the A600 from cells that express the wild type CAT gene without intron. Grey columns denote the A600 resulting from plasmids containing a CAT gene with an intron and wild-type spliceozymes. White columns denote the A600 from the same constructs as in the grey columns but with six mutations in the catalytic core of the spliceozymes. The black column on the right denotes the A600 resulting from cells containing the pUC19 plasmid without spliceozyme or CAT gene. Two different intron sequences with a length of 64 nucleotides were inserted into plasmid pUCSz2 for this assay (64-CiC2 and 64-SiC3; note that both of these introns were selected for their efficient removal by the spliceozyme; see materials and methods). Note the logarithmic scale for the A600. Error bars are standard deviations from three biological samples.

Sequence requirements of the intron

To determine whether the spliceozyme required specific sequences within the intron of the *CAT* pre-mRNA, two selection experiments were performed with a 100-nucleotide long intron. In the first selection experiment, the internal region of the intron between the substrate recognition sequences was randomized (N_{64}; figure 3.5A; clone names contain an i for 'internal'). A library of plasmid constructs was generated that was identical to the construct described above but differed in the length of the intron (100 nucleotides) and the randomization of the

64 internal nucleotides of the intron sequences. *E. coli* cells were then transformed with this plasmid library. Ten randomly chosen clones from this *E. coli* library were tested for growth on LB plates with chloramphenicol (10 µg/mL). Eight out of ten clones mediated detectable growth, and three of them mediated moderate growth (figure 3.5B, white columns; clone names contain an L for 'library') with efficiencies similar to an additional ten clones that were selected from the library for efficient growth on LB-chloramphenicol medium (figure 3.5B, grey columns; clone names contain an S for 'selected'). When the sequences of all 20 introns were correlated with the mediated chloramphenicol resistance, no sequence pattern could be identified that was associated with increased activity, based on a comparison of their sequences, predicted secondary structures, and predicted self-folding energies (30) (figure 3.5C). A similar selection was performed with a 64-nucleotide long intron containing a randomized sequence of 28 nucleotides. This experiment also gave rise to highly efficient introns, and similarly did not reveal favored patterns for the intron sequences (data not shown). Although neither of these experiments covered the complete sequence space of the randomized N_{64} and N_{28} sequences, this was not necessary to estimate the frequency of beneficial intron sequences within random libraries. The results revealed that the spliceozyme efficiency varied significantly with the internal sequence of the pre-mRNA intron, and that approximately 30% of sequences from a random library constituted a favorable internal intron sequence for processing by the spliceozyme.



C

100-SiC10	GUUCCUGGGAAGGCCUUCGGGUCACAGUAAACUGAGAGGCUUUGGUGUAAGCUCUCCCCUUGCA	1.0 ± 0.4
100-SiC9	AAGGUGCUAGCUUGACUACAGUUCGGUUAUUUGGCUCGGUUAUUUAGGUCACACCAUGGCCUGC	0.9 ± 0.4
100-SiC4	UAAUGACUAGACGGUUCCAAAUUCGAGCUGCGGUCCGGCCGGCAAGAAUACCUAUGCUUAUGC	0.7 ± 0.2
100-SiC8	AGCUACGUAGGGUCAGAUAUGCGACCGCGACGUGGUAGGCCCGGGCGGCCUAAGGUUGGCAUGU	0.67 ± 0.11
100-SiC2	GGUGUUUGUCCAGUGCGCGCUAACCGGGUCGAACUCCAAGGCCGCACGAGAGUCCAGCUUUUAA	0.62 ± 0.07
100-SiC3	GUUGACGUUGAUGUGCCCUUCUCUAACGACCACGGCGACGGCGUGGAAGAUAGUAGGAGGCACU	0.58 ± 0.06
100-LiC8	GAUCCAAGGUAGGGGUAGCAUCUUCCGAUUAGAGGGUAGCGCCCCACCGUUACGUGUUGGGUAA	0.53 ± 0.09
100-SiC5	CCUUGGAAUCAGUUCGGAGGGUGAGUCACAGAGGUGUAGGAUAUUGAUGCAGUCAAUCGGUCAC	0.31 ± 0.07
100-SiC1	ACAGACCUUUUCGGGAUCGCUCUGAGCCCCCUCAUGAGGGCUUCUAGCGUGCUUAAGGCGAUC	0.30 ± 0.15
100-SiC6	UAAAAACAGUUGGCAGUCUCUACUUUCAAGGUCUGCUGGUAGGUUGCACCCUGGAACCUAUGUC	0.28 ± 0.04
100-LiC9	UACUACGCGUGAUCGAUAGCCGUGUGUCCCUAAAAAUAGGGCCAUGCCCGGCGCCCGGAAUGGC	0.20 ± 0.05
100-SiC7	UUAAACUGGAUAGCUUAUGCGCAGGUGUCGUUGCGCUAGGGGCGCCAUGCAUCAGGCGCGCCAU	0.183 ± 0.001
100-LiC12	ACCUGCCAGGCCAAUCCGGAGCCCGGUGACGUGUAGCCCUUCUCCUAACUGCAGGAGGGACCGC	0.18 ± 0.00
100-LiC1	UAACGGGUCGUUGCUCCGUGUAAGGCAGGAGCGCGCUCAGGGAGAGACGCAUCCGGGAUCGCGG	0.05 ± 0.04
100-LiC4	AGAUCGUAACGCUGAGUGUGGGGCUGACGUAGUGCGAGCGUCCGGCCAGUAGUGAAGGGUUACG	0.04 ± 0.02
100-LiC7	UUGAAGAGGCGCUCUCUGUCAUUUUUCGCGUUUCUAUGGGUUCUAGGGUAUUUAUGCGAUGCGG	0.029 ± 0.019
100-LiC11	UGAUGAUUCAUUCUAUUCUGCGGCGGUAGCCGGCAAUUUCUACCGCAGGGAGGGUGUCGAAGCA	0.016 ± 0.008
100-LiC6	GGGGCGGCAGCUCGGCGUUCUUUAAUACUUUCAAUUGCGCUCGGAUGAGCGUUGUGCUUUGUCU	0.011 ± 0.006
100-LiC10	GAAAAGUACGCACGCGACUGAUGGCAACUGCGACGACGCGUGUGGCCUUACACGCCUUACGUGC	0.0006 ± 0.0002
100-LiC5	CCUUGGAAUCAGUUCGGAGGGUGAGUCACAGAGGUGUAGGAUAUUGAUGCAGUCAAUCGGUCAC	0.0005 ± 0.0009

Figure 3.5 Effect of the internal intron sequence on spliceozyme activity in *E. coli* cells. (A) Secondary structure of the spliceozyme (black), with the substrate exons in red and the position of the randomized N64 sequence indicated in the internal region of the intron (blue). (B) Quantitation of E. coli cell growth on LB-agar plates containing 10 µg/mL chloramphenicol. The resulting A600 is given for 20 clones, ten of which were chosen before selection on LB-chloramphenicol plates (white columns), and ten of which were chosen after this selection step (grey columns). The name of each clone is given below, with 'L' indicating clones from the unselected library and 'S' indicating selected clones, 100 denoting the length of the intron, the letter i indicating that the internal region of the intron was randomized, and the number after 'C' denoting the clone number. Error bars are standard deviations from three experiments. (C) Sequences of 20 cloned intron sequences, sorted according to their activity in cells. The left column lists the clone names, using the same nomenclature as in (B). The middle column shows the internal sequence of that clone that resulted from the N64 library, which was inserted between the constant regions of the intron (see (A)). The right column shows the growth activity, measured as A600 of cell suspensions that resulted from washing LB-growth plates containing chloramphenicol, relative to growth on medium without chloramphenicol.

The second selection experiment with an intron of 100 nucleotides tested the sequence requirement for the ten nucleotides between the 3'-splice site and the P9.2 helix (N_{10}; figure 3.6A). These ten nucleotides were randomized similar to the previous selection, and the corresponding plasmid library was transformed into *E.coli* cells. When ten randomly chosen clones from this *E.coli* library were tested for growth on medium containing 10 µg/mL chloramphenicol, none showed detectable growth (figure 3.6B, clone names contain an L for 'library'). Growth was also measured for ten clones selected from the small fraction of the library that mediated growth on LB-chloramphenicol medium (figure 3.6B, clone names contain an S for 'selected'). When the sequences of these clones were sorted according to the level of mediated growth a correlation between splicing activity and the identity of nucleotides 1, 2, and 10 within this region became apparent (figure 3.6B, right). The five most efficient sequences (figure 3.6B, column graph) were used to characterize the sequence requirement in this region. At position 1, four of these five sequences carried a U, consistent with the wild type ribozyme sequence (compared to 3 / 10 sequences from the selected library). At position 2, all five sequences carried a G, consistent with the wild type, and in contrast to the pre-selected pool (4 / 10 sequences). This is consistent with the role of nucleotide 2 in base pairing to a C adjacent to the P9.0 helix (31) in the wild type ribozyme. At position 10, all active sequences carried a C or a U, consistent with the wild type C, and different from the pre-selected pool (6 / 10 sequences). This is consistent with the role of nucleotides 9 and 10 in forming the P9.0 helix (32,33,34). At position 9, the presence of four U's in the 7

most active clones would be consistent with a weak selection at this position. No significant enrichment was observed at other positions. Together, these results suggested that the 10-nucleotide sequence between the P9.2 helix and the 3'-splice site was constrained to specific nucleotides in at least 3 positions.

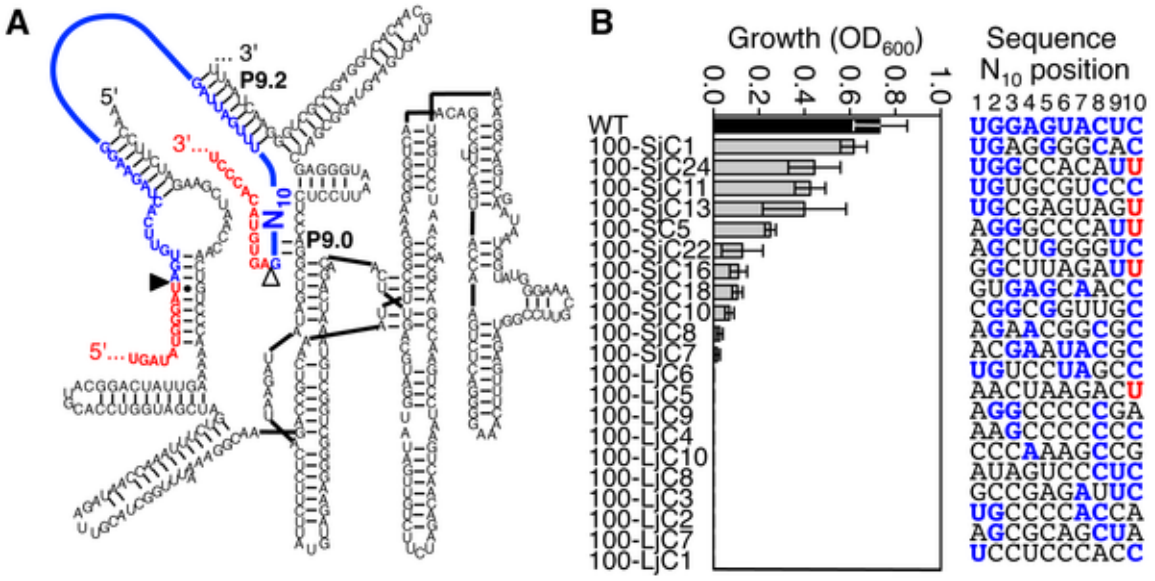


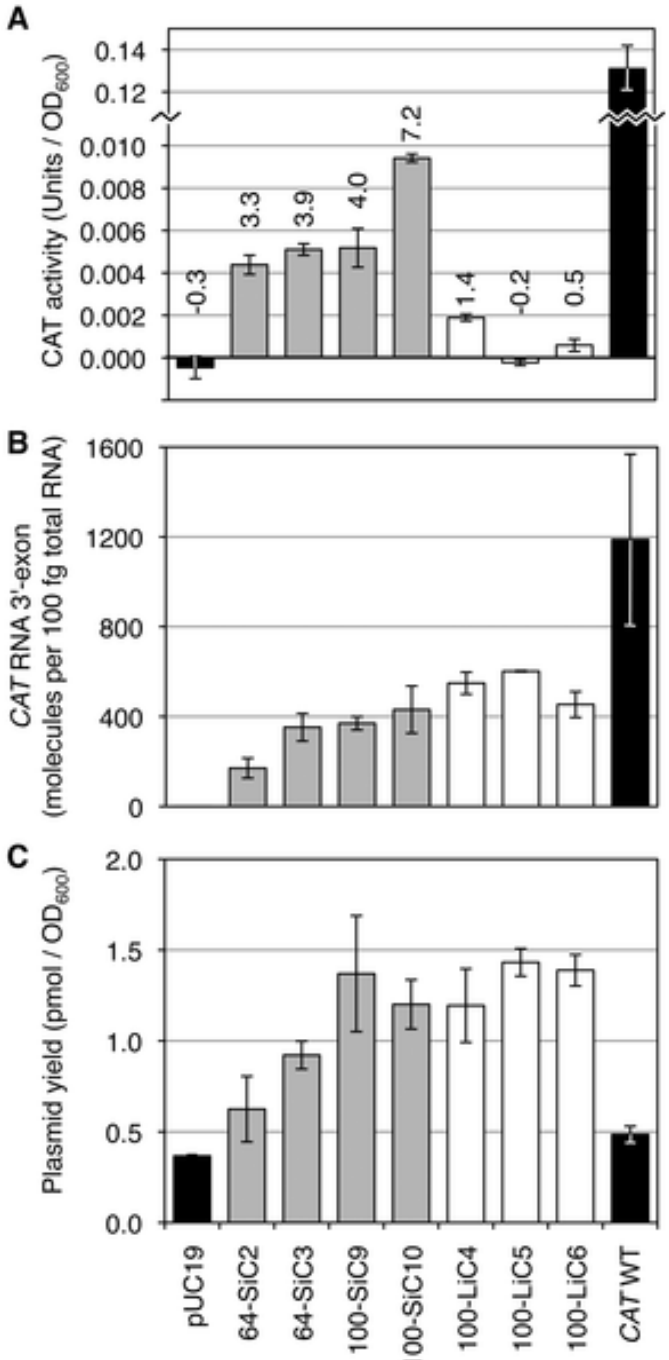
Figure 3.6 Effect of the 3'-terminal intron sequence on spliceozyme activity in *E. coli* cells. (A) Secondary structure of the spliceozyme (black), with the substrate exons in red, and the position of the randomized N10 sequence in the 3'-terminal region of the intron (blue). The helices P9.0 and P9.2 are labeled. (B) List of 20 analyzed clones, sorted according to the resulting activity in cells. The clone name is given on the left, the growth on LB-agar plates containing chloramphenicol is shown as horizontal columns, and the N10 sequence is given on the right. Clone names containing 'L' indicate the unselected library, while 'S' indicates a selection step with chloramphenicol. The letter 'j' denotes that the randomized region is located near the 3'-terminus of the CAT mRNA intron. The growth was measured on plates containing 10 µg/mL chloramphenicol and normalized to growth on medium without chloramphenicol. The sequence at the top (WT) corresponds to the wild type sequence of the Tetrahymena ribozyme. Nucleotides that are identical with the wild type sequence are colored in blue. The position 10 tolerates a U, which is consistent with its base-pairing role, and is colored in red. Errors are standard deviations from three experiments.

Efficiency of the spliceozyme in cells

To obtain a measure for the efficiency of the spliceozyme in *E. coli* cells

the amount of active CAT enzyme in *E. coli* cell extract, the formation of which requires the correct processing of *CAT* pre-mRNA, was quantified (figure 3.7A). CAT activity was measured by an assay utilizing Acetyl-CoA and DTNB (5,5-Dithio-bis(2-nitrobenzoic acid)) (35). Four constructs expressing spliceozyme and *CAT* pre-mRNA with introns that had previously mediated efficient growth on chloramphenicol containing medium resulted in CAT levels between 0.004 and 0.010 units of CAT per OD_{600} of cells (grey columns). This corresponded to 3.3% and 7.2% of the activity generated by a construct expressing *CAT* mRNA without an intron, in the absence of spliceozyme expression (0.13 units of CAT per OD_{600} of cells; black column on the right). In contrast, three constructs expressing spliceozyme and *CAT* pre-mRNA introns that mediated only poor growth or no growth showed significantly lower CAT activity in the bacterial extract (white columns). This confirmed that spliceozyme-facilitated CAT activity mediated the bacterial chloramphenicol resistance observed in the previous experiments.

Figure 3.7 Quantitation of CAT enzyme activity, *CAT* RNA level, and plasmid level in *E. coli* cells. (A) Units of CAT activity detected in E. coli cell extract, containing different plasmid constructs, as indicated below the column graphs. No detectable CAT activity was found when E. coli cells contained the plasmid pUC19 (left black column; -0.00045 ± 0.00053 units per OD600). Grey columns show the CAT activity of constructs expressing a spliceozyme and a CAT pre-mRNA containing an intron that mediated efficient bacterial resistance to chloramphenicol. White columns show the CAT activity mediated by introns that led to poor or no bacterial resistance to chloramphenicol. Construct names starting with 64 or 100 label the length of the intron in the CAT pre-mRNA gene, 'S' denotes that the clone was selected for activity on medium with chloramphenicol, 'L' denotes that the clone was chosen from an un-selected library, 'i' denotes that the library for this selection was randomized in the internal region of the intron, and the number after C is the clone number. CAT WT denotes a construct without spliceozyme, and in which the CAT gene did not contain an intron, resulting in 0.13 ± 0.01 units of CAT activity per OD600. The percent of CAT activity relative to the CAT WT construct is noted above each column. (B) Level of CAT RNAs estimated by RT-qPCR of the CAT RNA 3'-exon. This assay measures the sum of CAT pre-mRNA and CAT mRNA. (C) Molar amount of plasmid isolated from logarithmically growing E. coli cells. Note that the unit OD600 is used as an absolute value, corresponding to the cells in a volume of 1 mL cell suspension with OD600 = 1. For all graphs shown in figure 3.7, error bars denote standard deviations from three biological replicates.


A
CAT activity (Units / OD600)
0.14
0.12
0.010
0.008
0.006
0.004
0.002
0.000
-0.3
3.3
3.9
4.0
7.2
1.4
-0.2
0.5
B
CAT RNA 3'-exon
(molecules per 100 fg total RNA)
1600
1200
800
400
0
C
Plasmid yield (pmol / OD600)
2.0
1.5
1.0
0.5
0.0
pUC19
64-SiC2
64-SiC3
100-SiC9
100-SiC10
100-LiC4
100-LiC5
100-LiC6
CAT WT

To test whether the variation in CAT activity between different intron sequences was mediated by changes in the *CAT* pre-mRNA/mRNA expression level, total RNA was isolated from each of the bacterial constructs, and *CAT* pre-mRNA/mRNA 3'-exons were quantified by RT-qPCR (figure 3.7B). Resulting 3'-exon levels indicated that RNA expression levels were not higher for efficient introns (grey columns) than for inefficient introns (white columns), thereby excluding the possibility that the CAT activity differences were caused by differences in the *CAT* RNA expression levels. Similarly, plasmid expression levels (figure 3.7C) of efficient introns (grey columns) were not higher than those of inefficient introns (white columns), excluding the possibility that these introns acted by affecting the plasmid copy numbers. It is interesting to note that the lowest plasmid copy number was observed with pUC19 (no expression of pre-mRNA or spliceozyme), with a slightly higher expression for the construct that expressed the *CAT* mRNA, and even higher copy numbers for plasmids that expressed both *CAT* pre-mRNA and spliceozyme. This finding, however, does not affect the conclusion regarding modulation of plasmid copy numbers or RNA expression levels the *CAT* pre-mRNA intron sequences.

Together, these results confirm that catalytic activity of the spliceozymes mediated bacterial resistance to chloramphenicol by processing *CAT* pre-mRNA to *CAT* mRNA, thereby facilitating the translation of active CAT enzyme in bacterial cells. Future improvements may result in spliceozymes that process a larger portion of substrate RNAs, and with less intron sequence dependence, for possible uses as tools in research and therapy.

3.4 Discussion

In this study we demonstrated that variants of the *Tetrahymena* group I intron ribozyme can remove introns in *trans* from pre-mRNAs by using their 5'-terminus and 3'-terminal P9.2 helix for recognition of 5'- and 3'-splice sites, respectively. These spliceozymes can accurately remove 100-nucleotide long introns from a pre-mRNA, *in vitro* and in *E. coli* cells, thereby facilitating the expression of functional protein and mediating an antibiotic resistance phenotype. The spliceozymes require specific recognition sequences near the 5'- and 3'-splice sites but tolerate many sequences in the internal region of the intron.

Two short substrate recognition elements in the spliceozyme of five to nine base pairs were sufficient to specify the location of pre-mRNA splicing in *E. coli* cells despite the fact that in principle, any uridylate residue among the many RNAs in *E.coli* cells that is flanked by the complementary sequences could serve as substrate (11,12). This observation, however, is consistent with the finding that only a small fraction of all possible splice sites can be used efficiently by *trans*-splicing ribozymes. This fraction is in the range of ~3% for the *CAT* mRNA, as estimated by the computation of the total free energy of substrate binding, a good predictor of splice site efficiency (22). The fraction of efficient splice sites in a cellular environment may be even lower for highly structured and chemically modified RNAs such as the ribosome and tRNAs, which constitute the bulk of cellular RNAs. The precise frequency of off-target effects with the spliceozymes is currently unclear.

Quantitation of CAT activity in *E. coli* cell extract suggested that approximately 5% of the *CAT* pre-mRNA were converted to functional *CAT* mRNA by the spliceozyme, when compared to *CAT* mRNA lacking an intron (figure 3.7A). This efficiency is similar to the estimated 3-4 % of *CAT* mRNA repaired in a similar bacterial system by a *trans*-splicing group I intron using only a single splice site (21,36). This suggests that the spliceozymes described in this study were similarly efficient as the previously used *trans*-splicing ribozymes.

Spliceozymes could potentially be employed for several applications in research and therapy. Research applications could include the use of spliceozymes to create and study different splicing isoforms and phenomena such as alternative splicing, exon skipping, or recursive splicing (37,38,39,40). Spliceozymes could also be used to model specific biochemical steps in the evolution of the spliceosome, which likely originated from a common ancestor with group II intron ribozymes (41,42,43,44,45,46). It is currently unclear how the spliceosome recruited proteins during its evolution, fragmented into multiple RNAs, and developed multiple turnover reactions. The current spliceozyme could therefore serve as a model system to study analogous steps in the evolution from a ribozyme to a dynamic RNP.

Spliceozymes may also have applications in therapy, for example in the correction of aberrant splicing. At least 10% but perhaps 60% of all disease-causing genetic mutations result in aberrant splicing (47,48,49). It may be possible to use spliceozymes for the treatment of these genetic disorders by processing incompletely or incorrectly spliced pre-mRNAs to functional mRNAs.

The procedure for this approach would be similar to that envisioned for *trans*-splicing ribozymes for the replacement of RNA 3'-portions in human cells (13,50,51,52,53,54,55). For this application to be feasible it would first be necessary to improve the delivery (or in situ expression) of spliceozymes, their efficiency in cells, and minimize potential off-target splicing effects (10,13,21,51,56).

3.5 Materials and methods

Construction of spliceozyme / *CAT* pre-mRNA expression plasmids

The plasmid expressing the spliceozyme and the *CAT* pre-mRNA (figure 3.3A) was based on the previously described pUC19-based expression system (21), which includes a multiple cloning site with SacI, BamHI, SphI and HindIII. The ribozyme expression cassette was inserted between restriction sites SacI and BamHI, and the *CAT* expression cassette between SphI and HindIII, in three steps. First, the *CAT* gene was PCR amplified from the plasmid pLysS (Novagen) as described (21), and prepared for the insertion of an intron at position 258 by introducing three restriction sites through PCR mutagenesis. Specifically, the 3'-PCR primer for amplifying the 5'-exon of the *CAT* gene introduced the silent mutations G243A and G246A to generate a SpeI restriction site, and added a 3'-terminal XmaI site near position 258. The 5'-PCR primer for amplifying the 3'-exon introduced the silent mutation T264A to generate the BsrGI site, and added a 5'-terminal XmaI site near position 258. The PCR products of the two exons were digested with SphI and XmaI (5'-exon) and HindIII and XmaI (3'-exon), then both fragments were ligated into the SphI /

HindIII cassette of pUC19b (21) to create plasmid pUCSz1. The sequence with the silent mutations is visible in the sequence of the correctly spliced products (figure 3.6B).

Second, the spliceozyme cassette, including the *trc*2 promoter (57) and a hairpin terminator, was generated in successive PCR reactions analogous to another *trans*-splicing ribozyme (21). By successively modifying the 5'-terminus of the *Tetrahymena* group I intron gene with 5'-PCR primers, the ribozyme 5'-recognition sequence was adjusted to the new splice site 258 (see figure 3.1A), and the *trc*2 promoter and a BamHI restriction site were added. By successively modifying the 3'-terminus by 3'-PCR primers, the 3'-end of the spliceozyme was truncated (see figure 3.1A), a hairpin terminator (5'-GCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTG-3'; stem underlined) was added, and a SacI restriction site was added. The PCR products were then cloned into the BamHI - SacI sites of pUCSz1 to generate pUCSz2.

Third, the intron sequences were inserted into pUCSz2 using the SpeI and BsrGI restriction sites introduced into the *CAT* sequence. Oligonucleotides 5'-gaactagtgatatgggat AGTGTTCACTAGAAGG - N_{28} - GATTAGTTTTGGAGTACTCG agtgtacacc-3' and 5'-gaactagtgatatgggat AGTGTTCACTAGAAGG - N_{64} - GATTAGTTTTGGAGTACTCG agtgtacacc-3' were converted to double strands by PCR, digested with SpeI and BsrGI, and inserted into the corresponding sites of pUCSz2 (exon sequences are in lower case, intron sequences in upper case, restriction sites are underlined). This created introns with a total length of 64 and 100 nucleotides, and randomized

regions of 28 and 64 nucleotides, respectively. These two plasmid libraries were used in selections for efficient internal intron sequences.

For the selection of intron sequences between the P9.2 helix and the 3'-terminal guanosine (G414), plasmid libraries with ten randomized nucleotides were generated by PCR amplification with a primer that carried the randomized sequence. As PCR template, a clone with an intron length of 100 nucleotides (100-SiC10, figure 3.5C) was used. The 3'-PCR primer (5'-ccttgtacactc - N_{10} - AAACTAATCTGCAAGGGGAG-3') randomized the region of interest, and added the BsrGI site (underlined). All other steps were analogous to the selection of internal intron sequences.

Mutagenic inactivation of the spliceozyme at positions 261-266 was completed by site-directed mutagenesis, using two oligonucleotides that covered 19 bases on each flank of the mutated sequence. The resulting mutations were A261U, C262A, A263G, G264A, A265U, and C266A. The procedure used PrimeSTAR GXL DNA Polymerase (Invitrogen) and the QuickChange™ site-directed mutagenesis protocol developed by Stratagene (La Jolla, CA).

Selection of efficient intron sequences

Plasmid libraries containing partially randomized introns were transformed into electrocompetent *E. coli* DH5α cells, and the cells were plated onto LB-agar containing 100 µg/mL ampicillin. After incubation at 37°C for 16 hours, the plates were washed with LB medium to generate the bacterial libraries, which could be frozen as glycerol stocks. Bacterial libraries were diluted to an A_{600} of 0.015 with liquid LB medium, and induced with 1 mM IPTG by shaking for 1 hour at 37°C.

For the selective step, 100 µL of this culture were plated onto LB-agar plates containing chloramphenicol of the noted concentration, and allowed to grow for 16 hours at 37°C. Individual colonies were selected and analyzed for the sequences and efficiencies of individual introns.

Spliceozymes and pre-mRNAs for in vitro reactions

Spliceozymes and the *CAT* pre-mRNA substrate were generated by run-off *in vitro* transcription from PCR products with T7 RNA polymerase, essentially as described (22). The template for PCR of the spliceozyme, and of *CAT* pre-mRNA, was the plasmid clone pUCSz2_100-SiC10. The PCR of the spliceozyme used 5'-primers that contained the T7 RNA polymerase promoter and truncated 5'-recognition sequences, and 3'-primers that truncated the spliceozyme 3'-termini, as shown in the figures. For PCR amplification of the template for T7 transcription of *CAT* pre-mRNA, the 5' primer GCGTAATACGACTCACTATAGCAGGAGCTAAGGAAGCTAAAATG and the 3'-primer CGCCCCGCCCTGCCACTCATC were used (the T7 promoter is underlined). Transcriptions were performed at 37°C for 3 hours in 40 mM Tris/HCl pH 7.9, 26 mM $MgCl_2$, 5 mM DTT, 0.01% Triton X-100, 2.5 mM spermidine and 2 mM of each NTP. The transcription of *CAT* pre-mRNA substrates additionally included α(^{32}P)-ATP for internal radiolabeling. Transcribed spliceozyme and substrate *CAT* pre-mRNA were purified by 7 M urea 5% polyacrylamide gel electrophoresis (PAGE), eluted in 0.01% (w/v) SDS and 300 mM NaCl, ethanol precipitated, and redissolved in water. RNA concentrations were measured by their absorption at 260 nm.

In vitro reactions

In vitro reactions were done essentially as described (22). The final spliceozyme concentration was 1 μM or 100 nM, as noted, and *CAT* pre-mRNA substrate concentrations were 100 nM. Reactions were incubated in 5 mM $MgCl_2$, 135 mM KCl, 50 mM MOPS/NaOH pH 7.0, 20 μM GTP, and 2 mM spermidine, at 37°C, in reaction volumes of 20 μL. Spliceozyme and substrate were pre-incubated separately in the reaction buffer for 10 minutes at 37°C before combining the solutions to start the reaction. Samples were taken in 2 μL aliquots after reaction times of 0, 1, 2, 5, 10, 20, 30, and 60 minutes. *In vitro* reaction products were separated on 7 M urea 6% PAGE and visualized by phosphoimaging. Bands were quantified on a phosphorimager (PMI, Bio-Rad) using the software Quantity One. The background was substracted using the rolling disk method, with two different methods to minimize artifacts. To quantify substrate and products, disk sizes were set such that the background-signal flanking each peak was contacted by the background subtraction line. To quantify the signal of the 5'-exon side product, the disk size was chosen larger such that the peaks of substrate and product were not separated. The signal strengths were then normalized for two factors: to take into account that each internally labeled fragment contained a different number of A's, the signal strengths of each fragment were divided by their number of A's. To account for gel loading errors, signals were also normalized for the total radioactivity per lane. This resulted in the molar fraction of substrate converted to product, which was reported in the figures. The time courses of the signal intensities were fitted

for least-squares differences in Microsoft Excel using the Solver tool. All signals of products were fit well by single-exponential functions; all signals of substrates required double-exponential functions.

Activity measurements in *E.coli*

Fresh overnight cultures of *E. coli* cells in LB medium containing 100 µg/mL ampicillin were diluted with LB medium to an A_{600} of 0.025, induced with IPTG at a final concentration of 1 mM, and shaken at 37°C for 1 hour. Growth was then measured on LB-agar plates instead of LB liquid culture because plates prevented single false-positive clones from overtaking the population (21). Hundred µL of each IPTG-induced liquid culture were plated on one LB agar plate containing 100 µg/mL ampicillin. The same volume from the same cell suspension was plated on one LB agar plate containing the noted concentration of chloramphenicol and 1 mM IPTG. Plates were incubated at 37°C for 16 hours. Each plate was washed with 1.6 mL 1xPBS, and the A600 of each cell suspension was measured. The A_{600} obtained for LB-ampicillin plates was divided by the A_{600} from the corresponding LB-chloramphenicol plates to normalize the growth in the presence of chloramphenicol for the number of viable bacteria. The two introns used for the first activity measurements in *E. coli* cells (figure 3.4B) were two clones resulting from the selection with 28 random nucleotides in the internal sequence of the intron. These two N_{28} sequences were 5'-GGCCACAGGCCCCGCGTCGCGGTGGGGC-3' (clone 64-SiC2) and 5'-ATTCTTGATACTTTATTATTCAATTGTT-3' (clone 64-SiC3).

RT-PCR, sequence analysis, RT-qPCR, and plasmid yield

For the sequence analysis of *in vitro* splicing reactions, RNA was obtained by ethanol preparation after 60 minutes of incubation. For the analysis of *in vivo* splicing reactions, total RNA was isolated from logarithmically growing *E. coli* cultures using the Nucleospin RNA II kit (Machery-Nagel). In reaction volumes of 20 µL, RNA from the *in vitro* reaction containing 2 pmol *CAT* pre-mRNA and 20 pmol spliceozyme, or 3.2 µg of total RNA were incubated with 200 units of Superscript III reverse transcriptase (Invitrogen), for 60 minutes at 55 °C, using the primer 5'-ccgtaacacgccacatc-3' complementary to the *CAT* 3' exon. DNA sequences were amplified by PCR with primers 5'-cggcctttattcacattct-3' and 5'-gtgtagaaactgccggaa-3'. A nested, second PCR introduced restriction sites EcoRI and BamHI using primers 5'-gcatgaattccgcctgatgaatgctcat-3' and 5'-gcatggatccgtattcactccagagcgat-3' (restriction sites underlined). The products were cloned into plasmid pUC19b (21) and 10 clones were sequenced for each reaction.

To quantify the amount of *CAT* RNA 3'-exons in total RNA, reverse transcription products (see above) corresponding to 20 pg of total RNA were amplified in a volume of 10 mL using the two primers 5'-CCGTAACACGCCACATC-3' and 5'-TGTTACACCGTTTTCCATGAG-3', both of which anneal in the 3'-exon of the used *CAT* pre-mRNA sequence. The qPCR reactions were incubated with the Applied Biosystems qPCR master mix on the Fast 7500 RT-qPCR machine (Applied Biosystems), as described previously (21). Only a single PCR product was observed in melting profiles.

Plasmid yields were determined by isolating plasmids from 2 mL of

logarithmically growing *E. coli* cultures with an OD_{600} of 0.5, using the Nucleospin plasmid kit (Macherey Nagel). This was in the linear range of the assay, in which the plasmid yield from 2 mL bacterial culture correlated linearly with OD_{600} values in the range from 0.0 to 1.0 (data not shown).

CAT activity assays

The CAT activity in *E. coli* cell lysate was determined essentially as described (35) but used a 4-fold higher number of cells per assay to obtain a strong signal. Cells from fresh overnight cultures were induced for one hour with 1 mM IPTG, diluted to an A_{600} of 0.02, and grown to an A_{600} of 0.20. Note that the promoters for *CAT* pre-mRNA and *CAT* mRNA were the same. Cells from two mL of culture were concentrated to 200 µL and frozen. After thawing, the suspensions were mixed with 200 µL of 200 mM Tris/HCl pH 7.8, 10 mM Na_2EDTA, and 4 µL of toluene. Fifteen µL of this solution were diluted 10-fold with a buffer to obtain final concentrations of 1 mM DTNB, 0.2 mM Acetyl-CoA, and 0.2 mM chloramphenicol. The absorption at 412 nm was recorded in 15-second intervals. The slope of increase in the absorption at 412 nm was determined by linear least squares fitting to the readout between 6 minutes and 15 minutes. The lysate of cells expressing *CAT* mRNA without intron was diluted 10-fold to stay within the linear range of the assay. Units of CAT activity were determined using the extinction coefficient of 13,600 M^{-1} cm^{-1} for the reaction product, and the definition of one CAT unit, which corresponds to the acetylation of 1 µmol chloramphenicol per minute (35).

3.6 Acknowledgements

We thank Simpson Joseph, Gregory Dolan, and Janina Moretti for helpful discussions. This work was supported by the National Institutes of Health Molecular Biophysics Training Grant (grant number T32GM008326) to Z.N.A. and by the Hellman Family foundation with two fellowships (2011/2012 and 2012/2013) to U.F.M.

Chapter 3, in full, is a reprint of the material as it appears in PLoS One, **Amini Z.N.,** Olson K.E. and Muller U.F. (2014) Spliceozymes: ribozymes that remove introns from pre-mRNAs in *trans*. PLoS One. 9(7): e101932. The dissertation author is the first author on this paper.

3.7 Supporting Information

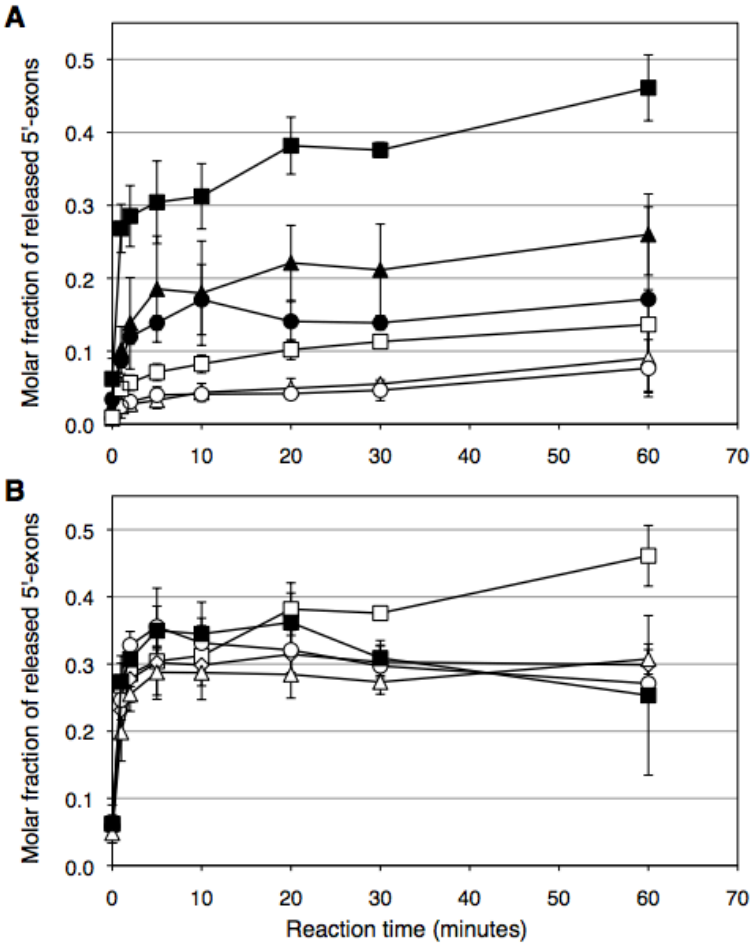


Figure 3.8 Fraction of 5′-exon side products during the *in vitro* reaction of the spliceozyme. The fraction is shown as the molar fraction of the CAT pre-mRNA at the start of the reaction (see Figure 3.2). (A) Influence of spliceozyme concentration and design at the ribozyme 5′-terminus. At the concentration of 1 mM (filled symbols) the 5′-exon was released 4-fold faster than at 100 nM (empty symbols). When the spliceozyme 5′-terminus ended in the P1 helix (squares) the release of the 5′-exon was 2-fold faster than when the spliceozyme 5′-terminus ended in the P1 extension (circles) or the 5′-duplex (triangles). (B) Influence of the length of the P9.2 duplex at the ribozyme 3′-terminus, with a 5′-terminal P1 duplex. The lengths of the P9.2 helix had no or a minor influence on the release of the 5′-exon, as judged by the comparison between P9.2 helices with a length of 9 base pairs (empty squares), 8 base pairs (empty diamonds), 7 base pairs (empty circles), 6 base pairs (empty triangles), and 5 base pairs (filled squares). Error bars show standard deviations from the average of triplicate experiments.

3.8 References

1. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and autocyclization of

the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31, 147-157

2. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., and Cheng, H. L. (1986) A self-splicing RNA excises an intron lariat. *Cell* 44, 213-223

3. van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., and Grivell, L. A. (1986) Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell* 44, 225-234

4. Cech, T. R. (1990) Self-splicing of group I introns. *Annu Rev Biochem* 59, 543-568

5. Stahley, M. R., and Strobel, S. A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol* 16, 319-326

6. Guo, F., Gooding, A. R., and Cech, T. R. (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16, 351-362

7. Golden, B. L., Kim, H., and Chase, E. (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12, 82-89

8. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons. *Nature* 430, 45-50

9. Inoue, T., Sullivan, F. X., and Cech, T. R. (1985) Intermolecular exon ligation of the rRNA precursor of Tetrahymena: oligonucleotides can function as 5' exons. *Cell* 43, 431-437

10. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* 371, 619-622

11. Been, M. D., and Cech, T. R. (1986) One binding site determines sequence specificity of Tetrahymena pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity. *Cell* 47, 207-216

12. Waring, R. B., Towner, P., Minter, S. J., and Davies, R. W. (1986) Splice-site selection by a self-splicing RNA of Tetrahymena. *Nature* 321, 133-139

13. Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes. *RNA* 9, 1254-1263

14. Jones, J. T., and Sullenger, B. A. (1997) Evaluating and enhancing ribozyme reaction efficiency in mammalian cells. *Nat Biotechnol* 15, 902-905

15. Alexander, R. C., Baum, D. A., and Testa, S. M. (2005) 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme. *Biochemistry* 44, 7796-7804

16. Dotson, P. P., 2nd, Johnson, A. K., and Testa, S. M. (2008) Tetrahymena thermophila and Candida albicans group I intron-derived ribozymes can catalyze the trans-excision-splicing reaction. *Nucleic Acids Res* 36, 5281-5289

17. Bell, M. A., Johnson, A. K., and Testa, S. M. (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs. *Biochemistry* 41, 15327-15333

18. Bell, M. A., Sinha, J., Johnson, A. K., and Testa, S. M. (2004) Enhancing the second step of the trans excision-splicing reaction of a group I ribozyme by exploiting P9.0 and P10 for intermolecular recognition. *Biochemistry* 43, 4323-4331

19. Baum, D. A., and Testa, S. M. (2005) In vivo excision of a single targeted nucleotide from an mRNA by a trans excision-splicing ribozyme. *RNA* 11, 897-905

20. Kohler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery. *J Mol Biol* 285, 1935-1950

21. Olson, K. E., and Muller, U. F. (2012) An in vivo selection method to optimize trans-splicing ribozymes. *RNA* 18, 581-589

22. Meluzzi, D., Olson, K. E., Dolan, G. F., Arya, G., and Muller, U. F. (2012) Computational prediction of efficient splice sites for trans-splicing ribozymes. *RNA* 18, 590-602

23. Guo, F., and Cech, T. R. (2002) In vivo selection of better self-splicing introns in Escherichia coli: the role of the P1 extension helix of the Tetrahymena intron. *RNA* 8, 647-658

24. Herschlag, D. (1991) Implications of ribozyme kinetics for targeting the cleavage of specific RNA molecules in vivo: more isn't always better. *Proc Natl Acad Sci U S A* 88, 6921-6925

25. Shaw, W. V. (1967) The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant Escherichia coli. *J Biol Chem* 242, 687-693

26. Tritton, T. R. (1979) Ribosome-chloramphenicol interactions: a nuclear magnetic resonance study. *Arch Biochem Biophys* 197, 10-17

27. Legault, P., Herschlag, D., Celander, D. W., and Cech, T. R. (1992) Mutations at the guanosine-binding site of the Tetrahymena ribozyme also affect site-specific hydrolysis. *Nucleic Acids Res* 20, 6613-6619

28. Nikolcheva, T., and Woodson, S. A. (1999) Facilitation of group I splicing in vivo: misfolding of the Tetrahymena IVS and the role of ribosomal RNA exons. *J Mol Biol* 292, 557-567

29. Hagen, M., and Cech, T. R. (1999) Self-splicing of the Tetrahymena intron from mRNA in mammalian cells. *EMBO J* 18, 6491-6500

30. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31, 3406-3415

31. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) New loop-loop tertiary interactions in self-splicing introns of subgroup IC and ID: a complete 3D model of the Tetrahymena thermophila ribozyme. *Chem Biol* 3, 993-1009

32. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) The guanosine binding site of the Tetrahymena ribozyme. *Nature* 342, 391-395

33. Burke, J. M. (1989) Selection of the 3'-splice site in group I introns. *FEBS Lett* 250, 129-133

34. Russell, R., and Herschlag, D. (1999) Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme. *RNA* 5, 158-166

35. Shaw, W. V. (1975) Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods Enzymol* 43, 737-755

36. Olson, K. E., Dolan, G. F., and Muller, U. F. (2014) In vivo evolution of a catalytic RNA couples trans-splicing to translation. *PLoS One* 9, e86473

37. Berget, S. M., Moore, C., and Sharp, P. A. (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* 74, 3171-3175

38. Chow, L. T., Gelinas, R. E., Broker, T. R., and Roberts, R. J. (1977) An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell* 12, 1-8

39. Hatton, A. R., Subramaniam, V., and Lopez, A. J. (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol Cell* 2, 787-796

40. Burnette, J. M., Miyamoto-Sato, E., Schaub, M. A., Conklin, J., and Lopez, A. J. (2005) Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. *Genetics* 170, 661-674

41. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) Splicing of messenger RNA precursors. *Annu Rev Biochem* 55, 1119-1150

42. Will, C. L., and Luhrmann, R. (2011) Spliceosome structure and function. *Cold Spring Harb Perspect Biol* 3

43. Sharp, P. A. (1985) On the origin of RNA splicing and introns. *Cell* 42, 397-400

44. Koonin, E. V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct* 1, 22

45. Dayie, K. T., and Padgett, R. A. (2008) A glimpse into the active site of a group II intron and maybe the spliceosome, too. *RNA* 14, 1697-1703

46. Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., and Piccirilli, J. A. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229-234

47. Hsu, S. N., and Hertel, K. J. (2009) Spliceosomes walk the line: splicing errors and their impact on cellular function. *RNA Biol* 6, 526-530

48. Ward, A. J., and Cooper, T. A. (2010) The pathobiology of splicing. *J Pathol* 220, 152-163

49. Hammond, S. M., and Wood, M. J. (2011) Genetic therapies for RNA mis-splicing diseases. *Trends Genet* 27, 196-205

50. Watanabe, T., and Sullenger, B. A. (2000) Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proc Natl Acad Sci U S A* 97, 8490-8494

51. Rogers, C. S., Vanoye, C. G., Sullenger, B. A., and George, A. L., Jr. (2002) Functional repair of a mutant chloride channel using a trans-splicing ribozyme. *J Clin Invest* 110, 1783-1789

52. Kastanos, E., Hjiantoniou, E., and Phylactou, L. A. (2004) Restoration of protein synthesis in pancreatic cancer cells by trans-splicing ribozymes. *Biochem Biophys Res Commun* 322, 930-934

53. Ayre, B. G., Kohler, U., Goodman, H. M., and Haseloff, J. (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc Natl Acad Sci U S A* 96, 3507-3512

54. Ryu, K. J., Kim, J. H., and Lee, S. W. (2003) Ribozyme-mediated selective induction of new gene activity in hepatitis C virus internal ribosome entry site-expressing cells by targeted trans-splicing. *Mol Ther* 7, 386-395

55. Song, M. S., and Lee, S. W. (2006) Cancer-selective induction of cytotoxicity by tissue-specific expression of targeted trans-splicing ribozyme. *FEBS Lett* 580, 5033-5043

56. Shin, K. S., Sullenger, B. A., and Lee, S. W. (2004) Ribozyme-mediated induction of apoptosis in human cancer cells by targeted repair of mutant p53 RNA. *Mol Ther* 10, 365-372

57. Weiss, D. S., Chen, J. C., Ghigo, J. M., Boyd, D., and Beckwith, J. (1999) Localization of FtsI (PBP3) to the septal ring requires its membrane anchor, the Z ring, FtsA, FtsQ, and FtsL. *J Bacteriol* 181, 508-520

58. Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994) Representation of the secondary and tertiary structure of group I introns. *Nat Struct Biol* 1, 273-280

Chapter 4: Decreased Side Product Formation in Evolved Group I Intron Spliceozymes

4.1 Abstract

The group I intron ribozyme from *Tetrahymena* was recently re-engineered into a *trans*-splicing variant that is able to remove 100-nucleotide introns from pre-mRNA, analogous to the spliceosome. These spliceozymes were improved in this study by evolution in *E. coli* cells over ten rounds of evolution. One clone with strongly increased activity in *E. coli* cells was analyzed in detail. Three of its 10 necessary mutations extended the substrate binding duplexes, which led to increased product formation and reduced cleavage at the 5'-splice site. Mutation U271C in the conserved core of the spliceozyme led to a further reduction of cleavage at the 5'-splice site but an increase in cleavage side products at the 3'-splice site. The latter were partially reduced by the six additional mutations. These results show the generation of an improved spliceozyme and provide deeper insights into the interdependent function of nucleotides within group I intron ribozymes. Implications for the possible use of spliceozymes as tools are discussed.

4.2 Introduction

Group I introns are catalytic RNAs (ribozymes) that are encoded as intervening sequences in pre-mRNAs. In contrast to most introns, group I introns do not require the spliceosome for their removal from pre-mRNA. Instead, they fold into structures that catalyze their own excision from the pre-mRNA, and the

joining of the flanking exons (1). Group I introns from different species have a conserved structure around the catalytic site, but differ in the peripheral regions (2-4). The most detailed biochemical characterization is available for the group I intron from *Tetrahymena thermophila* (5-7). This ribozyme is very robust with regard to sequence modifications, and is active *in vitro*, in bacteria, and eukaryotes including yeast, mouse and human cells.

Variants of group I introns have been engineered that use the same catalytic mechanism as the native ribozymes but show different functions due to conversion from the natural *cis*-action to *trans*-action at the two splice sites. Group I introns that act in *trans* at their 5'-splice site are able to bind to a substrate RNA and replace the substrate's 3'-terminus with the ribozyme's own 3'-exon (8). Similarly, *trans*-activity at the 3'-splice site allows replacement of the substrate 5'-terminus with the ribozyme's 5'-exon (9). *Trans*-splicing at both splice sites allows the removal of single nucleotides of a substrate RNA in cells, and up to 28 nucleotides *in vitro* (10,11). Using a different design principle for the substrate recognition at both splice sites in *trans*, a *Tetrahymena* group I intron was recently shown to excise 100-nucleotide long introns from pre-mRNAs *in vitro* and in bacterial cells, efficient enough to mediate antibiotic resistance (12). These ribozymes were termed 'spliceozymes' because their action is analogous to that of the spliceosome (Fig. 4.1).

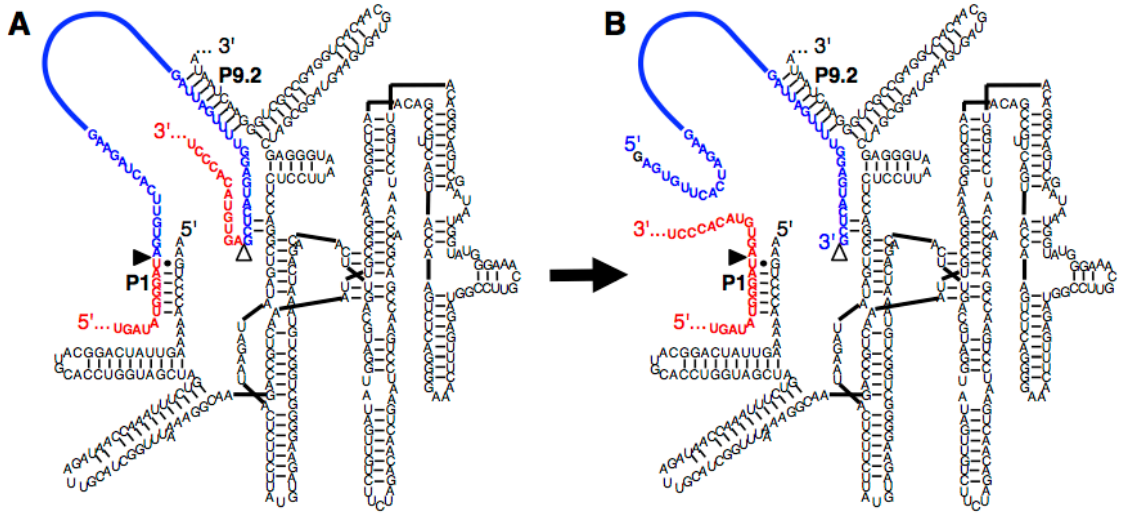


Figure 4.1 Secondary structure of spliceozyme used as parent in the evolution. The spliceozyme (black) base pairs with the CAT pre-mRNA (red/blue) by forming duplexes P1 and P9.2. (A) Before the splicing reaction, the pre-mRNA contains the 5'-exon (red), a 100- nucleotide intron (blue), and the 3'-exon (red). (B) After the splicing reaction, the 5'-exon and 3'-exon are joined to form CAT mRNA (red), and the intron (blue) is excised. The 5'- splice site is indicated by a filled triangle; the 3'-splice site is indicated by an open triangle. Note that the intron 5'-terminus is extended by an external guanosine during the reaction.

Spliceozymes may have the potential for applications as tools in therapy and in research. This system could be used to correct mis-splicing resulting from aberrant splicing diseases, to switch between splicing isoforms for the study of spliceosomal splicing, and as a model system to study the biochemical steps that likely occurred in the evolution of the spliceosome from a ribozyme ancestor. For the use of spliceozymes as tools, however, several hurdles need to first be overcome. For use in therapy, efficient delivery methods need to be developed. There are current efforts to develop adeno-associated virus vectors (13), cationic peptides (14), nanoparticles (15,16), and modified *Salmonella* strains (17,18) as delivery vehicles, yet improvements to minimize toxicity and immune responses

and increase delivery to target tissues are still needed. Additionally, for the use as a tool in therapy or research, improvements are required to increase splicing efficiency in cells, and reduce off-target effects. Currently, however, spliceozymes are able to serve as model systems to study the biochemical steps in the evolution of the spliceosome. While the spliceozyme evolved from a common ancestor with group II introns, group I introns can be used as analogs for the same biochemical steps. The existing *trans*-splicing group I intron spliceozymes show that *cis*-splicing ribozymes can make the evolutionary step to *trans*-splicing ribozymes (12), analogous to perhaps the first step in the evolution of the spliceosome from a *cis*-splicing ancestor. Among the steps still required to obtain an analog of the spliceosome, the existing spliceozymes need (i) an increase in splicing efficiency, (ii) a reduction in side products and off-target effects, (iii) the ability to recognize different splice sites, (iv) the recruitment of proteins for structural and catalytic roles, and (v) fragmentation into multiple RNA-protein particles.

Here we show the improvement of spliceozymes by directed evolution in *E. coli* cells. Over ten rounds of evolution, spliceozymes were challenged to remove an intron from the pre-mRNA of chloramphenicol acetyltransferase (CAT) under increasingly stringent conditions. Analysis of one of the fittest spliceozymes, containing 10 mutations relative to the parent spliceozyme, showed increased *trans*-splicing efficiency and reduced side product formation. Three mutations extended the substrate recognition sequences, resulting in increased product formation and reduced side product formation. Additionally, a

mutation in the conserved core of the spliceozyme (U271C) appeared to adjust activity at 5'-splice 3' splice sites, resulting in reduced activity of the 5'-splice site and increased activity at the 3'-splice site. A structural model is presented for a possible mechanism of this core mutation. A number of additional mutations were able to mitigate the increase in side product formation resulting from increased activity at the 3' splice site. The implications for possible uses of spliceozymes as tools and as model systems for ribozyme evolution are discussed.

4.3 Results

To improve the function of spliceozymes in cells, this system was subjected to directed evolution in *E. coli* cells (Fig. 4.2) using an evolution system established in earlier studies (19,20). As starting points for the evolution, two spliceozyme libraries were generated by mutagenic PCR, using as template either a single spliceozyme gene (Fig. 4.1; (12)) or a pre-evolved library that already contained genetic diversity, which, in some circumstances, can increase the speed of evolution (21). The spliceozyme libraries were cloned into the spliceozyme expression cassette of a library plasmid that also contained the expression cassette of the pre-mRNA of chloramphenicol acetyl transferase (CAT). After transformation into *E. coli* cells, the cells were plated on medium containing chloramphenicol. Only cells expressing functional spliceozymes, able to remove the intron from the *CAT* pre-mRNA, generated functional CAT enzyme, and formed colonies. The plasmids from these colonies were isolated to complete one round of evolution.

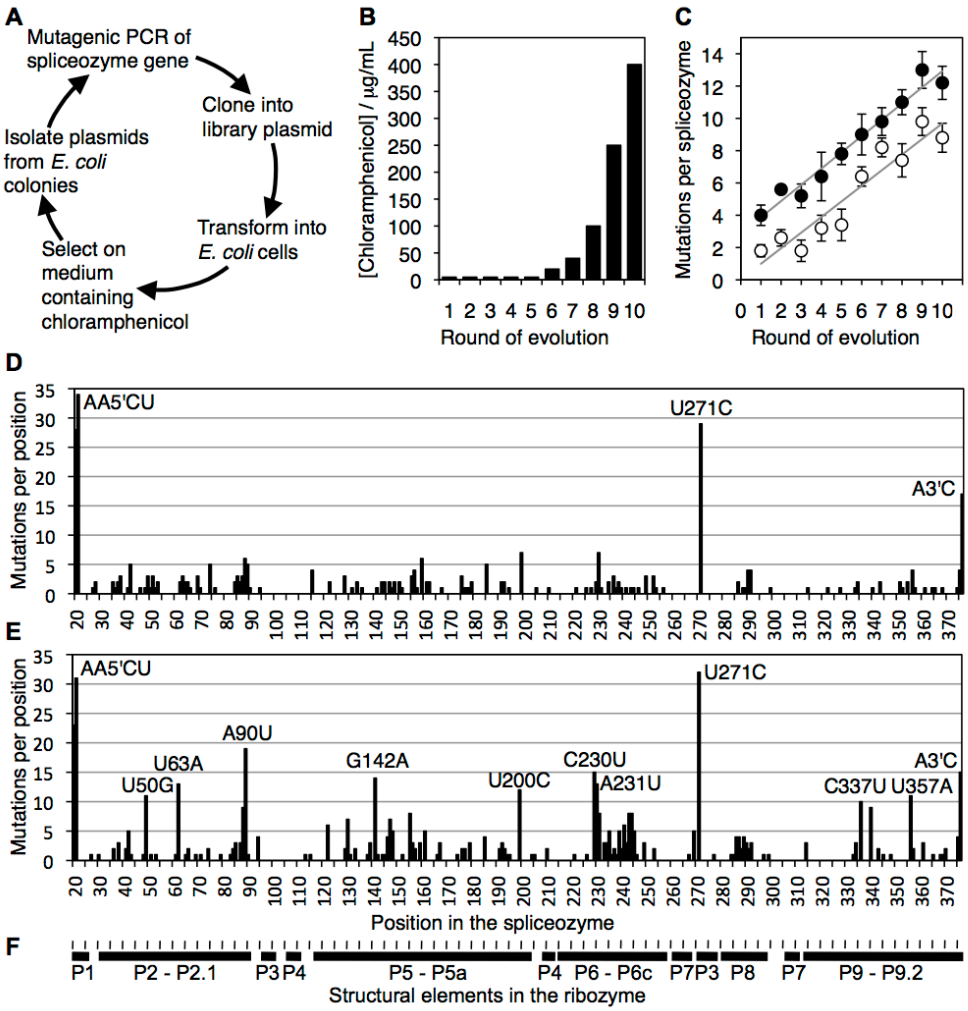


Figure 4.2 Evolution of spliceozymes in *E. coli* cells. (A) Simplified schematic of the evolution procedure. (B) Selection pressure as function of the rounds of evolution. The concentration of chloramphenicol was 5 µg / mL for the first 5 evolution rounds. Both strains of evolution were subjected to the same profile. (C) Average number of mutations per spliceozyme sequence over 10 rounds of evolution. Empty circles show the mutations in the evolutionary line starting from the wild-type sequence (W), and filled circles the mutations in the evolutionary line starting from a pre-evolved pool (P). Linear fits resulted in slope of 0.97 for line W, and 1.01 for line P. The y-axis offset was 0.0 for line W, and 2.9 for line P. Error bars are the uncertainties of the mean from five sequences (rounds 1-9), or 10 sequences (round 10). (D) Positions of the mutations in the spliceozyme sequence for line W, summed over 10 rounds of evolution. The height of each column is the number of mutations per position within 55 sequences. Mutations that occurred at least ten times are annotated with the mutation that occurred most frequently. Mutations at the spliceozyme 5'-terminus and 3'-terminus are labeled with 5' and 3', respectively. (E) as in (D) but for line P. (F) Secondary structure elements of the spliceozyme, aligned with the nucleotide positions in the graphs of (D) and (E).

A low selection pressure was applied for the first five rounds of evolution, improving the sampling of sequence space (22). Selection pressure was successively increased over the last five cycles of evolution to isolate the most efficient ribozyme variants generated (Fig. 4.2B). Sequencing of 5 -10 spliceozyme clones in each evolution round, for each of the two lines of evolution, showed that the number of mutations increased by ~1 mutation per spliceozyme per evolution round (Fig. 4.2C). Four mutations were highly enriched (Fig. 4.2D, 4.2E): Two mutations at the spliceozyme 5'-terminus (AA5'CU), one mutation in the core of the ribozyme (U271C), and one mutation at the ribozyme 3'-terminus (A3'C). Interestingly, enrichments of the U271C and 5'-terminus mutations were observed concurrently, both being enriched above 20% by round 4 or 5, while equivalent enrichment of the 3' terminus mutation was not observed until round 8 (Fig. 4.10). The enriched U271C mutation in the conserved core of the ribozyme was surprising; mutations in this region generally lead to severe functional deficiencies (5).

To identify the most active sequences resulting from the evolution, the spliceozyme genes of 20 clones were re-cloned and tested for their ability to mediate bacterial growth on medium containing chloramphenicol (data not shown). Of the two clones mediating most efficient growth, one clone (termed W11) contained the central U271C mutation and was analyzed in more detail. Clone W11 originated from the line of evolution that started with an individual sequence, and contained 10 mutations relative to the parent spliceozyme. To identify the mutations necessary for maximum activity, each mutation was

individually reverted to the parent sequence and the resulting activity was measured (Fig. 4.3). Surprisingly, all ten mutations were necessary for full activity. The strength of effects correlated with the enrichment during evolution: The strongest effects were mediated by the spliceozyme mutations at the 5'- and 3'-terminus (which extend the substrate recognition duplexes by two and one nucleotide, respectively) followed by the mutation U271C. The six additional mutations were also required for full activity (C37U, G175A, 234Gins., A286G, A334G, and G368D). These six mutations were not specifically enriched during the evolution (Fig. 4.2D) or associated with other mutations (data not shown). To simplify the analysis of mutations, these six mutations (termed ‘helper mutations’) were grouped together and studied in combination.

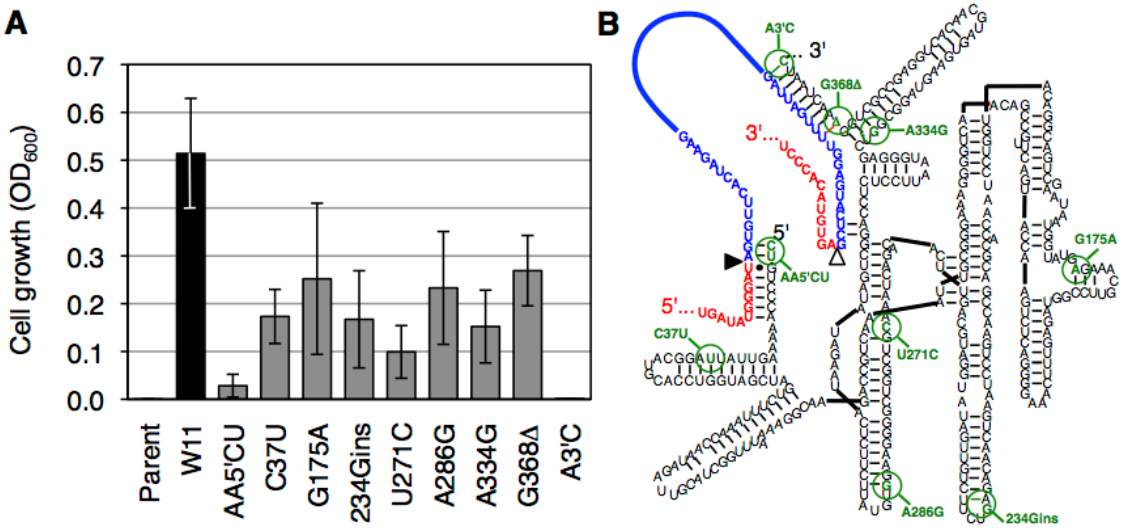


Figure 4.3 Identification of mutations necessary for full activity in the evolved spliceozyme W11. (A) Growth of E. coli cells expressing different spliceozyme variants on plates containing chloramphenicol. In addition to the parent ribozyme and the evolved ribozyme W11, variants of W11 with mutations reverted to the parental sequence were analyzed. The mutations are given below the graph, with AA5'CU describing the mutation of two 5'-terminal nucleotides to CU, and A3'C describing the mutation of the 3'-terminal nucleotide to C. The OD600 was measured from growth on plates, and was normalized to the OD600 from plate cultures containing ampicillin (i.e. viable cells containing plasmid). Error bars are standard deviations from biological triplicates. (B) Secondary structure of the evolved spliceozyme W11, with evolved mutations indicated in green and with green circles. New base pairs created by the mutation are highlighted in green, and base pairs deleted by the mutation are highlighted in red.

To analyze the effect of the three groups of mutations on activity in cells, two assays were performed (Fig. 4.4). Bacterial growth on plates containing chloramphenicol was increased dramatically with the 5'- and 3'-terminal mutations present (Fig. 4.4A). While the addition of the U271C mutation or the six helper mutations individually led to a decrease in growth, their combined addition led to maximum growth, showing that U271C acted cooperatively with at least a sub-set of the helper mutations. To test whether the improved growth of the evolved ribozyme W11 on chloramphenicol containing medium was due to an

increased production of the enzyme chloramphenicol acetyl transferase (CAT) the activity of this enzyme was measured in bacterial extract (Fig. 4.4B). Surprisingly, the U271C & helper mutations did not mediate an increase in CAT activity over the spliceozyme variant with only 5'- and 3'-terminal mutations. To reconcile the improved growth in the presence of chloramphenicol with reduced CAT activity, we hypothesized that in addition to facilitating CAT synthesis, spliceozyme activity may also impose a burden on the cell that reduces growth. The evolved mutations may have reduced this burden to mediate improved growth, possibly by a change in the pattern of product and side product formation, thereby reducing the production of and interference by nonfunctional RNA.

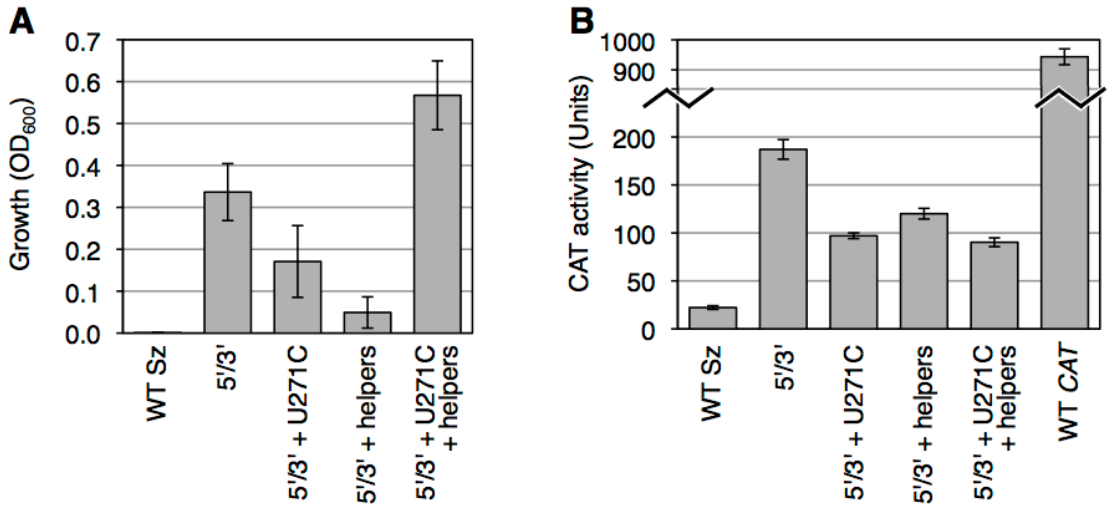


Figure 4.4 Activities of evolved spliceozyme mutations in *E. coli* cells. (A) Growth mediated by the spliceozyme variants in E. coli cells on medium containing 100 ug/mL chloramphenicol. Error bars are standard deviations from three experiments. (B) CAT enzyme activity of bacterial extract. The units correspond to 10^{12} cells. Error bars are standard deviations of three experiments.

Two experiments were conducted to determine whether spliceozyme variants produced an effect on cellular growth independent of the antibiotic resistance mediated by splicing. First, bacterial growth in liquid culture was measured in the absence of chloramphenicol (Fig. 4.11). Indeed, growth inhibition was observed in bacterial cells expressing both the cat pre-mRNA and either the parent spliceozyme, spliceozyme with 5-' and 3'-terminus mutations or spliceozyme with 5-' and 3'-terminus and helper mutations, as compared with cells expressing only the spliceozyme constructs. Similar growth inhibition was not observed in cells expressing the evolved W11 spliceozyme. Second, the splicing pattern of the parent and evolved W11 spliceozymes were compared when incubated with three essential E. coli mRNAs *in vitro* (Fig. 4.12). Although no significant difference in the amount of cleavage products was detected, a modulation of cleavage patterns was observed. It is therefore currently unclear whether the effect of the spliceozyme is mediated by the cleavage of CAT pre-mRNA, off target cleavage of endogenous *E. coli* mRNA, or both.

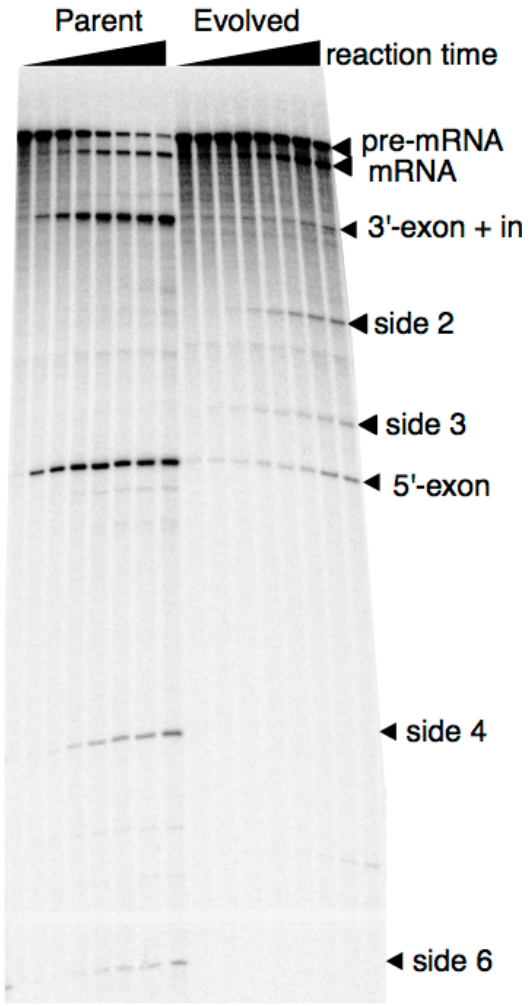


Figure 4.5 Effect of evolved spliceozyme mutations on the product pattern of *in vitro* splicing reactions. Shown is a phosphorimage of radiolabeled splicing products with two different spliceozymes, separated by denaturing polyacrylamide gel electrophoresis. The left time course shows the products from the parent spliceozyme; the right time course shows the products from the evolved spliceozyme W11. The time points are 0, 1, 2, 5, 10, 20, 30, and 60 minutes. The identity of the bands is given on the right. Unknown side products were labeled as 'side X'.

To test whether side product formation was reduced by the evolved spliceozyme, we performed *in vitro* splicing reactions with radiolabeled substrate RNA. Splicing products were separated by denaturing polyacrylamide gel electrophoresis (Fig. 4.5A) and the pattern of product bands was quantified (Fig. 4.5B). One striking difference between the parent and evolved spliceozyme was

the consumption of substrate; while the parent spliceozyme consumed ~80% of substrate within 10 minutes of reaction time, the evolved spliceozyme W11 reacted only with 46% of substrate during the same time. Importantly, formation of the major two side products of the splicing reaction formed by cleavage at the 5'-splice site, was significantly lower with the evolved spliceozyme (blue and red in Fig. 4.5B). Four additional unidentified bands were observed as side products during the *in vitro* reaction (side 2, 3, 4, and 6). Side products 2 and 3 were increased in the evolved spliceozyme W11 while bands 4 and 6 were decreased. The helper mutations did not produce a significant effect when combined with the 5'- and 3' mutations alone (compare 5'3' with 5'3' + helper in Fig. 4.5B). In contrast, the helper mutations caused a significant reduction of side product 2 when added to the U271C mutation (compare 5'3' + U271C with 5'3' + U271C + helpers in Fig. 4.5B). This is consistent with the cooperative effect between U271C and the helper mutations observed in the growth of *E. coli* on chloramphenicol containing medium (Fig. 4.4A, 4.4B). Similarly, the amount of product formation *in vitro* correlated well with the amount of active CAT enzyme produced in *E. coli* cells (Fig. 4.6). This suggested that the observed changes in splicing product patterns could be used to understand the phenotype in *E. coli*.

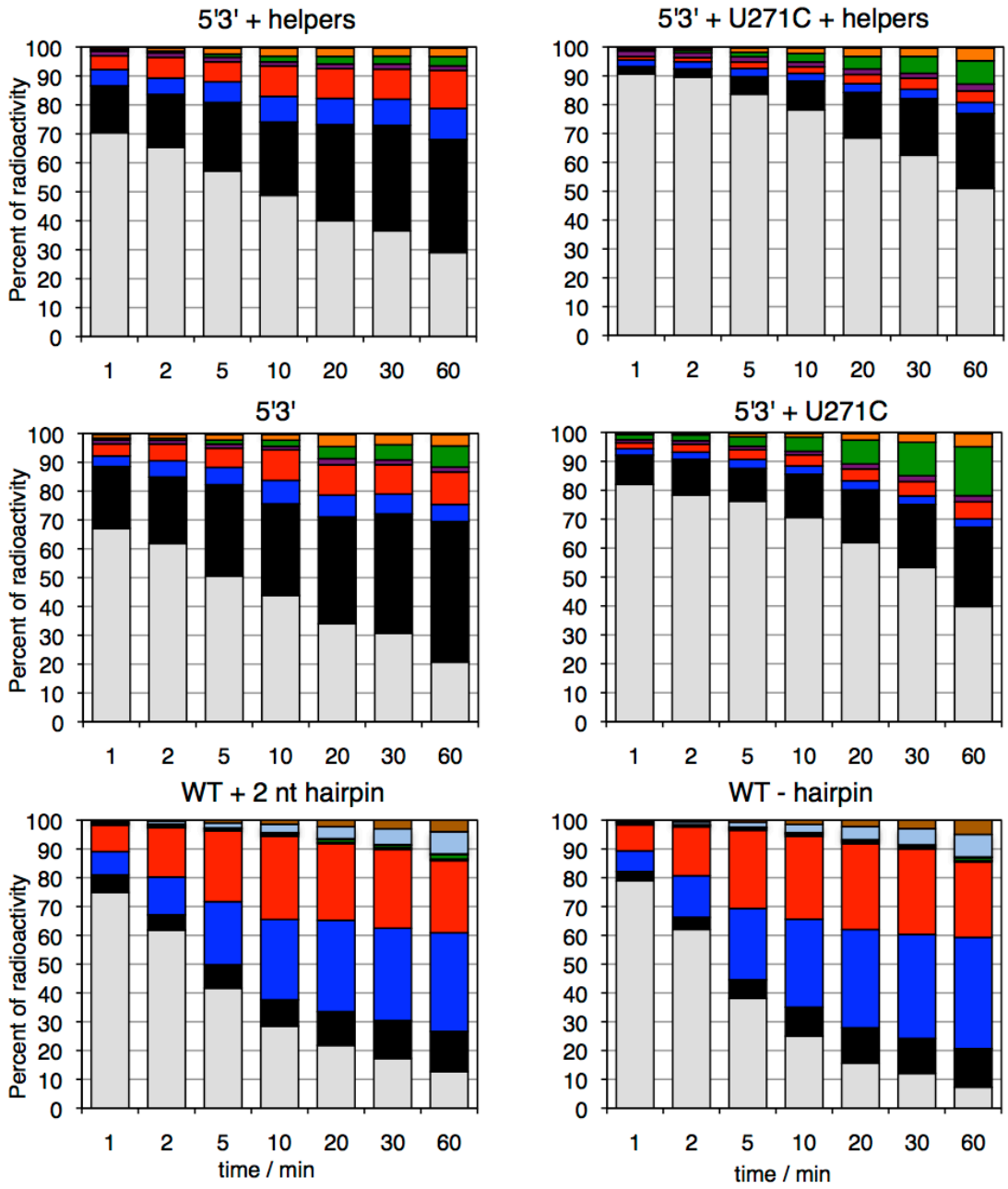


Figure 4.6 Effect of evolved mutations in the spliceozyme on the product pattern of *in vitro* splicing reactions. The percentage of radioactivity in specific bands is plotted as a function of the reaction time. Unreacted pre-mRNA (grey) is converted to mRNA (black). Cleavage products at the 5'-splice site are colored in red (5'-exon) and blue (3'-exon with intron). Additional side products are Unk1 (purple), Unk2 (green), Unk3 (orange), Unk4 (light blue), and Unk6 (brown). The values are the averages of three reactions. Error bars were not included for clarity and are given in Fig. 4.11 in plots of individual products.

To identify the RNA fragments that caused the observed changes in splicing product patterns *in vitro* (Fig. 4.5A), an RNase H assay was performed (Fig. 4.7). RNase H digests only RNAs that form double strands with DNA (23). Products of the splicing reaction with radiolabeled substrate were annealed with DNA oligonucleotides designed to base pair with segments along the length of the *CAT* pre-mRNA and incubated with RNase H. When these digestion products were separated by denaturing PAGE alongside undigested splicing products (Fig. 4.7A), the bands that contained sequences complementary to the DNA oligonucleotides shifted down or disappeared. By employing more than a dozen DNA oligonucleotides, it was possible to identify the location of side products 2, 3, 4, and 6 along the length of the *CAT* pre-mRNA (Fig. 4.7B). Side product 2 consisted of the 5'-exon connected to the intron, therefore corresponding to the 5'-fragment of cleavage at the 3'-splice site. Side product 3 contained the 5'-exon and a portion of the intron, corresponding to the 5'-fragment of a cleavage event within the intron. Side products 4 and 6 corresponded to two complementary fragments of the 5'-exon.

The limits of the RNase H assay, however, do not allow for single nucleotide resolution. To identify the cleavage sites with single nucleotide resolution, truncated 5'-radiolabeled substrates were generated, incubated with the spliceozyme, and analyzed by denaturing PAGE (Fig. 4.7C). As marker, the same radiolabeled substrates were incubated with DNAzymes (24) designed to cleave the substrates near the expected spliceozyme cleavage site. The results of this single-resolution assay confirmed that side product 2 was cleaved

precisely at the 3'-splice site, side product 3 had a 3'-terminus within the intron between nucleotides 317/318, and side products 4 and 6 were cleavage products of the 5'-exon between nucleotides 129/130, with perhaps a minor cleavage product at position 140/141 (see labels below Fig. 4.7B).

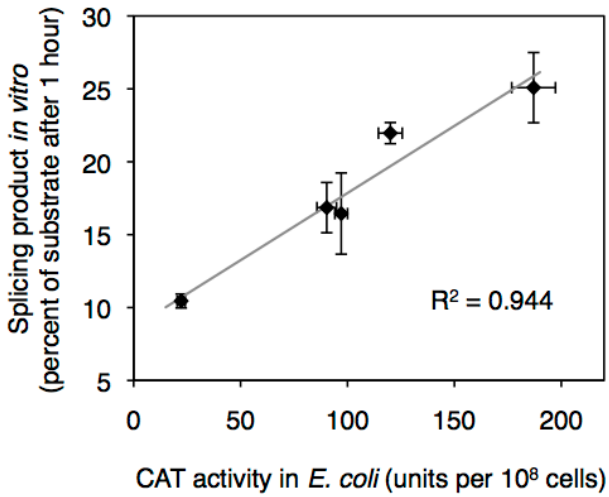
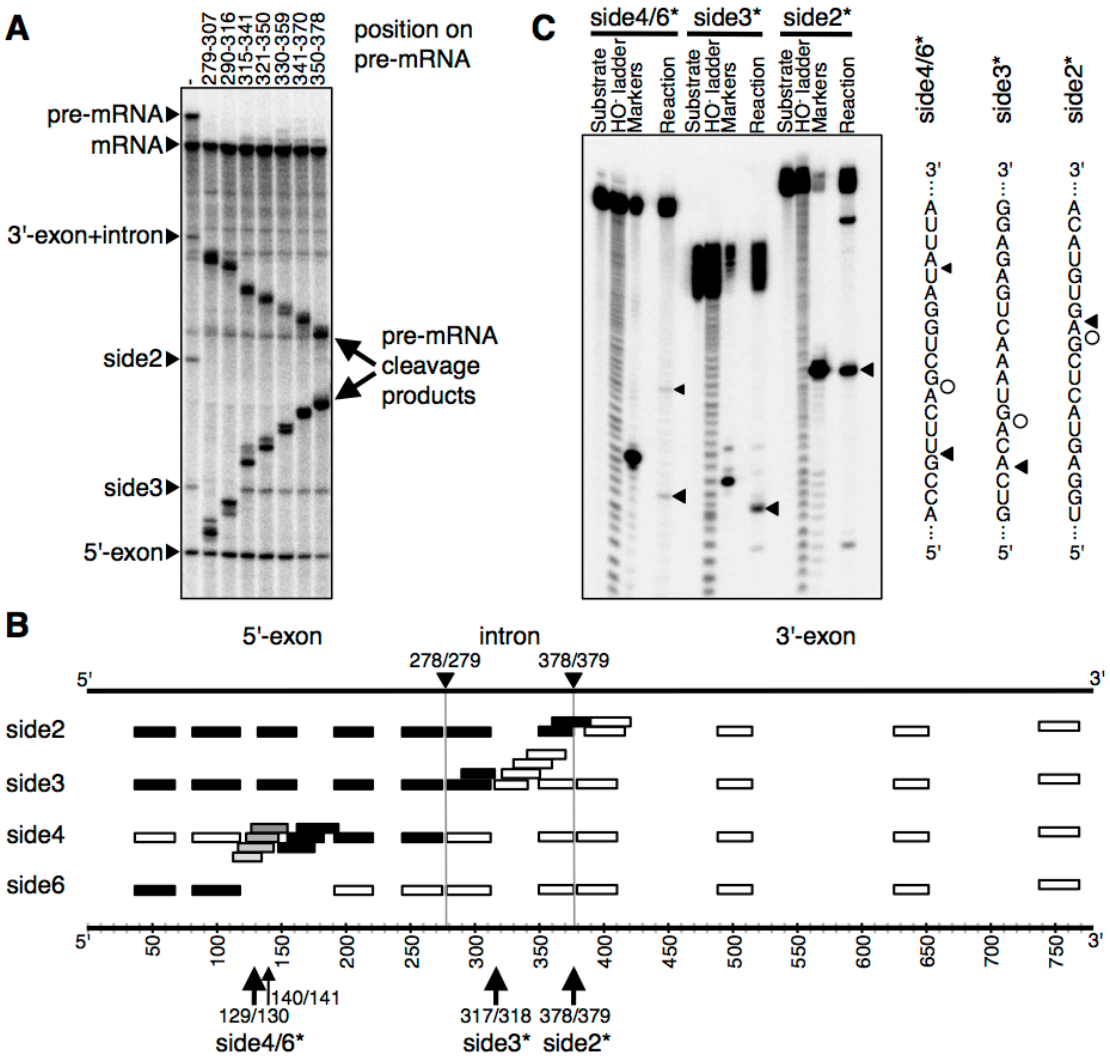


Figure 4.7 Correlation of product formation *in vitro* with CAT activity in *E. coli* cells. None of the other *in vitro* products was correlated similarly well with CAT activity or cell growth on chloramphenicol containing medium. Error bars are standard deviations from three experiments.

These results suggested that the mutations at the spliceozyme 5'- and 3'-terminus, which extended the substrate recognition duplexes, reduced premature release of cleavage products at the 5'-splice site by 6.1-fold (after 1 hour of splicing *in vitro*) and increased product (*CAT* mRNA) formation 3.5-fold. Cleavage products at the 5'-splice site were further reduced 2.0-fold by mutation U271C, which also increased cleavage at the 3'-splice site by 2.3-fold (side product 2). The six helper mutations reduced cleavage products at the 3'-splice site back by 2.1-fold. The overall effect of the evolved 5-‘ and 3’-terminus and U271C mutations was that of a re-balancing of activity between the 5'-splice site

(reduction) and 3'-splice site (increase), with additional 'helper' mutations to mitigate side effects from the increase in 3’-splice site activity.

Figure 4.8 Identification of side products from the *in vitro* splicing reaction. **(A)** Phosphorimage of a denaturing PAGE separation of spliceozyme reaction products with internally ^{32}P labeled *CAT* pre-mRNA. After the reaction, the products were annealed with DNA oligonucleotides and treated with RNase H. The disappearance of a band showed that the DNA oligonucleotides had complementarity to the band's RNA. The positions covered by seven representative DNA oligonucleotides are shown above the gel image. These positions extend over the entire intron sequence. **(B)** Schematic of *CAT* pre-mRNA with the position of DNA oligonucleotides indicated. Effects of the DNA oligonucleotides on side product bands side2, side3, side4, and side6 are indicated by filled rectangles (band disappears) or empty rectangles (band is unaffected). Grey rectangles indicate partial disappearance. The position of 5'-exon, intron, and 3'-exon in the pre-mRNA are indicated on top. The nucleotide position is given on the bottom. The two splice sites are denoted by black triangles and grey, vertical lines. Cleavage sites with single nucleotide resolution (see (C)) are given on the bottom. **(C)** Identification of the cleavage sites at single nucleotide resolution, using short substrate fragments with 5'-radiolabel. Each short substrate was incubated with an excess of spliceozyme to generate the cleavage products. Note that hydroxyl ladders and the marker created by a DNAzyme carry 3'-phosphates. This difference to the 3'-hydroxyl groups created by the spliceozyme makes the spliceozyme cleavage products migrate slower, corresponding to ~0.4 nucleotides. The sequences at which cleavage occurred are shown on the right, with triangles denoting cleavage sites by the spliceozyme and circles denoting the cleavage position by the DNAzyme to generate the Marker. The resulting cleavage sites are annotated at the bottom of sub-figure (B).


A
position on pre-mRNA
279-307
290-316
315-341
321-350
330-359
341-370
350-378
pre-mRNA
mRNA
3'-exon+intron
side2
side3
5'-exon
pre-mRNA cleavage products
C
side4/6*
side3*
side2*
Substrate
HO⁻ ladder
Markers
Reaction
B
5'-exon
intron
3'-exon
278/279
378/379
side2
side3
side4
side6
129/130
140/141
317/318
378/379
side4/6*
side3*
side2*

4.4 Discussion

The activity of group I intron spliceozymes was improved by evolution in *E. coli* cells. Three types of mutations were identified that improve their performance in cells. First, mutations extending the substrate recognition helices (P1 and P9.2) increased product formation and decreased side product formation. Second, mutation U271C in the conserved core of the spliceozyme further reduced side products generated by cleavage at the 5'-splice site but increased side products generated by cleavage at the 3'-splice site. Third, a group of six 'helper' mutations reduced the increase in side products from increased activity at the 3'-splice site by mutation U271C. Together, the resulting spliceozymes show stronger product formation and weaker side product formation.

The effect of mutations extending the 5’-substrate recognition sequence can likely be explained by previous observations that extending the P1 helix can increase *trans*-splicing efficiency, possibly through an improved conformational change between the two catalytic steps (19,25). Similarly, it is expected that mutations at the 3’ terminus, resulting in extension of the P9.2 helix, lead to better retention of the splicing intermediates due to increased duplex stability.

It is interesting that the U271C mutation, located in the conserved core of the group I intron, resulted in beneficial effects because mutations in the conserved core usually have detrimental effects of group I intron activity (5). The crystal structure of the *Tetrahymena* group I intron (2), provides a possible explanation as to how this mutation may have led to a decrease in activity of the 5'-splice site and an increase in activity of the 3'-splice site. This crystal structure

shows nucleobase U271 stacked upon the end of the P3 helix, oriented towards A103 at the same end of the helix (Fig. 4.8A). This crystal structure, however, lacks the substrate recognition helix P1 and therefore the 5'-splice site. The P1 helix is present in the crystal structures of two related group I introns from the species *Azoarcus* and *Twort* (Fig 4.8B, C) (3,4). Here, the nucleobases analogous to U271 are flipped out of helix P3 and stack upon a G in the junction J7/8. The junction J7/8 is important for binding the substrate recognition helix P1 through ribose zipper interactions and positioning the 5'-splice site to the catalytic site (26). The P1 helix is known to flip in and out of the active site frequently (27), therefore influences on stabilizing the P1 helix to the catalytic site can be expected to have a strong influence on 5'-splice site activity. In summary, modulation of splice site activity by mutation U271C may have been caused by an indirect effect on positioning of the P1 substrate recognition helix into the catalytic site.

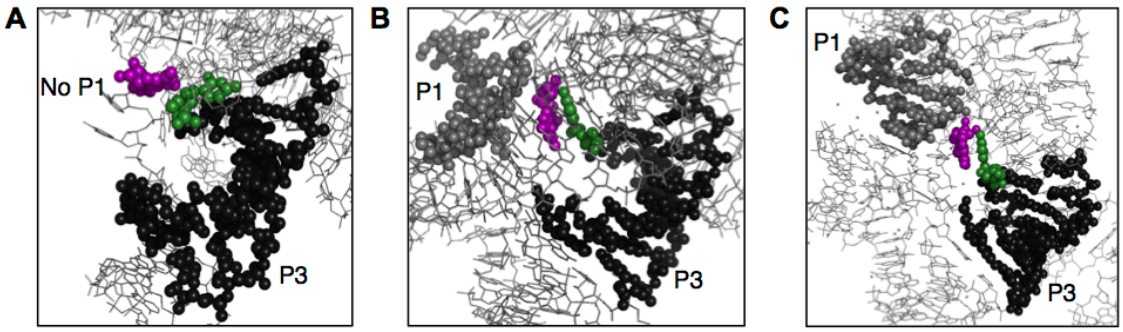


Figure 4.9 Crystallographic structures of group I introns that suggest a model for the function of mutation U271C. Shown are small portions of the structure that include the core helix P3 onto which U271 stacks. **(A)** The structure of a group I intron from *Tetrahymena* (1X8W) was made in the absence of the P1 helix. The nucleobase U271 (green spheres) is stacked onto helix P3 (black spheres) and points away from nucleotide G303 (purple spheres). The latter resides in the junction J7/8, which is known to assist the positioning of helix P1 into the catalytic site. **(B)** The structure of a group I intron from *Twort* (1Y0Q) was made in the presence of helix P1 (grey spheres). Here, the nucleotide analogous to U271 is stacked onto the nucleotide analogous to G303. **(C)** Similarly, the structure of a group I intron from *Azoarcus* (1ZZN) shows the stacking of the nucleobase analogous to U271 onto the base analogous to G303, in the presence of the P1 helix.

Two of the six 'helper mutations', A334G and G368D, are located at the base of the P9.2 helix at the entry point of the substrate intron into the catalytic site (Fig. 4.3B). It is possible that these two mutations serve to modulate the effect of increased activity at the 3'-splice site from the U271C mutation, thereby potentially explaining the cooperative effect between U271C and the helper mutations (Fig. 4.4A). The mechanism of the four remaining helper mutations is less clear. These mutations reside in the periphery of the spliceozyme, and not close to the catalytic site, the substrate binding sites, or the U271C mutation suggesting that these four mutations may exert their effect via long-distance interactions. The evolved spliceozymes may therefore be a good model system to study how long-range interactions are able to modulate the function of catalytic

RNAs.

It may be possible to develop spliceozymes that can treat certain types of mis-splicing disorders including intron and pseudoexon inclusion. In addition, certain disorders caused by mutations leading to premature termination or out of frame transcripts, can be treated by removal of non-essential portions of the mRNA; these too could be targeted by the spliceozyme. The evolved spliceozymes may be one step closer to possible applications in therapy because they now display an increase in product formation and decreased level of side product formation. The observation of these effects *in vitro* suggests that these results are not dependent on *E. coli* specific cellular factors, and therefore may be applicable to a broad range of cell types. A safe and effective delivery method is still needed before this system could be applied clinically.

Evolution experiments of spliceozymes in cells can also be seen as model system for the evolution of specific biochemical steps in the evolution of the spliceosome, which appears to have evolved from a common ancestor with group II intron ribozymes (28). While the 'group I spliceozyme' model system is a long way from resembling the spliceosome's impressive ability to recognize with high precision and very low error rate thousands of different splice sites, the results shown here are encouraging that at least some of the biochemical steps in the evolution of the spliceosome could be recapitulated by the further evolution of these *trans*-splicing group I intron ribozyme variants.

4.5 Materials and methods

Library plasmid

Library plasmid was constructed essentially as described (22) and expressed the indicated spliceozyme variant and the chloramphenicol acetyltransferase (*CAT*) pre-mRNA with a 100 nucleotide intron inserted at position 258. Spliceozyme expression is driven by the *trc2* promoter is inducible with IPTG, while the CAT pre-mRNA expression is driven by the constitutive promoter derived from the pLysS plasmid (Novagen). Both the spliceozyme and *CAT* pre-mRNA contain a 3' hairpin terminator sequence.

Evolution of spliceozymes in *E. coli* cells

The evolution was completed essentially as described (19) Briefly, spliceozyme constructs were randomized by mutagenic PCR and cloned into library plasmid. Library plasmids containing spliceozyme sequences were transformed into *E. coli* cells, plated on LB agar medium containing 100µg/mL ampicillin (amp) and incubated at 37°C for 16 hours to create pools. LB agar plates containing pools of *E. coli* cells were washed using liquid LB medium. The resulting medium was diluted to an OD600 of 0.0015, induced with IPTG to a final concentration of 1mM and shaken at 37°C for 1 hour. *E. coli* cells were then plated on LB agar medium containing the indicated concentration of chloramphenicol (cam) and incubated at 37°C for 16 hours to select for spliceozyme constructs able to mediate antibiotic resistance. LB agar plates were washed using liquid LB medium and plasmids were isolated by miniprep (5 Prime). 5 or 10 plasmids were chosen for sequencing from each round of evolution.

Generation of spliceozyme constructs by site directed mutagenesis

Internal mutations were inserted into spliceozyme constructs by site directed mutagenesis (29). Briefly, each 50µL reaction contained 8 ng template plasmid, 1.25 pmol forward primer, 1.25 pmol reverse primer, 2.5 nmol each dNTP, 2.5 units PrimeSTAR GXL DNA Polymerse (Takara), and 1x PrimeSTAR GXL buffer. PCR consisted of 5:00/95°C,18 cycles of 0:50/95°C, 0:50/60°C, 4:45/68°C, followed by 7:00/68°C. After PCR 20 units of DpnI restriction enzyme were added, the reaction was incubated at 37°C for 1 hour then purified (Zymogen DNA Clean and Concentrate). Reactions were transformed into DH5α by electroporation.

PCR mutagenesis

Mutations were randomly introduced into spliceozyme sequences using PCR mutagenesis (30). Each 100µL reaction contained 25ng template, 1µM forward primer, 1µM reverse primer, 10mM Tris-Cl pH 8.3, 50mM KCl, 7mM $MgCl_2$, 0.05% gelatin, 0.2mM dATP, 1.0mM dCTP, 0.2mM dGTP, 1.0mM dTTP, 0.5mM $MnCl_2$ and 1µL Taq polymerase. PCR consisted of 0:30/94°C,and 10 cycles of 0:30/94°C, 0:30/50°C, 1:00/72°C.

Measurement of bacterial growth on LB agar plates

Measurement of bacterial growth on LB agar plates was completed essentially as described (22). 5mL of liquid LB medium containing 100µg/mL ampicillin was inoculated and grown at 37°C for 5 hours. The OD600 of each culture was measured, diluted to an OD600=0.0025, induced with IPTG to a final concentration of 1mM and incubated, shaking at 37°C for 1 hour. 100µL of each culture was plated on an LB agar plate containing the indicated chloramphenicol

concentration and an LB agar plate containing 100μg/mL ampicillin. Plates were incubated at 37°C for 16 hours then washed with 1.6 mL liquid LB medium. The OD600 of each wash was measured; measurements from ampicillin plates were used for normalization. Experiments were completed in triplicate.

Measurement of bacterial growth in liquid culture

Measurement of bacterial growth in liquid culture was completed essentially as described (19). Overnight cultures of *E. coli* containing the library plasmid with the indicated spliceozyme variant were induced with IPTG to a final concentration of 1mM and incubated at 1 hour shaking at 37°C. Each culture was diluted to an OD600 of 0.05 with LB medium containing 100 μg/mL ampicillin, induced with IPTG to a final concentration of 1mM and incubated, shaking at 37°C. OD600 measurements were taken after 1 hour, and every following 30 minutes until an OD600 of 1.0 was reached. Doubling times were determined by least squares fitting.

Assay for CAT activity from *E. coli* extracts

The measurement of CAT activity from *E. coli* extracts was completed essentially as described (31). Briefly, overnight cultures of *E. coli* containing the library plasmid with the indicated spliceozyme variant were induced with IPTG to a final concentration of 1mM and incubated 1 hour shaking at 37°C. Each culture was diluted to an OD600 of 0.02 with LB medium containing 100 μg/mL ampicillin, induced with IPTG to a final concentration of 1mM and incubated, shaking at 37°C. When the OD600 reached 0.20, 2.0 mL of cells were concentrated to 200uL and frozen. During the reaction, cells were thawed and

mixed with 200 µL 200mM Tris-Cl pH 7.8, 10mM Na_2EDTA and 2uL toluene. 15 µL of this solution was diluted with 135 µL buffer to a final concentration of 0.2 mM chloramphenicol, 0.2 mM Acetyl-CoA and 1mM DTNB and A_{412} was measured every 15 seconds. The extinction coefficient of the reaction product (13,600 $M^{=1}$ cm^{-1}) was used to determine the units of CAT, with one unit of CAT being able to acetylate 1µmol chloramphenicol/minute. Note that measurements from cells containing wild type CAT used only 200µL of cell culture.

In vitro trans-splicing assays

In vitro trans-splicing assays were completed essentially as described (12). Spliceozyme and *CAT* pre-mRNA transcripts were generated by run off transcription of PCR products and purified by PAGE. *CAT* pre-mRNA transcripts were internally labeled with α(^{32}P)-ATP. 20 µL Splicing reactions contained 1µM spliceozyme, 100 nM *CAT* pre-mRNA, 5 mM $MgCl_2$, 50 mM MOPS-NaON pH 7.0, 2 mM spermidine, 135 mM KCl, and 20 µM GTP. Spliceozyme and *CAT* pre-mRNA were incubated separately for 10 minutes at 37°C then combined and incubated at 37°C. 2µL samples were taken at 0, 1, 2, 5, 10, 20, 30 and 60 minutes and separated by 6% PAGE. Gels were imaged on a phosphoimager (Bio-Rad) and bands were quantified (Quantity One). Signal strengths were normalized by the number of A's contained within each sequence and total radioactivity per lane.

Identification of in vitro splicing side products by RNase H (Fig. 4.7A)

In vitro trans-splicing reactions were run for 60 minutes (as described above) and quenched with EDTA to a final concentration 1.2 fold higher than

$MgCl_2$. 8 µL of each reaction was added to 2uL of the indicated DNA oligonucleotide, to a final concentration of 1 µM. The oligonucleotide was allowed to anneal to the products of the splicing reaction by heating to 80°C for 2 minutes, followed by cooling to 25°C. 2 µL $MgCl_2$ was added to the reaction in stoichiometric equivalent to the EDTA present. 50 µL RNase H reactions contained the 12 µL annealing reaction, 1x RNase H Reaction Buffer and 2.5 units RNase H (New England Biolabs) and were incubated at 37°C for 1 hour. Reactions were purified by ethanol precipitation and separated by 6% PAGE. Gels were imaged on a phosphoimager (Bio-Rad). DNA oligonucleotides were designed to bind to the following nucleotide postions along the *CAT* pre-mRNA: 35-68, 81-108, 112-139, 117-144, 122-149, 127-154, 132-162, 147-177, 153-182, 158-190, 162-194, 191-221, 245-276, 279-307, 290-316, 315-341, 321-350, 330-359, 341-370, 350-378, 360-391, 370-401, 380-411, 385-416, 390-421, 487-516, 624-653, 739-770.

Identification of in vitro splicing side products with single nucleotide resolution (Fig. 4.7C)

Substrate RNA oligonucleotides (below) that contained the suspected splice sites as shown by the RNAse H assays, were generated by run off transcription and purified by 10% PAGE. Substrates were dephosphorylated at 37 °C for 1 hour. 10 µL dephosphorylation reactions contained 2ng substrate oligonucleotide, 1X Antarctic Phosphatase Reaction Buffer and 5 units Antarctic Phosphatase (New England Biolabs). Substrates were then 5' radiolabeled at 37 °C for 1 hour. 20 µL labeling reactions consisted of 20 pmol RNA substrate,

20μCi γ(^{32}P)-ATP, 1x T4 Polynucleotide Kinase Reaction Buffer and 10 units T4 Polynucleotide Kinase (New England Biolabs). Radiolabeled substrates were purified by PAGE. DNAzymes (below) were designed to cleave substrate oligonucleotides between the underlined AG position to create markers. 20μL DNAzyme reactions consisted of traces of labeled substrate oligonucleotides, 1μM DNAzyme, 150mM NaCl, 100mM $MgCl_2$ and 50mM Tris-Cl pH 8.3. Reactions were heated to 80°C for 2 minutes, then incubated at 37 °C for 1 hour. The addition of $MgCl_2$ and Tris-Cl was after reactions reached 37 °C. Reactions were separated by 6% PAGE and gels were imaged on a phosphoimager (Bio-Rad). The following oligonucleotide and DNAzyme sequences were used:

Side 2 oligonucleotide:

GGUGUAAGCUCUCCCCUUGCAGAUUAGUUUUGGAGUACUCGAGUGUACA

CCCUUGUUACACCGUUUUCCAUGAGCAAACUGAAA

Side 2 DNAzyme: AACAAGGGTGTACACTCCGAGCCGGACGAGAGTACTC

Side 3 oligonucleotide:

GAAGGGUUCCUGGGAAGGCCUUCGGGUCACAGUAAACUGAGAGGCUUUG

GUG

Side 3 DNAzyme: CACCAAAGCCTCTCCGAGCCGGACGAAGTTTACT

Side 4 and side 6 oligonucleotide:

AGUUGCUCAAUGUACCUAUAACCAGACCGUUCAGCUGGAUAUUACGGCCU

UUUUAAAGACCGUAAAGAAAAAUAAGC

Side 4 and 6 DNAzyme:

AAAAAGGCCGTAATATCCGAGCCGGACGACAGCTGAA

In vitro splicing reactions on three essential *E. coli* mRNAs (Fig. 4.13)

E. coli genomic DNA was prepared by proteinase K digest of the cell pellet from a logarithmically grown culture, and two consecutive ethanol precipitations. The desired gene fragments were amplified from this preparation of genomic DNA by PCR. Transcripts of the first 602, 586, and 633 nucleotides of *E. coli* EF-Tu, Pyruvate Dehydrogenase Subunit E2, and DNA Polymerase III Subunit α genes, respectively, were generated by run off transcription from PCR products and purified by PAGE. The nucleotides GGGAG were added to the 5' end of each substrate to enhance transcription efficiency. Transcripts were internally labeled with α(^{32}P)-ATP. Splicing reactions with a volume of 20 µL contained 1µM spliceozyme, 100 nM of the indicated RNA substrate, 5 mM $MgCl_2$, 50 mM MOPS-NaOH pH 7.0, 2 mM spermidine, 135 mM KCl, and 20 µM GTP. Spliceozyme and RNA substrate were incubated separately for 10 minutes at 37°C then combined and incubated at 37°C. Samples with a volume of 2µL were taken at 0, 1, 2, 5, 10, 20, 30 and 60 minutes and separated by 6% PAGE. Gels were imaged on a phosphoimager (Bio-Rad) and bands were quantified (Quantity One). Signal strengths were normalized by the total radioactivity per lane.

4.6 Acknowledgements

We thank Simpson Joseph, Terry Hwa, and Jens Lykke-Andersen for helpful discussions.

Chapter 4, in full, is currently being prepared for submission for publication, **Amini Z.N.** and Muller U.F. Decreased side product formation in

evolved group I intron spliceozymes. The dissertation author is the first author on this paper.

4.7 Supporting Information

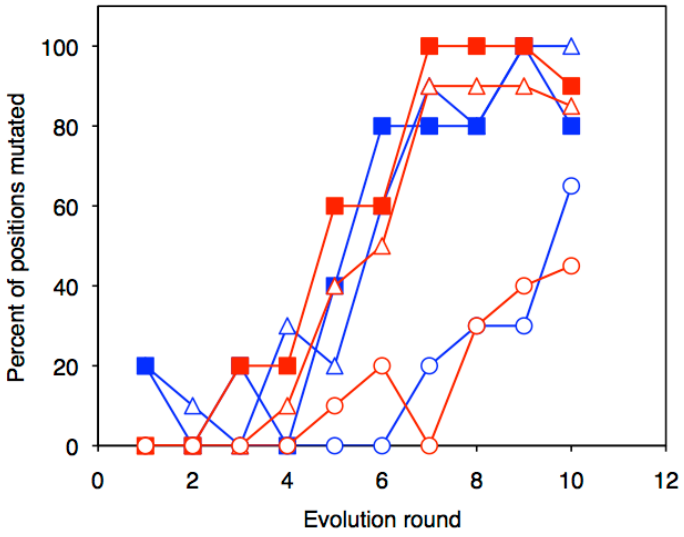


Figure 4.10 Enrichment of key mutations during evolution. Shown is the percentage of spliceozyme clones that were mutated at a given position. Data from the line starting at a single sequence are in blue; data from the pre-evolved line are in red. Mutations at the two 5'-terminal positions are labeled as open triangles, the U271C mutation is labeled with filled squares, and mutations at the two 3'-terminal positions are shown as open circles. In rounds 1-9 of the evolution, five clones were sequenced; in round 10, ten clones were sequenced.

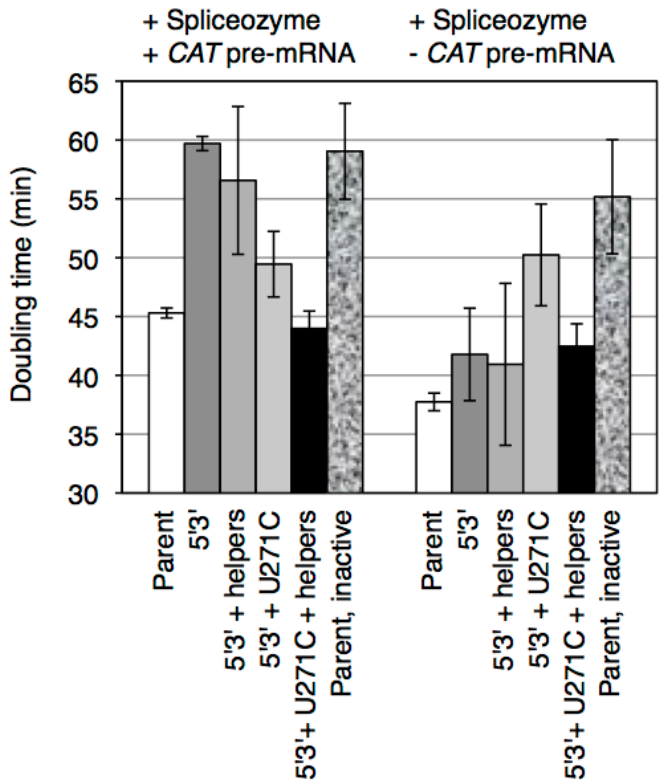


Figure 4.11 Growth inhibition of *E. coli* via the expression of spliceozyme variants and CAT pre- mRNA. Growth was measured by the increase in OD_{600} in liquid medium in the absence of chloramphenicol. Six spliceozyme variants were tested, in the presence (left) or absence (right) of CAT pre-mRNA. The spliceozyme variants are labeled below the graph. Error bars are standard deviations from biological triplicates.

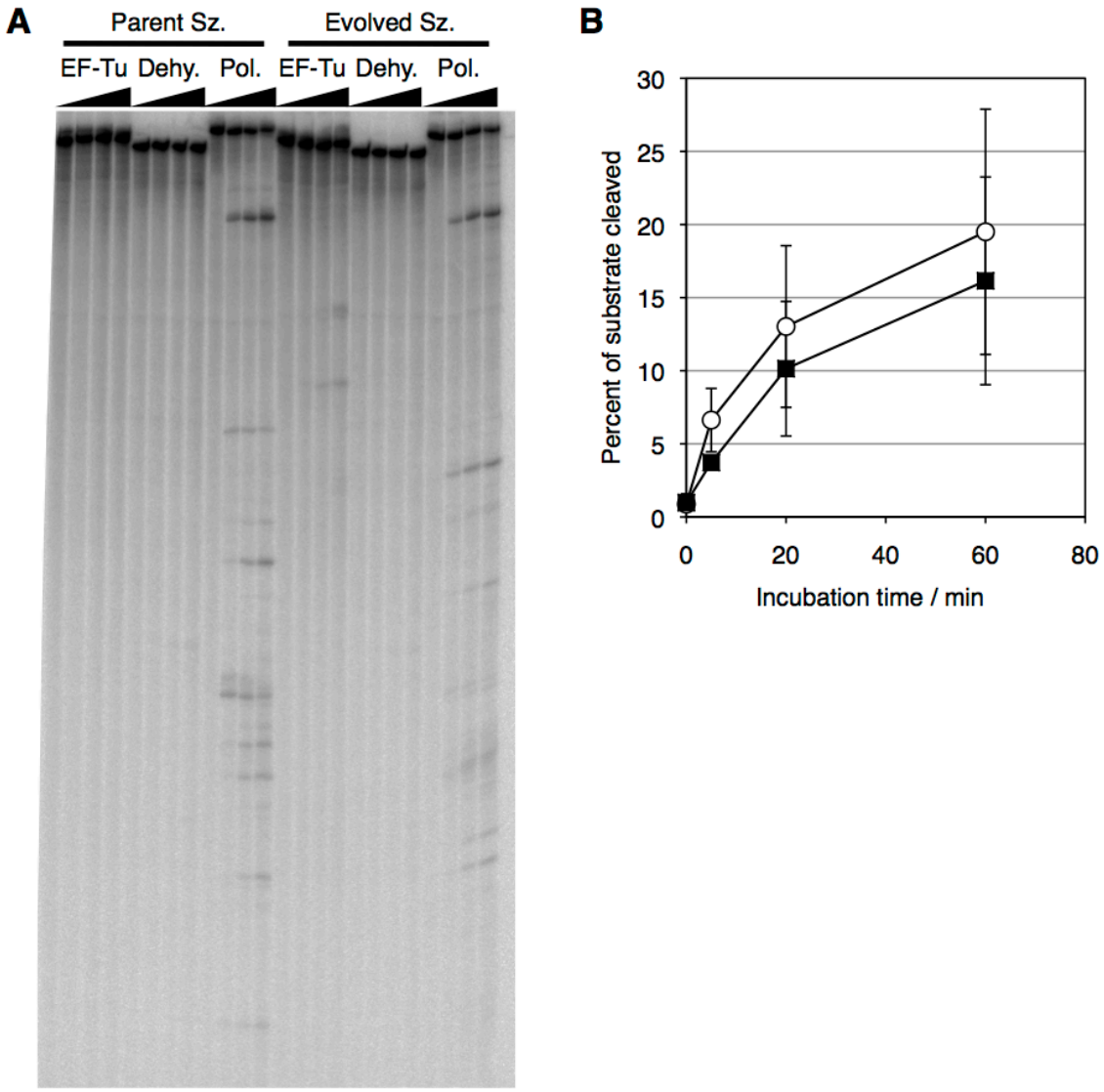


Figure 4.12 Off-target effects of the spliceozyme on three essential mRNAs in *E. coli*. (A) Phosphorimage of internally radiolabeled RNAs after incubation with spliceozyme. The sequences are from the mRNAs of E. coli elongation factor Tu (EF-Tu), E. coli pyruvate dehydrogenase sub-complex E2p (Dehy.), and E. coli DNA polymerase 3 (Pol.). The black triangles illustrate increasing incubation times (0 min., 5 min., 20 min., and 60 min.) with parent spliceozyme (Sz) or evolved spliceozyme. (B) Quantitation of the cleavage activity of DNA polymerase mRNA. The intensity of all bands visible for the parent spliceozyme (empty circles) and the evolved spliceozyme (black squares) is plotted as a function of incubation time with the spliceozyme. Error bars are standard deviations from three experiments.

4.8 References

1. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. (1982) Self-splicing RNA: autoexcision and

autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31, 147-157

2. Guo, F., Gooding, A. R., and Cech, T. R. (2004) Structure of the Tetrahymena ribozyme: base triple sandwich and metal ion at the active site. *Mol Cell* 16, 351-362

3. Golden, B. L., Kim, H., and Chase, E. (2005) Crystal structure of a phage Twort group I ribozyme-product complex. *Nat Struct Mol Biol* 12, 82-89

4. Stahley, M. R., and Strobel, S. A. (2005) Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 309, 1587-1590

5. Michel, F., and Westhof, E. (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216, 585-610

6. Cech, T. R. (1990) Self-splicing of group I introns. *Annu Rev Biochem* 59, 543-568

7. Stahley, M. R., and Strobel, S. A. (2006) RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr Opin Struct Biol* 16, 319-326

8. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing. *Nature* 371, 619-622

9. Alexander, R. C., Baum, D. A., and Testa, S. M. (2005) 5' transcript replacement in vitro catalyzed by a group I intron-derived ribozyme. *Biochemistry* 44, 7796-7804

10. Bell, M. A., Johnson, A. K., and Testa, S. M. (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs. *Biochemistry* 41, 15327-15333

11. Dotson, P. P., 2nd, Johnson, A. K., and Testa, S. M. (2008) Tetrahymena thermophila and Candida albicans group I intron-derived ribozymes can catalyze the trans-excision-splicing reaction. *Nucleic Acids Res* 36, 5281-5289

12. Amini, Z. N., Olson, K. E., and Muller, U. F. (2014) Spliceozymes: ribozymes that remove introns from pre-mRNAs in trans. *PLoS One* 9, e101932

13. Ketzer, P., Haas, S. F., Engelhardt, S., Hartig, J. S., and Nettelbeck, D. M. (2012) Synthetic riboswitches for external regulation of genes transferred

by replication-deficient and oncolytic adenoviruses. *Nucleic Acids Res* 40, e167

14. Ewert, K. K., Ahmad, A., Bouxsein, N. F., Evans, H. M., and Safinya, C. R. (2008) Non-viral gene delivery with cationic liposome-DNA complexes. *Methods Mol Biol* 433, 159-175

15. Chen, Y., Lian, G., Liao, C., Wang, W., Zeng, L., Qian, C., Huang, K., and Shuai, X. (2012) Characterization of polyethylene glycol-grafted polyethylenimine and superparamagnetic iron oxide nanoparticles (PEG-g-PEI-SPION) as an MRI-visible vector for siRNA delivery in gastric cancer in vitro and in vivo. *J Gastroenterol*

16. Namiki, Y. (2013) Synthesis of lipidic magnetic nanoparticles for nucleic acid delivery. *Methods Mol Biol* 948, 243-250

17. Yuhua, L., Kunyuan, G., Hui, C., Yongmei, X., Chaoyang, S., Xun, T., and Daming, R. (2001) Oral cytokine gene therapy against murine tumor using attenuated Salmonella typhimurium. *Int J Cancer* 94, 438-443

18. Bai, Y., Gong, H., Li, H., Vu, G. P., Lu, S., and Liu, F. (2011) Oral delivery of RNase P ribozymes by Salmonella inhibits viral infection in mice. *Proc Natl Acad Sci U S A* 108, 3222-3227

19. Olson, K. E., and Muller, U. F. (2012) An in vivo selection method to optimize trans-splicing ribozymes. *RNA* 18, 581-589

20. Olson, K. E., Dolan, G. F., and Muller, U. F. (2014) In vivo evolution of a catalytic RNA couples trans-splicing to translation. *PLoS One* 9, e86473

21. Hayden, E. J., Ferrada, E., and Wagner, A. (2011) Cryptic genetic variation promotes rapid evolutionary adaptation in an RNA enzyme. *Nature* 474, 92-95

22. Amini, Z. N., and Muller, U. F. (2013) Low selection pressure aids the evolution of cooperative ribozyme mutations in cells. *J Biol Chem* 288, 33096-33106

23. Davis, R. E., Davis, A. H., Carroll, S. M., Rajkovic, A., and Rottman, F. M. (1988) Tandemly repeated exons encode 81-base repeats in multiple, developmentally regulated Schistosoma mansoni transcripts. *Mol Cell Biol* 8, 4745-4755

24. Santoro, S. W., and Joyce, G. F. (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry* 37, 13330-13342

25. Kohler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999) Trans-splicing ribozymes for targeted gene delivery. *J Mol Biol* 285, 1935-1950

26. Silverman, S. K., and Cech, T. R. (1999) Energetics and cooperativity of tertiary hydrogen bonds in RNA structure. *Biochemistry* 38, 8691-8702

27. Shi, X., Mollova, E. T., Pljevaljcic, G., Millar, D. P., and Herschlag, D. (2009) Probing the dynamics of the P1 helix within the Tetrahymena group I intron. *J Am Chem Soc* 131, 9571-9578

28. Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P., and Piccirilli, J. A. (2013) RNA catalyses nuclear pre-mRNA splicing. *Nature* 503, 229-234

29. Weiner, M. P., and Costa, G. L. (1994) Rapid PCR site-directed mutagenesis. *PCR Methods Appl* 4, S131-136

30. Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 2, 28-33

31. Shaw, W. V. (1975) Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods Enzymol* 43, 737-755

Chapter 5: Future Directions

5.1 Introduction

The work in this dissertation has used *trans*-splicing group I introns to move forward in two distinct, yet related, areas of research. First, the group I intron was used as a model system to study the effect of evolutionary parameters on evolving populations of RNA in cells. This study, which compared high vs. low selection pressure applied in a controlled lab evolution experiment, found that low selection pressure is able to aid the evolution of cooperative mutations, while high selection pressure inhibits the emergence of cooperative mutations. In order to understand this observation, investigation of the fitness landscape connecting the parent clone to the evolved clone, containing the set of cooperative mutations, was conducted. It was shown that low selection pressure facilitates greater genetic diversity by allowing the survival of variants that would be prevented at a higher selection pressure.

The second body of work contained in this dissertation involved the development and evolutionary improvement of a new variant of the *trans*-splicing group I intron, termed the 'spliceozyme'. Sequence alterations to both the 5' and 3' end of the group I intron allow the intron to catalyze the removal of internal sequences from pre-mRNA substrates in *trans,* analogous to the reaction performed by the spliceosome. This reaction was found to be efficient both *in vitro* and in cells, and could be applied to a wide variety of internal sequences. The evolutionary improvement of the spliceozyme performed in *E. coli* cells led to

a spliceozyme construct with increased product formation and decreased side product formation. The detailed analysis of this effect was done *in vitro*, suggesting that the observed effect is not dependent on *E. coli* specific cellular factors.

5.2 Future Spliceozyme Applications

It is estimated that 10-60% of disease causing mutations result in aberrant splicing (1,2); spliceozymes could be designed to treat certain cases of aberrant splicing, including intron inclusion or pseudoexon inclusion. Niemann-Pick disease is one example of a disease involving pseudoexon inclusion, resulting in a premature termination codon and, subsequently, transcript degradation (3). In addition, certain mis-splicing events leading to out of frame transcripts or premature termination codons can be treated by removing non essential segments of the mRNA (1); these too could be targeted by the spliceozyme. One example of such a disorder is Duchenne muscular dystrophy (DMD) is a disease caused by mutations and deletions in the dystrophin gene, leading to out of frame transcripts and premature termination codons, depending on the exact mutation. Removing certain portions of the mRNA allows the protein to retain most of its function, thus lessening the severity of DMD. In order for the spliceozyme to be used in a clinical setting, it is necessary that the spliceozyme be able to efficiently remove internal fragments in mammalian cells, and that an effective and safe delivery method to be developed.

Optimization in a mammalian system. The spliceozyme evolution performed in *E. coli* cells was an excellent system for initial experiments and

optimization of the system owing to a previously established, time efficient evolutionary scheme (4). This evolution in *E. coli* cells resulted in of increased product formation and decreased side product formation that does not appear to be dependent on *E. coli* specific factors. Several long term natural evolutions of group I introns, however, have resulted in dependence upon proteins specific to the particular cell type. Examples of this include the dependence on the Cbp2 protein by the bI5 group I intron found in *Saccharomyces cerevisiae* (5) and the dependence on tyrosyl-tRNA synthetase in group I introns found in *Neurospora Crassa* (6). In addition, an artificial evolution of a group I intron in *E. coli* resulted recruitment of the transcription termination factor Rho (4). The cellular environment, including available proteins, differs greatly between *E. coli* and mammalian cells. In order to be implemented clinically, further optimization of the spliceozyme should be completed in a mammalian environment to ensure that the spliceozyme is able to efficiently catalyze intron removal in this environment and does not develop a dependence on factors specific to *E. coli* cells. Experiments in mammalian cells could challenge the spliceozyme to remove an intron from the puromycin resistance gene, selecting for ones capable of conferring growth of cells in puromycin containing growth media. This concept could be applied to an evolution, as is likely that the spliceozyme would benefit from further optimization to increase efficiency and specificity in mammalian cells. Once activity has been established and enhanced through evolution in mammalian cells, the spliceozyme could then be designed to repair clinically relevant mis-splicing events or mutations in mRNA substrates.

Multiple turnover spliceozymes. Although a spliceozyme variant has been evolved with higher product formation and lower side product formation, evidence of performing multiple turnover reactions has not yet been observed. Unlike several other *trans*-splicing group I intron variants, the spliceozyme is not limited by design to single turnover reaction as it does not transfer its own exon, nor is its sequence modified, during the splicing reaction. The possibility of multiple turnover reactions therefore still remains. The ability to perform multiple turnover reactions has the potential to not only increase product formation, but to also decrease the necessary cellular concentration of spliceozyme, which may serve to reduce adverse side effects of spliceozyme expression in cells. Evolution in the presence of shorter 5' and 3' recognition sequences or decreased promoter strength may help develop a ribozyme capable of performing multiple turnover reactions. *E. coli* cells would be an excellent system to use for such an evolution because the evolutionary scheme has already been established and has been used several times. In addition, one round of evolution in *E. coli* cells takes approximately 4 days, which makes this system much more time efficient than an evolution in mammalian cells. One drawback of performing an evolution in *E. coli*, however, is that the spliceozyme may evolve to utilize cellular factors specific to *E. coli*. This would not be applicable for use therapeutically in a mammalian system.

Methods for clinical delivery. In addition to developing and further improving spliceozyme efficiency in a mammalian system, a safe and effective method for clinical delivery must also be developed before the spliceozyme can

be used therapeutically. Delivery could come in the form of sustained expression or periodic administration but must be able to achieve tissue specific delivery while minimizing immune responses and other deleterious side effects. Several RNA delivery systems are currently being developed, with viral vectors, liposomes and nanocapsules being a few of the most common approaches.

Lentiviral and adenoviral vector delivery systems have been explored as potential options for delivery of therapeutic RNA. Adeno-associated viral vectors are attractive option with the potential to transduce dividing and quiescent cells, genomes that can persist in episomal form or integrate into specific sites and low immunogenic responses (7,8). Despite the potential advantages of adeno-associated viral vectors, deleterious effects of genomic integration, as well as unwanted immune responses have been observed for these and other viral vectors and have limited the their clinical use (7–9). Although methods to improve adeno-associated viral vectors are being developed, these harmful side effects have led to an increased interest in many non-viral delivery vehicles including liposomes and nanocapsules, requiring periodic administration.

Liposomes, able to fuse with cell membranes and release compounds contained within their core or attached to their surface, are potential delivery vehicles for therapeutic RNA. Neutral liposomes offer the advantage of low toxicity, but also come with the disadvantage of low cellular uptake. Cationic liposomes have the advantage of more efficiently facilitating cellular uptake of therapeutic RNA. An advantage conferred by both types of liposomes is the ability to protect RNA from degradation before entering the cells (10).

Immunoliposomes, liposomes with antibodies attached to the surface targeting cell specific receptors, are being developed to deliver RNA to targeted tissues (11,12). Unfortunately, potential negative side effects of liposomes have been observed, including aggregation in the blood and inflammatory reactions (13).

Several nanomaterials have demonstrated low toxicity, as well as favorable biological degradation profiles, and are therefore another promising candidate for clinical RNA delivery (9). In addition, nanospheres have the ability to cross the blood-brain, blood-testicle, and placental barriers and have shown fast cellular uptake (14–16). Like liposomes, nanoparticles containing peptides for targeting specific tissues are being developed (17). As it currently stands, however, the biosafety of nanoparticles remains unclear; research has shown that nanoparticles may be taken up by the phagocytic system leading to their accumulation in various organs (18,19). Unfortunately the transfection efficiency of non-viral delivery methods is lower than that of viral vectors. In order for these non-viral approaches to be used in a clinical setting, modifications are still necessary to increase cellular uptake and minimize toxic side effects (20).

Administration to target organs is another factor that must be taken into account with RNA based therapeutics. For most diseases, therapy must either be administered systemically and have the ability to target specific tissue types, or employ an effective local delivery method (21,22). The accessibility of target tissues, however, greatly impacts the ease of delivery. For the lungs, therapeutic delivery through inhalation of aerosols or intranasal installation, has proven to be a promising non-invasive method often with little systemic exposure (22–24).

When using this delivery method, factors such as size, flow rate, charge and density must be optimized to allow aerosols to avoid clearance from the lungs, efficiently access target cells and avoid systemic absorption (25,26). Direct injection is a method being developed for several organs, with methods including intravitreal injection for treatment of ocular diseases (22,27,28), intracerebroventricular, intrathecal and intraparenchymal injection for diseases of the nervous system (28–31) and intratumoral injection in oncological diseases (22,28,32,33).

In contrast to direct delivery methods, intravenous systemic delivery has been used to target blood cells, certain tissues that are not easily accessible, or diseases that require systemic application, such as metastasized cancer. Studies have applied systemic delivery methods for the treatment of pulmonary, oncological, liver and joint diseases (23,28,34). Although many of these delivery methods are being tested for RNAi therapies, they hold promise for clinical delivery of *trans*-splicing group I introns as well.

In addition to the development of delivery vehicles and methods, chemical modifications to RNA have been used to reduce immune responses, increase cellular stability and reduce the dose required, which in turn can also reduce biological toxicity (35). Phosphorothioate and boranophosphonate modifications have been made to the RNA backbone (36), along with a number of alterations to the ribose sugar (37) and terminal nucleotide (9). Chemical modifications may also prove to be a useful tool when developing the spliceozyme for clinical applications.

5.3 Future Applications for the Analysis of Evolutionary Parameters

High throughput sequence analysis. Much can be learned about RNA evolution from experiments conducted in a lab, because, in contrast to those occurring in nature, evolutionary parameters can be carefully measured and controlled and evolutionary intermediates can be analyzed. These evolutionary studies can provide insight into how RNA can evolve to recruit host proteins, how ribonucleoprotein complexes develop, the process by which RNA develop new structures with novel functions, and the effect on sequence space exploration by differing evolutionary parameters. In studying the effect of evolutionary parameters on an evolving population of *trans*-splicing group I introns, 5-10 clones were sequenced during each round of evolution from each evolutionary line. This data was used to describe the path taken by evolving ribozyme populations in each evolutionary line, but was quite limited in sample size. High throughput sequence analysis could be performed on each round of evolution and would allow for a quantitative measurement of the specific paths and sets of mutations that were explored under specific evolutionary parameters. Future, longer-term evolutions of *trans*-splicing ribozymes, including the spliceozyme could potentially result in the development of novel structural features, the ability to perform multiple turnover reactions, or recruitment of other cellular proteins.

Applications of our results on low selection pressure in future RNA evolution experiments. The finding, that a low selection pressure can be beneficial to the evolution of populations of RNA, has been applied to the evolution of a more efficient spliceozyme and can also be applied to the evolution

of other RNA species. There have been several attempts to utilize a *trans*-splicing group I intron for therapeutic purposes (38–43), yet these have been limited, in part, by low efficiency. Although this study did not generate ribozymes with increased therapeutic potential, due to the recruitment of the *E. coli* transcription termination factor Rho, evolutions performed in human cells under the appropriate conditions would aid in developing group I introns capable of more efficiently catalyzing these *trans*-splicing reactions in specific contexts.

In addition, selections of several RNA species have been performed in cellular environments. Selections of randomized ribozymes have been used to identify ribozyme targets in HIV (44), construct artificial riboswitches in bacteria (45), and identify efficient *cis*-cleaving hammerhead ribozyme variants (46). With very short RNA species, sequence randomization may allow for all possible sequence combinations to be contained within the starting population and employing a high selection pressure allows for the selection of the fittest species. For many RNA species, however, the length of the RNA does not allow for all possible sequence combinations to be contained within an initial pool, and optimization through successive rounds of mutagenesis and selection is necessary to sample a greater sequence space. The application of low selection pressure in future evolutions of RNA species may be beneficial to allow a population to sample a greater sequence space therefore yield more efficient RNA species.

5.4 References

1. Hammond S, Wood M (2011) Genetic therapies for RNA mis-splicing diseases. *Trends Genet* 27(5):196–205.

2. Wang G, Cooper T (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8(10):749–61.

3. Rodriquez-Pascau L, Coll M, Vilageliu L, Grinberg D (2009) Antisense oligonucleotide treatment for a pseudoexon-generating mutation in the NPC1 gene causing Niemann-Pick type C disease. *Hum Mutat* 30(11):E993–E1001.

4. Olson K, Dolan G, Muller U (2014) In Vivo Evolution of a Catalytic RNA Couples Trans-Splicing to Translation. *PLoS One* 9(1):e86473.

5. Lewin A, Thomas J, Tirupati H (1995) Cotranscriptional splicing of a group I intron is facilitated by the Cbp2 protein. *Mol Cell Biol* 15(12):6971–8.

6. Akins R, Lambowitz A (1987) A Protein Required for Splicing Group I Introns in Neurospora Mitochondria Is Mitochondrial Tyrosyl-tRNA Synthetase or a Derivative Thereof. *Cell* 50:331–45.

7. Deyle D, Russell D (2009) Adeno-associated virus vector integration. *Curr Opin Mol Ther* 11(4):442–7.

8. Kotterman M, Schaffer D (2014) Engineering adeno-associated viruses for clinical gene therapy. *Nat Rev Genet* 15:445–51.

9. Gao Y, Liu X, Li X (2011) Research progress on siRNA delivery with nonviral carriers. *Int J Nanomedicine* 6:1017–25.

10. Spagnoe S, Miller A, Keller M (2004) Lipidic carriers of siRNA: differences in the formulation, cellular uptake, and delivery with plasmid DNA. *Biochemistry* 43(42):13348–13356.

11. Pirollo K, Rait A, Zhao Q, Hwang S, Dagata J, Zon G, Hogrefe R, Palchik G, Change E (2007) Materializing the potential of small interfering RNA via a tumor-targeting nanodelivery system. *Cancer Res* 67(7):2938–43.

12. Rothdiener M, Beuttler J, Messerschmidt S, Kontermann R (2010) Antibody targeting of nanoparticles to tumor-specific receptors: immunoliposomes. *Methods Mol Biol* 624:295–308.

13. Stewart M, Plautz G, Del Buono L, Yang Z, Xu L, Gao X, Huang L, Nabel E, Nabel G (1992) Gene transfer in vivo with DNA-liposome complexes: safety and acute toxicity in mice. *Hum Gene Ther* 3(3):267–75.

14. Kim B, Cho M, Kim S, Kim Y, Woo H, Kim D, Sohn H, Li H (2007) Dry sol-gel polycondensation of hydrosilanes to organosilicas catalyzed by colloidal nickel nanoparticles. *J Nanosci Nanotechnol* 7(11):3964–8.

15. Kim D, Kang J, Kim T, Kim E, Im J, Kim J (2007) A polyol-mediated synthesis of titania-based nanoparticles and their electrochemical properties. *J Nanosci Nanotechnol* 7(11):3954–8.

16. Howard K, Rahbek U, Liu X, Damgaard C, Glud S, Andersen M, Hovgaard M, Schmitz A, Nyengaard J, Besenbacher F, Kjems J (2006) RNA interference in vitro and in vivo using a novel chitosan/siRNA nanoparticle system. *Mol Ther* 14:476–84.

17. Dickeroson E, Blackburn W, Smith M, Kapa L, Lyon L, McDonald J (2010) Chemosensitization of cancer cells by siRNA using targeted nanogel delivery. *BMC Cancer* 10:10.

18. Tang J, Xiong L, Wang S, Liu L, Li J, Yuan F, Xi T (2009) Distribution, translocation and accumulation of silver nanoparticles in rats. *J Nanosci Nanotechnol* 9(8):4924–32.

19. Xie G, Sun J, Zhong G, Shi L, Zhang D (2010) Biodistribution and toxicity of intravenously administered silica nanoparticles in mice. *Arch Toxicol* 84:183–190.

20. Guo P, Coban O, Snead N, Trebley J, Hoeprich S, Guo S, Shu Y (2010) Engineering RNA for targeted siRNA delivery and medical application. *Adv Drug Deliv Rev* 62:650–66.

21. Kullberg M, McCarthy R, Anchordoguy T (2013) Systemic tumor-specific gene delivery. *J Control Release* 173(2):730–6.

22. Vicentini F, Borgheti-Cardoso L, Depieri L, de Macedo Mano D, Abelha T, Petrilli R, Bentley M (2013) Delivery Systems and Local Administration Routes for Therapeutic siRNA. *Pharm Res* 30(4):915–31.

23. Kelly C, Yadav A, McKierman P, Greene C, Cryan S (2013) RNAi in Respiratory Diseases. *Advanced Delivery and Therapeutic Applications of RNAi*, pp 391–416.

24. Thomas M, Lu J, Chen J, Kilbarnov A (2007) Non-viral siRNA delivery to the lung. *Adv Drug Deliv Rev* 59(2-3):124–33.

25. Heyder J (2004) Deposition of inhaled particles in the human respiratory tract and consequences for regional targeting in respiratory drug delivery. *Proc Am Thorac Soc* 1(4):315–20.

26. Carvalho T, Peters J, RO W (2011) Influence of particle size on regional lung deposition - what evidence is there? *Int J Pharm* 406(1-2):1–10.

27. Shen J, Samul R, Silva R, Akiyama H, Liu H, Saishin Y, Hackett S, Zinnen S, Kossen K, Fosnaugh K, Vargeese C, Gomez A, Bouhana K, Aitchison R, Pavco P, Campochiaro P (2006) Suppression of ocular neovascularization with siRNA targeting VEGF receptor 1. *Gene Ther* 13(3):225–34.

28. Bumcrot D, Manoharan M, Koteliansky V, Sah D (2006) RNAi therapeutics: a potential new class of pharmaceutical drugs. *Nat Chem Biol* 2:711–719.

29. Dorn G, Patel S, Wotherspoon G, Hemmings-Mieszczak M, Barclay J, Natt F, Martin P, Bevan S, Fox A, Ganju P, Wishart W, Hall J (2004) siRNA relieves chronic neuropathic pain. *Nucleic Acids Res* 32(5):e49.

30. Thaker D, Natt F, Husken D, Maier R, Muller M, van der Putten H, Hoyer D, Cryan J (2004) Neurochemical and behavioral consequences of widespread gene knockdown in the adult mouse brain by using nonviral RNA interference. *Proc Natl Acad Sci USA* 101(49):17270–5.

31. Thakker D, Natt F, Husken D, van der Putten H, Maier R, Hoyer D, Cryan J (2005) siRNA mediated knockdown of the serotonin transporter in the adult mouse brain. *Mol Psychiatry* 10(8):782–9.

32. Minakuchi Y, Takeshita F, Kosaka N, Sasaki H, Yamamoto Y, Kouno M, Honma K, Nagahara S, Hanai K, Sano A, Kato T, Terada M, Ochiya T (2004) Atelocollagen mediated synthetic small interfering RNA delivery for effective gene silencing in vitro and in vivo. *Nucleic Acids Res* 32(13):e109.

33. Kim W, Christensen L, Jo S, Yockman J, Jeong J, Kim Y, Kim S (2006) Cholesteryl oligoarginine delivering vascular endothelial growth factor siRNA effectively inhibits tumor growth in colon adenocarcinoma. *Mol Ther* 14(3):343–50.

34. Shankar P, Manjunath N, Lieberman J (2005) The prospect of silencing disease using RNA interference. *JAMA* 293(11):1367–73.

35. Bramsen J, MB L, Nielsen A, Hansen T, Bus C, Langkjaer N, Babu B, Højland T, Abramov M, Van Aerschot A, Odadzic D, Smicius R, Haas J, Andree C, Barman J, Wenska M, Srivastava P, Zhou C, Honcharenko D, Hess S, Müller E, Bobkov G, Mikhailov S, Fava E, Meyer T, Chattopadhyaya J, Zerial M, Engels J, Herdewijn P, Wengel J, Kjems J (2009) A large-scale chemical modification screen identifies design rules

to generate siRNAs with high activity, high stability and low toxicity. *Nucleic Acids Res* 37(9):2867–81.

36. Gaglione M, Messere A (2010) Recent progress in chemically modified siRNAs. *Mini Rev Med Chem* 10(7):578–95.

37. Chiu Y, Rana T (2003) siRNA function in RNAi: a chemical modification analysis. *RNA* 9(9):1034–48.

38. Ryu K, Seong-Wook L (2004) Comparative analysis of intracellular trans-splicing ribozyme activity against Hepatitis C Virus internal ribosome entry site. *J Microbiol* 42(4):361–4.

39. Lan N, Howrey R, Lee S, Smith C, Sullenger B (1998) Ribozyme-mediated repair of sickle β-globin mRNAs in erythrocyte precursors. *Science (80-)* 280(5369):1593–6.

40. Ayre B, Kohler U, Goodman H, Haseloff J (1999) Design of highly specific cytotoxins by using trans-splicing ribozymes. *Proc Natl Acad Sci USA* 96(7):3507–12.

41. Ban G, Song M, Lee S (2009) Cancer cell targeting with mouse TERT-specific group I intron of Tetrahymena thermophila. *J Microbiol Biotechnol* 19(9):1070–1076.

42. Ryu K, Kim J, Lee S (2003) Ribozyme-mediated selective induction of new gene activity in hepatitis C virus internal ribosome entry site-expressing cells by targeted trans-splicing. *Mol Ther* 7(3):386–95.

43. Shin K, Sullenger B, Lee S (2004) Ribozyme-mediated induction of apoptosis in human cancer cells by targeted repair of mutant p53 RNA. *Mol Ther* 10(2):365–72.

44. Unwalla H, Li J, Li S, Abad D, Rossi J (2008) Use of a U16 snoRNA-containing ribozyme library to identify ribozyme targets in HIV-1. *Mol Ther* 16(6):1113–9.

45. Wieland M, Hartig J (2008) Improved Aptazyme Design and In Vivo Screening Enable Riboswitching in Bacteria. *Angew Chem Int Ed Engl* 47(14):2604–7.

46. Chen X, Denison L, Levy M, Ellington A (2009) Direct selection for ribozyme cleavage activity in cells. *RNA* 15(11):2035–45.