IN VITRO AND IN VIVO CHARACTERIZATION OF A TRANS EXCISION-SPLICING RIBOZYME

Dana Ann Baum

University of Kentucky

Recommended Citation
Baum, Dana Ann, "IN VITRO AND IN VIVO CHARACTERIZATION OF A TRANS EXCISION-SPLICING RIBOZYME" (2005). University of Kentucky Doctoral Dissertations. 284.
https://uknowledge.uky.edu/gradschool_diss/284

This Dissertation is brought to you for free and open access by the Graduate School at UKnowledge. It has been accepted for inclusion in University of Kentucky Doctoral Dissertations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
IN VITRO AND IN VIVO CHARACTERIZATION OF A TRANS EXCISION-SPLICING RIBOZYME

ABSTRACT OF DISSERTATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Dana Ann Baum
Lexington, Kentucky

Director: Dr. Stephen M. Testa, Professor of Chemistry
Lexington, Kentucky

2005
Copyright © Dana Ann Baum 2005
Abstract of Dissertation

In Vitro and In Vivo Characterization of a Trans Excision-Splicing Ribozyme

Group I introns are catalytic RNAs with the ability to splice out of RNA transcripts, often without the aid of proteins. These self-splicing introns have been reengineered to create ribozymes with the ability to catalyze reactions. One such ribozyme, derived from a Pneumocystis carinii group I intron, has been engineered to sequence specifically remove a targeted segment from within an RNA substrate, which is called the trans excision-splicing reaction.

The two catalytic steps of the trans excision-splicing reaction occur at positions on the substrate known as the 5’ and 3’ splice sites. Strict sequence requirements at these sites could potentially limit the target choices for the trans excision-splicing ribozyme, so the sixteen possible base pair combinations at the 5’ splice site and the four possible nucleotides at the 3’ splice site were tested for reactivity. All base pair combinations at the 5’ splice site allow the first reaction step (5’ hydrolysis) to occur and several combinations allow the second step to occur, resulting in trans excision-splicing product formation. Moreover, we found that non-Watson-Crick base pairs are important for 5’ splice site recognition and prevent product degradation via hydrolysis at other sequence positions. The sequence requirement at the 3’ splice site is absolute, as guanosine alone produced complete product.

To date, the experiments with the trans excision-splicing ribozyme have been conducted in vitro. The further development of this ribozyme as a biochemical tool and as a potential therapeutic agent requires in vivo reactivity. Thus, a prokaryotic system was designed and tested to assess the catalytic potential of the trans excision-splicing ribozyme. We show that the ribozyme successfully excised a single, targeted nucleotide from a mutated green fluorescent protein transcript in Escherichia coli. On average, 12% correction was observed as measured by fluorescence and approximately 1.2% correction was confirmed through sequence analysis of isolated transcripts.

We have used these studies to further characterize trans excision-splicing ribozymes in vitro and to pave the way for future development of this ribozyme.
reaction in vivo. These results increase our understanding of this ribozyme and advance this reaction as a biochemical tool with potential therapeutic applications.

Keywords: Group I intron; ribozyme; trans excision-splicing; splice site sequence requirements; in vivo targeted nucleotide excision
IN VITRO AND IN VIVO CHARACTERIZATION OF A TRANS EXCISION-SPLICING RIBOZYME

By
Dana Ann Baum

Stephen M. Testa, Ph.D.
Director of Dissertation

Mark Meier, Ph.D.
Director of Graduate Studies

August 2, 2005
RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor’s degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgements.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.
DISSERATION

Dana Ann Baum

The Graduate School
University of Kentucky
2005
IN VITRO AND IN VIVO CHARACTERIZATION OF A TRANS EXCISION-SPLICING RIBOZYME

DISSE rtATION

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Arts and Sciences at the University of Kentucky

By
Dana Ann Baum
Lexington, KY
Director: Dr. Stephen M. Testa, Professor of Chemistry
Lexington, Kentucky
2005
Copyright © Dana Ann Baum 2005
ACKNOWLEDGEMENTS

I would not have been able to complete the work presented in this dissertation without the support and guidance of many people. First, I want to thank my research advisor, Dr. Stephen Testa. He entrusted me with a research project that was challenging, high risk, and important to future projects in the lab and I am grateful for the opportunity to make that project a success. He gave me the room necessary to figure out the details and to work through the problems I encountered along the way, but he also pushed me when he felt it was necessary. I learned so much along the way and I gained valuable experience in manuscript writing, grant writing, and other tasks involved in running a lab. I will take these experiences with me as I go off to pursue my goal of becoming a professor at a research university. Thank you, Steve, for introducing me to a research area that is challenging and exciting. I appreciate your advice and the lively discussions on numerous topics, especially hockey. I have truly enjoyed my time in your lab and even when times were tough, I loved the work I was doing.

My time in Dr. Testa’s lab was made even more enjoyable because of my lab mates, both past and present. Dr. Michael Bell went out of his way to make me feel welcome and I learned so much from him during our morning coffee breaks. He is one of the most unique individuals I have ever met. Dr. Rashada Alexander has been a wonderful friend and always offered support and a sympathetic ear when I needed to vent. Dr. Ashley Johnson was also a source of support during my graduate school career. I also want to thank “the guys” in the lab. I had the pleasure of working with Joy Sinha on the splice site project presented in this dissertation. I’m still amazed at where that project started and how much more interesting it ended up being. I want to thank Patrick Dotson for his insights on GFP and the discussions about the trials and tribulations of working in vivo. I know the in vivo project is in good hands. Nick Tzouanakis was a lively addition to the lab and always had a way of making things interesting. Dustin Lafferty was a great help to me when I was completing the in vivo project and I appreciate his hard work and friendship.
I would also like to thank the members of my committee: Dr. Leonidas Bachas, Dr. Martha Peterson, Dr. Brian Rymond, and Dr. John Selegue. I appreciate their time spent serving on my committee. I am indebted to many members of the Chemistry Department. I especially want to thank Joyce Cambron, Geri Gerke, Jamie Torp, and June Smith in the Main Office for their help with various tasks. My thanks also goes out to Yuvonne Queen in the Graduate Program Office for her help in getting the appropriate paperwork filled out and submitted. I also appreciate the interactions I have had with various faculty members during my time at UK. I truly enjoyed my time here.

I would not be where I am today without the love and support of my family. First and foremost, I would like to thank my mom, Trish Baum. Mom, I know it was hard for you to let me go when I decided to move to Kentucky to go to grad school, but you stood by me as I pursued my goals. Thank you so much for everything you have done for me. I also want to thank my sister, Dawn, who is not only my sister, but also my best friend. Dawn, I know it wasn’t always easy having me as a sister, but I am so proud of you and of the person you are. You have been a great source of strength for me and have been so supportive and encouraging, especially when I was stressed. I would like to thank my dad, Richard Baum, for his support of my decision to go to graduate school. I also want to thank my grandma, Mary Williams, for the notes and the phone calls to see how I was doing, even when she wasn’t feeling the greatest. I am indebted to my aunt and uncle, Sandy and “Uncle” Dan Trobaugh, for letting me visit them for much needed vacations. My thanks also goes out to my neighbor back home, the Honorable Joseph Koval. While many people know him as Judge, to me he’s just Joe. Thank you so much for everything you have done for me and my family.

Dealing with the stresses of graduate school has been easier thanks to my wonderful friends. I especially want to thank Andrea Belko for her friendship. She was one of the first people I met who understood why I liked science so much because of her love of science and her desire to go to medical school. It has been so nice to have someone to talk to who understands how a research lab works. My thanks also goes out to Judith McCart for her ever optimistic outlook on life. To the
other friends I have made while growing up in Illinois and during my time at Washington University in St. Louis, Missouri, thank you for the experiences you have brought into my life. Even if we have lost touch, you are still on my mind and in my heart.

Last, but certainly not least, I want to thank God for all his gifts. I feel truly blessed by the talents he has given me and by the people he has brought into my life. Whenever I felt overwhelmed, faith and prayer saw me through. I would not be writing this work today if it was not for Him and for that I am so thankful.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ iii

TABLE OF CONTENTS ........................................................................................................ vi

LIST OF TABLES ................................................................................................................ ix

LIST OF FIGURES ............................................................................................................. x

CHAPTER 1 – INTRODUCTION .................................................................................. 1

Investigation of the Sequence Requirements for the TES Ribozyme
at the 5’ and 3’ Splice Sites ......................................................................................... 1

Investigation of the Ability of the TES Ribozyme to Excise a Single, Targeted Nucleotide from an mRNA In Vivo ......................................................... 2

CHAPTER 2 – BACKGROUND ................................................................................. 4

Nucleic Acids ........................................................................................................... 4

*Deoxyribonucleic acid (DNA)* ........................................................................ 4

*Ribonucleic acid (RNA)* .................................................................................... 5

Catalytic RNA ............................................................................................................ 6

*Self-splicing group I intron* ............................................................................... 7

*Trans excision-splicing reaction* ....................................................................... 8

Green Fluorescent Protein .................................................................................... 9

Methodologies ........................................................................................................... 10

*Gel electrophoresis* ............................................................................................. 10

*Autoradiography* ................................................................................................. 11

*Sub-cloning* .......................................................................................................... 11

*PCR and RT-PCR* ............................................................................................... 12

*Site-directed mutagenesis* .................................................................................. 13

*Fluorescence spectroscopy* ................................................................................. 14

CHAPTER 3 - DETERMINATION OF THE SPLICE SITE SEQUENCE REQUIREMENTS FOR THE TRANS EXCISION-SPLICING RIBOZYME .......... 30
CHAPTER 4 - IN VIVO EXCISION OF A SINGLE TARGETED NUCLEOTIDE FROM AN RNA TRANSCRIPT BY A TRANS EXCISION-SPlicing RIBOZYME...66

Proposed Project.................................................................66
Materials and Methods......................................................67
Oligonucleotide synthesis and preparation..........................67
Plasmid construction ..........................................................67
Preparing competent JM109(DE3) .........................................71
Transformation of competent JM109(DE3).............................71
In vivo TES reactions..........................................................72

Molecular recognition at the 5’ splice site........................................34
Source of cryptic products ......................................................37
Molecular recognition at the 3’ splice site....................................39
Discussion ..............................................................................40
Molecular recognition at the 5’ splice site....................................40
Relative effectiveness of each base in each position of the 5’ splice site.................................................................43
Molecular recognition at the 3’ splice site ....................................44
P9.0 and P10 are not required for TES reactions .........................44
A mechanism for ribozyme-mediated TES product degradation ....46
Non-Watson-Crick base pairs at the 5’ splice site can help determine the binding register of reaction substrates ....46
ωG can help determine the binding register of reaction substrates ....47
Implications ..........................................................................48
LIST OF TABLES

Table 3.1: Relative Effectiveness of Each Base in Each Position of the 5’ Splice Site in TES Reactions .............................................................. 50
LIST OF FIGURES

Figure 2.1.  Deoxyribonucleotide ................................................................. 15
Figure 2.2.  The Nitrogenous Bases in DNA ................................................. 16
Figure 2.3.  Structure of a DNA Strand ....................................................... 17
Figure 2.4.  Double Helix Structure of DNA ............................................... 18
Figure 2.5.  Watson-Crick Base Pairing in DNA .......................................... 19
Figure 2.6.  Central Dogma of Molecular Biology ....................................... 20
Figure 2.7.  Ribonucleotide ........................................................................ 21
Figure 2.8.  Uracil ...................................................................................... 22
Figure 2.9.  Examples of RNA Secondary Structure .................................. 23
Figure 2.10. Self-Splicing of Group I and Group II Introns ............................ 24
Figure 2.11. Predicted Secondary Structure of the Pneumocystis carinii
Self-Splicing Group I Intron ...................................................................... 25
Figure 2.12. Self-Splicing Reaction of a Group I Intron ................................. 26
Figure 2.13. Predicted Secondary Structure of the Pneumocystis carinii
Ribozyme ............................................................................................... 28
Figure 2.14.  Trans Excision-Splicing Reaction ........................................... 29
Figure 3.1.  The u-G Wobble Pair ............................................................... 51
Figure 3.2.  Conserved Splice Site Sequences ............................................. 52
Figure 3.3.  5' Splice Site Test System .......................................................... 53
Figure 3.4.  Results for the 5’ Splice Site During 1 Hour Reaction Times ...... 54
Figure 3.5.  Time Studies Investigating the Formation of Cryptic Products .... 56
Figure 3.6.  Results for the 5’ Splice Site During 15 Minute Reaction Times ... 58
Figure 3.7.  Results for the 5’ Splice Site Using 3’ End Labeled Substrates .... 60
Figure 3.8.  Time Studies to Identify the Source of the Cryptic Products ...... 61
Figure 3.9.  Competition Studies Investigating Product Dissociation and
Rebinding ............................................................................................... 62
Figure 3.10. Results for the 3’ Splice Site for 1 Hour Reaction Times .......... 63
Figure 3.11. Results for Reactions Where no P10 Formation Is Possible .... 65
Figure 4.1.  Design of the GFP Target System ........................................... 89
Figure 4.2.  The Trans Excision-Splicing Ribozyme with a T7 Terminator
Sequence ............................................................................................... 90
Figure 4.3.  Plasmid Construction ................................................................. 91
Figure 4.4.  GFP Ribozyme Base Pairing With Its Target ............................ 93
Figure 4.5.  Schematic of the Assay to Selectively Isolate Corrected
Transcripts ............................................................................................. 94
Figure 4.6.  Fluorescence Levels for Normal, Mutant and Expected Product
Forms of GFP .......................................................................................... 95
Figure 4.7.  TES Correction of Mutant GFP ................................................ 96
Figure 4.8.  Sequencing Results for Non-Fluorescent and Fluorescent Colonies . 97
CHAPTER 1 – INTRODUCTION

Ribonucleic acid, or RNA, is a biomolecule with an interesting history. It plays a pivotal role in the central dogma of molecular biology, yet it was initially thought to be a relatively passive molecule. The discovery of RNA molecules with the ability to catalyze chemical reactions in the early 1980’s (1, 2) changed this view of RNA and led to the development of new areas of research. These areas of research have focused on characterizing these catalytic RNAs, or ribozymes, and utilizing their properties to catalyze new reactions (3-30).

One such catalytic RNA, a group I intron-derived ribozyme from the opportunistic pathogen *Pneumocystic carinii*, has been the focus of ongoing work in the Testa lab (11, 22, 30-34). We have previously reported that this ribozyme can catalyze a reaction called the trans excision-splicing (TES) reaction (22). In this reaction, the ribozyme binds an RNA substrate and excises from that substrate an internal segment. This reaction has potential use as a biochemical tool for the sequence specific modification of RNA molecules. This ribozyme also has potential therapeutic applications in that it could be used to remove mutations (at the RNA level) that are implicated in a host of genetic diseases. The trans excision-splicing reaction also serves as a unique model system for studying the structure and function of group I introns. The research presented in this work further characterizes the TES ribozyme *in vitro* by elucidating molecular recognition components of the 5’ and 3’ splice sites (33). This ribozyme was also characterized by demonstrating, for the first time, the *in vivo* catalytic abilities of the TES ribozyme in the bacterium *Escherichia coli* (30).

Investigation of the Sequence Requirements for the TES Ribozyme at the 5’ and 3’ Splice Sites

TES ribozymes recognize their targets initially through base pairing. These base pairing interactions also help define the sites of catalysis, which are called splice sites. In work with the *P. carinii* intron (11, 31) and the ribozyme derived from this intron (22, 32), the highly conserved nucleotide sequences that define the 5’ and 3’ splice sites in group I introns were maintained. Specific sequence requirements at
these critical positions would limit the sequences that could be targeted by TES ribozymes, so in conjunction with another member of the Testa lab, Joy Sinha, we investigated all possible sequences for those that allowed catalytic activity at these splice sites (33). These studies were also undertaken to gain insight into the molecular recognition of the splice sites by this ribozyme.

The results show that the sequence requirement at the 5’ splice site is not stringent, however, certain sequences should be avoided to prevent degradation of the subsequent TES products (33). This lax sequence specificity indicates the molecular recognition of the 5’ splice site is dependent on structure, and not sequence (i.e., base pairing). Conversely, the sequence requirement at the 3’ splice site is absolute, with only guanosine allowing for TES product formation (33). These results allow us to establish guidelines for new TES target systems. The results also indicate that changing the 3’ splice site sequence to nucleotides other than guanosine will require an as yet unknown modification of the ribozyme.

Investigation of the Ability of the TES Ribozyme to Excise a Single, Targeted Nucleotide from an mRNA In Vivo

While TES ribozymes are potentially useful biochemical tools for the in vitro modification of RNA, their applicability could be greatly expanded by demonstrated in vivo reactivity. These ribozymes could be used as tools for the sequence specific modification of RNA transcripts in a cell. As a therapeutic agent, TES ribozymes could be used to target genetic mutations at the RNA level that are known to lead to disease. The benefit of this type of approach to treatment is that the transcript used for protein production is being fixed. This repair reduces the amount of mutant transcript, while increasing the amount of functional transcript. This differs from the approach of many recent RNA-based therapeutics, which simply destroy the mutant transcripts (18, 24, 35-40). To this end, a test system was designed to assess the ability of a TES ribozyme to excise a single, targeted nucleotide from an mRNA transcript in E. coli (30).

This testing showed that the P. carinii ribozyme is able to catalyze the trans excision-splicing reaction in a cellular environment. The ribozyme was successfully
reengineered to target and remove a single-base insertion mutation engineered into a green fluorescent protein gene. On average, 12% correction was observed as measured by fluorescence, and approximately 1.2% correction was confirmed through sequence analysis (30). The greatest amount of TES reactivity (as measured by the largest increase in restored fluorescence) was seen in tests involving a ribozyme with increased base pairing interactions with the substrate. This ribozyme-mediated repair of the transcript occurs with no observed detrimental effects on the cells. Corrected transcripts were selectively isolated and sequenced to confirm the removal of the targeted nucleotide (30). These results represent the first example of a catalytic RNA specifically excising a targeted nucleotide from within an RNA substrate in vivo.

The work presented here helps further characterize the trans excision-splicing ribozyme in vitro (33) and in vivo (30). The results from the splice site study establish a framework for designing new target systems, while the demonstrated activity of the TES ribozyme in a prokaryotic system lays the groundwork for further development of TES ribozymes in vivo, including tests of TES ribozyme reactivity in mammalian systems. These studies demonstrate the usefulness of the TES reaction as a way to study the structure and function of group I intron-derived ribozymes. These results also play a vital role in expanding the potential uses of this novel ribozyme, including those applications with potential therapeutic value.
Nucleic Acids

Nucleic acids are polymers of nucleotides, with each nucleotide consisting of a phosphate group, a five-carbon sugar and a nitrogenous base. There are two forms of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

Deoxyribonucleic acid (DNA)

DNA is the genetic storage material in a cell, and it gets its name from the deoxyribose sugar in its nucleotides (Figure 2.1). There are four different nucleotides that make up DNA. The identity of the nitrogenous base attached to the sugar is what makes each nucleotide distinct. There are 4 nitrogenous bases (or nucleobases) found in DNA (Figure 2.2): adenine (A), guanine (G), cytosine (C), and thymine (T). The nucleotides are commonly referred to by the single letter abbreviations of the attached nucleobases, shown in parentheses. Adenine and guanine are purines, while cytosine and thymine are pyrimidines.

In nucleic acids, the nucleotides are polymerized. The connection is made between the sugars and the phosphate groups of adjacent deoxyribonucleotides. Specifically, the linkage is a phosphodiester bond between the 3' hydroxyl group of one nucleotide and the 5' phosphate group of the next nucleotide (Figure 2.3). This series of connections is called the sugar-phosphate backbone or the phosphodiester backbone. The first and last nucleotides of the polymer (or strand) have available 5' (first nucleotide) or 3' (last nucleotide) functional groups. Thus, a strand has directionality, and the order of the nucleotides is read from either the 5' end to the 3' end or vice versa. This directionality is denoted when representing a DNA strand by the single letter abbreviations of the nucleobases. As an example, the strand $5'$GCAT$^3'$ is different from the strand $3'$GCAT$^5'$ because the nucleotides are being read from opposite directions.

In DNA, two strands come together to form a double helix. In the double helix, the nitrogenous bases of the opposing strands are oriented into the center of the helix and form base pairs through hydrogen bonding. These base pairing
interactions hold the strands together and look like the rungs of a ladder. The bases also stack upon each other, which lends further stability to the structure (41). The sugar-phosphate backbones twist around the central axis in a right-handed direction (Figure 2.4). The base pairing interactions found in DNA are the Watson-Crick base pairs. They are stable interactions and are specific, with adenine pairing with thymine via two hydrogen bonds and guanine pairing with cytosine via three hydrogen bonds (Figure 2.5). The Watson-Crick base pairs are isosteric, meaning exchanging one pair (A-T) with another pair (G-C) does not significantly change the structure of the sugar-phosphate backbones (42). This allows for the constancy of the double helical structure, regardless of the order of the base pairs. The backbones of each strand run in the opposite directions, so one strand runs in the 5'-to-3' direction, while its complement runs in the antiparallel direction of 3'-to-5'. DNA is stable because of this double helix structure. This stability is dependent on the base pairing interactions, as well as the stacking of the base pairs.

Ribonucleic acid (RNA)

In the central dogma of molecular biology (Figure 2.6), RNA acts as the intermediate between the genetic material (DNA) and the functional product (protein). Research shows, however, that RNA is much more than a passive intermediate (1, 2, 43, 44). RNA, like DNA, is a polymer of nucleotides. There are two key differences between RNA and DNA. First, the sugar in RNA is a ribose (with a hydroxyl group on the 2' carbon of the sugar) rather than a deoxyribose (Figure 2.7), thus the name ribonucleic acid. This 2' hydroxyl group gives RNA an additional functional group that is available for structural interactions and chemical reactions. Also, the nitrogenous base thymine, which is found in DNA, is replaced by uracil (U). Uracil differs from thymine by the replacement of a methyl group with a hydrogen atom (Figure 2.8).

Unlike DNA, RNA is synthesized in a single stranded form. It is capable, however, of forming secondary structure through intramolecular or intermolecular base pairing interactions. Thus, RNA can form double stranded regions, similar to the DNA double helix structure, as well as other structures that leave nucleobases
available for interactions (Figure 2.9). These secondary structures, as well as tertiary structures involving long range interactions, allow RNA transcripts to form 3-dimensional structures, some with potential catalytic structures.

**Catalytic RNA**

In the early 1980’s, the view of the role of RNA in the cell changed with the discovery of catalytic RNAs (1, 2). The first catalytic RNA discovered was a self-splicing group I intron from the ciliate *Tetrahymena thermophila* (1). This finding was soon followed by the discovery that the RNA component of RNase P (which is responsible for the processing of tRNA molecules) is the catalytic subunit of the complex (2). Introns are intervening sequences found in RNA transcripts that make functional sequences, called exons, discontinuous. Introns are removed from RNA transcripts via splicing. A group I intron is capable of catalyzing its own excision from an RNA transcription (1). Group I introns are widespread in nature and they have been found in mRNA, tRNA, and rRNA transcripts from many organisms, including algae, fungi, and other unicellular eukaryotic organisms (45). They share a similar secondary structure and a common splicing mechanism. Another type of catalytic intron, the self-splicing group II intron, was later discovered (46). Group I and group II introns differ in their secondary structures and in their splicing mechanisms (Figure 2.10). Thus, catalytic RNA molecules can utilize different mechanisms to achieve the same result: excision of the intron and ligation of the flanking exons.

Since the initial discoveries of these large catalytic RNAs, other naturally occurring catalytic RNAs have been found, including the small cleaving ribozymes. These ribozymes include the hairpin, hammerhead, Hepatitis Delta Virus (HDV), and Neurospora Varkud Satellite (VS) ribozymes (18, 37, 43, 47, 48). These small ribozymes are associated with various satellite and virus RNAs (47, 49), and they cleave the phosphodiester backbone of an RNA molecule to produce 2’,3’ cyclic phosphates. Group I introns, group II introns, and RNase P cleave the phosphodiester backbone in a different manner to produce 3’ OH groups. The small cleaving ribozymes are of great interest for the sequence specific cleavage of RNA
targets because of their small size and ease of adaptability, as they recognize their targets primarily through simple base pairing.

**Self-splicing group I intron**

The trans excision-splicing reaction studied in this work is based on the self-splicing reaction of group I introns. In this reaction, the intron folds into its catalytic form through extensive base pairing and tertiary interactions. In this folding pathway, the intron also base pairs with its 5’ and 3’ exons, which will ultimately be ligated together (Figure 2.11), via three important base pairing interactions. The intron sequences responsible for these key targeting interactions are termed recognition elements (abbreviated as RE) (22). Recognition element 1 (RE1) base pairs with the 3’ end of the 5’ exon (shown in blue; Figures 2.11 and 2.12) to form a base paired helix called P1. The 5’ end of the 3’ exon (shown in red; Figures 2.11 and 2.12) is recognized by RE3 and forms the P10 helix. Recognition elements 1 and 3 are side by side (and in some cases overlapping) in group I introns and together compose the internal guide sequence or IGS (6, 31, 50). A third important contact is made between two segments of the intron itself. RE2 (which is near the catalytic core of the intron) base pairs with sequences near the 3’ end of the intron (near the 3’ splice site) to form the P9.0 helix (Figures 2.11 and 2.12). These interactions allow for recognition of the correct splice sites and align the exons for the splicing event.

The first catalytic step of self-splicing is 5’ cleavage (Figure 2.12). An exogenous guanosine nucleotide binds a particular site in the intron known as the guanosine binding site [also known as the G-binding site or the GBS, shown in the gray oval in Figure 2.11] (51, 52). The 3’ hydroxyl of this guanosine monophosphate serves as a nucleophile and attacks the phosphodiester backbone of the 5’ exon at the 5’ splice site. The backbone is cleaved, resulting in a 3’ hydroxyl on the terminal uridine nucleotide at the splice site (the last base of the 5’ exon). The guanosine nucleophile is attached to the end of the intron [Figure 2.12] (1, 53, 54). The second reaction step is exon ligation, in which the 3’ hydroxyl of the terminal uridine of the 5’ exon attacks the phosphodiester backbone at the 3’ splice site [Figure 2.12] (55, 56). This nucleophilic attack ligates the exons together and releases the intron. The
excised intron still maintains catalytic activity and goes on to circularize via an intramolecular attack (1, 57, 58).

The group I intron is a single-turnover catalyst in terms of the self-splicing reaction. The intron can only self-splice out of the transcript once. Removing the exon sequences (which are the substrates in the self-splicing reaction) from the intron produces a ribozyme, or an RNA enzyme. These ribozymes can now function in a multiple-turnover fashion and catalyze reactions using exogenous substrates (Figure 2.13). These ribozymes have been used to characterize the steps and the sequence requirements of the self-splicing reaction (16, 59-76). They are also being reengineered to perform new catalytic reactions (3, 4, 7, 22, 34, 56, 77-79).

Trans excision-splicing reaction

The trans excision-splicing (TES) reaction was developed by the Testa lab using a ribozyme (~330 nucleotides in length) derived from a group I intron in an rRNA gene from the opportunistic pathogen *Pneumocystis carinii* (11, 80, 81). This ribozyme can catalyze the excision of a targeted sequence from within an RNA substrate (22). First, the ribozyme folds into its catalytic form (Figure 2.13). The three molecular recognition elements used by the self-splicing intron (described above) are utilized by the TES ribozyme to identify its target. RE1 base pairs with the 5' end of the substrate, forming the P1 helix (Figure 2.14). RE2 base pairs with the segment targeted for excision (when longer than a single nucleotide) to form the P9.0 helix, while RE3 base pairs with the 3' end of the substrate to form the P10 helix (Figure 2.14). Thus, the ribozyme recognizes its target through simple base pairing. Not all of these base pairing interactions are required for reactivity, as test systems that do not allow P9.0 formation still produce appreciable amounts of TES product (22, 33).

After the ribozyme has bound an RNA substrate containing the sequence targeted for excision, the first reaction step occurs (Figure 2.14). In this 5' cleavage reaction, a hydroxyl (presumably from a water molecule bound to the ribozyme) attacks the phosphodiester backbone of the substrate specifically at the 5' splice site, generating 5' and 3' exon intermediates (22). This differs from the first step of
the self-splicing reaction, which utilizes guanosine as the nucleophile. The result, however, is the same, as the phosphodiester backbone of the substrate is cleaved to produce a 3' hydroxyl group on the terminal nucleotide of the 5' exon (a uridine). In the second reaction step (exon ligation), the 3' hydroxyl on the terminal nucleotide of the 5' exon attacks a specific base within the 3' exon intermediate, simultaneously ligating the exon intermediates and excising the targeted internal segment. This internal segment can vary in size, as sequences from 1 to 28 nucleotides have been successfully excised (22). The ribozyme recognizes its target primarily through base pairing, so changing the sequence of the recognition elements of the ribozyme can direct the ribozyme to different substrates (22), while increasing the length of these interactions can improve reactivity (32).

**Green Fluorescent Protein**

The studies investigating the ability of the *P. carinii* ribozyme to catalyze the TES reaction *in vivo* utilize a green fluorescent protein gene as the target (30). Green fluorescent protein (GFP) originates from the jellyfish *Aequorea victoria* (82). Excitation of the protein with blue or UV light results in the emission of fluorescence in the green region of the visible spectrum. The protein has a highly stable structure, known as a β-can structure (83, 84), which is resistant to a wide variety of denaturing conditions (85). Fluorescence is dependent on the formation of the chromophore by cyclization and oxidation of three particular amino acids (Serine\(^{65}\)-Tyrosine\(^{66}\)-Glycine\(^{67}\)) in the core of the protein (82-86). The chromophore formation for GFP is autocatalytic (87) and does not require species-specific cofactors, as GFP has been successfully expressed in a wide range of hosts and cell lines (85, 88-91). The highly stable structure of the protein and the autocatalytic formation of the chromophore (seeming to only require molecular oxygen), make GFP an attractive tool for labeling molecules for various studies *in vivo*.

The wild type form of GFP has two excitation peaks: a major peak at 395 nm and a minor peak at 475 nm. These excitations correspond with emissions at 508 nm and 503 nm, respectively (92). Analyses of randomly mutated forms of the gene have led to the engineering of GFP proteins with altered spectral properties to
improve the usability of these proteins in biological applications (82, 93). The form of GFP utilized in this work was developed by Quantum (distributed by QBioGene; Carlsbad, CA) and is a red-shifted version of GFP (rsGFP). In this form of GFP, Serine\textsuperscript{65} has been mutated to Cysteine\textsuperscript{65}. This mutation changes the excitation spectrum from a double peak (described above) to a single peak at \(~473\) nm (94). This change in the excitation maximum is desirable, because this and other red-shifted variants exhibit greater photostability than wild type GFP (88, 94). The emission peak for the form of GFP used in these studies is at \(~509\) nm, essentially the same as an emission peak for the wild type form of GFP (94).

Methodologies

\textit{Gel electrophoresis}

Gel electrophoresis is a method for separating biomolecules, generally by mass (95). While gel electrophoresis is also utilized for separating proteins, the descriptions outlined here will focus on the separation of nucleic acids (DNA or RNA). There are two types of gels used for these separations: agarose and polyacrylamide. Agarose, which is derived from seaweed, is best suited for separating large segments of nucleic acids, from as small as 50 base pairs to several thousand base pairs in length (95). The percentage of agarose used in the gel matrix determines the size of the molecules that can be effectively separated. Larger, unstructured molecules will be impeded in their migration through the matrix, while smaller or structured molecules will migrate faster. Agarose gels are used in the work presented here to estimate the size of large DNA or RNA molecules (relative to controls) and to separate and purify large pieces of DNA for applications such as plasmid construction.

Polyacrylamide gels are used for higher resolution of molecules and can allow for the separation of molecules differing by a single nucleotide (95). Polyacrylamide gels are composed of polymerized acrylamide crosslinked with bis-acrylamide to form a polyacrylamide lattice. The percentage of acrylamide and the ratio of acrylamide to bis-acrylamide determine the size range of molecules that can be effectively separated. Polyacrylamide gels can be native (non-denaturing) or
denaturing. Native polyacrylamide gels maintain the structure of the molecules being separated, allowing for different forms of the same molecule to be separated. Denaturing gels contain a denaturant, such as urea, to disrupt any self-structure (for example via base pairing) and allow for the separation of molecules based on size (or mass) alone. Native polyacrylamide gels are used in this work to purify radiolabeled substrates for use in in vitro studies. Denaturing polyacrylamide gels are used to separate reaction substrates, products, and intermediates after conducting in vitro reactions. The presence of the denaturant (urea) promotes the dissociation of the product from the ribozyme.

**Autoradiography**

Autoradiography is a method used to visualize biomolecules based on the decay of a radioactive label (96). In this work, radioactive phosphorus ($^{32}\text{P}$) is used to label DNA or RNA oligonucleotides on either the 5' end or the 3' end of the molecule. For labeling the 5' end of an oligonucleotide, the reaction involves the transfer of the radioactive phosphorus from ATP to the desired molecule by a kinase. To label the 3’ end of a molecule, a cytidine monophosphate nucleotide that has been labeled with a radioactive phosphate (*pCp) is ligated to the end of the molecule by an RNA ligase (97). These labeled molecules are purified by polyacrylamide gel electrophoresis (PAGE) and are then used as substrates in in vitro reactions. After a reaction is complete, the resulting products are separated on polyacrylamide gels. The gels are then dried under vacuum to trap the molecules in place, and the dried gels are exposed to a PhosphorImager screen, which detects the energy from the decay of the radioactive phosphorus. After an overnight exposure, the screen is scanned using a PhosphorImager. The resultant gel image allows us to determine the size of products of the reaction, based on the migration of the bands relative to size controls. The intensities of the bands are quantified, and these intensities allow us to determine the extent of a reaction.

**Sub-cloning**

In sub-cloning, a DNA fragment of interest is ligated into a cloning vector, also known as a plasmid. Plasmids are self-replicating, double-stranded, circular DNA
molecules (96). For molecular cloning purposes, a plasmid will typically contain an origin of replication specific for the organism the plasmid will be introduced into; a selectable marker, such as an antibiotic resistance gene, to allow for selection of cells containing the plasmid; and a multiple cloning site to allow for the insertion of the sequence of interest. This multiple cloning site consists of sequences that can be cleaved by restriction enzymes. Restriction enzymes are bacterial endonucleases, or enzymes that cut DNA molecules internally at specific base pair sequences (96). Digesting DNA molecules with restriction enzymes produces DNA fragments with specific, known ends. This allows for the joining of DNA molecules in a sequence-specific manner. The cloning vector with the DNA sequence of interest is then introduced into cells. For bacteria, this process is known as transformation, while the same process is termed transfection for mammalian cells. There are various methods for transforming bacterial cells (96). The work presented here utilizes chemically competent cells, meaning the cells have been treated with chemicals to stimulate them to accept plasmids. Once introduced into bacteria cells, the plasmid is replicated numerous times. Thus, cloning can be used to produce a large amount of the desired plasmid. The plasmid can also serve as a template for transcription of a gene of interest. This allows for production of a desired protein in a cell, without modifying the genome of the cell.

**PCR and RT-PCR**

PCR and RT-PCR stand for Polymerase Chain Reaction (98) and Reverse Transcription-Polymerase Chain Reaction (95), respectively. These reactions allow for the exponential amplification of DNA and RNA molecules. A typical PCR reaction mix contains a template (a DNA molecule that contains a sequence of interest), two oligonucleotide primers (short DNA molecules that are complementary to the sequences flanking the region of interest), the four deoxyribonucleotides (called dNTPs, where N stands for A, C, T, or G) and a thermostable DNA polymerase (an enzyme that catalyzes the polymerization of deoxyribonucleotides). A thermostable DNA polymerase maintains its activity at higher reaction temperatures, and these DNA polymerases have made PCR automatable. The most common thermostable
DNA polymerase was originally isolated from the bacterium *Thermus aquaticus* and is commonly known as *Taq* polymerase.

PCR is a 3-step, cyclic process. The first step is denaturation. The reaction mix is heated to 95 °C for about 1 min to separate the two strands of the DNA template. The second step is annealing, or renaturation. In this step, the reaction mixture is cooled to a temperature that allows the oligonucleotide primers to anneal to the DNA template (typically around 55 °C). The third step is elongation (or synthesis). In this stage, the reaction mixture is heated to around 75 °C, which is the optimum reaction temperature for *Taq* DNA polymerase. The polymerase then catalyzes the polymerization of the DNA strand, using the oligonucleotide primers as a starting point and progressing in the 3' direction. The polymerase uses the template DNA to direct what nucleotide is added to the growing chain. PCR allows for exponential amplification of a template \(2^n\), where \(n\) equals the number of cycles, with just 30 rounds of PCR resulting in a billion copies of the desired sequence.

RT-PCR is a similar process, with a few key differences. The initial template for RT-PCR is RNA. Before PCR can amplify the sequence of interest, the RNA template must be converted into a DNA template. This is done using an enzyme known as reverse transcriptase. Reverse transcriptases catalyze the polymerization of deoxyribonucleotides to form a DNA strand, but they utilize RNA as their template (98). After the initial reverse transcription reaction, the newly formed DNA template is used as the template for a standard PCR reaction. In this work, RT-PCR was used to amplify the GFP transcripts from the total RNA isolated from *in vivo* test reactions.

**Site-directed mutagenesis**

Site-directed mutagenesis is a method for altering the sequence of a DNA molecule in a sequence-specific manner (96). There are various methods available, but the procedure used in this work is similar to PCR (outlined above). Mutagenic oligonucleotides containing the desired changes are used as primers, and the DNA template is most commonly a plasmid. The reaction mixture is subjected to the PCR steps of denaturation, annealing, and polymerization. To isolate selectively the mutated plasmid, the template plasmid (with no mutations) is selectively degraded.
with a restriction enzyme. The reaction mixture is then used to transform *E. coli*. The cells will circularize the mutant PCR products, producing the desired mutant plasmids.

*Fluorescence spectroscopy*

The measurement of the concentration of green fluorescent protein relies on fluorescence. Fluorescence refers to a pathway by which an excited atom or molecule relaxes to its ground state. This relaxation results in the emission of radiant energy (99, 100). For GFP, a conjugated pi-system results from the cyclization of the side chains of Ser\textsuperscript{65}-Tyr-Gly\textsuperscript{67} (82, 93). In terms of biological systems, fluorescence results from exciting small molecules with these types of extended conjugated pi-systems. Their chemical resonance frequencies are in the visible spectrum (101). Fluorescence spectroscopy involves exciting a sample at a particular wavelength (the excitation wavelength) and then detecting the emission at a different wavelength. For the *in vivo* studies outlined in this work, measurements of GFP fluorescence were made utilizing a CytoFluor. This instrument uses filters to select the wavelength band of the excitation beam and the wavelength band of the fluorescence emission. Thus, cells containing GFP in these studies were excited using an excitation filter of 485 ± 20 nm (as the maximum excitation wavelength for the GFP used in these studies is 475 nm), and fluorescence was detected using an emission filter of 508 ± 20 nm (as the maximum emission wavelength of this GFP is 509 nm).
Deoxyribonucleotides are the monomers that make up a DNA strand. Each deoxyribonucleotide consists of a phosphate group, a deoxyribose sugar and a nitrogenous base. The identity of the nitrogenous base distinguishes each nucleotide.
There are four nitrogenous bases found in DNA. Adenine and guanine are purine bases, while thymine and cytosine are pyrimidine bases. Each base is typically referred to by its single letter abbreviation: A=Adenine, G=Guanine, T=Thymine, and C=Cytosine. These single letter abbreviations are also used to designate the nucleotides of DNA.
Figure 2.3. Structure of a DNA Strand

A DNA strand results from the polymerization of deoxyribonucleotides. The linkage is a phosphodiester bond between the 3’ hydroxyl group of one deoxyribonucleotide and the 5’ phosphate group of an adjacent deoxyribonucleotide. These linkages are also known as the phosphodiester backbone or the sugar-phosphate backbone. The formation of linkages in this manner gives the strand directionality. As shown, the strand runs in the 5’-to-3’ direction.
Figure 2.4. Double Helix Structure of DNA

Two strands of DNA come together and interact via hydrogen bonding between the bases, known as base pairing. The phosphodiester backbones of each strand twist around a central helix with the bases of each strand interacting in the middle. This gives the double helix the appearance of a ladder, with the base pairs acting as the “rungs”. The two strands run in the opposite directions and are called antiparallel.
The normal base pairing interactions (as shown) are the Watson-Crick base pairs and result from adenine base pairing with thymine via two hydrogen bonds (top pair) and guanine base pairing with cytosine via three hydrogen bonds (bottom pair). The Watson-Crick base pairs are isosteric, meaning they are the same shape. This allows for the continuity of the double helix, regardless of the order of the nucleotides. Note the antiparallel orientation of the backbones.
DNA, the genetic storage material in a cell, is replicated to pass genetic information to progeny. DNA also serves as the template for transcription to produce RNA. RNA serves as the template for translation to produce protein (a functional gene product).
Ribonucleotides are the monomers that make up an RNA strand. Each ribonucleotide consists of a phosphate group, a ribose sugar, and a nitrogenous base. The identity of the nitrogenous base distinguishes each nucleotide. A key difference between the nucleotides found in DNA and the nucleotides found in RNA is the presence of a hydroxyl group on the 2’ position of the ribose sugar (boxed in red).
The base thymine in DNA is replaced by uracil in RNA. The difference between thymine and uracil is the absence of a methyl group in uracil. Uracil is typically referred to by its single letter abbreviation of U.
RNA is synthesized single stranded (top left); however, it has the ability to base pair and to form structures. Examples include double stranded RNA, a hairpin loop, and a single base bulge. Double stranded RNA results from two RNA strands interacting via base pairing, similar to the DNA double helix structure. The formation of a hairpin results from a single strand folding on itself to form a base paired stem connected by a loop of unpaired nucleotides. A single base bulge results from two strands forming a double stranded interaction, but a base in one strand lacks its base pairing partner in the opposing strand. This causes the unpaired base to bulge out.
Figure 2.10. Self-Splicing of Group I and Group II Introns

General schematics of the self-splicing reactions of group I (left) and group II (right) introns. For both diagrams, the 5’ exon is blue, the 3’ exon is red and the catalytic introns are gray. For group I introns, the first cleavage step is mediated by an exogenous guanosine cofactor (shown in green) that binds a particular site in the intron. The 3’ hydroxyl of the guanosine cofactor performs a nucleophilic attack at the 5’ splice site, resulting in cleavage of the phosphodiester backbone and attachment of the G cofactor to the free end of the intron. For group II introns, the first cleavage step is mediated by the 2’ hydroxyl of an adenosine in the intron and results in the formation of a lariat structure. For both introns, the 3’ hydroxyl on the end of the 5’ exon then performs a second nucleophilic attack at the 3’ splice site. This attack ligates the exons together and releasing the introns. The free group I intron can undergo a cyclization reaction via a nucleophilic attack by the terminal guanosine.
Figure 2.11. Predicted Secondary Structure of the *Pneumocystis carinii* Self-Splicing Group I Intron

Predicted secondary structure of the self-splicing group I intron from an rRNA from the opportunistic pathogen *Pneumocystis carinii*. The above structure was adapted from (11). The 5′ exon sequence is shown in lowercase, blue lettering; the 3′ exon sequence is shown in lowercase, red lettering; and the intron sequences are in uppercase, black lettering. The recognition elements of the intron (RE1, RE2, and RE3) and their base-pairing partners (which form the P1, P9.0, and P10 helices, respectively) are enclosed in gray. The guanosine binding site, in helix P7, is indicated by the gray oval. The cleavage sites for the first and second reaction steps of self-splicing are designated by numbered arrows.
Figure 2.12. Self-Splicing Reaction of a Group I Intron

Detailed schematic of the group I intron self-splicing reaction. The 5’ exon is shown in blue, the 3’ exon is red, and the catalytic intron is gray. The intron initially folds on itself and makes contact with the 5’ and 3’ exons. Three important targeting interactions result from base pairing between sequences in the intron (termed recognition elements, RE) and the 5’ exon (RE1; to form the P1 helix), the 3’ exon (RE3; to form the P10 helix) and another sequence within the intron (RE2; to form P9.0). An exogenous guanosine nucleotide (shown in green) binds a particular site in the intron (known as the guanosine binding site) and the 3’ hydroxyl group of this nucleotide performs a nucleophilic attack at the 5’ splice site in the first reaction step. The phosphodiester backbone is cleaved at the 5’ splice site, resulting in a 3’ hydroxyl on the terminal nucleotide of the 5’ exon (typically a uridine; shown as a blue circle). The guanosine cofactor becomes attached to the end of the intron.
In the second reaction step, the 3’ hydroxyl on the terminal uridine of the 5’ exon performs a nucleophilic attack at the 3’ splice site, immediately after a guanosine that defines the end of the intron (referred to as $\omega G$; shown as a gray circle). This attack results in ligation of the two exons and release of the intron.
Predicted secondary structure of the *Pneumocystis carinii* ribozyme base pairing with exogenous substrate mimics of the native 5' and 3' exon sequences. The above structure was adapted from (22). The recognition elements (RE1, RE2 and RE3) are highlighted in gray and are shown base pairing to the 5' exon (shown in blue), a segment mimicking a portion of the native intron (shown in lowercase, black lettering), and the 3' exon (shown in red) to form the P1, P9.0 and P10 helices, respectively. The G-binding site in helix P7 is enclosed in a gray oval. The arrow designates the nucleophilic attack that results in exon ligation (the second reaction step of self-splicing).
Figure 2.14. Trans Excision-Splicing Reaction

Detailed schematic of the trans excision-splicing reaction (22). The 5’ exon is in blue, the 3’ exon is in red, and the TES ribozyme is in gray. The sequence targeted for excision is a broken line. The recognition elements from the self-splicing intron are used for target recognition. RE1 base pairs with the 5’ exon (to form the P1 helix), RE3 base pairs with the 3’ exon (to form the P10 helix) and RE2 base pairs with the insert (when longer than a single nucleotide) targeted for excision (to form P9.0). A hydroxyl group (most likely from a water molecule bound to the ribozyme) performs a nucleophilic attack at the 5’ splice site in the first reaction step. The phosphodiester backbone is cleaved, resulting in a 3’ hydroxyl on the terminal nucleotide of the 5’ exon (typically a uridine; shown as a blue circle). In the second reaction step, the 3’ hydroxyl of the terminal nucleotide of the 5’ exon performs a nucleophilic attack on the 3’ splice site, immediately after a guanosine that defines the end of the insert region (termed ωG; shown as a gray circle). This results in exon ligation and release of the excised segment and the ribozyme.
CHAPTER 3 - DETERMINATION OF THE SPLICE SITE SEQUENCE REQUIREMENTS FOR THE TRANS EXCISION-SPlicing RIBOZYME

Proposed Project

Previous work with the trans excision-splicing ribozyme has demonstrated that the recognition elements of the ribozyme can be changed to target different substrates (22, 32). In those studies, as well as studies looking at the molecular recognition of the *Pneumocystis carinii* intron in the related suicide inhibition (11, 31) and reverse cyclization (31) reactions, two highly conserved elements from self-splicing group I introns were maintained. These elements are the formation of a u-G wobble pair (Figure 3.1) at the -1 position of the P1 helix to define the 5’ splice site (11, 22, 59, 68, 73, 102, 103) and a guanosine as the last nucleotide of the sequence to be excised (Figure 3.2). This guanosine corresponds to the last nucleotide of the intron (called ωG) in the self-splicing reaction and has been shown to be a critical component of the exon ligation reaction (60, 104-109). Note that throughout this chapter, lowercase nucleotide abbreviations refer to the substrate, uppercase lettering refers to the ribozyme, and the 5’ splice site refers to the base pair that forms between position 12 of the ribozyme and the -1 position of the substrate (11). If these sequence requirements at the 5’ and 3’ splice sites are absolute, trans excision-splicing ribozymes will be limited in what sequences they can target for excision.

Previous work with other group I introns and their derived ribozymes demonstrated that the sequence requirements at these splice sites may not be stringent. It had been shown that the c-A wobble pair can be a substitute for the u-G wobble pair at the 5’ splice site in the 5’ cleavage reaction (59), and c-G can act as a substitute in the self-splicing reaction (102, 110). In these cases, product yields were reduced and the reaction rates were slower than when u-G is present. It has also been reported that ωG can be changed to ωA, with either no change required to the catalytic core (72) or by modifying the guanosine binding site (GBS) of the ribozyme in a particular way to accommodate adenosine (60, 111). Most of the above
mentioned results were obtained from work on the well-studied self-splicing group I intron and ribozyme from *Tetrahymena thermophila* (1). Previous work with the *P. carinii* ribozyme indicated some differences in substrate recognition compared to the *Tetrahymena* ribozyme (11, 112-115). Therefore, I worked with another member of the Testa lab, Joy Sinha, to analyze the 16 possible base pair combinations at the 5' splice site and the four nucleotides found in RNA at the ω position at the 3' splice site in the context of the TES reaction. We utilized our simplest TES system in which a single nucleotide is excised from an RNA substrate. The excision of a single nucleotide does not utilize RE2 for targeting, so the number of recognition elements is limited to two. Also, the single nucleotide that is targeted for excision is analogous to the ω position in group I introns. Utilizing this simple system limits the number of reaction variables and simplifies the analyses. These studies were undertaken to provide a more thorough understanding of the sequence requirements for the TES reaction, specifically at the 5' splice site and for the ω position of the 3' splice site. This information will be useful in developing guidelines for what sequences the ribozymes can target, what sequences the ribozymes can excise, and how specific these reactions might be (relative to each other). These studies will also provide insight into the molecular recognition of the splice sites by this *P. carinii* ribozyme.

**Materials and Methods**

*Nucleic acid synthesis and preparation*

DNA oligonucleotides for site-directed mutagenesis were purchased from Integrated DNA Technologies (Coralville, IA) and used without further purification. RNA oligonucleotides were purchased from Dharmacon Research Inc. (Boulder, CO) and deprotected following the manufacturer's recommended protocol. The manufacturing process leaves a protecting group on the 2' hydroxyl groups of RNA oligonucleotides, which must be removed prior to using the molecules.

*Radiolabeling RNA oligonucleotides*

Designated RNA oligonucleotides were 5’ end-radiolabeled with T4 polynucleotide kinase (New England Biolabs; Beverly, MA) and \([γ^{32}P]\)-ATP
(Amersham Pharmacia Biotech; Piscataway, NJ) and purified on a 20% native polyacrylamide gel, as previously described (32). Designated RNA oligonucleotides were also 3’ end-radiolabeled using a modified version of a two-step procedure previously described (97). Cytidine monophosphate (Cp) was radiolabeled at the 5’ terminus by incubating 10 μM Cp, 0.85 μM [γ-32P] ATP (Amersham Pharmacia), 70 mM Tris-HCl (pH 7.5), 50 mM KCl, 10 mM MgCl2, 5 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.1 mM ATP, 10% glycerol and 10 units T4 polynucleotide kinase (New England Biolabs) in a reaction volume of 10 μL for 1.5 h at 37 °C. The kinase was heat-inactivated at 65 °C for 15 min. Next, the prepared *pCp was ligated to the 3’ end of the designated RNA substrates in a reaction containing 2 μM RNA substrate, 0.28 μM *pCp, 50 mM Tris-HCl (pH 7.4), 50 mM KCl, 10 mM MgCl2, 10 mM DTT, 1 mM ATP, 0.1 mM EDTA, 10% dimethyl sulfoxide (DMSO), and 20 units of T4 RNA ligase (New England Biolabs). Ligation reactions proceeded at 4 °C for at least 16 hours. The labeled substrates were then gel purified by the same method as the 5’ end-labeled substrates listed above.

Ribozyme preparation

The *P. carinii* ribozyme plasmid precursor, P-8/4x, was previously constructed as described (11). Modifications to alter the guanosine in the 12 position of the ribozyme that is involved in the wobble pair at the 5’ splice site and to delete the RE3 sequences that are involved in P10 formation were made using site-directed mutagenesis. The following pairs of primers were used for altering the ribozyme at the 12 position (underlined base represents the altered bases as compared to P-8/4x):

5’CGACTCACTATAGAGCGTCATGAAAGCGGC3’ and
5’GCCGCTTTTCATGACCATCTCTATAGTGAGTCG3’ to create P-8/4x-5’C;
5’CGACTCACTATAGAGACATGAAAGCGGC3’ and
5’GCCGCTTTTCATGACATTCTCTATAGTGAGTCG3’ to create P-8/4x-5’A
5’CGACTCACTATAGAGTGTCAAGAAAGCGGC3’ and
5’GCCGCTTTTCATGACACTCTCTATAGTGAGTCG3’ to create P-8/4x-5’U. The following primer pair was used to create P-8/4x-noP10:
5’CGACTCACTATAGGTCAAGAAAGCGGC3’ and
5’GCCGCTTTCATGACCTATAGTGAGTCG3’. The site-directed mutagenesis reactions were performed as previously described (22), with some modifications. The reaction mixtures were first subjected to denaturation at 95 °C for 30 s, followed by 16 temperature cycles of 95 °C for 30 s, either 50 °C or 60 °C for 2 min, and 68 °C for 6 min. The parental plasmids were then digested with 20 units of DpnI (Invitrogen; Carlsbad, CA) in 4.2 μL of manufacturer’s buffer for at least 2 h at 37 °C. A 3-μL aliquot of this mixture was then used to transform Escherichia coli DH5α chemically competent cells (Invitrogen). The resultant plasmids were purified using a QIAprep Spin Miniprep kit (QIAGEN; Valencia, CA), and sequenced to confirm the changes (Davis Sequencing; Davis, CA).

Prior to transcription, plasmids were linearized by restriction enzyme digest with XbaI and purified from the reaction mixture using a QIAquick PCR Purification kit (QIAGEN). Transcription and purification of the ribozymes proceeded as previously described (22).

Trans excision-splicing reactions

Trans excision-splicing reactions were conducted under conditions previously optimized for the reaction containing a u-G wobble pair at the 5’ splice site (22). This involved running the reactions at 44 °C in H10Mg buffer, which consists of 50 mM Hepes (25 mM Na+), 135 mM KCl, and 10 mM MgCl2 (from 0 mM to 15 mM when using the rP-8/4x-noP10 ribozyme) at pH 7.5. Prior to each reaction, 166 nM ribozyme in 5.0 μL of H10Mg buffer was preannealed at 60 °C for 5 min and then slow cooled to 44 °C. The reactions were initiated by adding 1.0 μL of an 8 nM solution of 5’ end radiolabeled or 3’ end radiolabeled substrate in H10Mg buffer. The Kd of the substrate is expected to be approximately 5.2 nM, similar to the Kd of the 6-mer 5’ exon mimic (11). Reaction times for the TES reactions investigating the 5’ splice site were 15 min and 1 h. The 3’ splice site studies and the rP-8/4x-noP10 reactions were allowed to proceed for 1 h. Time dependence assays to determine the source of cryptic products were run from 1 min to 120 min. All reactions were terminated by adding an equal volume of stop buffer (10 mM urea, 0.1X TBE, and 3 mM EDTA). The products and reactants were denatured at 90 °C for 1 min prior to
loading on a 12.5% polyacrylamide/8M urea gel for separation. Following electrophoresis, the gel was transferred to chromatography paper (Whatman 3MM CHR) and dried under vacuum for 1 h at 70 °C. The dried gels were exposed to phosphorescence screens overnight. Reaction products were then visualized and quantified using a Molecular Dynamics Storm 860 PhosphorImager. All the data reported are the average of at least two independent assays.

Competition assays

166 nM ribozyme was preannealed in H10Mg buffer for 5 min at 60 °C. The ribozyme was then slow cooled to 44 °C, at which point 1.3 nM of the radiolabeled substrate in H10Mg buffer was added to initiate the reaction. After 5 min, 1000-fold excess of unlabeled TES product competitor (over substrate) in H10Mg buffer was added to the reaction mixture. Periodically, an aliquot was removed and added to an equal volume of stop buffer over a period of 90 min, starting 10 min after addition of the competitor (15 min after reaction initiation). The substrates and products were denatured by heating at 90 °C for 1 min prior to gel loading, and the products were separated, visualized, and quantified as described above.

Results

Molecular recognition at the 5’ splice site

Four different 10-mer substrates, 5’augacygcuc3’ (where y stands for u, g, c, or a), were utilized with four different ribozymes, each containing a different nucleotide at position 12 of the ribozyme (shown in white lettering in Figure 3.3). This allowed us to test all 16 possible base pair combinations at the 5’ splice site. We chose to use our simplest TES target system, in which a single nucleotide (analogous to ωG) is excised from within the substrates. Using this system means that one of the recognition elements, RE2, is not utilized in the reaction as it cannot base pair with the substrate to form P9.0 due to a lack of complementary bases (Figure 3.3). The reaction conditions used were previously optimized for the TES reaction using the 10-mer substrate and ribozyme that would reconstitute the conserved u-G wobble pair found at the 5’ splice site (22). This initially included running the reactions for 1
Under these conditions, all 16 base pair combinations gave first-step product (5’ cleavage via hydrolysis), although the extent of reaction varies significantly as a function of sequence (Figure 3.4). It should be noted that the yield of the first step is the combined total of the first and second steps of the reaction as the product of the second reaction step must necessarily have undergone the first reaction step. This total is represented by the sum of the black bars and the white bars of the graph in Figure 3.4.

Unexpectedly, multiple sequence combinations gave an appreciable amount (>10%) of complete TES product (black bars on the graph; Figure 3.4). There were also cases (u-A, a-U, c-G, g-C, and c-C) that produced appreciable amounts of cryptic products which were shorter than the 9-mer TES product. The total percentage of all cryptic products formed for a base pair are shown as gray bars on the graph in Figure 3.4. We decided to investigate the 16 base pair combinations further by running time dependent assays to determine when the cryptic products appear. These assays allowed us to find a reaction time that minimized cryptic product formation while still producing appreciable amounts of TES product.

The results of the time dependence assays fell into three classes. In the first class, which consisted solely of those combinations that put Watson-Crick base pairs at the 5’ splice site, cryptic products appear after 15 min (represented by c-G; Figure 3.5A). In the second class, consisting only of c-C, cryptic products appear much earlier (after 1 min) and to a greater extent than the other combinations (Figure 3.5B). In the third class, consisting of all other non-Watson-Crick combinations, appreciable cryptic products do not appear, even in cases where substantial TES products form (represented by u-G; Figure 3.5C). These results indicate there are two possible routes to cryptic product formation. For the Watson-Crick pairs, cryptic products appear to be formed after TES product formation. For c-C, cryptic product formation appears to occur shortly after substrate binding. As c-C never produced appreciable amounts of TES product, it was not investigated in further detail. Based on these studies, we looked at the 5’ splice site sequences in the TES reaction using 15 min as the reaction endpoint.
When comparing the results from the 15 min and 1 h reactions, the trends with regard to the yield of TES products formed were very similar (compare Figures 3.4 and 3.6). There was, however, a noticeable difference in the production of cryptic products, as they were effectively removed for the Watson-Crick pairs at the shorter reaction time. Again, multiple sequence combinations gave an appreciable amount (>10%) of complete TES product (Figure 3.6). These include the u-G and c-A wobble pairs, all four Watson-Crick base pairs, and the a-G combination. The conserved u-G wobble pair was the most effective, producing 68% TES product, with the c-A wobble pair being the next highest (43%). These results were not surprising as it was previously known that these wobble combinations were acceptable at this position in other introns (59). What was surprising was that all four Watson-Crick base pairs produce a relatively high amount of product (>25%). These results were in contrast to results for *Tetrahymena* ribozymes in terms of 5’ cleavage (59, 102). Studies looking at exon ligation in *Tetrahymena* indicated other combinations were allowable (102), but not all nucleotides worked in the -1 position of the 5’ exon, as was the case in our studies.

While the TES product from reactions with the 10-mer substrate with a u-G at the 5’ splice site had previously been enzymatically sequenced to ensure it was the expected TES product (22), it was important to ensure that all the observed TES products resulted from the excision of the targeted internal guanosine. If the substrate bound the ribozyme in a misaligned register, a 9-mer product could form if the 3’ c of the substrate was cleaved off in the first reaction step. To confirm that this misalignment was not occurring in any of the sixteen 5’ splice site cases, TES reactions were run using substrates that contain a radiolabeled nucleotide added to the 3’ end of the 10-mer substrates (to give 11-mer substrates). The radiolabel is now on the opposite end of the substrate and an alternate pathway to the 9-mer “product” involving the cleavage of the 3’ c would result in the production of a radiolabeled 2-mer (the 3’ c with the additional radiolabeled nucleotide on the 3’ end). As seen in Figure 3.7, no 2-mer products are formed. However, we get 10-mer bands, representing the TES product (the 9-mer product with the additional radiolabeled nucleotide on the 3’ end). When combining our results from the 5’
labeled substrates and the 3’ labeled substrates, we see that the reaction end-products all contain both the 5’ and 3’ ends of their substrates and their products are a single nucleotide shorter. This indicates the ribozymes are excising an internal segment. As the intermediates do not vary in size and are the expected 6-mers (for the 5’ labeled substrates), a different base is not being utilized as the ω base.

The results showed that the sequence of both the exon and the ribozyme (at the 5’ splice site) is important for each reaction step. Guanosine is favored at the 12 position of the ribozyme, with all four possible combinations giving over 60% first step product (5’ cleavage) and only the g-G combination producing less than 10% TES product during the 1 h reaction. These results are perhaps not unexpected as a guanosine in this position is the native form of the intron. Cytidine in either the ribozyme or the substrate is unfavorable for TES product formation, unless it is involved in a Watson-Crick base pair or a c-A wobble pair. We also found that a high level of first step (hydrolysis) product does not necessarily lead to a high level of TES product. This is seen when adenosine is present at the –1 position in the substrate. Adenosine in the substrate leads to at least 30% first step reactivity with all four ribozymes in the 15 min reactions, but only a-U and a-G lead to significant amounts of TES product (Figure 3.6). By looking at the thermodynamic stability of the different base pair combinations (116), those base pairs expected to be more stable, such as the Watson-Crick base pairs and the wobble pairs, produce significant amounts of TES product. These results indicate a stable base pair is required for the second reaction step (exon ligation) to proceed efficiently.

Source of cryptic products

We found it interesting that of all the combinations that produced substantial TES product, only the Watson-Crick pairs led to cryptic products at extended reaction times. We wanted to determine the source of these cryptic products. To this end, we ran time dependent TES assays utilizing a 6-mer, which mimics the 5’ exon product of the 5’ cleavage reaction, and a 9-mer, which mimics the expected TES product, as the reaction substrates. Since the cryptic products do not appear until after the reactions have been running for 15 min, we know the cryptic degradation is
not occurring with the 10-mer substrates. If the degradation is occurring prior to exon ligation (the second reaction step), we expect the 6-mer to be degraded. If the 9-mer is degraded, then the TES products are being degraded after the second reaction step. For these assays, we utilized two representative ribozyme-substrate combinations, the u-A pair (Figure 3.8) and the c-G pair (data not shown). Only the 9-mer was degraded in these assays, indicating that the cryptic products are produced from cleavage of the TES products themselves, and not the 5’ exon intermediate of the first step of the TES reaction. This explains why cryptic products only occur at relatively long reaction times, as there must be production of TES product before cryptic hydrolysis occurs. The products are not only cleaved at the cryptic sites, but also at the original 5’ splice site (producing the 6-mer intermediate; Figure 3.8), showing that some amount of TES products for all base pair combinations are likely recleaved at the correct 5’ splice site.

Once we determined the source of the cryptic products, we decided to investigate the mechanism of cryptic product formation. There are two possible routes to cryptic product formation. The first route involves the newly formed TES product staying bound to the ribozyme, but changing binding registers. If there was a shift in the binding register, a new site could be positioned for 5’ cleavage within the catalytic core of the ribozyme. There are two ways this shift could occur. The 5’ exon could slip relative to the RE1 sequence of the ribozyme, leading to mispairing between the exon and RE1 and putting a different exon nucleotide at the -1 position for 5’ cleavage (102). The second way this shift could occur is that the helix formed by the product base pairing to RE1 and RE3 of the ribozyme, which is a continuous helix after the removal of the targeted nucleotide from the substrate and is called P1 extended [or P1ex (32)], translocates through the catalytic core of the ribozyme. This translocation would put a new base pair in position for 5’ cleavage (59, 63, 103). The second route to cryptic product formation involves dissociation of the product from the ribozyme. The product is then able to rebind the ribozyme, and, by either mispairing or by translocation (as described above), presents different positions on the product for the 5’ cleavage reaction.
To distinguish between these possible routes, we conducted competition assays. In these assays, a large excess of unlabeled TES product was added to TES reactions before cryptic products started to form. If the product dissociates and rebinds prior to cryptic product formation, the excess unlabeled product will outcompete the labeled product for rebinding free ribozyme. This competition would lead to a decrease in the amount of cryptic degradation seen in the labeled material. If the product does not dissociate prior to cryptic product formation, the excess competitor will have no effect and the amount of cryptic products formed will remain the same relative to product formation. Once again, we studied two representative pairs, the u-A pair (Figure 3.9) and c-G pair (data not shown) at the 5’ splice site. We utilized combinations that produce cryptic products instead of those combinations that did not produce cryptic products (such as the u-G or c-A wobble pairs) to allow us to differentiate between the 5’ exon intermediate produced from the TES substrates and the cryptic products produced from the TES products. As seen in Figure 3.9 (representing the u-A pair), by adding excess competitor, we almost completely eliminate the cryptic products. This result indicates the TES products are dissociating and then rebinding free ribozyme. Once the products bind the ribozyme, they are subjected to 5’ cleavage to produce cryptic products (as seen for the u-A pair in Figure 3.8, data not shown for c-G).

Molecular recognition at the 3’ splice site

To look at the sequence requirements at the 3’ splice site, we used another simple substrate-ribozyme system that was similar to the system we used to look at the 5’ splice site. Once again, a single nucleotide was targeted for excision. This nucleotide, which corresponds to the ω position within self-splicing introns, was altered in four different 10-mer substrates (5’augacuxcuc3’, where x=G, C, A, or U) to test each nucleotide in the ω position (Figure 3.10). These substrates were tested with the native ribozyme (containing a u-G wobble pair at the 5’ splice site) for TES reactivity. Typically, the ω position defines the 3’ splice site by binding to the guanosine binding site (GBS) of the catalytic core of the ribozyme (104, 109). For TES systems that excise a region longer than a single nucleotide, the ω position is
the last base of the excised region (22). The first step of the TES reaction (5’ cleavage) occurs in all four cases (white bars on the graph; Figure 3.10), with the greatest amount of 5’ cleavage occurring when guanosine is the ω base. For second step reactivity, and thus complete TES product formation, guanosine is absolutely required at the ω position (black bar on the graph; Figure 3.10). We tried to rationally redesign the GBS of the ribozyme based on work that was done with Tetrahymena ribozymes (60, 111), but these changes did not alter the specificity at the ω position (data not shown). These results indicate the mode of molecular recognition of the ω position base by the GBS may not be the same between the two ribozymes. Thus in its current incarnation, the TES ribozyme requires guanosine in the ω position.

The ωG is not the only determinant of the 3’ splice site. The P9.0 helix, which forms between elements of the intron itself in the self-splicing reaction, and the P10 helix, which forms between the intron and the 3’ exon, flank the ω position (Figure 3.2) and are also important elements for 3’ splice site recognition (60, 76, 104-109, 117-119). For the TES reactions where a single nucleotide is excised, P9.0 helix formation is not possible (Figure 3.3) or required (22). The binding of ωG to the GBS is predicted to be a weak interaction (52, 76, 118), so we expected that P10 formation would be critical for TES product formation as it would be involved in holding the 3’ intermediate of the first reaction step prior to exon ligation. To test this, we created a ribozyme that lacked the RE3 sequences, which prevents P10 formation (rP-8/4x-noP10). Surprisingly, TES product was formed in the absence of P9.0 and P10 formation, albeit in low yield (black bars on the graph in Figure 3.11). Therefore, ωG is sufficient as a 3’ molecular recognition element for TES reactions, although P10 and P9.0 are beneficial.

Discussion

Molecular recognition at the 5’ splice site

In nature, a u-G wobble pair is almost universally conserved at the 5’ splice site of self-splicing group I introns. The only known exceptions are c-G pairs that exist for translational coding purposes, as changing these c-G pairs to u-G pairs lowered the magnesium requirements of the self-splicing reaction and increased the
reaction rate (110). The TES ribozyme is derived from a group I intron, so one of the expected sequence limitations of the TES reaction is the requirement for either a u-G pair (11, 22, 59, 68, 73, 102, 103) or possibly a c-G base pair (110) at this position. In these cases, a guanosine is present at position 12 of the ribozyme (located in RE1), and uridine or cytidine is at the –1 position of the substrate (Figure 3.2). For group I intron-derived ribozymes, it has also been reported that the c-A wobble pair allows the 5’ cleavage reaction, although it is not nearly as effective as u-G pairs (59). Other base pairs and mismatch combinations do not work at all (59) or to a very small extent (102, 120). The c-A substitution is most likely allowable because c-A can form a wobble base pair with a spatial orientation and accessibility of free functional groups analogous to that of u-G pairings (42, 59, 70, 121). A wobble pair may also help to define the 5’ splice site by distorting the backbone structure in the P1 helix. X-ray crystallography studies on tRNA (122, 123) and an NMR structure of a model of a P1 helix (124) showed that a u-G wobble pair perturbs an RNA helix by positioning U closer to the helix axis and G away from it. It is believed the distorted backbone of the wobble pairs is accessible for nucleophilic attack in the 5’ cleavage reaction, which helps to determine the correct 5’ splice site.

Also, for the *Tetrahymena* ribozyme, the free exocyclic amine group of guanosine (at ribozyme position 12) in the minor groove is believed to be involved in tertiary interactions that help the ribozyme recognize the correct splice site (68, 70, 73). It must be noted, however, that there are significant differences in molecular recognition between different introns. For example, the measured contribution of the u-G wobble pair to tertiary interactions is greater for the *P. carinii* ribozyme than for *Tetrahymena* (113) and *Candida albicans* (114). Moreover, the 2’-hydroxyl groups of the ribose sugar of the nucleotides have been shown to take part in binding of *Tetrahymena* substrates (125-129), yet these interactions do not appear to be critical for *P. carinii*-derived ribozymes (11). Due to these differences, identifying and understanding the sequence requirements for *P. carinii* ribozymes is useful for developing effective and specific TES ribozymes. To this end, we have tested the effectiveness of each of the 16 possible base combinations at the 5’ splice site in the TES reaction.
We report that all 16 base pair combinations undergo the first reaction step (5' cleavage), while the u-G and c-A wobble pairs, all four Watson-Crick pairs, and the a-G pair go on to produce appreciable amounts of TES product (>10%). These results, particularly in terms of the first reaction step (5' cleavage), suggest that the interactions involved in 5' splice site determination are not stringent. This substrate promiscuity at the 5’ splice site was unexpected, and is in contrast to results reported for analogous reactions using a *Tetrahymena* ribozyme (59, 102). Apparently, for *P. carinii*-derived ribozymes, identification of the 5' splice site is not entirely dependent on having specific functional groups in the ribozyme at position 12 or the substrate at -1. That 7 out of 16 combinations give an appreciable amount of TES product, including all four Watson-Crick base pairs, suggests unexpectedly lax sequence and structural requirements at the 5’ splice site. Some general trends to note include a preference for a purine (G or A) in the 12 position of the ribozyme, a strong benefit for the u-G pair, and the appearance of substantial amounts of cryptic products at extended times for the Watson-Crick pairs. It is interesting that although u-G and c-A give substantial amounts of TES product, switching the location of the pairs (to produce g-U and a-C) did not produce significant amounts of TES product (Figures 3.4 and 3.6). Apparently, it is not just the presence of a wobble pair that is beneficial for the TES reaction, but the presence of a wobble pair with a purine in the ribozyme and a pyrimidine in the substrate.

Note that effective 5’ cleavage, however, does not necessarily translate into high levels of TES product. For example, base combinations a-A, a-C, and g-G produce reasonable amounts of 5’ cleavage product, yet they each produce only about 2% TES product (Figure 3.4). These base pairing combinations are expected to be thermodynamically weak, especially in relation to the wobble pairs and the Watson-Crick pairs (116). Therefore, these results indicate that after 5’ cleavage, a thermodynamically stable base pair at the 5’ splice site, although not required, is beneficial for the exon ligation reaction, perhaps to orient the newly created 3’ hydroxyl group on the end of the 5’ exon more favorably for the second reaction step. Previous work with the *Tetrahymena* ribozyme supports this idea (102).
Relative effectiveness of each base in each position of the 5’ splice site

Because multiple base pair combinations at the 5’ splice site give a substantial amount of TES product, and because all combinations give some amount of first step product, it is worth considering the relative specificity of targeting different substrate sequences with the different ribozymes. Table 3.1 (top) shows the relative specificity for each ribozyme construct and (bottom) the relative specificity for each base at the –1 position in the substrate. These results show that there are distinctive combinations of substrate and ribozyme that work much better in the TES reaction relative to similar combinations. In terms of the ribozyme, a cytidine is very specific for targeting a guanosine in the substrate, as virtually no TES product is formed with this ribozyme when targeting cytidine, adenosine, or uridine at the corresponding substrate position. Note that we would still get a substantial amount of hydrolysis products, however, ribozyme reconstruction methods have been developed that could be used to help overcome this problem (32). Uridine in the ribozyme has the next highest specificity, then guanosine, and lastly adenosine. Adenosine is interesting in that it will target substrates with uridine and cytidine approximately equally. The lower specificity of guanosine and adenosine in the ribozyme might not be surprising as they are effective in their Watson-Crick base pairing combinations (c-G and u-A), as well as in their wobble pair configurations (as u-G and c-A). In terms of the substrate position, to target a guanosine in the substrate, cytidine is the overwhelmingly best choice for the corresponding ribozyme position. When targeting an adenosine, uridine is the best choice and when targeting a uridine, guanosine is the best choice. When targeting a cytidine, adenosine is the best choice (even better than a guanosine), indicating the importance of forming a wobble pair at the 5’ splice site. If one has the flexibility to choose any substrate base, using a cytidine in the ribozyme to target a guanosine in the substrate would appear to give the most specific TES product (although not necessarily the highest yield). Note that this analysis does not take into account the amount of product subjected to degradation via cryptic hydrolysis. While a Watson-Crick base pair may be beneficial in terms of specificity (such as for g-C), such a base pair combination
would not be a good choice due to the high level of cryptic degradation of products (~40% in the 1 h reactions, Figure 3.4).

Molecular recognition at the 3’ splice site

We looked at the molecular recognition of the 3’ splice site by altering the base at the ω position, which is the base excised from the 10-mer substrate (Figure 3.10). Any of the four bases at the ω position allowed 5’ cleavage (the first reaction step); however, the second reaction step (exon ligation) was completely inhibited when ωG was replaced with the other bases (Figure 3.10). Thus, the ω position must be a guanosine for complete TES product formation, probably due to a requirement for specific binding of the ω position base with the GBS. Since any base can be utilized for the first reaction step, this indicates that the GBS only comes into play in the second reaction step. These results show that, in their current incarnation, *P. carinii*-derived TES ribozymes require a guanosine as the last (or only) base to be excised in the substrate. Alteration of this specificity most likely will require changes to the GBS of the ribozyme. The GBS of the *Tetrahymena* ribozyme has been altered to change the specificity from guanosine to adenosine (60, 111), however, these same mutations did not produce a change in specificity for the *P. carinii* ribozyme (data not shown). One possible explanation is that there is a difference in the recognition of the ωG by the *Tetrahymena* and *P. carinii* ribozymes.

*P9.0 and P10 are not required for TES reactions*

Of the three known elements that can aid in the molecular recognition of the 3’ splice site in group I intron self-splicing reactions (ωG, P10, and P9.0), only ωG and P10 are present in TES reactions that remove a single nucleotide (Figure 3.2). The P9.0 helix cannot form in these constructs, and its absence does not appear to have negative consequences on the TES reaction (22). Since ωG binding to the GBS is a relatively weak interaction (52), and because it is thought that this interaction only occurs after the first reaction step (see previous section), it was expected that the formation of P10 would play a crucial role in defining and binding the 3’ end of the substrate to the ribozyme (for example, as the first reaction step
intermediate). Surprisingly, however, we found that P10 and P9.0, although beneficial, are not required for either step of the TES reaction.

How is it that the 3’ exon intermediate does not readily dissociate from the ribozyme between the two reaction steps in this system? One possible explanation is that the 3’ exon is sterically constrained within the catalytic core of the ribozyme after the first reaction step. In other words, the 3’ exon is physically trapped within the ribozyme in a space for which no significant molecular interactions occur between the ribozyme and the 3’ reaction intermediate. Then, a conformational rearrangement takes place between the two reaction steps that places ωG in the GBS, in preparation for the second reaction step. An alternative explanation is that the interactions that hold the 3’ exon intermediate do not require base pairing. For example, it could be that the 3’ intermediate nucleotides themselves, perhaps in combination with the hydroxyl groups from the sugar-phosphate backbone, form tertiary interactions with the ribozyme. The recent report of a crystal structure of a group I intron with both exons trapped between the first and second splicing step argues against this alternative explanation (109). The reported structure indicates that the 3’ splice site is selected primarily by interactions with ωG as there were no tertiary contacts with the 3’ exon and only a single tertiary hydrogen bond between the intron and the P10 helix (109). A closer examination of the reported structure reveals that another portion of the intron could, in fact, help physically trap the 3’ exon in the catalytic core.

This ability of ωG alone allowing complete TES product formation was also surprising because previous studies with this same ribozyme indicated it was unable to catalyze the complete TES reaction when excising a 20-nucleotide segment in the absence of P10 (32). It is likely that dissociation of the 3’ intermediate may have been responsible for a lack of second-step product in this previous case, as including a longer P9.0 interaction partially restored the function of a TES ribozyme (32). Apparently, this problem of intermediate dissociation is not as overwhelming when excising a single nucleotide. This also indicates that there are functional differences between TES reactions where a single base is excised relative to excising larger sequences.
A mechanism for ribozyme-mediated TES product degradation

We found that at long reaction times (relative to product formation), TES products dissociate and rebind the ribozyme, at which point degradation occurs via the 5’ cleavage reaction (i.e., ribozyme-mediated hydrolysis) at one or more new 5’ splice sites. We show that product cleavage does not occur through translocation of a bound P1 helix after product formation, as the product must first dissociate. Note, however, that c-C is a special case in that the substrate appears to be degraded at cryptic sites. Nevertheless, these results show that TES ribozymes can bind the products of TES reactions and essentially destroy them. This degradation mechanism is probably greatly aided by the conditions under which the reactions were run (i.e. ribozyme excess and long reaction times). Shorter reaction times greatly decreased cryptic degradation of products (Figure 3.6), so it is expected that conducting TES reactions with lower ribozyme concentrations will also help prevent this cryptic hydrolysis of TES products.

Non-Watson-Crick base pairs at the 5’ splice site can help determine the binding register of reaction substrates

The activation of cryptic sites during the first step of self-splicing (5’ cleavage) has been investigated with *Tetrahymena* ribozymes (63, 68, 73, 102, 103). It was shown that cryptic cleavage sites can become activated when the substrate helix shifts from the original binding register to another register without dissociation (63). Such a mechanism is probably occurring in our reactions where a c-C base pair forms at the 5’ splice site. For our cases where Watson-Crick base pairs are present at the 5’ splice site, cryptic sites are activated only after correct TES products form, the products dissociate from the ribozyme, and then bind a new ribozyme. Since 5’ exon translocation was not detected in these assays as a source of cryptic cleavage sites, it seems apparent that these TES products are binding to the recognition elements (RE1 and RE3) of the ribozyme, with the subsequent P1 helix docking in the wrong helical register. In contrast, TES products that do not contain Watson-Crick pairs at the 5’ splice site are not degraded at cryptic sites. Apparently, non-Watson-Crick base pairs at the 5’ splice site are acting in concert with other factors.
to precisely determine the docking register of the P1 helix in the catalytic core of the ribozyme. The mechanism of this role likely includes the fact that non-Watson-Crick base pairs are causing a structural perturbation at the 5’ splice site, which would prevent the formation of a structurally uniform P1-P1ex helix (32). In the absence of this perturbation (which would be the case with Watson-Crick pairs at the splice site), the TES products bind RE1 and RE3 of the ribozyme and the subsequent P1-P1ex helix docks within the catalytic core of the ribozyme in multiple registers, activating various positions within the 5’ exon for subsequent 5’ cleavage reactions. Such a structural perturbation would not be needed for the original TES substrates that have Watson-Crick base pairs, as ωG in these substrates should also be able to perform this role (see next section). Note that the role of non-Watson-Crick base pairs in defining the correct 5’ splice site must be combined with a stable base pairing conformation (as mentioned above) in order to allow the second reaction step to proceed effectively.

Except for the case with c-C, cryptic products are created from TES products, indicating that the 1 h yield of TES products formed for those ribozyme-substrate complexes with Watson-Crick pairs at the 5’ splice site is higher than originally calculated. The yield of TES product created is actually the yield of TES product plus the yield of cryptic products. This increases the percent TES product formed to over 50% for all Watson-Crick pairs during the 1 h reactions. Since we also see degradation of the TES products to 6-mers (which is the same size as the expected TES intermediate), some of the hydrolysis products shown in all the TES reactions (as the 6-mer intermediates) could have stemmed from the degradation of TES products. Thus, the quantified TES products presented in this work at all time points represent a minimum.

ωG can help determine the binding register of reaction substrates

Cryptic site degradation does not occur with the substrate-ribozyme combinations that have a Watson-Crick base pair at the 5’ splice site, yet cryptic site degradation occurs readily for these dissociated TES products (which rebind ribozyme). The only difference between the substrates and the products is the ωG
found in the substrates. Thus, it appears that ωG can play a role in determining the binding register of TES reaction substrates in the catalytic core of the ribozyme. It is unlikely that the ω position imparts a substantial thermodynamic advantage on the first reaction step, so ωG is probably acting through some other mechanism. One possibility is that ωG, which is a single nucleotide bulge immediately 3' to the actual site of 5' cleavage, is situated ideally for disrupting the uniform sugar-phosphate backbone at the correct 5' splice site in the P1 helix. This structural perturbation, akin to that seen above with non-Watson-Crick base pairs, although subtle, could be recognized and exploited by the ribozyme for directing proper binding of the P1 helix in the catalytic core of the ribozyme. In fact, it is likely that the ω position does not even have to be a guanosine to fulfill this role in the 5' cleavage reaction, but it does have to be a guanosine for the second reaction step to proceed. Such a role for ωG in helping define the 5' splice site has not been previously reported in any context with any ribozyme construct. ωG probably does not perform this role in self-splicing reactions, as ωG is not only part of the intron (instead of the substrate), but it is also not immediately adjacent to and covalently attached to the 5' splice site. Whether the ribozyme has evolved to exploit this molecular recognition component, or it is fortuitous, is unclear. Nevertheless, in the context of the TES reaction, this unexpected function for ωG in aiding the fidelity of the reaction is important.

**Implications**

These results advance our knowledge of the molecular recognition involved in both steps of the TES reaction (and the analogous steps of self-splicing) and allow us to improve our design principles for developing target systems for these trans excision-splicing ribozymes. We are now able to target a wider range of TES substrates, and with more sequence specificity, particularly with regard to the 5' splice site. This information will help us to further develop TES ribozymes as potential biochemical tools. Examples include RNA repair ribozymes (22, 30, 34), which would include strategies that are complementary to trans-splicing ribozymes (7, 21, 25), and transcript cleaving ribozymes, which would exploit the first reaction step, to produce transcripts ending in 3' OH groups [which is in contrast to
hammerhead and hairpin ribozymes, which leave 2’, 3’-cyclic phosphates (43, 48)]. Lastly, exploiting the TES reaction, in contrast to self-splicing or suicide inhibition, has allowed us to observe new and unexpected molecular recognition principles exploited by group I intron-mediated catalytic reactions.
Table 3.1: Relative Effectiveness of Each Base in Each Position of the 5’ Splice Site in TES Reactions

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>U</th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>U</th>
<th>G</th>
<th>C</th>
<th>A</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>When the ribozyme</td>
<td>42.29</td>
<td>18.29</td>
<td>6.71</td>
<td>0.87</td>
<td>26.74</td>
<td>30.14</td>
<td>0.48</td>
<td>0.71</td>
<td>1.51</td>
<td>1.27</td>
<td>0.82</td>
<td>44.16</td>
</tr>
<tr>
<td>contains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>u</td>
<td>c</td>
<td>a</td>
<td>g</td>
<td>u</td>
<td>c</td>
<td>a</td>
<td>g</td>
<td>u</td>
<td>c</td>
<td>a</td>
<td>g</td>
</tr>
<tr>
<td>When the substrate</td>
<td>3.58</td>
<td>32.48</td>
<td>1.13</td>
<td>1.72</td>
<td>21.72</td>
<td>0.55</td>
<td>0.63</td>
<td>11.47</td>
<td>0.60</td>
<td>0.70</td>
<td>33.86</td>
<td>26.50</td>
</tr>
<tr>
<td>contains</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>U</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>U</td>
<td>C</td>
<td>A</td>
<td>G</td>
<td>U</td>
<td>C</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>

Top: Effectiveness of each ribozyme construct in TES reactions as a function of base identity at position 12 using each of the four 10-mer substrates that differ at position –1. Bottom: Effectiveness of each of the four 10-mer substrates that differ at position –1 in TES reactions as a function of each of the four ribozyme constructs. These numbers were obtained by taking the extent of TES product at 15 min of a particular case (for example; u-G) and dividing that by all products formed for the other three combinations (for example; TES, 5’ cleavage, and cryptic products for u-A, u-C, and u-g) plus the nonproductive products of the particular case (for example; 5’ cleavage and cryptic products for u-G). These theoretical numbers, therefore, represent percent product formation of one particular base pair combination in relation to all other products that would be produced in a 1:1:1:1 mixture of each ribozyme (Top of table) or substrate (Bottom of table).
Structure showing the hydrogen bonding interactions of the u-G wobble pair commonly found in RNA structures. This base pairing interaction is highly conserved at the 5’ splice site of group I introns. The formation of this wobble pair causes a structural perturbation within a helical structure. The hydrogen bonding interactions also free two functional groups: an exocyclic amine group on the guanosine and a carboxyl group on the uracil. These functional groups could be recognized by proteins or the catalytic core of the ribozyme.
Figure 3.2. Conserved Splice Site Sequences

Conserved sequence elements at the 5' and 3' splice sites of group I introns and their derived ribozymes. The diagram shows the native recognition elements of the *Pneumocystis carinii* ribozyme base pairing with an RNA substrate with three nucleotides targeted for removal. The 5' splice site is defined by a highly conserved u-G wobble pair (shown in blue). The 3' splice site is defined by two structure elements (P9.0 and P10) as well as an absolutely conserved guanosine as the last base of the sequence targeted for excision (shown in yellow). This guanosine is equivalent to the ωG of the self-splicing group I intron.
Figure 3.3. 5' Splice Site Test System

The rP-8/4x ribozyme is in uppercase lettering, the 10-mer substrate is in lowercase lettering, and the single guanosine nucleotide to be excised is circled and in italics. The base to be excised corresponds to the \( \omega \) position of self-splicing introns and is referred to as the substrate \( \omega \) position. The ribozyme recognition elements RE1, RE2, and RE3 base pair with the substrate to form the P1, P9.0 and P10 helices, respectively. Note that the P9.0 helix does not form in this system because of a lack of complementary bases. The sites of catalysis for the first step (5' cleavage) and the second step (exon ligation) are shown with large bold arrows. The P10 helix is boxed. The –1 position of the substrate (designated as \( y \), where \( y = u, g, c, \text{ or } a \)) and the 12 position of the ribozyme (designated as \( X \), where \( X = G, C, A, \text{ or } U \)) are shown in white lettering and define the native 5’ splice site. Every combination of the four nucleotides at \( X \) and \( y \) was analyzed. Note that the diagram shows only the recognition elements of the ribozyme.
Figure 3.4. Results for the 5’ Splice Site During 1 Hour Reaction Times

Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in 10 mM MgCl₂. (Top) A representative polyacrylamide gel using all 16 base pair combinations at the 5’ splice site (33). Each complete substrate sequence and the base at ribozyme position 12 (in uppercase lettering) is shown above its corresponding lane. The migration position on the gel of the 10-mer substrates, the 9-mer products, and the 6-mer intermediates are labeled. All other bands represent cryptic sites. Note that there is some sequence-dependent migration variability between these lanes. The lanes marked “buffer” had substrate augacugcuc incubated as a typical reaction in the absence of ribozyme, both with (+) and without (-) added buffer.
Graph of the percent of all products formed in 1 h in the TES reactions as a function of 5' splice site sequence (33). The black bars on the graph represent 9-mer TES products, the white bars represent 6-mer 5' cleavage products, and the gray bars represent all the cryptic products formed. The results are the average of three independent assays, and the standard deviation in all cases is less than 10%. Note that the order of the data in the graph does not correspond to the loading order of the representative gel.
Figure 3.5. Time Studies Investigating the Formation of Cryptic Products

Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM MgCl₂. Aliquots were removed at the times listed above each lane. The migration position on the gel of the 10-mer substrates, the 9-mer products, the 6-mer intermediates, and the cryptic products are labeled. The lane labeled “(+) Buffer” contains a 120 min reaction in the absence of added ribozyme. (A) TES reaction utilizing the substrate augaccgcuc and the ribozyme rP-8/4x, which places a c-G pair at the 5’ splice site (33). Note that cryptic products begin to appear after the 15 min time point. This combination is representative of all four Watson-Crick base pair combinations. (B) TES reaction utilizing the substrate augaccgcuc and the ribozyme rP-8/4x-5’C, which places a c-C pair at the 5’ splice site (33). Note that cryptic products begin to appear after 1 min.
(C) TES reaction utilizing the substrate augacuguc and the ribozyme rP-8/4x, which places a u-G wobble pair at the 5’ splice site (33). Note that no cryptic products form in this case. This combination is representative (in terms of cryptic product formation) of all the other base pair combinations (excluding the Watson-Crick base pairs and the c-C pair).
Figure 3.6. Results for the 5’ Splice Site During 15 Minute Reaction Times

Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 15 min at 44 °C in 10 mM MgCl₂. (Top) A representative polyacrylamide gel using all 16 base pair combinations at the 5’ splice site (33). Each complete substrate sequence and the base at ribozyme position 12 (in uppercase lettering) is shown above its corresponding lane. The migration position on the gel of the 10-mer substrates, the 9-mer products, and the 6-mer intermediates are labeled. All other bands represent cryptic sites. Note that there is some sequence-dependent migration variability between these lanes. The lanes marked “buffer” had substrate augacugcuc incubated as a typical reaction in the absence of ribozyme, both with (+) and without (-) added buffer.
(Bottom) Graph of the percent of all products formed in 15 min in the TES reactions as a function of 5’ splice site sequence (33). The black bars on the graph represent 9-mer TES products, the white bars represent 6-mer 5’ cleavage products, and the gray bars represent all the cryptic products formed. The results are the average of four independent assays and the standard deviation in all cases is less than 10%. Note that the order of the data in the graph does not correspond to the loading order of the representative gel. Data has been ordered according to percent TES product formation.
Figure 3.7. Results for the 5’ Splice Site Using 3’ End Labeled Substrates

Representative gel of the sequence analysis of the 5’ splice site of TES reactions at 15 min using 3’ end radiolabeled substrates (33). TES reactions were conducted the same as the reactions utilizing the 5’ radiolabeled substrates with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM MgCl₂. Each complete substrate sequence and the base at ribozyme position 12 (in uppercase lettering) is shown above its corresponding lane. Note that the final c on each substrate is the radiolabeled c added by the 3’ radiolabeling procedure. The migration position of the 11-mer substrates and the 10-mer products are labeled. The lanes marked “buffer” had substrate augacugcucc incubated as a typical reaction in the absence of ribozyme, both with (+) and without (-) added buffer. The lane marked “(+) control” is a TES reaction using 5’ radiolabeled substrate and the conserved u-G wobble pair at the 5’ splice site run under the same conditions as the 3’ radiolabeled substrates.
Figure 3.8. Time Studies to Identify the Source of the Cryptic Products

Time-dependent polyacrylamide gel of TES reactions utilizing 9-mer TES product (left) and 6-mer intermediate (right) as reaction substrates (33). Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM MgCl$_2$. The ribozyme used in each case was rP-8/4x-5'A, which when paired with the substrate will create a u-A base pair at the 5' splice site. Aliquots were removed at the times listed above each lane. The migration positions on the gel of the 9-mer and 6-mer starting material are labeled. Unlabeled bands are cryptic products. The lanes marked “buffer” were incubated as a typical reaction for 120 min in the absence of ribozyme, both with (+) and without (-) added buffer. Note that cryptic products only occur when using the 9-mer TES product.
Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM MgCl₂. The ribozyme used in each case was rP-8/4x-5'A, which when paired with the substrate augacucuc will create a u-A base pair at the 5' splice site. After the reaction proceeded for 5 min, 1.3 μM of the unlabeled TES product (1000-fold excess over substrate) was added to one of the reactions. Aliquots were removed starting 10 min later (15 min after start of the reaction). Shown are graphs comparing reactions in the absence (left) and presence (right) of this unlabeled competitor (33). TES products are represented by black circles, 6-mer hydrolysis products by white circles, and all cryptic products by white triangles. The addition of competitor does reduce the amount of TES product formed because not all the labeled substrate has had a chance to bind the ribozyme prior to this addition. Of the labeled substrate that does bind, however, the addition of competitor eliminates the formation of cryptic products. This indicates the products dissociate and, in the presence of excess unlabeled product, are unable to rebind ribozymes for cryptic cleavage.
Figure 3.10. Results for the 3’ Splice Site for 1 Hour Reaction Times

Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in 10 mM MgCl₂. (Top) Diagram of the model TES reaction used (33). The four substrates utilized were 10-mers (shown in lowercase lettering), where x is one of each of the four nucleotides. Note that x represents the substrate position analogous to the ω position of self-splicing introns. The recognition elements from ribozyme rP-8/4x are shown in uppercase lettering.
Graph of the percent of all products formed in the TES reactions as a function of 3’ splice site sequence (33). The black bars represent 9-mer TES products and the white bars represent 6-mer 5’ cleavage products. The results are the average of two independent assays and the standard deviation in all cases is less than 10%. Notice that only ωG produces TES product.
Figure 3.11. Results for Reactions Where no P10 Formation Is Possible

Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C at the MgCl₂ concentrations listed on the graph. (Top) Diagram of the model TES reaction used (33). The substrate 10-mer is shown in lowercase lettering and the recognition elements of the ribozyme are shown in uppercase lettering. In this system, P10 formation is not possible. (Bottom) Graph of the percent of all products formed in the TES reactions as a function of MgCl₂ concentration. The black bars represent 9-mer TES products and the white bars represent 6-mer 5’ cleavage products. The results are the average of two independent assays, and the standard deviation in all cases is less than 10%.
CHAPTER 4 - IN VIVO EXCISION OF A SINGLE TARGETED NUCLEOTIDE FROM AN RNA TRANSCRIPT BY A TRANS EXCISION-SPLICING RIBOZYME

Proposed Project

To date, the studies on TES ribozymes have been performed in vitro (22, 32, 33). While these studies have shown the potential of TES ribozymes as viable biochemical tools in the test tube, the usefulness of these ribozymes will be enhanced by demonstrating in vivo reactivity. These TES ribozymes could be used as biochemical tools for the sequence specific modification of RNA transcripts in vivo. The ability of TES ribozymes to specifically remove sequences from within RNAs has potential therapeutic value in that they can be used to remove mutations associated with diseases at the RNA level. These therapeutic applications depend on demonstrated in vivo reactivity. Other group I intron-derived ribozymes have demonstrated reactivity in non-native cellular contexts, including therapeutically relevant systems (7, 10, 12-15, 19, 21, 23, 25, 26, 28, 29), so it is feasible that the P. carinii ribozyme can catalyze the TES reaction in vivo. However, there are differences between the TES ribozyme and other group I intron-derived ribozymes that may result in a lack of activity in a non-native cellular setting. Thus, it is vitally important to determine if the TES ribozyme is active in vivo.

To this end, I have designed a test system to assess the ability of the P. carinii ribozyme to catalyze the TES reaction in a non-native cellular environment, specifically Escherichia coli. Green fluorescent protein (GFP) was chosen as the target for this system because it can be expressed in many different cellular environments (85, 88) and simple changes to the coding region can completely destroy its fluorescent properties (90, 130). A single base insertion mutation was engineered into the GFP gene, which causes a frameshift in the coding region of the mRNA and produces a non-fluorescent protein product. The P. carinii ribozyme was reengineered to target this mutation, so a successful TES reaction will reestablish the reading frame of the protein and will produce a fluorescing protein (Figure 4.1). A single base was targeted for excision because this represents the simplest TES target system and limits the number of recognition elements utilized by the ribozyme.
to two. Cell cultures were monitored for an increase in GFP fluorescence upon induced ribozyme transcription to determine ribozyme reactivity \textit{in vivo}. These studies were undertaken to demonstrate the ability of the TES ribozyme \textit{in vivo} and to establish the groundwork for future studies of these ribozymes in more complex, and therapeutically relevant, systems.

\textbf{Materials and Methods}

\textit{Oligonucleotide synthesis and preparation}

DNA oligonucleotides for site-directed mutagenesis were purchased from Integrated DNA Technologies (Coralville, IA). Oligonucleotides for the insertion of the T7 terminator sequence into the P-8/4x ribozyme precursor plasmid (P3X) were also purchased from Integrated DNA Technologies and were 5’ phosphorylated and purified via polyacrylamide gel electrophoresis (PAGE) by the company.

\textit{Plasmid construction}

The \textit{P. carinii} ribozyme precursor plasmid, P-8/4x, was constructed as previously described (11). Note the P-8/4x plasmid will be referred to as the P3X plasmid in this chapter. To stop transcription of the ribozyme \textit{in vivo}, a T7 terminator sequence (131) was inserted immediately 3’ to the ribozyme sequence (Figure 4.2). This insertion was made using the \textit{Xba}I and \textit{HindIII} restriction enzyme sites present in the P3X plasmid (Figure 4.3). The P3X plasmid was digested with \textit{HindIII} and the linear plasmid was purified from the reaction mixture using a QIAquick PCR Purification kit (QIAGEN; Valencia, CA). The linear plasmid was then digested with \textit{Xba}I and the double-cut plasmid was purified from the reaction mixture using a QIAquick PCR Purification kit. The following oligonucleotides containing the T7 terminator sequence were ligated into the double cut P3X plasmid to create P3X+T7T:

\begin{align*}
5’\text{CTAGATAGCATAACCCCTTGGGCGCTCTAAACGGGTCTTGAGGGGTTTTTGA} & \text{ and} \\
3’\text{AGCTTCAAAAAACCCCTCAAGAAGCGGTTTAGAGGCCCCCAAGGGGTATGCTAT} & \text{. The ligation reaction consisted of 300 ng double-cut P3X plasmid, } \sim 245 \text{ ng}
\end{align*}
double-stranded T7 terminator insert, 30 mM Tris-HCl (pH 7.8), 10 mM MgCl₂, 10 mM dithiothreitol (DTT), 1 mM ATP, and 3 U T4 DNA ligase (Promega; Madison, WI) in a 10-μL reaction volume. All components, except the ligase, were mixed and incubated at greater than 80 °C for 1 minute. The mixture was cooled to room temperature and then the ligase was added. The ligation reaction proceeded at room temperature overnight. A 3-μL aliquot of the overnight ligation reaction was then used to transform *Escherichia coli* DH5α competent cells (Invitrogen; Carlsbad, CA). The resulting plasmid (P3X+T7T) was purified using a QIAprep Spin Miniprep kit (QIAGEN) and sequenced for confirmation (ACGT, Inc.; Northbrook, IL).

Two unique restriction enzyme sites, *Bgl*II and *Sph*I, were engineered into the newly created P3X+T7T plasmid to allow for the excision of the ribozyme construct (Figure 4.3). The plasmid, P3X+T7TBS, was created by two successive rounds of site-directed mutagenesis. The following primer pair was used to create a *Bgl*II restriction site upstream of the T7 promoter:

\[
5'\text{GGAAACAGATCTGACATGATTACGAATTGG}3' \quad \text{and} \quad 5'\text{CCAAATTCGTAATCATGTCGATCTGTTTCC}3'.
\]

A *Sph*I restriction site was created downstream of the T7 terminator with the following primer pair:

\[
5'\text{GCTTACTAGTGATGCATGCTCTATAGTGTCACC}3' \quad \text{and} \quad 5'\text{GGTGACACTATAGGAGCATGCATCACTAGTAAGC}3'.
\]

The created restriction sites are underlined. The site-directed mutagenesis reactions were conducted as previously described (22), with the changes that follow. The reaction mixtures were subjected to denaturation at 95 °C for 30 s, followed by 15 temperature cycles of 95 °C for 30 s, 55 °C for 2 min, and 68 °C for 6 min. The parental plasmids were digested with 20 units of *Dpn*I (New England Biolabs; Beverly, MA) in 4.2 μL of manufacturer’s buffer for at least 1 h at 37 °C. A 3-μL aliquot of this digest was then used to transform *E. coli* DH5α competent cells. The resultant plasmids were purified using a QIAprep Spin Miniprep kit and were sequenced for confirmation (Davis Sequencing; Davis, CA).

The plasmid containing the GFP target, pQBI T7-GFP, was purchased from QBIOGENE (Carlsbad, CA). The base plasmid for testing, pQBI GFP + P3X, was created by inserting the ribozyme construct into the pQBI T7-GFP plasmid using the
BglII and the Sphl restriction sites present in the pQBI T7-GFP plasmid (Figure 4.3). The pQBI T7-GFP plasmid was prepared by first digesting with Sphl. The linear plasmid was purified from the reaction mixture using a QIAquick PCR Purification kit. The linear plasmid was then digested with BglII. The double-cut plasmid was purified on a 1.2% agarose gel and the plasmid band was excised. The plasmid was extracted from the gel matrix using a QIAquick Gel Extraction kit (QIAGEN). The gel-purified plasmid was then ethanol precipitated twice prior to ligation. The ribozyme construct was isolated from the P3X+T7TBS plasmid by the following process. The P3X+T7TBS plasmid was first digested with Sphl and the linear plasmid was purified from the reaction mixture using a QIAquick PCR Purification kit. The linear plasmid was then digested with BglII and the resulting bands were purified on a 1.2% agarose gel. The ribozyme construct band was excised and extracted from the gel matrix using a QIAquick Gel Extraction kit. The gel-purified product was ethanol precipitated twice prior to ligation. The ligation reaction consisted of ~50 ng double-cut pQBI T7-GFP plasmid, ~125 ng ribozyme construct, 30 mM Tris-HCl (pH 7.8), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, and 3 U T4 DNA ligase (Promega) in a 13-μL reaction volume. All components, except the ligase, were mixed and incubated at greater than 90 ºC for 1 minute. The mixture was cooled to room temperature and then the ligase was added. The ligation reaction proceeded at room temperature overnight. A 3-μL aliquot of the overnight ligation reaction was then used to transform E. coli DH5α competent cells. The resulting plasmid (pQBI GFP + P3X) was purified using a QIAprep Spin Miniprep kit and sequenced for confirmation (Davis Sequencing).

Test plasmids were generated via site-directed mutagenesis of the base plasmid pQBI GFP + P3X. The following primers were used to modify the GFP gene to contain a single base insertion mutation to destroy GFP fluorescence (Mut GFP):

\[ 5'\text{GCAGATTGTGTGGACAGGTAATGTTGTCTGG}^3' \quad \text{and} \quad \text{and} \]
\[ 5'\text{CCAGACAACCATTACCTGTCCACACAATCTGCG}^3' \]

The underlined base represents the insertion mutation. The expected product from a successful TES reaction (Corr GFP) was created with the following primers:

\[ 5'\text{GCAGATTGTGTGGAAAGGTAATGTTGTCTGG}^3' \quad \text{and} \quad \text{and} \]
The underlined base represents the silent mutation expected from removal of the targeted guanosine by the TES ribozyme. These mutations alter codon 201 of the GFP gene. Site-directed mutagenesis was performed as described above, with the following modifications. In various reactions, the annealing temperature was lowered to 50 ºC, the elongation time was increased from 6 min to 8 min to account for the larger plasmid, and/or the number of temperature cycles was increased to 18 cycles. Resulting plasmids were purified as above and sequenced for confirmation (Davis Sequencing).

The ribozyme portion of the test plasmid was modified in the following ways using site-directed mutagenesis. First, two of the three recognition elements of the ribozyme (RE1 and RE3) were modified to recognize the GFP target by two successive rounds of site-directed mutagenesis. A third recognition element (RE2) is not utilized or required in TES reactions excising a single base (22, 33), so it was not modified. The primers for the first round of sequence changes were:

5'CGACTCAGTACCCACTCATACACCAGACAGGGGTAGAAAGCGGC3'
and
5'GCCGCTTTCTACCCCTCTCTATAGTGAGTGC3'.

The second round of changes was made with the following primer pair:

5'CGACTCAGTACCCACTCATACACCAGACAGGGGTAGAAAGCGGC3'
and
5'GCCGCTTTCTACCCCTCTCTCTATAGTGAGTGC3'.

The length of RE3 was increased to 5 bases using the following primers:

5'CGACTCAGTACCCACTCATACACCAGACAGGGGTAGAAAGCGGC3'
and
5'GCCGCTTTCTACCCCTCTCTCTCTATAGTGAGTGC3'.

An RE3 containing 7 bases was created using the following primer pair:

5'CGACTCAGTACCCACTCATACACCAGACAGGGGTAGAAAGCGGC3'
and
5'GCCGCTTTCTACCCCTCTCTCTCTCTATAGTGAGTGC3'.

The underlined bases represent the bases used to increase the length of the RE3 region. A mutant form of the ribozyme was created by deleting four bases from the Guanosine Binding Site (GBS) using the following primers: 5'GGATGCAGTTCCACTAGATCGGCAGTG3' and 5'CCACTGACCCTATCTAGTGAATGCTCC3'. This removes bases 250-253 from the ribozyme sequence (11). Site-directed mutagenesis reactions were performed as above, with annealing temperatures of 50 ºC or 55 ºC, elongation times of 8 min,
and 18 temperature cycles. Resulting plasmids were sequenced to confirm the changes (Davis Sequencing).

Preparing competent JM109(DE3)

In vivo testing of the TES ribozyme was conducted in E. coli strain JM109(DE3). This strain was chosen because it contains the gene encoding T7 RNA polymerase under the control of an inducible promoter. Competent JM109(DE3) cells were prepared using the rubidium chloride protocol reported by Promega (132). JM109(DE3) cells from a glycerol stock (Promega) were streaked on an M-9 minimal agar plate containing 1mM thiamine-HCl and the plate was incubated at 37 °C overnight. An overnight culture of JM109(DE3) was created using 5 mL of room temperature LB media and colonies from the streaked JM109(DE3) plate. The culture was then incubated at 37 °C with shaking at 225 rpm overnight. The next day, an aliquot of the overnight culture of JM109(DE3) was used to inoculate LB media containing 20 mM MgSO₄ to give a 1:100 dilution. The dilution culture was grown at 37 °C with shaking at 225 rpm until the A₆₀₀ (absorbance of the bulk culture at 600 nm) reached 0.4–0.6. Cells were pelleted by centrifuging at 4,500 x g for 5 min at 4 °C. The supernant was removed from the pellets and the cell pellets were resuspended in 0.4 volume (of the original culture) of ice cold TFB1, which consists of 30 mM potassium acetate, 10 mM CaCl₂, 50 mM MnCl₂, 100 mM RbCl and 15% glycerol at pH 5.8. The resuspended cells were incubated on ice for 5 min and the cells were then pelleted by centrifuging at 4,500 x g for 5 min at 4 °C. The supernant was removed and the cells were resuspended in 1/25 volume (of the original culture volume) of ice cold TFB2, which consists of 10 mM PIPES (pH 6.5), 75 mM CaCl₂, 10 mM RbCl, and 15% glycerol at pH 6.5. The resuspended cells were then incubated on ice for 1 h.

Transformation of competent JM109(DE3)

Following the 1 h incubation on ice, the competent cells were transformed with the test plasmids using a modified version of the procedure outlined by Promega (132). 100 μL of competent cells was used in each transformation reaction. Approximately 10 ng of test plasmid was swirled into the competent cells and the
tubes were incubated for 30 min on ice. The cells were then heat shocked to allow transformation by heating the tubes at 44 °C for 50 sec and placing the tubes immediately back on ice for 2 min. 1.5 mL of room temperature LB media was then added to each tube and the tubes were incubated at 37 °C with shaking at 225 rpm for 45 min. After incubation, 200 µL of each transformation reaction was plated on LB plates containing ampicillin and the plates were incubated at 37 °C overnight.

In vivo TES reactions

The colonies resulting from transformation of the competent JM109(DE3) cells were used for *in vivo* testing of the TES ribozymes. The colonies were picked and used to inoculate 3 mL of LB media containing ampicillin. The cultures were incubated at 37 °C with shaking at 225 rpm overnight. The following day, the overnight cultures were used to inoculate fresh LB media without antibiotic to create 1:20 dilution cultures. Each 1:20 dilution culture consisted of 8 mL of LB media and 400 µL of overnight culture. The dilution cultures were incubated at 37 °C with shaking at 225 rpm for 3 h to allow growth to reach log phase ($A_{600} > 0.4$). After 3 h, aliquots were removed from each culture for cell density analysis (500 µL) and fluorescence analysis (1 mL). The rest of the dilution culture was then subdivided. The inducer, isopropylthiogalactoside (IPTG), was added to one of the cultures to a final concentration of ~1.3 mM, while the other culture served as a non-induced control. IPTG is a galactose mimic which induces, or activates, the lac promoter that controls production of T7 RNA polymerase in JM109(DE3). IPTG cannot be hydrolyzed by the cell, so the concentration of IPTG stays constant throughout the course of the experiment. Cultures were incubated at 37 °C with shaking at 225 rpm for 5 h. After 5 h, aliquots were removed for cell density analysis (500 µL), GFP fluorescence analysis (1 mL) and for select cultures, total RNA isolation (1 mL).

Analysis

Cell density was determined by measuring the absorbance of the culture at 600 nm ($A_{600}$). For GFP fluorescence analysis, the culture aliquots were centrifuged at 6,000 x g for 12 min at 4 °C to pellet the cells. The media was removed and the cell pellets were resuspended in 300 µL of phosphate buffered saline (PBS) at pH
7.4. The LB media used in these studies has fluorescent properties that could interfere with GFP fluorescence detection, so the cells were resuspended in PBS, which is not fluorescent. The resuspended pellets were then loaded into a FluoroNunc™ MaxiSorp™ 96-well plate (Nalge Nunc International; Rochester, NY). GFP fluorescence was measured in a CytoFluor (PerSeptive Biosystems; Framingham, MA), using a 485 ± 20 nm filter for excitation and a 508 ± 20 nm filter for emission, as the excitation wavelength for this form of GFP is 474 nm and the emission wavelength is 509 nm. Use of the CytoFluor was kindly provided by the lab of Dr. Sylvia Daunert (University of Kentucky). The raw fluorescence measurements are adjusted to correct for differences in cell culture growth by dividing the raw fluorescence values by the $A_{600}$ readings. These adjusted fluorescence values were then normalized to the non-induced values to correct for “leaky” T7 RNA polymerase activity. This correction was made by subtracting the adjusted fluorescence for the non-induced culture from the adjusted fluorescence for the induced culture. These adjusted values are the normalized fluorescence values reported in the graphs. Note that omitting the corrections for differences in cell density does not significantly affect the trends obtained (data not shown). Error limits for the estimated percentages of RNA repair were calculated as for multiplicative expressions (133).

**Total RNA isolation**

Total RNA was isolated from 1 mL aliquots from designated cultures using the Ambion RiboPure™-Bacteria Kit (Ambion, Inc.; Austin, TX). Isolated RNA was treated with RQ1 RNase-Free DNase (Promega) to completely remove the DNA prior to RT-PCR. RNA was isolated from the DNase reaction mixture by acid phenol (pH 4.3) extraction, followed by phenol/chloroform extraction and ethanol precipitation. DNase treatment was repeated as necessary to ensure the DNA had been removed from the samples (as confirmed by PCR).

**RT-PCR reactions**

The GFP transcripts were amplified from the total RNA by RT-PCR using the following primers: $^{5'}$GTTGTACAGTTCATCCATGCC$^{3'}$ and $^{5'}$GGAGAAGAACTCTTCACTGG$^{3'}$. RT-PCR reactions were performed using the
Access RT-PCR System (Promega) and consisted of 50 μL reactions containing ~1 μg total RNA, 1 mM MgSO₄, 45 pmol of each primer, 0.2 mM dNTPs, 5 U AMV Reverse Transcriptase, and 5 U Tfl DNA Polymerase in the provided reaction buffer. PCR reactions were performed as a control to ensure the DNA had been removed from the samples. The reactions were subjected to 45 °C for 45 minutes for first strand cDNA synthesis, followed by 2 minutes at 94 °C to inactivate the AMV Reverse Transcriptase. The reactions then underwent 40 temperature cycles consisting of 94 °C for 30 s, 54 °C for 1 min, and 68 °C for 2 min. After cycling was complete, the reactions underwent a final extension cycle at 68 °C for 10 min. The RT-PCR products were separated on a 2% agarose gel and the GFP band was excised from the gel. The band was extracted from the gel matrix using a QIAquick Gel Extraction kit. The gel-purified products were ethanol precipitated twice prior to use in ligation reactions.

**Assay for corrected transcripts**

The GFP RT-PCR products were ligated into the pDrive cloning vector using the QIAGEN PCR Cloning kit (QIAGEN). Ligation reactions consisted of 50 ng pDrive Cloning Vector, 200 ng of RT-PCR product, 1 μL 50% polyethylene glycol (PEG; 5% final w/v), and 2X Ligation Master Mix (provided with the kit) in a final volume of 10 μL. Ligations proceeded for 1.5 h at 4 °C. Immediately after ligation was complete, a 3-μL aliquot of the ligation reaction was used to transform E. coli DH5α competent cells. Transformations were plated on LB media plates containing kanamycin and the plates were incubated at 37 °C overnight. The following day, the plates were placed at 4 °C for at least 4 h to enhance GFP fluorescence for visualization. After incubation at 4 °C, the plates were exposed to UV light and colonies showing GFP fluorescence were isolated. The isolated colonies were used to inoculate LB media containing kanamycin and the cultures were incubated at 37 °C with shaking at 225 rpm overnight. The following day, the resulting plasmids were isolated from the cultures using a QIAprep Spin Miniprep kit and sequenced for identification (Davis Sequencing).
System Design

Design of the in vivo test system

In order to evaluate the activity of the TES ribozyme in vivo, a reporter gene with activity that could be easily measured was chosen as the target. Green fluorescent protein (GFP) was a good candidate because it has been expressed in many cell types and it does not require species-specific cofactors to fluoresce (85, 88). Also, simple changes to the coding region completely eliminate fluorescence (90, 130). Thus, creating a simple insertion mutation in the coding region of GFP will destroy fluorescence and the TES ribozyme can be used to remove the insertion and reestablish fluorescence.

The design of the in vivo TES test system was based on previous work involving the trans-splicing ribozyme in E. coli (7, 15), as well as design principles garnered through studies of the TES ribozyme in vitro (22, 32). The simplest TES reaction system involves the removal of a single nucleotide from a substrate. This single nucleotide is equivalent to the $\omega$ position in the self-splicing reaction (60, 104-109, 134). We had previously shown that the only single nucleotide that can be excised by the TES ribozyme is a guanosine (33), so the nucleotide targeted for removal is a guanosine. A GFP test system was designed in which a single base insertion mutation was created in the coding region of GFP (Figure 4.1). This mutation causes a deleterious shift in the reading frame of the transcribed RNA, which alters the amino acid sequence of the resultant protein product, resulting in a loss of fluorescence. To create this mutation, a uridine was inserted 5′ to a guanosine in the RNA transcript (Figure 4.1). In the context of the TES reaction, this uridine will serve to define the 5′ splice site by forming the highly conserved u-G wobble pair with a guanosine in the ribozyme (11, 22, 33, 59, 68, 73, 102, 103). Insertion of a uridine ensures that the ribozyme will target the mutant form of GFP and not the normal form, as the formation of a u-G wobble pair will be favored over a g-G pair at the 5′ splice site, although 5′ cleavage is still possible (33). The guanosine following the inserted uridine represents $\omega$G and it is the base targeted for excision (Figure 4.4). After a successful TES reaction, the inserted uridine
replaces the guanosine in the coding region. Thus, the site for insertion was limited to those codons where replacement of guanosine by uridine results in a silent mutation and the amino acid sequence of the normal and corrected protein are the same. Codon 201, which codes for leucine, fits these criteria and, based on RNA structure prediction estimates, was accessible in the transcript. Thus, a successful TES reaction will put the GFP transcript back in frame, will produce the correct amino acids, and will restore fluorescence to the resultant protein product, while creating a G-to-U transversion in the RNA sequence (Figure 4.1). This G-to-U transversion can later be identified via sequencing of isolated transcripts.

Changes to modify the ribozyme were kept to a minimum for these initial *in vivo* studies. A T7 terminator sequence (131) was added to the 3' end of the ribozyme sequence to stop transcription of the ribozyme *in vivo*. Addition of this terminator increases the length of the ribozyme by ~50 nucleotides. This T7 terminator sequence (shown in red in Figure 4.2), slows down the T7 RNA polymerase (via the formation a hairpin structure) and promotes dissociation of the polymerase from the transcript (via a string of uridines), thus stopping transcription (131). Two of the three recognition elements used by the ribozyme for target identification were modified to target the GFP transcript at the site of the insertion mutation (Figure 4.4). The third recognition element, RE2, is not utilized in TES reactions involving the removal of a single nucleotide, so it was not modified. In all the *in vivo* test systems, the length of RE1 was maintained at the native length of 6 nucleotides. Previous work on the TES ribozyme indicated that increasing the length of RE3 can be beneficial to the reaction (32), while work with the trans-splicing ribozyme showed a requirement for P10 and increased 3' exon interactions for *in vivo* activity (14, 15, 19, 21, 23, 25-29), so ribozymes were tested with RE3 lengths of 3 nucleotides (native length), 5 nucleotides, and 7 nucleotides. The choice of increasing RE3 to 5 nucleotides and 7 nucleotides was made primarily on the sequence of the target transcript. T7 RNA polymerase prefers that the initiation base for transcription be a guanosine (135, 136). As the first nucleotide of the ribozyme that is transcribed is the 5' nucleotide of RE3, this base must be a guanosine. This also means that the target must contain a cytidine (the base pairing partner for
guanosine), so the GFP sequence was analyzed for downstream cytidines to define the end of the P10 helix. Increasing the helix length to 5 nucleotides and 7 nucleotides utilizes the next two cytidines in the GFP sequence (Figure 4.1). As a control for ribozyme activity, mutant ribozyme constructs were constructed by deleting four nucleotides from the guanosine binding site of the ribozyme (enclosed in the gray oval in Figure 4.2). This mutation completely inactivates the ribozyme when tested in vitro (data not shown).

In designing a plasmid for testing, I chose to utilize a single plasmid system to introduce the ribozyme and the target into the cells. A single plasmid ensures that both components for the reaction (substrate and ribozyme) are efficiently transformed into the cells for testing.

**Experimental design for the assay for corrected transcripts**

Selectively isolating the corrected transcripts from a mixture of mutant and corrected transcripts presents unique challenges. The TES reaction removes a targeted sequence from within an RNA, so the 5' and 3' ends of the transcript are the same in the mutant and corrected forms. Moreover, in this simple system, the difference between the mutant GFP transcript and the corrected GFP transcript is a single nucleotide. Thus, the corrected transcript cannot trivially be selectively amplified by RT-PCR. Therefore, an assay was designed to isolate and analyze the pool of transcripts after the reaction for those that are corrected and thus produce fluorescent protein.

In this assay, a pool of GFP transcripts is amplified from total RNA isolated from the *in vivo* reactions (Figure 4.5). The primers used to amplify the transcripts are targeted to the ends of the GFP transcript, thus amplifying the entire GFP coding region. The pool of GFP RT-PCR products will represent the pool of GFP transcripts, with mutant GFP and corrected GFP products being present. The resulting GFP RT-PCR products are purified on an agarose gel and the mixture of the RT-PCR products are ligated into the pDrive cloning vector. This vector was chosen for a few reasons. The pDrive cloning vector is pre-cut and contains U overhangs, which allow for cloning of the RT-PCR products, which have A overhangs. Thus, one RT-PCR
product ligates into one plasmid. The placement of the overhangs in the pDrive vector allows for the insertion of the GFP RT-PCR product in frame with the lac Z gene present in the plasmid, thus creating a fusion between the lac Z and the GFP. Expression of this fusion protein allows for the identification of the GFP product. Cloning of the ligation products into DH5α cells leads to the introduction of one plasmid per transformed cell. When the transformations are plated on agar plates, each transformed cell produces a colony. Each colony on the plate represents an individual ligation event and thus a single RT-PCR product that resulted from amplification of a GFP transcript (either mutant or corrected). As each colony will only contain one plasmid, the lac Z-GFP fusion protein produced will be the same in all the cells that make up the colony. Those colonies containing the corrected GFP products (with the insertion mutation removed) will produce fluorescent GFP at a high level, which can be visualized on the plate. On the other hand, those colonies containing the mutant GFP products will produce non-fluorescent protein products and will not fluoresce when exposed to UV light. This assay allows for the screening of numerous RT-PCR products and selective isolation of corrected GFP products.

Results

The expected product form of GFP is fluorescent

Prior to testing the targeted ribozymes with the mutant GFP target, I tested to see that the mutation introduced into the GFP gene abolishes fluorescence. I also tested to see if the expected product form of GFP (with the G-to-U transversion) fluoresces. As seen in Figure 4.6, introducing the insertion mutation into the GFP gene abolishes GFP fluorescence. The expected product fluoresces, although at a level somewhat below that for the normal form of GFP, indicating the G-to-U transversion is not largely disruptive to the translation of the GFP protein. These tests were conducted using the different forms of GFP in the same plasmid as the non-specific ribozyme (rP3X), so the results are obtained in the presence of active (although non-targeted) ribozyme. The reported values are the result of 4 independent assays and the standard deviations are below 10%.
The TES ribozyme is active in vivo

The ribozymes engineered to target the mutation in GFP were tested for in vivo activity by pairing the ribozymes with the mutant form of GFP (Mut GFP). Targeted ribozymes (GFP rP3X) were tested with RE3 lengths of 3 nucleotides, 5 nucleotides and 7 nucleotides. Increasing the length of RE3 increases the interaction between the ribozyme and the 3’ exon of the substrate. These increased base pairing interactions help prevent 3’ exon dissociation prior to the second reaction step and can improve TES product formation in vitro (32). The native form of the ribozyme (rP3X) served as a non-GFP-targeted control for ribozyme activity, while inactive forms of the ribozymes (containing a 4-nucleotide deletion in the guanosine binding site) were used as negative controls. As seen in Figure 4.7, increasing the length of RE3 to 5 nucleotides for the targeted ribozyme (GFP rP3X RE3=5) significantly increases the fluorescence over the inactive and the non-specific ribozyme controls, indicating TES reactivity. The reported values are the result of 4 independent assays and the standard deviations are below 10%. Note the level of fluorescence for the negative controls is not zero due to a low level of inherent fluorescence in the samples. I obtained an estimate of TES reactivity by comparing the data from the constructs with mutant GFP (Figure 4.7) and the same ribozyme constructs with the expected product form of GFP. For the expected product, normalized fluorescence values of 16265.99 ± 388.75 (ribozyme with an RE3=3), 3105.83 ± 1940.59 (ribozyme with an RE3=5), and 5971.97 ± 2272.05 (ribozyme with an RE3=7) were obtained from at least 3 independent tests. Note the large standard deviations result from greater variability in reactions conducted in vivo when the RE3 length is increased from the native length of 3 nucleotides in constructs containing the expected product form of GFP. The estimated percentage of RNA repair was low for the targeted ribozymes with RE3 lengths of 3 nucleotides (1.1 ± 0.1%) and 7 nucleotides (3 ± 1%), while utilizing a targeted ribozyme with an RE3 length of 5 nucleotides increased the estimated percentage of repair to 12 ± 8%.
The targeted guanosine is removed from the corrected transcripts

Total RNA was isolated from reactions involving 3 different constructs: the mutant GFP paired with the targeted ribozyme containing an RE3=5 (the repair construct that showed the greatest increase in GFP fluorescence), the mutant GFP paired with the inactive targeted ribozyme with an RE3=5 (a negative control), and the expected product form of GFP paired with the targeted ribozyme with an RE3=5 (a positive control to assess the insertion efficiency). The total RNA was subjected to RT-PCR to isolate the GFP transcript and the resulting RT-PCR products were ligated into a cloning vector and were assayed for their ability to produce fluorescent protein (Figure 4.5). In the screening assays, twelve fluorescent colonies (out of ~2000 screened colonies) were obtained from products from 2 independent *in vivo* tests with the active repair construct, indicating the presence of corrected transcripts. Assays involving products from three independent *in vivo* tests with the mutant ribozyme construct produced no fluorescent colonies in ~2000 screened colonies, indicating that no TES reaction had occurred. The twelve fluorescent colonies from the active repair construct, one non-fluorescent colony from the active repair construct and eighteen non-fluorescent colonies from the negative control construct were isolated and the resulting plasmids were sequenced. All twelve sequences from the fluorescent colonies showed the removal of the targeted guanosine, resulting in the G-to-U transversion in the corrected transcript (Figure 4.8). If the cell randomly fixed the transcript and produced fluorescent protein, the loss of a uridine would be anticipated to be as likely as the loss of a guanosine. As this loss of a uridine was not observed, the fluorescent protein was produced as a result of a successful TES reaction. There was also no evidence of a deletion that could serve as a compensatory mutation to restore fluorescence. The non-fluorescent colonies were confirmed to contain the mutant GFP transcript, which still contains the targeted guanosine (Figure 4.8). Thus the removal of the targeted guanosine was TES ribozyme-mediated (without being deleterious to the cell) and dependent on active ribozyme. Based on this method, ~1.2% of the mutant transcripts are corrected by the TES ribozyme.
Discussion

The arsenal of molecular-based RNA tools is rapidly growing; however, the majority of these tools have focused on the destruction of the RNA message rather than the modification of the message (24, 37, 39). Destroying the RNA message is useful for studying the effects of shutting off genes and can be useful in therapeutic applications to prevent the production of mutant proteins that lead to disease. On the other hand, modifying the RNA message can restore the function to the transcript, thus reducing mutant protein and producing normal protein. With this in mind, this work reports the development of a catalytic RNA that can sequence specifically target a single nucleotide within a cellular transcript \textit{in vivo} and excise that targeted nucleotide from the transcript, with little or no apparent toxicity to the cell (30). Combined with the ease of adaptability, this provides for a wide variety of new \textit{in vivo} experimental strategies.

As TES ribozymes recognize their targets initially and primarily through base pairing, engineering the ribozymes to target new transcript regions simply requires changing the sequence of the recognition elements to base pair with the desired target. Other considerations for ribozyme targeting include reconstituting a u-G wobble pair at the 5’ splice site and a guanosine at the 3’ splice site as the last (or only) base of the sequence targeted for excision (33). In this report, a single nucleotide was targeted for excision, however, previous work with this ribozyme demonstrated that segments larger than a single nucleotide can be excised \textit{in vitro} (22, 32). It is anticipated that larger regions could be excised \textit{in vivo}, and investigations into these types of targets would utilize the third recognition element (RE2) of the ribozyme. This added interaction could improve the reactivity and specificity of the ribozyme \textit{in vivo}, as it adds to the interactions between the ribozyme and its target.

Comparison to \textit{in vitro} TES results

In the simplest TES test system, which involves the excision of a single, targeted nucleotide, the ribozyme is able to produce 70% product \textit{in vitro} under optimized conditions of ribozyme excess and 10 mM MgCl$_2$ (22). For the \textit{in vivo} tests
presented here, the ribozyme restores approximately 12% of the fluorescence (30). This decrease in reactivity is not unexpected and is likely due to a number of factors. Nevertheless, TES ribozymes are adaptable to the cellular environment and are able to produce detectable amounts of repaired RNA.

There are several possible explanations for the difference between the *in vitro* results and the *in vivo* results. First, the sequences targeted by the ribozymes differ. The *in vitro* system utilizes a substrate that mimics the native exon sequences of the *P. carinii* intron (22), while the *in vivo* system utilizes a different sequence in the GFP gene (30). The differences in the sequences could have an effect on the efficiency of the reaction. Target accessibility may also play a role in the levels of correction observed. RNA structure prediction models indicated the region containing the insertion mutation would be accessible, so it was anticipated that the ribozyme would be able to base pair with the target without structural interference. As TES correction was observed, the ribozyme was able to recognize the target *in vivo*. It is possible that other factors, such as proteins binding the target transcript, are affecting the ability of the ribozyme to effectively recognize its target.

The differences between the controlled *in vitro* reaction conditions and the cellular environment could also be affecting the structure and the reactivity of the ribozyme. The optimum MgCl\(_2\) concentration for *in vitro* TES ribozyme activity for the test system excising a single nucleotide is 10 mM (22). The concentration of free magnesium in the cell is expected to be much less (137, 138), so other factors may be required for the proper folding of the ribozyme. *E. coli* is a non-native system for the *P. carinii* ribozyme, so species-specific, beneficial cellular cofactors may not be available. *In vitro*, the concentrations of the ribozyme and of the substrate can be controlled and the reactions are conducted under ribozyme excess conditions (22, 32, 33). This level of control is not possible in this *in vivo* system and the concentrations are expected to be closer to a 1:1 ratio. It is possible that cellular proteins are binding the ribozyme and are preventing the ribozyme from folding into its optimum catalytic conformation. This would also lower the relative concentration of active ribozyme. *In vitro*, guanosine cofactor is not required for the TES reaction. As TES reactivity is observed *in vivo*, it appears that the TES reaction occurs in the
presence of free guanosine. The lower levels of TES product formation do not appear to be a result of a detrimental effect of transcribing the TES ribozyme in the cell, as there was no observed significant difference in growth for cultures containing active or inactive ribozymes.

Another explanation for the lower amount of in vivo TES product formation is reduced exon ligation (the product of the second reaction step). Previous work with the TES ribozyme in vitro showed that TES product formation could be increased by decreasing 3’ exon dissociation after the first reaction step. This dissociation is mediated by the presence of sequences downstream of the P10 helix (32). The simple substrate (10 nucleotides) used in the in vitro studies (22, 33) does not have additional sequences downstream of the P10 helix, while the transcript target in these in vivo studies has many nucleotides downstream of the P10 helix (30). The presence of these additional sequences could increase the dissociation of the 3’ exon. One way to decrease the dissociation is by increasing the length of RE3, which increases the length of the P10 helix. A crystal structure of a group I intron showed there are no tertiary interactions between the intron and the 3’ exon (109, 134), so increased base pairing is required to hold the 3’ exon during the reaction. Previous in vitro work showed that an RE3 of 5 nucleotides led to a dramatic improvement in TES reactivity, while a further increase to 10 nucleotides did not improve the reaction over the RE3 of 5 (32). As the greatest increase in GFP fluorescence was observed with a ribozyme containing an RE=5 (Figure 4.7), it appears that an RE3 of 5 nucleotides may represent an optimum length for reactivity in vitro, as well as in vivo. Note RE3 lengths greater than 7 nucleotides were not tested in these in vivo studies. The effect of RE3 may not depend solely on length, but also on sequence, so longer RE3 lengths may further improve the reaction in other systems.

Comparison to trans-splicing

The TES ribozyme offers a different approach to RNA repair compared to the Tetrahymena-derived trans-splicing ribozyme (7). The trans-splicing ribozyme replaces the 3’ exon of a targeted transcript with a new version of the exon carried
by the ribozyme. Thus trans-splicing works best for mutations located near the 3’
end of the transcript as the ribozyme is required to carry the replacement exon and
any required regulatory elements contained in the 3’ region of the transcript. The
TES ribozyme excises an internal segment, so no additional sequences are carried
by the ribozyme. This lack of extra sequences in the TES ribozyme eliminates
potential delivery issues that could arise from the size of the repair ribozyme. Also,
the internal segment targeted for excision by the TES ribozyme can, in theory, be
located anywhere on the transcript. The excision of an internal segment maintains
the regulatory elements present in the target transcript, thus eliminating concerns
about post-transcriptional regulation of the corrected transcript. While the trans-
splicing ribozyme is single turnover, the TES ribozyme has the potential to be a
multiple-turnover catalyst, as it is not changed in the reaction.

In *E. coli*, the level of activity restored by the trans-splicing ribozyme was
approximately 1% (7). In the study presented here, a higher level of repair in *E. coli*
was observed, as the level of activity restored by the TES ribozyme was
approximately 12% (30). That trans-splicing ribozymes (8, 10, 13, 15, 23, 25, 26, 28,
29) work fairly well in mammalian cells (upwards of 50%) is encouraging and it is
anticipated that TES ribozymes will also be active in mammalian systems. As is the
case with trans-splicing (10, 23, 25, 29), we do not anticipate that TES modification
will have to be 100% to have measurable effects as a biochemical tool. A low level
of TES activity may produce a relevant amount of modified transcript for protein
production.

**Comparison to SMaRT**

RNA transcripts have also been modified by exon replacement using a
method called spliceosome-mediated RNA *trans*-splicing, or SMaRT (139, 140).
SMaRT utilizes a pre-trans-splicing molecule (PTM), which contains the replacement
exon, and the spliceosomes present in nucleated cells (141). This method aims to
set up a competition between *cis*-splicing (occurring between two exons in the same
transcript) and *trans*-splicing (occurring between the PTM and one of the exons in
the target transcript). In terms of RNA repair, SMaRT is more akin to trans-splicing in
that it modifies RNA transcripts by replacing a portion of the transcript with a new RNA sequence and thus faces some of the same challenges. While the spliceosome is a multiple turnover catalyst, the PTM is single turnover since it donates its replacement sequence during the reaction. This means SMaRT is not a multiple turnover reaction. While SMaRT has been used to replace the 3’ (139, 142, 143) and the 5’ (144) ends of mutant transcripts, the splicing event must occur at an intron/exon junction. This requirement precludes the use of SMaRT in lower organisms. The efficiency of SMaRT is comparable to the repair rates obtained with trans-splicing ribozymes (139, 142-145) at the RNA transcript level. SMaRT produces measurable amounts of correct protein product, once again indicating that 100% modification of RNA transcripts is not always necessary to have beneficial results.

Comparison to RNA cleavage methods

Other approaches that aim to reduce the presence or the effects of mutant RNA transcripts in a cell involve cleaving ribozymes, such as the hairpin, hammerhead and HDV ribozymes (18, 20, 24, 37, 39, 43, 48, 146, 147); RNA interference, or RNAi (24, 36, 39); and antisense oligonucleotides (24, 39, 148). The cleaving ribozymes reduce the number of mutant transcripts in a cell by binding a target transcript (via base pairing) and cleaving the target in a sequence specific manner (24, 39). RNAi utilizes small double-stranded RNA molecules in a RNA-induced silencing complex (RISC) to identify target transcripts via base pairing. The identified targets are then degraded, thus preventing protein synthesis from that target (36, 39). Antisense also utilizes small oligonucleotides, but it does not utilize the RISC. The base pairing of these antisense RNAs to the target can serve as a block to prevent translation, or can mark the transcripts for destruction (24, 39). These RNA tools reduce the number of mutant transcripts in a cell and knock out the function of the target transcript. TES ribozymes are able to remove defective elements of a transcript, without destroying the transcript. Thus, TES ribozymes can be used to restore the function to the target transcript. The correction of mutant transcripts, as opposed to the destruction of the transcripts, reduces the amount of
mutant protein produced while increasing the amount of correct protein being made. The TES ribozyme also offers a new approach to mutation analysis. Instead of knocking out the function of a gene by destroying the entire target transcript, TES ribozymes allow for the removal of specific segments of a transcript, including promoters, regulatory elements and binding domains, while maintaining the rest of the transcript. Thus, deleterious effects of a mutant transcript can be assigned to specific regions of the transcript, which could lead to more directed therapeutic approaches.

Issues

While the work presented here demonstrates that the TES ribozyme is active in *E. coli*, there are issues that must be overcome to make this ribozyme a generally applicable biochemical tool and a potential therapeutic agent. One major obstacle is the low level of TES correction. As mentioned above for trans-splicing (10, 23, 25, 29) and SMaRT (142, 143), TES modification will not necessarily have to be 100% to have measurable effects as a biochemical tool. Nevertheless, a higher rate of repair is desirable. It should be noted that the ribozyme used in these studies was minimally altered from its native state. The possibility of mutating the ribozyme in regions other than the recognition elements to improve *in vivo* activity has not yet been investigated. Random mutagenesis of the *Tetrahymena* group I self-splicing intron allowed for selection of introns that were more active in non-native *in vivo* systems (149), so it is anticipated that a similar selection process will allow for the selection of TES ribozymes with greater activity *in vivo*.

The low levels of TES correction could also be related to poor specificity and targeting. These effects are difficult to assess for the TES reaction *in vivo*. The TES ribozyme removes an internal segment from within an RNA and leaves the flanking ends intact. The TES ribozyme could be performing only the first step (5' cleavage) on non-targeted transcripts, which would lead to non-productive products. No new sequence from the ribozyme is transferred to the targeted transcript, so “tagged” transcripts cannot selectively isolated, as was done for trans-splicing, to determine optimum reaction sites (12, 17, 19, 23, 26, 29) or to analyze the specificity (8). As no
detrimental affect on cell growth from the presence of the active ribozyme was observed, it appears that the ribozyme is not attacking crucial transcripts. The sequencing results presented in this work showed that the targeted guanosine is removed from the fluorescent transcripts, indicating a sequence specific reaction (Figure 4.8).

Targeting is an inherent issue for the TES ribozyme. Internal segments are being excised in a sequence specific manner, so one cannot necessarily search for accessible sites for targeting. The target choice is determined by the segment targeted for removal and its flanking sequences. Thus it is possible that sequences that could be targeted for removal in vitro will not be viable targets in vivo due to structure or other factors. There may be ways to utilize antisense approaches to assess the accessibility of the region of interest at a base pairing level (150), but this may not reflect the ability of the ribozyme to bring the target into proper alignment in the catalytic core for reactivity.

Note that the challenges we face in improving the TES reaction in vivo and adapting TES ribozymes to a mammalian system are the same challenges faced by others developing ribozymes in mammalian systems (24, 39, 151). Common issues in E. coli and mammalian cells include increasing the level of ribozyme activity, improving ribozyme specificity, assessing target accessibility, and effectively delivering the ribozyme. Work is ongoing to address these issues and is encouraging (24, 37, 39, 151). Advances in these areas will allow for the further development of TES ribozymes as biochemical and therapeutic tools.

Implications

The ability of the TES ribozyme to removal internal segments from within RNA transcripts offers many potential applications. As biochemical tools, TES ribozymes could be used as inducible tools to modulate the production of proteins in vivo. As demonstrated in this report, TES ribozymes can be targeted to remove sequences to restore the reading frame of a transcript, thus producing active protein. The removal of nucleotides to shift the sequence out of frame is also possible, so inactive or mutant proteins could also be produced in an inducible manner.
The applications of TES ribozymes could also include therapeutic applications. Targeting genetic mutations at the DNA level is difficult due to the stable structure of DNA. RNA transcripts (which contain the mutations found in the DNA sequence) are more accessible and more open to modification. As a therapeutic agent, TES ribozymes could be used to remove insertion mutations and premature stop codons from transcripts, thus restoring the reading frame of the transcript for protein production. We have previously demonstrated the \textit{in vitro} ability of a TES ribozyme to target a small model mimic of the triplet expansion implicated in Myotonic Dystrophy and to excise from that mimic a short triplet expansion (22). Thus the repair of transcripts involved in triplet expansion diseases, such as Muscular Dystrophy and Huntington's Disease, is another potential application for TES ribozymes \textit{in vivo}.

Demonstrating TES activity in a cellular context advances the potential uses of this ribozyme. As outlined in this work, the TES ribozyme can target a transcript and modify it by excising an internal single nucleotide \textit{in vivo}. The results presented here show the potential of the TES ribozyme as a tool for mRNA modification and repair. Targeting the GFP gene demonstrates that the recognition elements of the ribozyme are mutable in an \textit{in vivo} setting, as was previously demonstrated \textit{in vitro} (22). The ability to change the recognition elements and to successfully target the complement to those recognition elements demonstrates that the TES ribozyme can be rationally designed to target and modify transcripts in a sequence specific way \textit{in vivo}. 

88
Figure 4.1. Design of the GFP Target System

Design of the GFP target system (30). The top line shows the sequence of interest of the GFP transcript. The designated codon codes for leucine (Codon 201). The boxed codons base pair with recognition element 1 (RE1) of the ribozyme. Site-directed mutagenesis was used alter the GFP coding region to insert a uridine (in red) 5’ to a guanosine in the mRNA transcript, which causes a frameshift and a premature stop codon (not shown). The TES reaction removes the targeted guanosine and restores the correct reading frame of the GFP gene (bottom line). Note that the leucine codon is CUU in the corrected transcript, resulting from a G-to-U transversion from the normal transcript.
Figure 4.2. The Trans Excision-Splicing Ribozyme with a T7 Terminator Sequence

The predicted secondary structure of the trans excision-splicing ribozyme (rP3X) with the addition of a T7 terminator sequence. The T7 terminator sequence (shown in red) stops transcription of the ribozyme in vivo (131). The above structure was adapted from (22).
Figure 4.3. Plasmid Construction
Construction of the *In Vivo* TES Test Plasmid (30). The original TES ribozyme plasmid (P3X) was modified by the insertion of a linker containing a T7 terminator sequence 3’ to the ribozyme sequence to create P3X + T7T. Site-directed mutagenesis was used to create a unique *Bgl*II site 5’ to the T7 promoter of the ribozyme and a unique *Sph*I site 3’ to the T7 terminator of the ribozyme (P3X + T7TBS). The ribozyme construct was isolated from the plasmid by a double restriction enzyme digest with *Bgl*II and *Sph*I. The GFP-containing plasmid, pQBI T7-GFP (QBIOGENE), was prepared for insertion of the ribozyme construct by a double restriction enzyme digest with *Bgl*II and *Sph*I. The ribozyme construct was ligated into the cut pQBI T7-GFP plasmid to create pQBI GFP + P3X. Modifications to the GFP gene to create the mutant and corrected forms and modifications to the ribozyme to change the recognition elements and to create the inactive forms were all created via site-directed mutagenesis.
The GFP rP3X ribozyme is in uppercase lettering, the GFP target is in lowercase lettering, the insertion mutation (uridine) is in red, and the guanosine to be excised (ωG) is encircled. The ribozyme recognition elements RE1, RE2, and RE3 base pair with the substrate to form the P1, P9.0, and P10 helices, respectively. Note that the P9.0 helix is not formed in this system. The sites of catalysis for the first step (5' cleavage) and the second step (exon ligation) are shown with large bold arrows. The 3-base P10 helix is boxed, with brackets indicating the extensions to 5 and 7 bases. Note that the diagram only shows the recognition elements of the ribozyme. This ribozyme is the same as rP-8/4x (11), except for the sequences of RE1 and RE3 and the addition of a T7 terminator (131) on the 3’ end (Figure 4.2).
Figure 4.5. Schematic of the Assay to Selectively Isolate Corrected Transcripts

Schematic of the assay developed to selectively isolate corrected GFP transcripts (30). Total RNA is isolated from cells that have undergone in vivo testing. RT-PCR is used to amplify the GFP coding region from the isolated RNA. The resulting RT-PCR products are ligated into the pDrive PCR cloning vector (Qiagen). Note that the RT-PCR products can be ligated into the vector in the forward direction (which puts the GFP gene in frame with the LacZ α-peptide of the plasmid) or the reverse direction, with each direction assumed to be equally possible. Only products ligated in the forward direction will produce fluorescent GFP. The colonies resulting from transformation of the ligated plasmids into E. coli cells are exposed to UV light to identify fluorescent colonies. Selected colonies are grown up in bulk culture. The resulting plasmids are isolated and sequenced to confirm the expected sequence change in the corrected transcripts.
Comparison of the normalized GFP fluorescence levels for the normal, mutant and expected product forms of GFP in the presence of active, non-targeted ribozyme (30). All points utilize constructs with the non-targeted ribozyme (rP3X) and the GFP form listed on the graph. Normalized fluorescence was determined by correcting the raw fluorescence for differences in cell density (determined by the absorbance of the bulk culture at 600nm) and normalizing to non-induced controls. Each data point represents 4 independent assays and the standard deviations are below 10%.
Figure 4.7. TES Correction of Mutant GFP

Normalized fluorescence data for constructs pairing mutant GFP with the TES ribozymes (30). All points utilize the mutant GFP target with the ribozyme listed above the graph. The green bars (+) indicate active ribozymes, while the black bars (−) indicate the ribozymes have been inactivated by a deletion in the guanosine binding site. Normalized fluorescence was determined by correcting the raw fluorescence for differences in cell density (determined by the absorbance of the bulk culture at 600nm) and normalizing to non-induced controls. Each data point represents 4 independent assays and the standard deviations are below 10%.
Figure 4.8. Sequencing Results for Non-Fluorescent and Fluorescent Colonies

Sequencing confirmation of the *in vivo* TES reaction (30). Sequencing results for uncorrected GFP (left) and corrected GFP (right) transcripts. Total RNA isolated from *in vivo* tests involving the construct Mut GFP + GFP rP3X RE3=5 was utilized as the template for RT-PCR to isolate GFP transcripts. The uncorrected GFP sequence represents plasmids isolated from non-fluorescent colonies. The corrected GFP sequence represents plasmids isolated from fluorescent colonies.
CHAPTER 5 - CONCLUSIONS

The trans excision-splicing reaction is a potentially useful biochemical tool for the sequence specific modification of RNAs. While the ability of the *Pneumocystis carinii* ribozyme to catalyze this reaction using small RNA substrates *in vitro* had been demonstrated previously (22, 32), many questions remain about the interactions between the ribozyme and its target, as well as about the abilities of this ribozyme in a cellular setting. The work presented in this dissertation further characterizes the *P. carinii* ribozyme *in vitro* and gives insight into the molecular recognition of the 5’ and 3’ splice sites by this ribozyme. These studies also demonstrate, for the first time, that this ribozyme-catalyzed reaction can occur *in vivo*.

**The Sequence Requirements for the TES Ribozyme at the 5’ and 3’ Splice Sites**

The work investigating the sequence requirements of the trans excision-splicing ribozyme at the 5’ and 3’ splice sites has provided further insight into the molecular recognition of this ribozyme. The sequence requirements at the 5’ splice site are not stringent, as all base pair combinations allow some level of 5’ cleavage (the first reaction step). Sequences other than the highly conserved u-G wobble pair can produce TES product in appreciable yields. The greatest TES product yields required a stable base pair (a Watson-Crick or wobble pair) at the 5’ splice site, indicating stricter sequence requirements for the second reaction step (exon ligation). The degradation of TES products containing Watson-Crick base pairs at the 5’ splice site demonstrated that the ribozyme recognizes a structural perturbation in the helix formed between the recognition elements of the ribozyme and the substrate. Non-Watson-Crick base pairs and \( \omega \)G can play a role in 5’ splice site determination by creating this perturbation. This structural perturbation had not been demonstrated previously and these results are the first to implicate \( \omega \)G in 5’ splice site determination (33). This role of \( \omega \)G may be limited to these simple target systems where a single nucleotide is removed due to its location immediately adjacent to the 5’ cleavage site.
The sequence requirements at the 3’ splice site are absolute, with only guanosine in the ω position allowing TES product formation. The specificity of the ribozyme for ωG is determined by the G-binding site of the ribozyme. Attempts to rationally redesign this site to accept a different nucleotide were unsuccessful, so in its current state, the TES ribozyme is only able to remove single guanosine insertions or sequences ending in guanosine from RNA substrates.

The results of this study gave us insight into the molecular recognition of the splice sites by this ribozyme. We are able to develop guidelines for new TES target systems, particularly in terms of the 5’ splice site. The highly conserved u-G wobble pair can be replaced with other base pairs, which expands the pool of potential substrates for TES reactions. The results obtained from the studies presented here give an indication of sequence combinations that will provide the most specific ribozyme for a particular target.

**The TES Ribozyme Can Excise a Single, Targeted Nucleotide from an mRNA In Vivo**

The studies outlined in this work demonstrate that the TES ribozyme is active in a cellular setting. The ribozyme was successfully reengineered to excise a single, targeted nucleotide from within an RNA transcript in *E. coli*. The transcription of active ribozyme produced no observed negative effect on cell growth, indicating that production of the ribozyme was not detrimental to the cell. The ribozyme was able to correct the mutant, targeted transcripts, resulting in a measurable restoration of protein activity. The ribozyme catalyzes the successful TES reaction in a sequence specific manner, as the sequenced corrected transcripts showed the targeted guanosine had been removed.

These results begin the characterization of the TES ribozyme in the cellular environment. For the first time, we have demonstrated that the ribozyme is able to target and remove a sequence from within an RNA transcript. An RNA transcript is more complex than the small RNA substrates utilized in the previous *in vitro* studies, as there are other sites that the ribozyme could potentially base pair with. RNA transcripts could also contain structure that would interfere with ribozyme binding.
The ribozyme was able to overcome these potential issues and successfully excised the targeted nucleotide from GFP transcripts.

These results are also the first demonstration of a catalytic RNA sequence specifically removing a targeted segment from within an mRNA \textit{in vivo}. This targeted segment can be as small as a single nucleotide. The demonstration of this \textit{in vivo} reactivity paves the way for the future development of this ribozyme-catalyzed reaction as a biochemical tool and as a potential therapeutic.
References


Vita

Dana Ann Baum was born March 8, 1977 in Staunton, Illinois. In 1999, she graduated cum laude from Washington University in St. Louis, Missouri, with a Bachelor of Arts degree in Chemistry with a concentration in Biochemistry. After graduation, she took a position at the Genome Sequencing Center at the Washington University School of Medicine in St. Louis, Missouri. She began as a lab technician and was later named a senior lab technician. Dana spent a year and a half at the Genome Sequencing Center and was a part of the Human Genome Project. Her efforts earned her the Genome Sequencing Center Employee of the Month award in June of 2000. In the spring of 2001, Dana began her graduate studies in the Chemistry Department at the University of Kentucky and conducted her dissertation research in the lab of Dr. Stephen M. Testa.

Publications


International Human Genome Sequencing Consortium (2001) “Initial sequencing and analysis of the human genome” Nature 409, 860-921. (Member of the Washington University Genome Sequencing Center)