GROUP I INTRON-DERIVED RIBOZYME REACTIONS

Ashley Kirtley Johnson
University of Kentucky

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation
https://uknowledge.uky.edu/gradschool_diss/286

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.
ABSTRACT OF DISSERTATION

Ashley Kirtley Johnson

The Graduate School
University of Kentucky
2005
GROUP I INTRON-DERIVED RIBOZYME REACTIONS

ABSTRACT OF DISSERTATION

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

By
Ashley Kirtley Johnson
Lexington, Kentucky

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

2005

Copyright © Ashley Kirtley Johnson
ABSTRACT OF DISSERTATION

GROUP I INTRON-DERIVED
RIBOZYME REACTIONS

Group I introns are catalytic RNAs capable of self-splicing out of RNA transcripts. Ribozymes derived from these group I introns are used to explore the molecular recognition properties involved in intron catalysis. New ribozyme reactions are designed based on the inherent ability of these ribozymes to perform site-specific nucleophilic attacks. This study explores the molecular recognition properties of group I intron-derived ribozyme reactions and describe a new ribozyme reaction involving molecular recognition properties previously not seen.

We report the development, analysis, and use of a new combinatorial approach to analyze the substrate sequence dependence of suicide inhibition, cyclization, and reverse cyclization reactions catalyzed by a group I intron from the opportunistic pathogen Pneumocystis carinii. We demonstrate that the sequence specificity of these Internal Guide Sequence (IGS) mediated reactions is not high, suggesting that RNA targeting strategies which exploit tertiary interactions could have low specificity due to the tolerance of mismatched base pairs.

A group I intron-derived ribozyme from P. carinii has been previously shown to bind an exogenous RNA substrate, splice-out an internal segment, and then ligate the two ends back together (the trans excision-splicing reaction). We now report that a group I intron derived ribozyme from the ciliate Tetrahymena thermophila can also perform the trans excision-splicing reaction, although not nearly as well as the P. carinii ribozyme.

In addition, we discovered a new ribozyme reaction called trans insertion-splicing where the P. carinii ribozyme binds two exogenous RNA substrates and inserts one directly into the other. Although this reaction gives the reverse products of the trans excision-splicing reaction, the trans insertion-splicing reaction is not simply the reverse reaction. The ribozyme recognizes two exogenous substrates through more complex molecular recognition interactions than what has been previously seen in group I intron-derived ribozyme reactions. We give evidence for this new reaction mechanism composed of three steps, with intermediates attached to the ribozyme.
KEY WORDS: Ribozymes, Group I Intron, Molecular Recognition, Pneumocystis carinii, Trans Insertion-Splicing Reaction

Ashley Kirtley Johnson

March 11, 2005
GROUP I INTRON-DERIVED RIBOZYME REACTIONS

By

Ashley Kirtley Johnson

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

Mark Meier, Ph.D.
Director of Graduate Studies

March 11, 2005
Date
RULES FOR 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 Graduate School of the University of Kentucky.
DISSERTATION

Ashley Kirtley Johnson

The Graduate School
University of Kentucky
2005
GROUP I INTRON-DERIVED RIBOZYME REACTIONS

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
Ashley Kirtley Johnson

Lexington, Kentucky

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

Lexington, Kentucky

2005

Copyright © Ashley Kirtley Johnson
Even as a young child, I always wanted to be a scientist. I asked for a microscope for Christmas when I was seven. My career plans varied a bit up until my junior year of high school, when I took an AP chemistry class and knew I wanted to be a chemist. My parents lamented that I took a job right after finishing my undergraduate degree in chemistry, but I knew I was not ready, nor focused enough for the challenges of graduate school. Three years at a job that did not challenge me was enough to make me long to go back to school.

It is difficult to appreciate the highs and lows of graduate research without experiencing it first hand. Research in science is the exploration of the unknown. To figure something out and have a true “Eureka” experience makes all the hard work and frustrations of failed experiments worthwhile. You sift through stacks of journal articles and stare for hours at data in your notebook, then all of a sudden; a light bulb turns on. You have solved the question that has been nagging you for months. These special moments make you forget all the pain and hard work you have had to endure.

There are many people I would like to thank who have helped me become a better scientist, either directly or through friendship and support. I will begin with my research advisor, Dr. Stephen Testa. Dr. Testa has challenged me to think more deeply about my research projects. He has taught me how to be a good scientist, by planning thorough experiments, critically evaluating my data, and then clearly expressing my results and conclusions in writing. Thank you Steve. I will use these important skills the rest of my life. I would like to thank all the members of the Testa Lab, both past and present. My lab mates Dr. Michael Bell and Dr. Rashada Alexander helped show me the ropes when I first started in the lab. Also, it has been a pleasure to work with Dana Baum, Joy Sinha, Patrick Dotson, and Nick Tzouanakis. I want to especially thank Byron Stapleton for helping me run twice as many experiments as humanly possible.

I thank the members of my committee: Dr. Mark Lovell, Dr. Anne Francis Miller, Dr. Joseph McGillis, and Dr. Karen McDowell. I appreciate your time and energy to attend meetings, my oral exam, and my exit seminar. I thank you for agreeing to be on my committee and especially for reading this dissertation.
I thank my husband Corey for his love and support these past five years. Corey is a third-year medical student, and listens to me talk about my research highs and lows with an understanding ear. At the end of every day, he asks about the reactions I ran or the results I got. Corey is my best friend and has made the past five years of graduate school a fun and exciting time that I will never forget.

I thank my family, especially my parents, Bob and Pam Kirtley, for always encouraging me to succeed. Thank you for always asking about my research and especially for coming to my exit seminar. I also appreciate all the calls, cards, and visits. I also thank my sisters Allison and Amanda Kirtley and my grandmother Bessie Clark (Meme) for their love and support. I am lucky to have another wonderful family: my husband’s parents Ron and Bev Johnson and sister Keeli. Thank you for coming to my exit seminar and all of the calls of support. I also want to thank Corey’s grandparents Phil and Bernie Blose for their love and generosity.

I want to thank my life-long friend Mica Evans. Even though she lives in Washington D.C., we see each other several times a year and talk almost weekly. We met when we were 11 and have been great friends ever since. Whenever I need support, Mica is there.

I have made many wonderful friends over the past five years in Lexington. I would like to thank Joanna (Aska) Koziara, Chandra Patel, Jigna Patel, and Roxana Ciochina for all the great times we had watching the Cats play at Rupp Arena. I will definitely miss the student tickets when I am in the real world. I thank Aaron Urbas and Brandon Tackett for all the fun we had brewing beer, and of course the even better time we had drinking it. Being married to a medical student, I have enjoyed the company and support of many medical student wives. Some of my dear friends from this group Alyssa Averill, Kim Wilson, and Katie Sachlaben, I truly thank for your friendship.

I will always treasure the four wonderful years I spent at Agnes Scott College. I loved balancing chemistry and long hours in the lab with literature and political science. The small size of the chemistry department meant that I got plenty of time with my professors. I thank my undergraduate research advisor Dr. T. Leon Venable for all his infinite wisdom and excellent teaching. I thank Dr. Lilia Harvey for her friendship and advice on life outside the lab. I also want to thank all my life-long friends from Agnes
Scott: Audra Brecher, Martha Bailey, Priya Sivanesan, Ruth White, and Sasha Mandic. Emails, phone calls, and visits with these ladies are always enlightening and a pleasure.
# TABLE OF CONTENTS

**ACKNOWLEDGMENTS** .......................................................................................................................... iii  
**LIST OF TABLES** .............................................................................................................................. ix  
**LIST OF FIGURES** .......................................................................................................................... x  

## CHAPTER ONE - Introduction .............................................................................................................. 1  
Combinatorial Approach to Analyze Molecular Recognition in the *P. carinii* Ribozyme ........................................ 1  
A Tetrahymena Ribozyme Performs the Trans Excision-Splicing Reaction ............................................. 2  
The *P. carinii* Ribozyme Performs a New Reaction, Trans Insertion-Splicing .................................................. 3  

## CHAPTER TWO - Background on Catalytic RNA .............................................................................. 5  
Central Dogma of Molecular Biology .................................................................................................... 5  
Deoxyribonucleic Acid ......................................................................................................................... 5  
Ribonucleic Acid .................................................................................................................................................. 6  
RNA Splicing .................................................................................................................................................. 6  
Catalytic RNA .................................................................................................................................................. 7  
Group I Intron Self-Splicing ....................................................................................................................... 7  
Group I Intron-Derived Ribozymes ............................................................................................................. 8  
The Trans Excision-Splicing Reaction ........................................................................................................... 9  
Suicide Inhibition and Reverse Cyclization Reactions ................................................................................. 10  

## CHAPTER THREE - Molecular Recognition Properties of IGS-Mediated Reactions 
Catalyzed by a *Pneumocystis carinii* Group I Intron ............................................................................. 21  
**Introduction** ............................................................................................................................................... 21  
**Materials and Methods** ......................................................................................................................... 23  
Synthesis and Purification of Nucleic Acids ......................................................................................... 23  
MgCl₂ Concentration Dependence of Group I Intron Reactions ............................................................. 24  
Combinatorial Assays ............................................................................................................................... 25  
Product Isolation and Identification ......................................................................................................... 26  
Determination of Binding Constants ........................................................................................................ 27  
Optical Melting Curves .............................................................................................................................. 28  
Kinetics ....................................................................................................................................................... 28  
Modified IGS Studies ............................................................................................................................... 29  
**Results** ................................................................................................................................................... 30  
Combinatorial Assay Rationale .................................................................................................................. 30  
Suicide Inhibition and Reverse Cyclization Assay Development ............................................................. 30  
Sequence Specificity of Suicide Inhibition ................................................................................................. 31  
The MgCl₂ Dependence of Suicide Inhibition ............................................................................................ 32  
Sequence Specificity of Reverse Cyclization .............................................................................................. 32  
Identification of Cyclization Intermediates ............................................................................................... 33  
Binding of Representative 5’ Exon Mimics to the rPC Ribozyme ............................................................. 33
CHAPTER FOUR - A Tetrahymena Group I Intron-Derived Ribozyme Catalyzes the Trans Excision-Splicing Reaction ........................................................................... 56

Introduction................................................................................................................... 56

Methods......................................................................................................................... 58
   Synthesis and Purification of Oligonucleotides ........................................................ 58
   TES Reaction and Optimization ............................................................................... 59
   Kinetic Studies ............................................................................................................ 60

Results ................................................................................................................................................................. 61
   A Tetrahymena Group I Intron-Derived Ribozyme Catalyzes the TES Reaction... 61
   The Tetrahymena Ribozyme does not Perform TES as well as the P. carinii Ribozyme ........................................................................................................ 62
   Hybrid Tetrahymena and P. carinii Ribozymes.......................................................... 63
   Optimization of the Hybrid P. carinii Ribozyme with the Tetrahymena IGS and
   Comparison to the P. carinii Ribozyme .................................................................. 64
   Optimization of the Hybrid Tetrahymena Ribozyme with the P. carinii IGS and
   Comparison to the P. carinii Ribozyme .................................................................. 65

Discussion ................................................................................................................................. 67
   The Tetrahymena Ribozyme Catalyze the TES Reaction......................................... 67
   Comparison of the Four Ribozymes ........................................................................... 67
   The Tetrahymena Substrate is Problematic in the TES Reaction............................. 68
   The TES Reaction Depends on Ribozyme Mediated 5'-Site Cleavage.................... 69
   Implications .................................................................................................................... 69

CHAPTER FIVE - Trans Insertion-Splicing: A Novel Group I Intron-Derived Ribozyme Mechanism for Inserting One RNA Sequence into Another .................... 78

Introduction......................................................................................................................... 78

Materials and Methods ........................................................................................................ 79
   Synthesis and Preparation of Oligonucleotides ........................................................ 79
   Ribozyme Preparation ................................................................................................... 79
   The Trans Insertion-Splicing Reaction ..................................................................... 80
   Product Isolation and Identification .......................................................................... 81

Results ................................................................................................................................. 82
   The P. carinii Group I Intron Catalyzes the TIS Reaction ......................................... 82
   Step 1: The ωG on the Ribozyme Performs a Nucleophilic Attack on the Insert...... 83
Step 2: The $\omega$ Gi on the Insert Performs a Nucleophilic Attack on the Starting Material................................................................. 85
Step 3: The 3'-U on the 5'-half of the Starting Material Attacks the $\omega$ G of the Ribozyme, Forming TIS Product........................................ 87
P9.0 Formation Inhibits the TIS Reaction ......................................................... 88

Discussion............................................................................................................. 89
The P. carinii Group I Intron-Derived Ribozyme Catalyzes the TIS Reaction ...... 89
The TIS Mechanism.............................................................................................. 89
Factors that Influence the TIS Reaction.............................................................. 90
Comparison with Previous Studies ................................................................. 92
Implications: Mechanism.................................................................................. 93
Implications: Potential Applications............................................................... 94
Implications: Cellular Function ...................................................................... 94

CHAPTER SIX - Conclusions ............................................................................... 106
Molecular Recognition in the Suicide Inhibition and Reverse Cyclization Reactions ................................................................. 106
Molecular Recognition in the TES Reaction ....................................................... 107
Molecular Recognition in the TIS Reaction ....................................................... 107
References ........................................................................................................ 109
Vita ...................................................................................................................... 115
LIST OF TABLES

Table 3.1 Suicide Inhibitors of the Self-Splicing Reaction…………………………..43
Table 3.2 Reverse Cyclization Products…………………………………………………44
Table 3.3 Cyclization of the rP-h Intron…………………………………………………45
Table 3.4 Thermodynamic Parameters for Representative Hexamers Binding to rPC
   and to r(GGUCAU) in H15Mg Buffer………………………………………………46
Table 3.5 Kinetics of Representative Exons in the Suicide Inhibition and Reverse
   Cyclization Reactions as a Function of MgCl₂ Concentration…………………47
Table 3.6 Thermodynamic Parameters for Representative Exon Mimics Binding to
   Ribozymes and to IGS Mimics in H15Mg Buffer…………………………….48
Table 4.1 Tetrahymena, P. carinii, and Hybrid Ribozyme TES Optimum Reaction
   Condition and Results……………………………………………………………71
Table 5.1 Ribozyme, Starting Material, and Insert Sequences………………………95
LIST OF FIGURES

Figure 2.1. The Central Dogma of Molecular Biology ...........................................12
Figure 2.2. Translation of RNA ............................................................................13
Figure 2.3. DNA Nucleotides, Double Stranded DNA, and the Double Helix .........14
Figure 2.4. RNA Nucleotides, Single Stranded RNA, and RNA Secondary Structures.15
Figure 2.5. Watson-Crick versus Wobble Base Pairs ............................................16
Figure 2.6. RNA Undergoes Post-Transcriptional Modifications .........................17
Figure 2.7. Group I Intron Self-Splicing and Secondary Structure of a *P. carinii* Group I Intron .................................................................18
Figure 2.8. The Trans Excision-Splicing Reaction and Secondary Structure of a *P. carinii* Group I Intron-Derived Ribozyme ..................19
Figure 2.9. The Suicide Inhibition and Reverse Cyclization Reactions .................20
Figure 3.1 Schematic Diagram of Self-Splicing and Cyclization (Steps A → E), Suicide Inhibition (Steps F → G), and Reverse Cyclization (Steps H → I) .........49
Figure 3.2 Magnesium Dependence of the Suicide Inhibition (S.I.) and Reverse Cyclization (R.C.) Reactions .................................................................51
Figure 3.3 Competitive Binding Assay ...................................................................52
Figure 3.4 Kinetic Analysis of Suicide Inhibition and Reverse Cyclization Reactions...53
Figure 3.5 Schematics of 5’ Exon Mimics Base Pairing with the IGS of the *P. carinii* Ribozyme .................................................................54
Figure 3.6 Schematics of the 5’ End of the Intron Base Pairing with the IGS Before Cyclization ......................................................................................55
Figure 4.1. Schematic of the Trans Excision-Splicing Reaction .............................72
Figure 4.2. Diagrams of the *Tetrahymena*, *P. carinii*, and Hybrid Ribozymes Base Paired to their Complementary TES Substrate ................................................73
Figure 4.3. Representative Polyacrylamide Gel and Graphs of rT TES Optimization...74
Figure 4.4. Representative Polyacrylamide Gel and Graphs of rPC TES Optimization .................................................................75
Figure 4.5. Representative Polyacrylamide Gel and Graphs of rPC-T TES Optimization ......................................................................................76
Figure 4.6. Representative Polyacrylamide Gel and Graphs of rT-PC TES Optimization ......................................................................................77
Figure 5.1. The Trans Excision-Splicing Reaction and Trans Insertion-Splicing Reaction ......................................................................................96
Figure 5.2. Proposed Secondary Structure of a *P. carinii* Group I Intron-Derived Ribozyme ......................................................................................97
Figure 5.3. Proof for the First Nucleophilic Attack in the TIS Reaction .................98
Figure 5.4. Sequencing of 5’-end Radiolabeled TIS Product ..................................99
Figure 5.5. TIS Optimization Graphs..........................................................100
Figure 5.6. A Representative Time Study Gel and Observed Rate Constants for the
TIS Reaction..............................................................................................101
Figure 5.7. Proposed Mechanism for the Trans Insertion-Splicing Reaction.........102
Figure 5.8. Proof for the Second and Third Nucleophilic Attacks in the TIS
Reaction.....................................................................................................104
Figure 5.9. Secondary Structure at the 3’-end of the rPC-1 and rPC Ribozymes......105
CHAPTER ONE - Introduction

Since the discovery of catalytic RNA in the early 1980s (1, 2), the view of RNA as simply an intermediary between DNA and proteins has drastically changed. For example, many naturally occurring catalytic RNA reactions have been found (3, 4). Ribozymes derived from naturally occurring catalytic RNAs have been developed for use as therapeutics (5-9) and biochemical tools (10, 11). Even more recently, the discovery of RNA interference (12, 13) and riboswitches (14, 15) show that RNA is involved in the most basic levels of gene regulation.

Self-splicing group I introns are catalytic RNAs that splice out of cellular RNA transcripts, often without the need for proteins (1). Group I introns are defined as having a conserved secondary structure (a series of base paired helices named P1 through P10) and a similar self-splicing mechanism. The intron positions itself for catalysis by base pairing with its exons and forming tertiary interactions. These complex catalytic activities are self-contained within RNAs of relatively moderate size, often less than 400 nucleotides. Therefore, group I introns are frequently exploited as model systems in the continual pursuit of a more thorough understanding of RNA structure and function. As the self-splicing reaction is single-turnover, such studies often utilize group I intron-derived ribozymes, which are introns that lack the endogenous 5’ and 3’ exons (16, 17).

Ribozymes recognize their substrates (exon mimics) partially through base pairing with their internal guide sequence (IGS). The base paired IGS makes up the P1 and P10 helices that dock into the catalytic core of the ribozyme (positioned through tertiary interactions). The molecular recognition of the substrate by the ribozyme is dependent on these base pairing and tertiary interactions. Studying these critical molecular recognition interactions allows us to better understand the ribozyme and its ability to catalyze a reaction.

Combinatorial Approach to Analyze Molecular Recognition in the P. carinii Ribozyme

The sequence dependence of the 5’ exon-IGS interaction has been studied using a group I intron-derived ribozyme from the ciliate *Tetrahymena thermophila*, but the sequence dependence of this interaction for a *Pneumocystis carinii* group I intron-derived
A combinatorial approach, using randomized 5’ exon mimics, allowed us to assess the sequence specificity of the *P. carinii* IGS. Exogenous 5’ exon mimics can bind the IGS of the intron, forming a P1 helix. If the 5’-exon mimic binds the intron before it self-splices, it becomes ligated to the 3’-exon (Figure 3.1, F-G). This reaction disrupts self-splicing and is called suicide inhibition. If the 5’-exon mimic binds the intron that has already spliced out and cyclized, it will open the circle and become ligated to the 5’-end of the intron (Figure 3.1, H-I). This is called the reverse cyclization reaction. Both of these catalytic reactions rely on the molecular interaction of the IGS of the intron with an exon mimic, and allow us to analyze what sequences can bind and react.

The results show that additional 5’ exon mimics, other than the complementary mimic to the IGS, can be substrates in the suicide inhibition and the reverse cyclization reactions. Reactions run with ribozymes, changed in their IGS sequences, show that the IGS itself can be altered to target a diverse variety of substrate sequences. These alterations may give complementary base paired P1 helices, but the changes can affect the thermodynamics of P1 helix base pairing and tertiary interactions. The results from this combinatorial study have given us a greater understanding of the fundamental molecular recognition properties of *P. carinii* IGS-mediated reactions.

*A Tetrahymena Ribozyme Performs the Trans Excision-Splicing Reaction*

The trans-excision splicing reaction (TES) is a group I intron-derived reaction where the ribozyme base pairs with an exogenous RNA substrate, excises a specific
segment, and ligates the ends of the RNA substrate back together (Figure 4.1) (6). This TES reaction was developed with a group I intron derived ribozyme from the opportunistic pathogen \textit{P. carinii} and has been extensively studied (6, 20, 21). To show that the TES reaction is more general to group I intron-derived ribozymes, we tested whether another group I intron-derived ribozyme could perform the TES reaction. A group I intron from the ciliate \textit{Tetrahymena thermophila} has predominantly been used to study self-splicing and develop new ribozyme reactions. Therefore, we wanted to see if the \textit{Tetrahymena} ribozyme could perform the TES reaction and if so, compare it to the \textit{P. carinii} TES reaction.

The \textit{Tetrahymena}-derived ribozyme performs the TES reaction, indicating that the TES reaction is not specific to the \textit{P. carinii} derived-ribozyme. The \textit{Tetrahymena} TES reaction has lower optimized product yields and decreased observed rate constants for both the first and second steps of the reaction compared to the \textit{P. carinii} ribozyme, suggesting that the \textit{P. carinii} ribozyme might be better at the TES reaction. Two hybrid ribozymes, with the IGS of either ribozyme switched, were also tested. By exchanging the IGS of each ribozyme, we can compare separately the ability of either the ribozyme or the native substrates to perform the TES reaction. The results suggest that the TES reaction is a general group I intron-derived ribozyme reaction, although the \textit{P. carinii} ribozyme is substantially better at the TES reaction than the \textit{Tetrahymena} ribozyme. Both the specific base paired P1 and P10 helices and the catalytic core of the individual ribozyme appear to contribute to the accuracy and efficiency of the TES reaction.

\textit{The P. carinii Ribozyme Performs a New Reaction, Trans Insertion-Splicing}

A new group I intron-derived ribozyme reaction called trans insertion-splicing (TIS) was developed, where one RNA substrate is directly inserted into a second RNA substrate (Figure 5.7). The \textit{P. carinii} ribozyme that performs the TES reaction also performs the TIS reaction, although the TIS reaction is not the reverse of the TES reaction. The unique TIS mechanism proceeds through three nucleophilic attacks and involves formation of two separate P1 helices, showing that group I introns are capable of more dynamic reactions than previously thought.
The TIS reaction is the first example of the direct insertion of one exogenous RNA into a second exogenous RNA using a ribozyme essentially unaltered from its intronic form. The molecular recognition interactions between the ribozyme and the two TIS substrates were analyzed and this new ribozyme reaction mechanism was elucidated.

Copyright © Ashley Kirtley Johnson
Central Dogma of Molecular Biology

In the central dogma of molecular biology, DNA is transcribed into RNA and RNA is translated into protein (Figure 2.1) (22) Clearly, RNA has a central and very important role in protein synthesis (Figure 2.2). There are three main classes of RNA found in the cell: 1) Messenger RNA (mRNA) codes for amino acids (the building blocks of protein), and its sequence is recognized in three nucleotide long segments called codons; 2) Transfer RNA (tRNA) has an amino acid attached and recognizes the codon sequence of the mRNA (Figure 2.2); and 3) Ribosomal RNA (rRNA) is part of the ribosome complex where proteins are synthesized. The majority of RNA found in the cell is rRNA (80-85%). tRNA makes up 10-15% of the RNA in the cell, and the least stable of the three, mRNA, represents less than 5% (23).

Deoxyribonucleic Acid

DNA, or deoxyribonucleic acid, is a polymer made up of nucleotides. These nucleotides contain a deoxyribose sugar, a nucleobase, and a phosphate (Figure 2.3.A). DNA nucleotides form a polymer linked together through a phosphodiester backbone. This phosphodiester backbone has directionality from the 5’-carbon to the 3’-carbon on the sugar. DNA is typically double stranded and base pairs with a complementary strand that is anti-parallel (in a 3’ to 5’ direction). The base pairs are formed by hydrogen bonding between the nucleobases (Figure 2.3.B). Watson-Crick base pairs are adenine and thymine (forming two hydrogen bonds) and guanine and cytosine (forming three hydrogen bonds). The nucleotides are generally identified by their letter abbreviations (A, C, G, and T). The geometry of the two sets of base pairs are isosteric, so A-T pairs can replace C-G pairs and visa versa without changing the geometry of the helix. Essentially, the Watson-Crick base pairs allow for the unique double helix structure of DNA (Figure 2.3.C).
Ribonucleic Acid

RNA, or ribonucleic acid, is also a polymer made up of nucleotides. The main difference between RNA and DNA is that RNA contains a ribose sugar instead of a deoxyribose. The ribose sugar has a hydroxyl group on its 2’-carbon (Figure 2.4.A, highlighted in gray), essentially giving RNA nucleotides an extra functional group compared to DNA. This extra functional group makes RNA chemically less stable than DNA and allows for RNA to have catalytic function. Another difference between RNA and DNA is that RNA uses the nucleobase uracil and DNA uses thymine. Uracil differs from thymine by containing one less methyl group (compare Figures 2.3.B and 2.4.B).

RNA, like DNA, is a polymer with a phosphodiester backbone linked in a 5’ to 3’ direction (Figure 2.4.B), but RNA is typically single stranded (does not have a complementary strand like DNA). Therefore, it can base pair with complementary sections of itself, producing secondary structures such as hairpins and pseudoknots (Figure 2.4.C). The secondary structure of tRNA forms several hairpins and is often referred to as a cloverleaf pattern (shown in Figure 2.2). RNA secondary structures form A-U and C-G Watson-Crick base pairs (Figure 2.5). In addition, energetically less stable non-Watson-Crick base pairs, such as the G-U wobble pair, are also formed (Figure 2.5). The G-U wobble pair has different functional groups available for interactions (Figure 2.5, indicated with gray circles), compared to Watson-Crick base pairs.

Secondary structure often leaves unpaired nucleotides in the form of loops (Figure 2.4.C) and bulges that can interact with substrates or other regions of the RNA. A loop in tRNA called the anti-codon recognizes the codon in the mRNA through base pairing (Figure 2.2). These unpaired regions also add to the three-dimensional tertiary structure that gives large RNAs a globular shape, similar to proteins.

RNA Splicing

In eukaryotes, after RNA is transcribed and before it is translated, it is altered in a process called post-transcriptional modification. Transcribed RNA can be thought of as regions of exons (expressed sequences) and introns (intervening sequences) as depicted in Figure 2.6. Through a process called RNA splicing, introns are removed, resulting in
the exons being ligated together to form mature RNA. Phillip Sharp and Richard Roberts received the 1993 Nobel Prize in medicine for their discovery of RNA splicing. Through a process called alternative splicing, removal of various introns allows for different RNAs to be generated from the same gene.

**Catalytic RNA**

Catalytic RNAs have specific three-dimensional shapes that form active sites. The active sites recognize substrates and position them for catalysis, thus lowering the activation energy for the reaction. Catalytic RNA was discovered by Thomas Cech, as an intron from the ciliate *Tetrahymena thermophila* capable of self-splicing without the need for proteins (1) (Figure 2.7.A). Concurrently, Sidney Altman demonstrated that the catalytic portion of Ribonuclease P (an RNA and protein complex) was RNA (2). The idea that RNA could catalyze a reaction revolutionized the biological view of RNA and won Cech and Altman the 1989 Nobel Prize in Chemistry.

**Group I Intron Self-Splicing**

Introns are classified based on their conserved secondary structure, as well as their splicing reaction mechanism. The self-splicing mechanism and secondary structure for a group I intron from the fungal pathogen *Pneumocystis carinii* is depicted in Figure 2.7. Group I introns perform self-splicing through two nucleophilic attacks (Figure 2.7.A) (25, 26). The first attack occurs by a hydroxyl (OH) group on an exogenous guanosine (G) at the 5’exon-intron junction (see Figure 2.7.A, marked 1st). This attack causes release of the intron from the 5’exon and leaves a free 3’-OH on the 5’exon. In the second nucleophilic attack, this free 3’-OH on the 5’exon attacks at the intron-3’exon junction (see Figure 2.7.A, marked 2nd). The second attack results in ligation of the two exons and release of the intron.

The secondary structure of group I introns is defined by ten conserved base paired regions labeled P1-P10 (Figure 2.7.B). These base paired regions form helices and unpaired loops that are then involved in tertiary interactions. The secondary and tertiary structure of the group I intron helps position the exons for attack at the 5’ and 3’-splice
sites. The 5’-exon base pairs with the intron, forming the P1 helix (see Figure 2.7.B, 5’exon in pink). The first attack occurs at the 5’-exon/intron junction and is directed by formation of the P1 helix. A highly conserved G-U wobble pair helps define the 5’-splice site (27, 28). The G-U wobble pair has different functional groups available compared to Watson-Crick base pairs. One functional group in particular, an exocyclic amine (Figure 2.5), is directly involved in catalysis at the 5’splice site (29, 30). The first nucleophilic attack occurs by the 3’-hydroxyl group on an exogenous guanosine (G), attacking the 5’-splice junction (Figure 2.7.B).

To position the 3’ exon for the second step of self-splicing, the 3’exon forms a P10 helix with the intron (see Figure 2.7.B, 3’exon in light blue). The second nucleophilic attack is performed by a 3’-OH on the 5’-exon that was made available after the first nucleophilic attack. The second nucleophilic attack occurs at the junction between the intron and the 3’-exon. This attack occurs after the last nucleotide in the intron, a conserved omega G (ωG), and results in ligation of the two exons and release of the intron sequence.

The intron recognizes its substrates and positions them for catalysis through several important interactions. The internal guide sequence (IGS) is the sequence in the intron that forms the P1 and P10 helices (31). This sequence helps the intron recognize its substrate through base pairing. The P1 and P10 helices also form tertiary interactions with the intron by docking into the catalytic core (17, 18, 32, 33). The guanosine binding site (GBS) is a site in the intron that interacts with the exogenous G and positions it for the first nucleophilic attack of self-splicing (34). The GBS is highlighted in yellow in Figure 2.7.B. After the first nucleophilic attack, the ωG of the intron is positioned into the GBS to help direct the second nucleophilic attack (35-37). The 3’-end of the intron forms a P9.0 helix (Figure 2.7.B) that helps position the ωG into the GBS.

---

*Group I Intron-Derived Ribozymes*

Group I introns are frequently exploited as model systems to pursue a more thorough understanding of RNA structure and function. Since the self-splicing reaction is single-turnover, studies often utilize group I intron-derived ribozymes. These ribozymes are essentially introns that lack their endogenous 5’ and 3’ exons (16, 17). Therefore,
group I intron derived-ribozymes recognize exogenous substrates and utilize the same molecular recognition interactions that the intron uses in the self-splicing reaction. New ribozyme reactions are developed to exploit the molecular recognition interactions and the inherent catalytic properties of the intron. In addition, it has been shown that the IGS sequence of these ribozymes can be altered to target the ribozyme to different substrates, such that complementary base pairing is maintained within the P1 helix (5, 31, 38, 39). This characteristic has been exploited for engineering group I intron-derived ribozymes that target and react with specific sequences.

The Trans Excision-Splicing Reaction

The trans excision-splicing (TES) reaction results in removal of specific nucleotides from the center of an RNA substrate. Specifically, the TES ribozyme cuts open an RNA substrate, removes a section in the middle, and splices the ends of the substrate back together (6). The TES reaction was demonstrated using a P. carinii ribozyme and is related to the group I intron self-splicing reaction. The two nucleophilic attacks in the TES reaction and the secondary structure of the P. carinii-derived ribozyme are shown in Figure 2.8. The P. carinii ribozyme recognizes the TES substrate in part through base pairing (Figure 2.8.B). The 5’ and 3’ ends of the TES substrate are essentially 5’ and 3’ exon mimics (Figure 2.8, 5’exon mimic in pink and 3’exon mimic in light blue). The TES reaction results in removal of an internal segment (Figure 2.8, shown in green) and ligation of the two exon mimics (TES product).

Similar to self-splicing, the TES reaction proceeds through two nucleophilic attacks. The 5’-exon mimic in the substrate base pairs with the ribozyme, forming a P1 helix (Figure 2.8.B, 5’exon mimic in pink). The first site of attack occurs at the 5’-exon mimic/intron junction (5’-splice site) and is directed by formation of the P1 helix. The first nucleophilic attack occurs by a free hydroxyl group from the buffer at the 5’-splice site (unlike self-splicing that utilizes an exogenous G). To position the 3’ exon mimic for the second nucleophilic attack, the 3’exon mimic forms a P10 helix with the ribozyme (Figure 2.8.B, 3’exon mimic in light blue). The G being removed (Figure 2.8, in green) acts like the ωG from the intron, and probably interacts with the GBS in the ribozyme.
The second nucleophilic attack is performed by a 3’-hydroxyl on the 5’-exon mimic that was made available after the first nucleophilic attack. The second attack occurs at the junction between the G (Figure 2.8.B in green) and the 3’-exon mimic (Figure 2.8.B in light blue), thus resulting in ligation of the two exons mimics (Figure 2.8.A, TES product) and release of the G.

The TES reaction with the *P. carinii* ribozyme has been extensively studied. The removed section can be as little as one nucleotide (required to be a G as described) and as long as 28 nucleotides (6). The sequence of the IGS can also be changed to target the ribozyme to different substrates (6). The length of the P10 and P9.0 helices can be increased to enhance the second step of the TES reaction (20). Additionally, the TES reaction can still occur without a G-U wobble pair at the 5’splice site, although a G-U wobble pair at this position gives the greatest yield and rates for formation of TES product (21).

**Suicide Inhibition and Reverse Cyclization Reactions**

Introns in large RNA transcripts can self-splice and then the excised intron can react with itself in a cyclization reaction (40-42). Therefore, the self-splicing and cyclization reactions occur consecutively (Figure 2.9.A). Addition of an inhibitor to these reactions can prevent the self-splicing reaction and also open the circular intron. These reactions are called suicide inhibition and reverse cyclization.

In the suicide inhibition reaction, exogenous 5’ exon mimics can compete with the 5’-exon, bind the IGS of the intron, and form a P1 helix. 5’exon mimics that bind and react prevent the self-splicing reaction and become ligated to the 3’exon (Figure 2.9.B) (43-45). This reaction is called suicide inhibition because the inhibitor disrupts self-splicing and is actually used in the reaction.

The reverse cyclization reaction occurs concurrently with suicide inhibition. Some of the introns that self-splice react with themselves and form a cyclized product (Figure 2.9.A). These circular introns still contain their IGS. The same suicide inhibitor added to the reaction can base pair with the IGS of the cyclized intron, forming a P1
helix. If the inhibitor binds and reacts, it opens up the cyclized intron and becomes ligated to the 5’-end of the intron (Figure 2.9.C) (31, 46, 47).

In both of these reactions, the inhibitor becomes ligated to part of the RNA transcript, allowing for relatively easy product analysis.
Figure 2.1. The central dogma of molecular biology. DNA is the heredity material that codes for RNA in a process called transcription (22). RNA codes for protein in a process called translation.
Figure 2.2. Translation of RNA. DNA is transcribed into RNA, and RNA is translated into protein. In protein synthesis, RNA has several functions: as messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). mRNA codes for the protein sequence in three nucleotide segments called codons. The codon ACU codes for the amino acid threonine (blue). And likewise, UAC codes for tyrosine (green), and GUC codes for valine (red). tRNA brings amino acids to the ribosome. The anti-codon in the tRNA binds the codon in the mRNA. The protein synthesis occurs in the ribosome (yellow circle), which is an rRNA/protein complex. The protein grows longer as tRNAs bring more amino acids until a stop codon in the mRNA is reached (48).
Figure 2.3. DNA nucleotides, double stranded DNA, and the double helix.  A) Deoxyribonucleic acid (DNA) is made up of nucleotides.  DNA nucleotides consist of a phosphate, a deoxyribose sugar, and a nucleobase.  B) DNA forms a phosphodiester backbone in a 5’ to 3’ direction.  A complementary strand running 3’ to 5’ forms base pairs with the nucleobases.  Four different nucleobases are found in DNA; adenine, thiamine, guanine, and cytosine.  Adenine forms two hydrogen bonds with thymine and guanine forms three hydrogen bonds with cytosine.  C) The double helix of DNA forms by twisting of two complementary paired DNA strands.
Figure 2.4. RNA nucleotides, single stranded RNA, and RNA secondary structures. 

A) Ribonucleic acid (RNA) is made up of nucleotides. RNA nucleotides consist of a phosphate, a ribose sugar, and a nucleobase. RNA differs from DNA by having an additional functional group (the OH highlighted in gray). 

B) RNA forms a phosphodiester backbone in a 5’ to 3’ direction. Four different nucleobases are found in RNA; adenine, uracil, guanine, and cytosine. RNA is typically found single stranded, although it can base pair with itself. Adenine forms two hydrogen bonds with uracil and guanine forms three hydrogen bonds with cytosine. 

C) RNA forms secondary structures when it base pairs with itself. These structures typically have unpaired loops and bulges, as seen in the hairpin and pseudoknot.
Figure 2.5. Watson-Crick versus wobble base pairs. Watson-Crick base pairs have the same geometry, so either a G-C or an A-U pair gives the same dimensions. A G-U wobble base pair forms two hydrogen bonds and is less energetically stable than a Watson-Crick base pair. There are also different functional groups available when the G-U wobble pair forms (highlighted in gray).
DNA → RNA → Protein

Figure 2.6. RNA undergoes post-transcriptional modifications. Intervening sequences called introns are spliced out of the RNA transcript to form mature RNA.
Figure 2.7. Group I intron self-splicing and secondary structure of a P. carinii group I intron. A) Group I intron self-splicing proceeds through two nucleophilic attacks. The first attack occurs by a 3’-OH group on an exogenous G at the 5’-exon/intron junction. The second attack occurs by a free 3’-OH on the 5’-exon at the intron/3’-exon junction. B) The secondary structure of group I introns is defined by the conserved base paired regions labeled P1-P10. The 5’-exon is in pink, the 3’-exon is in light blue, and the rest of the sequence makes up the intron. The 5’ and 3’-exons base pair with the intron to help position them for catalysis. The first and second positions of nucleophilic attack are labeled. The first attack occurs at the end of the P1 helix at a highly conserved G-U wobble pair. The second attack occurs at the end of the P10 helix at the ωG in the intron. The exogenous G and the ωG interact with the G binding site (GBS) marked by the yellow circle.
Figure 2.8. The trans excision-splicing reaction and secondary structure of a *P. carinii* group I intron-derived ribozyme. The 5’-exon mimic is shown in pink. The 3’-exon mimic is shown in light blue. The region to be removed in the TES reaction is shown in green. Arrows labeled 1st and 2nd denote the two nucleophilic attacks. A) The TES reaction proceeds through two nucleophilic attacks. The first attack occurs by a free OH group from the buffer at the 5’-exon mimic/intron junction. The second attack occurs by a free 3’-OH on the 5’-exon mimic at the intron/3’-exon mimic junction. B) The secondary structure of group I introns is defined by the conserved base paired regions labeled P1-P10. The group I intron-derived ribozyme has its exon sequences removed. The 5’ and 3’-exon mimics base pair with the ribozyme to help position them for catalysis. The first and second positions of nucleophilic attack are labeled. The first attack occurs at the end of the P1 helix at a highly conserved G-U wobble pair. The second attack occurs at the end of the P10 helix at the ωG in the region being removed.
Figure 2.9. The suicide inhibition and reverse cyclization reactions. A) The group I intron self-splicing and cyclization reactions occur consecutively. B) Exogenous 5’ exon mimics can bind the IGS and react with the intron before self-splicing, thereby acting as suicide inhibitors of the self-splicing reaction. If the inhibitor binds and reacts, it becomes ligated to the 3’-exon and prevents self-splicing. C) The reverse cyclization reaction occurs concurrently with the suicide inhibition reaction. Introns that self-splice can react with themselves and form a cyclized product that still contains the IGS. If the inhibitor binds and reacts, it opens up the cyclized intron and becomes ligated to the 5’-end of the intron.
CHAPTER THREE - Molecular Recognition Properties of IGS-
Mediated Reactions Catalyzed by a *Pneumocystis carinii* Group I Intron

**Introduction**

Self-splicing group I introns are catalytic RNAs that splice out of cellular RNA transcripts (Figure 3.1, A-C), often without the apparent requirement for ancillary proteins (1). The excised introns can then react with themselves in a process called cyclization (Figure 3.1, D-E) (40-42). Group I introns recognize their exons through base pairing at their internal guide sequence (IGS). Both the intron with its exons attached (before self-splicing) and the cyclized intron contain an IGS and can react with substrates that bind to it. In the suicide inhibition reaction, an exogenous 5’ exon mimic can compete with the native exon, bind the intron, and react, becoming ligated to the 3’ exon in the process (Figure 3.1, F-G) (43-45). This reaction disrupts the self-splicing reaction and is therefore called the suicide inhibition reaction. Also, the same exogenous 5’ exon mimic can bind the IGS of the cyclized intron and react, opening up the cyclized intron and becoming ligated to its 5’-end (Figure 3.1, H-I) (31, 46, 47). This reaction opens up the circle and is called the reverse cyclization reaction. Both of these catalytic reactions rely upon the molecular interaction of the IGS with its exon mimic.

Although the sequence dependence of the 5’ exon-IGS interaction has been studied using *Tetrahymena* ribozymes (in trans-splicing and reverse splicing reactions), the sequence dependence of this interaction has not been analyzed for any *P. carinii* ribozyme reaction, or for the suicide inhibition reaction with any ribozyme. This is of interest because previous results have shown that the *P. carinii* and *Tetrahymena* ribozymes utilize a different array of tertiary interactions to dock the P1 helix into the catalytic core (17-19). Essentially, the *Tetrahymena* ribozyme has been shown to recognize the particular structure of the P1 helix, whereas the *P. carinii* ribozyme appears to recognize specific base pairs in the P1 helix. Furthermore, the *P. carinii* ribozyme catalyzes a unique trans excision-splicing reaction, which along with the suicide inhibition reaction has potential applications. These potential applications involve manipulating RNA sequences *in vitro* and *in vivo*. It was of interest then to analyze the
sequence dependence (and increase our understanding of this key molecular recognition interaction) for the *P. carinii* suicide inhibition reaction. Since the cyclization and reverse cyclization reactions occur concurrently with self-splicing and suicide inhibition, the results from the reverse cyclization reaction were also analyzed.

A combinatorial approach, using randomized 5’ exon mimics, allowed us to assess the sequence specificity of the *P. carinii* IGS. The results show that certain 5’ exon mimic chimeras, other than that complementary to the IGS, can be substrates in the suicide inhibition and the reverse cyclization reactions. The distribution of substrate sequences is not random, as only certain mismatches are permissible at each of the five randomized 5’ exon positions. Exon sequences that maintain Watson-Crick or wobble base pairs with the IGS, especially within the four 3’ nucleobases of the 5’ exon, are the most effective substrates in these reactions. The sequence specificity of the suicide inhibition reaction is MgCl$_2$ dependent, with a more diverse array of acceptable substrate sequences at higher MgCl$_2$ concentrations. Our results also indicate that the reverse cyclization reaction is substantially more sequence specific than the suicide inhibition reaction. Furthermore, we report that the reverse cyclization products stem from multiple cyclized intron intermediates, which consequently lead to multiple reverse cyclization products.

Using the *P. carinii* intron and a ribozyme derived from this intron (42, 49), the binding properties of a representative group of synthetic 5’ exon mimic chimeras were quantified. In these studies, we demonstrate the importance of a base pair (position –4 of the P1 helix) for tertiary docking of the P1 helix into the catalytic core of the intron. Finally, data shows that the IGS itself can be altered to target a diverse variety of substrate sequences. Such alterations, however, even though resulting in fully complementary P1 helices, can significantly affect the thermodynamics of P1 helix base pairing and tertiary docking. This combinatorial approach therefore has proven useful for gaining a deeper understanding of the fundamental molecular recognition properties of IGS-mediated reactions.
Materials and Methods

Synthesis and Purification of Nucleic Acids

The oligonucleotide \( \text{d(GCTCGTCGACAACGGCTCATGAC)rU} \) [named Pr2-d(ATGAC)rU] was synthesized on an Applied Biosystems 380B DNA/RNA synthesizer and purified by trityl-on reverse phase column chromatography (Sep-Pak, Milford, MA) as described (17). The oligonucleotides \( \text{d(GCTCACTCATTAGGCACC)} \) [named Pr1], \( \text{d(GCTCGTCGACAACGGGCTC)} \) [named Pr2], \( \text{d(GCTCGTCGACAACGGGCTC\text{NNNNN})rU} \) [named Pr2-d(NNNNN)rU], \( \text{d(CAACATTGTACTATACCCAGGGCT)} \) [named Pr3], \( \text{d(GACAGGTGTGCAATCTGAAT)} \) [named Pr4] and all the chimeric hexamers were synthesized, deblocked, and desalted by Integrated DNA Technologies (Coralville, IA). N represents machine randomization of all four bases at the designated positions. The all RNA oligonucleotides were synthesized by Dharmacon Research, Inc. (Lafayette, CO). All hexamers (RNAs and chimeras) were purified by thin-layer chromatography as described (17). Designated oligonucleotides were 5’ end radiolabeled and purified by polyacrylamide gel electrophoresis (PAGE) as previously described (17). Oligonucleotide concentrations were calculated based on UV-absorption measurements using a Beckman DU 650 UV-Spectrophotometer (Beckman Coulter, Inc., Fullerton, CA).

The ribozyme rPC, which was used in the binding assays, was synthesized by run-off transcription of \( \text{XbaI} \) linearized P-8/4x plasmid (17). The group I intron transcript rP-h, which was used for the reverse cyclization, cyclization, initial suicide inhibition, and the kinetic assays was synthesized by run-off transcription of \( \text{HindIII} \) linearized P-h plasmid (17). This produces an RNA transcript with an 82 base 5’ exon, 344 base intron, and 27 base 3’ exon. The group I intron transcript rP-V, used in the combinatorial suicide inhibition assay, was synthesized by run-off transcription of \( \text{VspI} \) linearized P-h plasmid. This produces an RNA transcript with an 82 base 5’ exon, 344 base intron, and 233 base 3’ exon. The rP-V transcript is an elongated version of rP-h with the 3’ exon extended such that it can serve as a primer-binding region for RT-PCR in the combinatorial suicide
inhibition assay. Note that the 206 base extended region is derived entirely from the P-h plasmid and is not native to the *P. carinii* intron or flanking exons.

For designated studies the IGS of the rPC ribozyme was modified from r(GGUCAU) to either r(GGUGAU) or r(GUGACG). These mutant ribozymes were named rPC-CtoG and rPC-oppo, respectively. Site-directed mutagenesis was performed on the plasmid precursor, P-8/4x, using the following mutagenic primer pairs: for rPC-CtoG, CGACTCAGTATAGGGTGATGAAAGCGGC and GCCGCTTTCACCCTCTATAGTGAGTCG; and for rPC-oppo, CGACTCAGTATAGGGTGACGAAAGCGGC and GCCGCTTTCCGTACCTCTATAGTGAGTCG. Each modified plasmid was generated in a 50 μL reaction consisting of 25 ng P-8/4x plasmid, 2.5 units Pfu DNA polymerase (Stratagene; La Jolla, CA), and 0.5 μM dNTPs in a buffer of 10 mM KCl, 10 mM (NH₄)₂SO₄, 20 mM Tris-HCl (pH 8.8), 2 mM MgSO₄, 0.1% Triton X-100, and 0.1 mg/mL BSA. The mixtures were initially denatured at 95 °C for 30 seconds, then underwent 15 cycles of 95 °C (30 sec), 50 °C (2 min), and 68 °C (6 min) in a PCRExpress Thermal Cycler (Hybaid; Ashford, Middlesex, UK). The parental plasmids were digested with 20 units of *Dpn*I (Gibco BRL, Rockville, MD) in 5.7 μL of the manufacturer’s buffer at 37 °C for 2 hours. A 3 μL aliquot was used to transform *E. coli* DH5α competent cells (Gibco BRL). Individual isolates were grown and the modified plasmids were purified using a QIAdex Spin Miniprep Kit (QIAGEN; Valencia, CA). The plasmids were sequenced for confirmation (ACGT, Inc; Northbrook, IL). The two mutant ribozymes were transcribed as described for rPC (17).

*MgCl₂ Concentration Dependence of Group I Intron Reactions*

Initially, as a test of reactivity, the suicide inhibition and reverse cyclization reactions were conducted as a function of MgCl₂ concentration using the radiolabeled 5’ exon mimic chimeras d(ATGAC)rU and Pr2-d(ATGAC)rU. In the experiments with d(ATGAC)rU, 3 μL of 667 nM r-P-h was preannealed at 55 °C in HXMg buffer [50 mM Hepes (25 mM Na⁺), 135 mM KCl, and X mM MgCl₂ (from 0 to 15) at pH 7.5] for 5 min and slow cooled to 37 °C. A 1 μL solution of 1 nM 5’ end radiolabeled d(ATGAC)rU at
37 °C was added. The final concentration of rP-h in the d(ATGAC)rU assays was 500 nM because the expected dissociation constant for d(ATGAC)rU was 60 to 100 nM (in 15 mM MgCl₂ at 37 °C). The reaction was allowed to proceed at 37 °C for 2 hours. The experiments utilizing Pr2-d(ATGAC)rU were conducted as above, except for using a final concentration of 5 μM rP-h in H4Mg or H15Mg buffer. The concentration of rP-h for the Pr2-d(ATGAC)rU reactions was 10-fold higher than that with d(ATGAC)rU, which was required due to less efficient reactivity with the former. The reactions were quenched by the addition of 4 μL stop buffer (0.1X TBE, 10 M urea, and 3 mM EDTA) and the reactants and products were separated on an 8% polyacrylamide, 7.5 M urea gel. The gel was dried under vacuum and the bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. Note that this reaction produces visible suicide inhibition and reverse cyclization products, whose ratio depends on the solution concentration of MgCl₂. Also note that these reactions were done in the absence of free guanosine monophosphate to enhance suicide inhibition relative to self-splicing (43). The high rate of ribozyme-mediated hydrolysis at the 5’ exon-intron junction, however, still allows a significant amount of self-splicing, and hence formation of the circularization intermediate, to occur (17, 43).

**Combinatorial Assays**

For the suicide inhibition reactions, 1 μM of rP-V precursor was preannealed at 55 °C in 5 μL of either H4Mg or H15Mg buffer for 5 min and slow cooled to 37 °C. A 5 μL solution of either 400 nM Pr2-d(ATGAC)rU or Pr2-d(NNNNN)rU in the appropriate buffer at 37 °C was added and the reaction was allowed to proceed for 1 h. The reverse cyclization and cyclization reactions were conducted essentially the same as the suicide inhibition reactions, except that the rP-h intron transcript was used and the assay was performed only in H15Mg buffer. In these combinatorial assays, the final concentration of rP-V or rP-h was 500 nM and the final concentration of Pr2-d(ATGAC)rU or Pr2-d(NNNNN)rU was 200 nM. A concentration of 200 nM of the randomized oligonucleotides gives approximately 0.2 nM of each particular sequence.
Product Isolation and Identification

The reaction products were first reverse transcribed using primer Pr1 for suicide inhibition, primer Pr3 for reverse cyclization, and primer Pr4 for the cyclization reaction. The binding sites for these primers are shown in Figure 3.1. In these reactions, 2 μL of the final non-radiolabeled reaction mixture was added to 3.2 μL sterilized H₂O, 0.5 μL of a 10 mM stock of all four dNTPs, as well as 0.3 μL of a 3.3 μM stock of the respective RT primer. This was incubated at 65 °C for 5 min and then quick cooled on ice. At this point, 2 μL of 5x reverse-transcription buffer [250 mM Tris-HCl (pH 8.3), 375 mM KCl, and 15 mM MgCl₂], 1 μL of 0.1 M DTT, and 0.5 μL (20 units) RNaseOut (GibcoBRL) was added. This was incubated at 42 °C for 1 minute, after which 0.5 μL (100 units) of Superscript II (GibcoBRL) was added and the reaction incubated at 42 °C for 50 minutes. The reaction was terminated by incubation at 70 °C for 15 minutes and then was quick cooled on ice. The reverse transcription products were then PCR amplified. Briefly, reactions contained PCR buffer [50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100], 0.1 μL of the reverse transcription reaction from above, 200 nM primer Pr2, 200 nM of the respective reverse transcription primer, 1.2 mM MgCl₂, 20 μM dNTP mix, and 1 μL of Taq Polymerase (5 units, GibcoBRL). An initial incubation of 2 minutes at 94 °C was followed by 30 cycles of 94 °C (45 sec), 55 °C (1 min), and 72 °C (1.5 min). The resulting PCR products were separated on a 2% agarose gel. The distinct bands at 150 base pairs (suicide inhibition reaction), 175 base pairs (cyclization reaction), and 200 base pairs (reverse cyclization reaction) were extracted and purified using QIAGen Gel Extraction Kits (Qiagen).

We exploited the natural 3’ A overhangs of the purified PCR products to ligate them into pGEM-T vectors (Promega; Madison, WI). The 4 °C overnight ligation protocol recommended by the manufacturer was followed. The resultant plasmids were transformed into E. coli DH5α cells and plated. Colonies were picked from the plates and grown overnight in LB-ampicillin media. The resultant amplified plasmids were purified using QIAprep Spin Miniprep Kits (Qiagen).

All plasmids were sequenced off the T7 promoter indigenous to the plasmid. Roughly half of the resultant plasmids were sequenced by Davis Sequencing (Davis, California). Otherwise, sequencing reactions (total volume of 6 μL) consisted of 250 ng
plasmid template, 1X sequencing buffer [50 mM Tris-HCl (pH 9.0) and 2 mM MgCl₂], 0.10 fmol 5’end radiolabeled T7 sequencing primer, 0.25 μL Taq Polymerase (1.25 units, GibcoBRL), and one of the following nucleotide mixtures: 300 μM ddATP and 13.3 μM of each of the four dNTPs; 400 μM ddCTP and 13.3 μM of each of the four dNTPs; 33.3 μM ddGTP and 13.3 μM of each of the four dNTPs; or 200 μM ddTTP and 6.6 μM of each of the four dNTPs. The sequencing reactions were run in a PCRExpress thermal cycler. An initial incubation of 2 min at 95 °C was followed by 30 cycles of 95 °C (30 sec), 46 °C (30 sec), and 70 °C (1 min). The reaction was quenched by adding 3 μL of sequencing stop buffer (10 mM EDTA in de-ionized formamide). The samples were heated for 2 min at 70 °C and immediately loaded onto a 6% denaturing polyacrylamide gel. The gels were transferred onto chromatography paper (Whatman 3MM CHR), dried under vacuum at 65 °C, and visualized on a Molecular Dynamics Storm 860 Phosphorimager.

Determination of Binding Constants

The dissociation constant, $K_d$, of representative 5’ exon mimics binding to the rP-h intron was approximated using the rPC ribozyme and the radiolabeled native 5’ exon mimic, r(AUGACU), in competition band-shift gel electrophoresis assays (17, 18). rPC (45 nM) was preannealed at 55 °C for 5 minutes in 5 μL of a solution of H15Mg buffer containing 4.5% glycerol (v/v). The solution was then slowly cooled to 37 °C, at which time approximately 1 nM radiolabeled r(AUGACU) and various concentrations of representative 5’ exon mimic competitors in a total volume of 2.5 μL (in H15Mg buffer at 37 °C) was added. The final concentration of ribozyme was 30 nM and the competitor concentrations varied from 5 to 1500 nM. A ribozyme concentration of 30 nM was used because this will bind essentially all of the r(AUGACU), which has a $K_d$ of 5.2 nM for this ribozyme under these conditions (17). The solutions equilibrated for 90 min. The fraction of 5’ exon mimic bound was partitioned from unbound on a 37 °C, 10% native polyacrylamide gel made with H15Mg buffer. The gel was dried under vacuum and the bands quantified with a Molecular Dynamics Storm 860 Phosphorimager. Dissociation constants were calculated by fitting the data to a mass-action equation describing competitive binding of two ligands to a peptide, as previously described (17, 50).
data reported is the average of at least two independent assays. Quantification of binding in H4Mg buffer using this assay was not accurate because binding was below detection limits ($K_d > 1.5 \, \mu M$).

Optical Melting Curves

The strength of base pairing between the representative 5’ exon mimics and an oligonucleotide mimic of the intron’s native IGS, r(GGUCAU), was analyzed through thermal denaturation experiments on a Beckman DU 650 UV-Spectrophotometer. Melts were conducted using an incident UV wavelength of 280 nM at 100 \, \mu M total strand concentration in H15Mg buffer. Thermodynamic parameters were quantified as previously described (17). The data reported is the average of at least two independent assays.

Kinetics

The observed rate constants for the suicide inhibition and reverse cyclization reactions using the representative 5’ exon mimics were obtained under single-turnover ‘intron excess’ conditions. In these experiments, 15 \, \mu L of an 833 nM solution of rP-h was preannealed at 55 \, ^\circ C in either H4Mg or H15Mg buffer for 5 min and slow cooled to 37 \, ^\circ C. A 10 \, \mu L solution of approximately 1 nM 5’ end radiolabeled 5’ exon mimic in the same buffer at 37 \, ^\circ C was added, bringing the final rP-h concentration in the sample to 500 nM. An rP-h concentration of 500 nM was used to permit hexamers with relatively high $K_d$s to bind the intron. A 2 \, \mu L aliquot was periodically removed and added to 2 \, \mu L stop buffer over typically 120 min. The reactants and products were separated on an 8% polyacrylamide, 7.5 M urea gel. The gel was dried under vacuum and the bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. The observed rate constants for the suicide inhibition and reverse cyclization reactions were calculated by fitting the data to a pseudo first-order rate equation as previously described (17). The data reported is the average of at least two independent assays.
The dissociation constants for IGS-modified ribozymes (rPC-CtoG and rPC-oppo) binding various 5’ exon mimics were determined by direct band-shift electrophoresis assays. These assays were conducted using various radiolabeled 5’ exon mimics; rPC, rPC-CtoG, or rPC-oppo ribozymes at various concentrations (from 50 to 1500 nM); and H15Mg as the binding and electrophoresis buffers. The ribozyme mixtures were preannealed at 55 °C for 5 minutes in 5 μL solutions of H15Mg buffer and 4.5% glycerol (v/v). The solutions were then slow cooled to 37 °C, at which time 2.5 μL of approximately 1 nM radiolabeled 5’ exon mimic (in H15Mg buffer at 37 °C) was added. The assay was completed as described for the competitive binding assays and quantified as described (17). Thermal denaturing experiments were conducted as described above, except modified 5’ IGS sequences [r(GGUGAU) and r(GUGACG)] with their complementary 5’ exon mimics [d(ATCAC)rU and d(CGTCArU)] were analyzed.
Results

Combinatorial Assay Rationale

It was previously demonstrated that the \(N3'\rightarrow P5'\) phosphoramidate 5’ exon mimic \(d(AnTnGnAnCn)rU\), where \(n\) represents a phosphoramidate linkage \((19, 43, 51-53)\), is a suicide inhibitor of the self-splicing reaction \(in vitro\) of a group I intron from the opportunistic pathogen \(P. carinii\) (Figure 3.1, F-G). In the suicide inhibition reaction, the exogenous 5’ exon mimic becomes ligated to the endogenous 3’ exon. This property makes the suicide inhibition reaction ideal for developing a combinatorial assay to systematically analyze the sequence dependence of the 5’ exon-IGS molecular interaction. In this combinatorial assay, a 5’ elongated version of the 5’ exon mimic (as seen in Figure 3.1) acts as a suicide inhibition substrate and the elongated base region is exploited as a primer-binding site for RT-PCR amplification of the suicide inhibitor sequence. The 3’ bases of the suicide inhibitor, which act as the 5’ exon mimic, are randomized, permitting the simultaneous assessment of all possible sequences to be substrates in the suicide inhibition reaction. In addition to the suicide inhibition product, a reverse cyclization product (previously described as the 350 product in reference \((43)\)) is concurrently generated in this reaction \((43, 46, 47, 54)\). Reverse cyclization substrates are also 5’ exon mimics in that they first base pair, at least in part, with the IGS of the intron (Figure 3.1, H). Therefore, this combinatorial assay can also analyze the sequence dependence of the reverse cyclization reaction.

Suicide Inhibition and Reverse Cyclization Assay Development

The 5’ exon mimic, \(d(ATGAC)rU\) \([K_d = 61 \text{ nM at } 37^\circ \text{C in } 15 \text{ mM MgCl}_2]\), binds to the \(P. carinii\) rPC ribozyme nearly as tightly as the \(N3'\rightarrow P5'\) phosphoramidate equivalent, \(d(AnTnGnAnCn)rU\) \([K_d = 16 \text{ nM at } 37^\circ \text{C in } 15 \text{ mM MgCl}_2]\) \((5I)\). The non-modified DNA-RNA chimeras were tested to assess whether they could serve as active substrates in the suicide inhibition and the reverse cyclization reactions. The reactions were run as a function of MgCl\(_2\) concentration using the 5’ exon mimic \(d(ATGAC)rU\), as well as the elongated 5’ exon mimic, \(Pr2-d(ATGAC)rU\) (Figure 3.2). Using \(d(ATGAC)rU\), products of approximately 31 nucleotides in length (the suicide inhibition
product) and approximately 350 nucleotides in length (the reverse cyclization product) were produced, as expected (49). In addition, the MgCl₂ dependence of these reactions is similar to that obtained with the phosphoramidate-modified 5’ exon mimics (43). When the elongated 5’ exon mimic, Pr2-d(ATGAC)rU, was used in the reaction, the product at approximately 31 nucleotides was replaced by a product at approximately 52 nucleotides (Figure 3.2, lanes A and B), as expected. For confirmation, the products of the reactions with the elongated 5’ exon mimic Pr2-d(ATGAC)rU (conducted at 4 mM MgCl₂ for suicide inhibition and 15 mM MgCl₂ for reverse cyclization) were RT-PCR amplified, cloned, and sequenced. The sequences obtained were those expected for the two reactions and are diagrammed in Figure 3.1 (with the randomized regions equal to d(ATGAC)). These results demonstrate that the elongated, non-modified chimeric 5’ exon mimic is a substrate in these reactions.

Sequence Specificity of Suicide Inhibition

The combinatorial suicide inhibition reaction was initially conducted in H4Mg buffer using the P. carinii rP-V intron and the non-radiolabeled, randomized 5’ exon mimic Pr2-d(NNNNN)rU. Suicide inhibition reactions were run in 4 mM MgCl₂ because it is at (43) or near (Figure 3.2) the optimum concentration for suicide inhibition reactivity. The 3’ uridine was not randomized, as it is nearly universally conserved. The resulting mixture of products was RT-PCR amplified, cloned, and the suicide inhibition junctions were sequenced. The results in Table 3.1 show that the native 5’ exon sequence, d(ATGAC)rU, occurs most often (22%), however 56% of the sequences (18 out of 32) obtained are unique. These 5’ exon mimics can effectively compete with the endogenous RNA exon and the exogenous chimeric native exon for the IGS of the intron. Although the sequence-specificity is relatively low for this interaction, the distribution of tolerated bases at individual positions within the exon is not random (Table 3.1, C). The exon positions are numbered according to the intron transcript (17) and are shown in Table 3.1. Position −2 can tolerate a dT, which presumably forms a dT-rG wobble base pair, however, a Watson-Crick base pair is strongly favored. At position −3, dA and dG were found in equal proportions, so a Watson-Crick pair (dA-rU) or a wobble pair (dG-rU) is either strongly favored or required. Position −4 allows a dC and, albeit rarely a dT, thus
strongly favoring a Watson-Crick base pair. Positions –5 and –6 are significantly less specific for Watson-Crick or wobble pairs than the other positions, although they are not random and they do prefer Watson-Crick pairing.

The MgCl₂ Dependence of Suicide Inhibition

The combinatorial assay was performed in 15 mM MgCl₂ to determine if the concentration of magnesium has an effect on the sequence specificity of the suicide inhibition reaction. The results at 15 mM MgCl₂ are similar to results at 4 mM MgCl₂ except there is significantly more sequence diversity (83% unique sequences as compared to 56%) at the higher magnesium concentration. A greater number of mismatches are tolerated at positions –2 through –4, which appear to be the most critical positions for docking of the resultant P1 helices ((19), this work). In addition, the native sequence is found almost half as often at the higher MgCl₂ concentration. Taken together these results indicate that the sequence-specificity of the suicide inhibition reaction decreases at higher MgCl₂ concentration.

Sequence Specificity of Reverse Cyclization

The proportion of substrates that form reverse cyclization products relative to those that form suicide inhibition products increases at higher MgCl₂ concentrations ((43), see also Figure 3.2), perhaps because high salt favors intron cyclization (31, 55). Therefore, the sequence specificity of reverse cyclization was analyzed at 15 mM MgCl₂. The results show that the native sequence occurs 58% of the time (with 42% unique sequences). Apparently, the reverse cyclization reaction at 15 mM MgCl₂ shows a substantially higher preference for the native 5’ exon mimic sequence than the suicide inhibition reaction at either 4 mM (22%) or 15 mM (13%) MgCl₂. The results also show that the 5’ exon mimics ligate to at least 4 different positions at the 5’ end of the intron (compare results of Table 3.2 with Figure 3.1). Evidently, multiple routes of reverse cyclization occur. It is possible that this stems from the prior formation of multiple cyclized intron intermediates.
Identification of Cyclization Intermediates

Amplification of the reverse cyclization reaction mixture with RT-PCR primers Pr2 and Pr4 resulted in a product whose sequence shows the identity of the intron’s cyclization junction. Table 3.3 shows the sequences of the cyclization junctions for 11 product sequences. There are two Gs at the 3’ end of the intron (Figure 3.1, D, shown in white lettering) that can each act as nucleophiles in the cyclization process. In addition, there are at least 3 different 5’ intron positions (Figure 3.1, D, shown in black lettering in dark gray background) that are the acceptor sites for the two possible G nucleophiles. Apparently, there are multiple routes to cyclization, which ultimately lead to a mixture of reverse cyclization products. Similar results arise from a *Tetrahymena* group I intron (42, 46, 55).

Binding of Representative 5’ Exon Mimics to the rPC Ribozyme

Representative 5’ exon mimics (most sequences were found in the combinatorial assays) were chemically synthesized and their binding strengths to the intron were approximated using competitive band-shift electrophoresis assays in H15Mg buffer with the rPC ribozyme. Quantification of binding in H4Mg buffer using this assay was not possible because binding was below detection limits ($K_d > 1.5 \, \mu M$). The competitive binding assays (Figure 3.3, Table 3.4) reveal that the hexameric 5’ exon mimics either bind the ribozyme with a $K_d$ less than 200 nM or they are above the detection limits. Furthermore, addition of the primer binding region to an exon results in a 2-fold reduction in overall binding. The results also show that a wobble base pair at positions –2, –3, or –6 (exon mimics #6, #5, and #3 in Table 3.4) changes the $K_d$ at most by approximately two-fold compared to the complementary sequence (exon mimic #1). Therefore, it appears that wobble base pairs, where possible, are well tolerated in terms of overall binding strength.

A mispair at position –4 of the exon (exon mimic #8) forms a dC-rC pair with the IGS and prohibits binding. From the combinatorial assay, in every exon where a mutation occurs at position –4, there is a T to G mutation at position –5. As seen in Table 3.4, this double mutant (exon mimic #7) binds the ribozyme reasonably tightly ($K_d = 137 \, nM$),
which is not due to the single G mutation (exon mimic #4) at position –5 ($K_d = 158$ nM).

Apparently, poor binding due to mismatch formation at position –4 can be rescued by the addition of a specific mutation at position –5.

**Base Pairing of Representative 5’ Exon Mimics with the IGS, $r$(GGUCAU)**

The stability of the representative 5’ exon mimics base pairing with the IGS of the *P. carinii* ribozyme was approximated (51) by thermal denaturation analysis with the IGS mimic, $r$(GGUCAU), in H15Mg buffer. The results of at least two experimental trials (Table 3.4) indicate that the thermodynamic stabilities ($-\Delta G^{\circ}_{37,\text{BP}}$) of the resultant hybrid helices are very similar, although significantly less than the all RNA helix. The most thermodynamically stable hybrid helix is formed with $d$(ATGAC)$r$U, which is the complementary 5’ exon sequence to the IGS. Helices containing one or more mismatches are only at most 0.3 kcal/mol less stable than the native exon hybrid helix, which translates into a two-fold decrease in base pairing strength. Therefore, none of the mismatched helices studied completely disrupts the ability of the exon mimics to base pair with the IGS. This indicates that exon mimics that do not bind to the ribozyme do so because of an inability to form critical tertiary interactions with the ribozyme.

**Kinetics of Representative 5’ Exon Mimics Reacting with the rP-h Intron**

The combinatorial assay exploits the ability of each of the 5’ exon mimics to not just bind, but also to react with the intron. Therefore, the observed rate constant, $k_{obs}$, of each representative oligonucleotide forming suicide inhibition and reverse cyclization products at 4 mM and 15 mM MgCl$_2$ was measured under single turnover ‘intron-excess’ conditions. Figure 3.4 shows a typical gel and graph using the 5’ exon mimic $d$(ATGAC)$r$U. Note that each individual assay produces both suicide inhibition and reverse cyclization products. Thus, the extent of reaction for these two products are not altogether independent of each other. The results are shown in Table 3.5. The observed rate constants are less for the chimeras than the all RNA mimic, indicating that the chimeras are less effective substrates in these reactions. For comparison, the $k_{obs}$ for the suicide inhibition and reverse cyclization reactions using $r$(AUGACU) are 5 to 6-fold less.
than that for the self-splicing reaction under the same conditions (17). For the suicide inhibition reaction using the chimeras, the magnitude of \( k_{\text{obs}} \) changes less than 3-fold as a function of either sequence or MgCl\(_2\) concentration, at least for those assays that were quantifiable. For the reverse cyclization reaction using the chimeras, the magnitude of \( k_{\text{obs}} \) changes less than 3-fold as a function of sequence, but changes up to 7-fold as a function of MgCl\(_2\) concentration. Therefore for the chimeras, increasing the MgCl\(_2\) concentration differentially enhances the reverse cyclization reaction relative to the suicide inhibition reaction. Note that the presence of the elongated P2 region disrupts reactivity under the conditions used. This is not problematic in the combinatorial assays, however, as a higher exon mimic concentration was used and the final products were ultimately RT-PCR amplified.

**Analysis of the Sequence Requirements at the –4 Position of the 5’ Exon**

Our results suggest that there is a critical requirement for a base pair at position –4. To directly test whether a G(exon)-C(IGS) pair at exon position –4 is required [in contrast to a C(exon)-G(IGS) pair], a ribozyme (rPC-CtoG) was synthesized with a G in position –4 of the IGS (instead of a C) and its binding to the exon mimics d(ATGAC)rU and d(AT\(\overline{C}\)AC)rU was analyzed (Table 3.6). Apparently, a C(exon)-G(IGS) base pair at position –4 does permit substrate binding, although binding strength decreases two-fold relative to the native G(exon)-C(IGS) base pair (274 nM versus 151 nM). This decrease is entirely at the level of base pairing, and thus does not result from a loss of tertiary interactions. Therefore, although the IGS can be changed to alter the sequence-specificity of the ribozyme, certain types of base pairs at certain positions are thermodynamically favored over other base pairs. In addition, the 5’ exon mimic d(AT\(\overline{C}\)AG)rU does not bind to the ribozyme with the native IGS sequence (which forms a dC-rC mismatch at position –4) and the native 5’ exon mimic d(ATGAC)rU does not bind the rPC-CtoG ribozyme (which forms a dG-rG mismatch at position –4). Apparently in these two later cases the P1 helix loses the ability to dock into the catalytic core of the ribozyme, most likely through disruption of base pair specific tertiary interactions.

As a preliminary test to see how much the IGS can be altered and still retain complementary exon mimic binding, the rPC-oppo ribozyme was synthesized. In this
ribozyme, every purine of a C-G pair in the IGS is replaced by a pyrimidine of a U-A pair, and vice-versa (except for the conserved terminal U-G pair). Results in Table 3.6 show that the exon mimic d(CGTCArU) does bind to this ribozyme, although the strength of binding is reduced approximately eight-fold (1130 nM versus 151 nM) compared with the native sequences. Interestingly, this decrease is entirely due to a loss of tertiary interactions, as the base pairing stability of the resultant 5’ exon-IGS helix increases (Table 3.4). This exon mimic does not bind the native 5’ IGS of rPC, nor did the native 5’ exon mimic bind the modified rPC-oppo ribozyme. Therefore, although changing the 5’ IGS can result in a change in the sequence that the ribozyme targets, this change can result in a drastic reduction in the binding strength of the exon mimic-catalytic RNA interaction (even when using complementary 5’ exons). Also note that these results show that a dT-rA base pair at position –4 is tolerated.
Discussion

We show that 5’ exon mimic chimeras can contain normal phosphodiester backbones and still be substrates in the suicide inhibition and reverse cyclization reactions. These non-modified chimeras, however, are less effective substrates in the suicide inhibition and reverse cyclization reaction, as compared to the N3’→P5’ phosphoramidates previously used (19, 43). This is not unexpected, as the phosphoramidate-containing duplexes are significantly more thermodynamically stable than the unmodified chimera-containing duplexes (53). In addition, the modified duplexes are expected to more closely resemble A-form RNA duplexes as compared with the unmodified duplexes (53), thus maintaining helix geometry more amenable to tertiary interactions.

The combinatorial approach developed in this work identifies substrate sequences that are able to effectively compete with a pool of all available sequences for being a substrate in the suicide inhibition and reverse cyclization reactions. The resultant sequences, therefore, are not a comprehensive list of all possible substrates, but rather those most able to compete for the 5’ IGS binding site for subsequent reactivity. Note, however, that this assay can detect very poor substrates as well, albeit rarely, as long as the rate of substrate dissociation is not substantially faster than the rate of chemistry.

For those exon mimics that were obtained in the combinatorial assay, we could find no correlation between the number of times the exon mimics occur and their observed rate constants. This indicates that the rate of chemistry is likely not the limiting factor in determining how often the 5’ exon mimics occur (for those that are obtained in the assay). The number of occurrences of each exon mimic does appear to correlate with the binding strength of the mimics, although there are exceptions. These exceptions likely stem from limitations of the binding assay itself. Analyzing exon mimic binding using the ribozyme approximates the actual binding that occurs when the exon mimic has to compete with the endogenous 5’ exon in the combinatorial assay. Note, however, that we do have one example of a representative exon mimic, d(AGCAC)rU), that binds the ribozyme but does not react in the assay. Therefore, it is expected that either the rate of
chemistry or the strength of substrate binding can be a limiting factor in whether a particular exon mimic is detected in the combinatorial assay.

Sequence Specificity of Suicide Inhibition

As expected, the native 5’ exon sequence, d(ATGAC)rU, occurred most often in the suicide inhibition reaction at 4 and 15 mM MgCl$_2$, indicating that the native sequence is the most effective suicide inhibitor and that the reactions were run under reasonably stringent conditions. However, a number of other sequences are also permitted. In fact, all five randomized exon positions tolerate nucleobases not present in the native 5’ exon. In general, the requirement for even moderate sequence specificity decreases as you get further away from the reaction site (position –1 of the 5’ exon), with the specificity decreasing considerably at positions –5 and –6. In addition, it appears that a wobble pair is an acceptable substitute for Watson-Crick base pairs at positions –2 and –3. The binding data (Table 3.4) and reactivity data (Table 3.5) using representative hexamers show that the 5’ exon sequences obtained in the combinatorial assay are indeed specific substrates in the suicide inhibition reaction. Therefore, similar to the specificity of trans-splicing and reverse splicing reactions for *Tetrahymena* ribozymes (56-60), the sequence specificity of the suicide inhibition reaction for the *P. carinii* intron is not high.

The physiological environment within which the group I intron-catalyzed self-splicing reaction takes place is largely unknown. Reports indicate that the amount of free magnesium in a cell is at or below 1 mM (61, 62). We have previously shown that this is not enough magnesium for the intron to fold and catalyze the self-splicing reaction *in vitro* (17, 43). There is evidence, however, that proteins can aid in the folding of introns (63), which diminishes the need for a high salt environment in the cell. In our assays, a greater diversity of sequences (and of positional mismatches) is permitted at higher salt. One possible explanation for this is that the differential stability of the various 5’ exon mimic-IGS helices decreases at increased MgCl$_2$ concentration. Extrapolating these results to physiological concentrations of MgCl$_2$ suggests that a low salt environment (perhaps containing RNA folding chaperones) would naturally facilitate IGS-mediated reactions with higher sequence specificity than reported here.
It has been noted that one possible way to enhance specificity is by weakening the strength of the 5’ exon-IGS interaction, for example by mismatch incorporation, such that the discrimination threshold increases between similar sequences \((54, 64)\). As we see here, mismatch incorporation does not always lead to weaker binding [compare \(d(\text{ATGAT})\text{rU}\) and \(d(\text{ATGAC})\text{rU}\) in Table 3.4]. Nevertheless, the \(\text{MgCl}_2\) concentration dependent studies do appear to confirm that weaker binding (at 4 mM relative to 15 mM \(\text{MgCl}_2\)) results in increased specificity.

Sequence Specificity of Reverse Cyclization

Using the combinatorial approach, we have shown that the reverse cyclization product is actually a combination of products. The sequence dependence of the reverse cyclization reaction was of interest because it was previously shown with a *Tetrahymena* intron system that reverse cyclization substrates, which are also 5’ exon mimics, can be as short as a dinucleotide, indicating that the 5’ exon mimics might base pair with only relatively short regions of the IGS \((32, 46, 47)\). The results (Table 3.2) for the *P. carinii* intron show there is a functional advantage to having all or most of each reverse cyclization substrate complementary to the corresponding IGS region. In addition, the sequence specificity for the suicide inhibition and the reverse cyclization reactions are substantially dissimilar. For example, 58% of the substrates in the reverse cyclization reaction are the native 5’ exon sequence, while in the suicide inhibition reaction just 13% of the substrates are the native sequence (each at 15 mM \(\text{MgCl}_2\)). In addition, unlike suicide inhibition substrates, reverse cyclization substrates do not require a dG at either position –4 or –5 of the exon. One possible explanation for this is that complementary base pairing at position –4 (and other positions) is beneficial in the reverse cyclization reaction because it enhances base pairing stability, while in the suicide inhibition reaction this position is additionally involved in critical tertiary interactions. Apparently, factors that govern the molecular recognition of the IGS for the 5’ exon in the reverse cyclization and suicide inhibition reactions are not identical, indicating a dynamic use of the IGS between the two reactions.
In all assays conducted, the permissibility of non-pairing mismatches increases the farther away from the reactivity center (exon position –1) that the mismatch occurs, with a marked decrease after position –4. This supports previous experimental data that shows nearly all of the tertiary stability of this molecular interaction occurs within the four 3’ bases of the 5’ exon-IGS interaction (19). Thus, the occurrence of base pairs at positions –5 and –6 may serve to simply increase the stability of the P1 helix in preparation for subsequent tertiary interactions during the docking stage(s) of folding. In addition, replacing Watson-Crick base pairs with wobble base pairs at positions –2, –3, and –6 does not significantly interfere with the overall binding of the mimic to the catalytic core of the intron. This was not anticipated because wobble pairs have some different accessible functional groups for tertiary interactions than Watson-Crick pairs (30). Apparently, these structural differences at these positions do not alter the ability of the resultant P1 helices to form tertiary interactions with the catalytic core of this intron.

The decrease in occurrence of dG at exon position –4 in the suicide inhibition assay as the MgCl₂ concentration increases directly correlates with an increase in the occurrence of dG at position –5. The binding assays, where a dG at position –5 rescues lost binding that occurs in the absence of dG at position –4, shows that this is not a random trend. In addition, P1 helices with dC-rC or dG-rG mismatches at position –4 are not able to stably dock into the catalytic core of the ribozyme. This suggests that there is an absolute requirement for a base pair at the –4 exon-IGS position. In cases where positional rescue of binding occurs, it is likely that the dG at position –5 base pairs with the corresponding position –4 of the IGS, leaving the nucleobase at exon position –4 as a single nucleotide bulge in the P1 helix (Figure 3.5). The overall stability that results from having a base pair at position –4 appears to be strong enough to dominate the disruption of base pairs at exon positions –3 and –5. It is interesting to note that the corresponding position in an Azoarcus intron likely participates in a structurally important A-minor tertiary motif (65). Perhaps in support of this, A-minor motifs appear to thermodynamically favor canonical base pairs over base pair mismatches (66), which is the same trend we see at the –4 position.
Interestingly, the exon mimic d(AGCAC)rU does not react in the suicide inhibition assays (exon mimic #7 in Table 3.5), but does bind the ribozyme (exon mimic #7 in Table 3.4). Apparently, binding alone is not enough to specify reactivity, as local and global structure of the P1 helix also plays a role in determining reactivity (27). This is further supported with the 5’ exon mimic d(ATGAT)rU (exon mimic #6), which binds very tightly to the ribozyme, but its extent of reaction is relatively low.

We show that IGS-modified P. carinii ribozymes do favor binding to 5’ exon mimics that are complementary to their particular IGS. However, the strength of the overall binding of the exon to the ribozyme can change considerably even though the 5’ exon-IGS helices are complementary at exon positions –2 through –6. In the case of the rPC-oppo ribozyme, this change in overall binding stability is primarily due to a reduction in tertiary stability (approximately 1.5 kcal/mol). This indicates that the ribozyme likely forms tertiary interactions with only specific base pairs at certain P1 helix positions, in contrast to that reported with the Tetrahymena ribozyme (27, 67). One possible source of base pair specific tertiary interactions is Type I A-minor motifs (68), although there is no direct evidence that such a motif occurs here.

It does not appear that the large sequence diversity of 5’ exons obtained in the combinatorial suicide inhibition and reverse cyclization assays can be explained by the 5’ exons misaligning (or slipping) on the IGS (55, 69), although in a few cases it could be occurring. The different cyclization products obtained (and hence different sites of reverse cyclization), however, can be explained by different regions of the 5’ end of the spliced intron base pairing with the IGS (Figure 3.1). Note that the cyclization reaction is similar to the self-splicing reaction in that the reaction substrates bind to the IGS (70). The forward and reverse cyclization data reveal that in three cases (Figure 3.6) the cyclization sites in the exons contain three contiguous pyrimidines that possibly pair with the G-rich end of the IGS. In these cases the cyclization junction is likely to be situated along the IGS at the same position as the self-splicing junction. The positioning of the exon cyclization junction along the IGS is not obvious for the fourth case, and so may not utilize the same positioning as the self-splicing junction. The data also reveals that substrate slipping occurs in the P9.0 helix, as either of the last two guanosines can act as cyclization nucleophiles. This is possible because either of these guanosines can bind the
G-binding site, while upstream bases are able to form a P9.0 helix. Implicit then is the possibility that either guanosine can act as the terminal guanosine of the intron during the second step of self-splicing under the reaction conditions used here.

All of the above results show that the combinatorial assay developed in this report has proven useful for gaining a more fundamental understanding of the molecular recognition properties of IGS-mediated reactions.

**Implications**

Previous reports have shown that oligonucleotides that mimic the 5’ exon of the *P. carinii* rRNA group I intron are suicide inhibitors of the self-splicing reaction *in vitro* (43). Such inhibitors are novel in that they specifically exploit the tertiary interactions of the structured RNA. This was coined Binding Enhancement by Tertiary Interactions (BETI) (45, 58). This strategy was developed to potentially alleviate the problem of non-specific binding, which appears to be a factor in using conventional antisense approaches (64, 71, 72). As a general rule, these results show that BETI (and more specifically suicide inhibition) likely will suffer from the same problem, although perhaps to a lesser degree: mismatches can be tolerated when short helices are stabilized by tertiary interactions (unless the mismatches severely disrupt tertiary interactions).

RNA targeting strategies such as BETI (45, 58) and mRNA repair (5, 6) rely on tertiary interaction formation between the catalytic RNA enzyme and its substrate (after base pairing). This report shows that the strength of these tertiary interactions can depend on the presence of certain base pairs at certain positions along the catalytic RNA-substrate helix, which could effectively limit the choice of targets that the catalytic RNA can act on. This could, however, also increase the specificity of the catalytic RNA, as the strength of the catalytic RNAs binding to similar, but undesired targets could be decreased.
### Table 3.1 Suicide Inhibitors of the Self-Splicing Reaction

#### A

<table>
<thead>
<tr>
<th>Number of Occurrences at each MgCl₂ Concentration</th>
<th>4 mM MgCl₂</th>
<th>15 mM MgCl₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>-6</td>
<td>1</td>
<td>4 (15%)</td>
</tr>
<tr>
<td>-5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1阳性  (22%)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2阳性  (13%)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3阳性  (4%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4阳性  (16%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7阳性  (32%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30 Total</td>
<td>30 Total</td>
<td></td>
</tr>
</tbody>
</table>

(B) Suicide inhibitors obtained from the combinatorial assays (written 5'→3'). The number of occurrences for each sequence is indicated in the left column for reactions at 4 mM MgCl₂ and in the right column for reaction at 15 mM MgCl₂. Note that position –1 is RNA and was not randomized, while positions –2 through –6 are DNA and were randomized. The numbering system for the inhibitor positions is shown above the sequences. Note that the inhibitors are 5' exon mimics. (B) Sequence of the IGS of the rP-h intron (and the rPC ribozyme). (C) The percent occurrence of the four nucleobases at each position. Note that the percent occurrences are different for the two MgCl₂ concentrations.

#### B

<table>
<thead>
<tr>
<th>3'</th>
<th>U</th>
<th>A</th>
<th>C</th>
<th>U</th>
<th>G</th>
<th>5'</th>
</tr>
</thead>
</table>

#### C

Percent Occurrence in 4 mM MgCl₂

<table>
<thead>
<tr>
<th>A⁷</th>
<th>T₂</th>
<th>G₄</th>
<th>A₉</th>
<th>C₄</th>
<th>rU₁₀₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₄</td>
<td>A₁</td>
<td>C₅</td>
<td>G₁</td>
<td>T₁</td>
<td></td>
</tr>
<tr>
<td>C₁</td>
<td>C₀</td>
<td>T₀</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Percent Occurrence in 15 mM MgCl₂

<table>
<thead>
<tr>
<th>A₄₀</th>
<th>T₀</th>
<th>Gₗ</th>
<th>A₃₅</th>
<th>C₅₀</th>
<th>rU₁₀₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₇</td>
<td>G₄₀</td>
<td>T₃</td>
<td>G₇₇</td>
<td>T₁₉</td>
<td></td>
</tr>
<tr>
<td>G₁₅</td>
<td>A₄₀</td>
<td>C₇</td>
<td>T₇</td>
<td>A₃</td>
<td></td>
</tr>
<tr>
<td>C₁₀</td>
<td>A₃</td>
<td>C₃</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.2 Reverse Cyclization Products

<table>
<thead>
<tr>
<th>Exon Substrate</th>
<th>Intron Splice Site</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>-6 -5 -4 -3 -2 -1</td>
<td></td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CCUUCUGAG--------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CCUUCUGAG--------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CCUUCUGAG--------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CCUUCUGAG--------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CCUUCUGAG--------</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(A T G A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
<tr>
<td>d(C A G G C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(C A G G C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(C A G G C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A A G A T) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A A G A T) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A A G A T) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(T G T A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(T G T A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(T G T A C) rU-</td>
<td>-UCUGAG-----------</td>
</tr>
<tr>
<td>d(A T A A C) rU-</td>
<td>-CACCUCUCUGAG-----</td>
</tr>
</tbody>
</table>

(B) Sequence of the IGS of the rP-h intron (and the rPC ribozyme). (C) The percent occurrence of each nucleobase at each exon position.

(A) Reverse cyclization products (written 5' → 3') of the combinatorial assay at 15 mM MgCl$_2$ (24 total). The dashed lines represent a single phosphodiester bond. The solid lines represent nucleobase continuation within the intron. The numbering system for exon substrate positions is indicated above the sequences. Note that position –1 is RNA and was not randomized, while positions –2 through –6 are DNA and were randomized.
Table 3.3 Cyclization of the rP-h Intron

<table>
<thead>
<tr>
<th>3’ end of intron ligated to 5’ end of intron</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>G - - - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>G - - - - - - - - - UCUGAG</td>
</tr>
<tr>
<td>GUG- - - - - - - - - UGAG</td>
</tr>
<tr>
<td>G - - - - - - - - CACCUUCUGAG</td>
</tr>
</tbody>
</table>

Shown are the cyclization junctions, which encompass the 3’ end and the 5’ end of the intron ligated together. The sequences are written in the 5’ → 3’ direction. The dashed lines represent a single phosphodiester bond. The solid lines represent nucleobase continuation within the intron.
Table 3.4 Thermodynamic Parameters for Representative Hexamers Binding to rPC and to r(GGUCAU) in H15Mg Buffer

<table>
<thead>
<tr>
<th>Exon Mimic</th>
<th>Binding to rPC</th>
<th>Binding to r(GGUCAU)</th>
<th>Tertiary Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(K_d,\text{TOTA}L)</td>
<td>(-\Delta G_{37,\text{TOTA}L})</td>
<td>(K_d,\text{BP})</td>
</tr>
<tr>
<td>t(AUGACU)</td>
<td>1.79 ± 0.27</td>
<td>12.41 ± 0.10</td>
<td>0.51 ± 0.06</td>
</tr>
<tr>
<td>#1 d(ATGAC)rU*</td>
<td>(5.21 ± 1.4)-overlay</td>
<td>(11.75 ± 0.19)</td>
<td>(0.32 ± 0.04)</td>
</tr>
<tr>
<td>#2 Pr2-d(ATGAC)rU*</td>
<td>62.6 ± 10.3</td>
<td>10.22 ± 0.11</td>
<td>2.35 ± 0.69</td>
</tr>
<tr>
<td>#3 d(GTGAC)rU*</td>
<td>(61.1 ± 7.0)-overlay</td>
<td>(10.2 ± 0.08)</td>
<td>(4.072 ± 1.208)</td>
</tr>
<tr>
<td>#4 d(GGGAC)rU*</td>
<td>139.5 ± 0.95</td>
<td>9.72 ± 0.01</td>
<td>NQ(^a)</td>
</tr>
<tr>
<td>#5 d(ATGAC)rU*</td>
<td>158.3 ± 26.9</td>
<td>9.65 ± 0.11</td>
<td>3.65 ± 1.73</td>
</tr>
<tr>
<td>#6 d(ATGAC)rU*</td>
<td>129.8 ± 23.5</td>
<td>9.77 ± 0.12</td>
<td>3.93 ± 0.62</td>
</tr>
<tr>
<td>#7 d(GGCAC)rU*</td>
<td>38.8 ± 11.5</td>
<td>10.5 ± 0.02</td>
<td>3.66 ± 0.58</td>
</tr>
<tr>
<td>#8 d(ATCAC)rU*</td>
<td>-</td>
<td>-</td>
<td>3.61 ± 1.02</td>
</tr>
<tr>
<td>#9 d(CGCAC)rU*</td>
<td>-</td>
<td>-</td>
<td>3.11 ± 0.63</td>
</tr>
<tr>
<td>#10 d(CGCTA)rU*</td>
<td>-</td>
<td>-</td>
<td>3.90 ± 1.13</td>
</tr>
</tbody>
</table>

\(^a\)H15Mg buffer consists of 50 mM Hepes (25 mM Na\(^{+}\)), 15 mM MgCl\(_2\), and 135 mM KCl at pH 7.5. Nucleobases that are bold and underlined represent deviations from the native 5' exon sequence d(ATGAC)rU. An asterisk (*) indicates oligonucleotides that occur in the combinatorial assay. \(^b\)\(K_d,\text{TOTA}L\) was measured by a competition band-shift electrophoresis assay and \(-\Delta G_{37,\text{BP}}\) was measured by thermal denaturation analysis. The error is the standard deviation of the measurements. \(^c\)Calculated from \(-\Delta G_{37} = RT\ln(K_d)\) where R = 0.001987 kcal mol\(^{-1}\) K\(^{-1}\) and T=310 K, using more significant digits than listed in this table. \(^d\)BETI represents Binding Enhancement by Tertiary Interactions. \(^e\)Free energy increment from tertiary interactions calculated from the difference in \(-\Delta G_{37}\) values \[\(\Delta\Delta G_{37,\text{BETI}} = \Delta\Delta G_{37,\text{TOTA}L} - \Delta\Delta G_{37,\text{BP}}\)\]. The \(-\Delta \Delta G_{37,\text{BETI}}\) error was calculated from the square root of the sum of the squares of each individual \(-\Delta G_{37}\) error. \(^f\)The \(K^2\) values were calculated by dividing \(K_d,\text{BP}\) by \(K_d,\text{TOTA}L\), using \(K_d\) values containing more significant digits than those listed in this table. \(^g\)Values in parentheses are from reference 6. \(^h\)Values were not quantifiable (NQ) due to multiphasic transitions. \(^i\)ND represents values that are not determinable. \(^j\)A dash (-) indicates no measurable binding (\(K_d > 1.5\) μM).
Table 3.5 Kinetics of Representative Exons in the Suicide Inhibition and Reverse Cyclization Reactions as a Function of MgCl$_2$ Concentration

<table>
<thead>
<tr>
<th>Exon Mimics</th>
<th>Suicide Inhibition</th>
<th>Reverse Cyclization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k_{obs}$ (min$^{-1}$)$^b$</td>
<td>Relative Extent of Reaction</td>
</tr>
<tr>
<td></td>
<td>4 mM MgCl$_2$</td>
<td>15 mM MgCl$_2$</td>
</tr>
<tr>
<td>r(AUGACU)</td>
<td>0.081 ± 0.031</td>
<td>100.0%</td>
</tr>
<tr>
<td>#1 d(ATGAC)rU*</td>
<td>0.013 ± 0.003</td>
<td>63.0%</td>
</tr>
<tr>
<td>#2 Pr2-d(ATGAC)rU*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#3 d(GTGAC)rU*</td>
<td>0.019 ± 0.003</td>
<td>37.5%</td>
</tr>
<tr>
<td>#4 d(AGGAC)rU*</td>
<td>0.012 ± 0.002</td>
<td>9.4%</td>
</tr>
<tr>
<td>#5 d(ATGCC)rU*</td>
<td>0.018 ± 0.001</td>
<td>62.5%</td>
</tr>
<tr>
<td>#6 d(ATGAT)rU</td>
<td>0.031 ± 0.005</td>
<td>25.0%</td>
</tr>
<tr>
<td>#7 d(GACAC)rU</td>
<td>NQ</td>
<td>&lt;3%</td>
</tr>
<tr>
<td>#8 d(TACAC)rU</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>#9 d(CGCCT)rU</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$^a$Buffers consist of 50 mM Hepes (25 mM Na$^+$), 4 mM or 15 mM MgCl$_2$, and 135 mM KCl at pH 7.5. Nucleobases that are bold and underlined represent deviations from the native 5' exon sequence d(ATGAC)rU. An asterisk (*) indicates oligonucleotides that occur in the combinatorial assay. $^b$The error is the standard deviation of the measurements. $^c$The extent of reaction for the suicide inhibition reactions are reported relative to that obtained with r(AUGACU) at 4 mM MgCl$_2$ (actual extent of reaction = 32%). Similarly, the extent of reaction for the reverse cyclization reactions are reported relative to that obtained with r(AUGACU) at 15 mM MgCl$_2$ (actual extent of reaction = 17%). The relative extent of reaction is calculated by dividing the actual extent of reaction for each oligonucleotide by that for r(AUGACU) and multiplying by 100. $^d$A dash (-) indicates no measurable activity. $^e$NQ represents measurements that are not quantifiable due to inconsistent results with low product formation.
Table 3.6 Thermodynamic Parameters for Representative Exon Mimics Binding to Ribozymes and to IGS Mimics in

<table>
<thead>
<tr>
<th>Exon Mimic</th>
<th>Binding to Ribozyme</th>
<th>Binding to IGS Mimic</th>
<th>Tertiary Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$K_d$, TOTAL (nM)$^b$</td>
<td>$-\Delta G^{37, \text{TOTAL}}$ (kcal/mol)$^c$</td>
<td>$K_d$, BP (mM)$^c$</td>
</tr>
<tr>
<td>d(ATGAC)rU$^g$</td>
<td>151.3 ± 10</td>
<td>9.67 ± 0.05</td>
<td>2.24 ± 0.66</td>
</tr>
<tr>
<td>d(ATCAC)rU$^b$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>d(CGTCArU)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rPC-CtoG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d(ATGAC)rU$^g$</td>
<td>274 ± 25.3</td>
<td>9.31 ± 0.06</td>
<td>3.69 ± 1.47</td>
</tr>
<tr>
<td>d(ATCAC)rU$^b$</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>d(CGTCArU)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>rPC-oppo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d(ATGAC)rU$^g$</td>
<td>1130 ± 47</td>
<td>8.43 ± 0.03</td>
<td>1.35 ± 0.20</td>
</tr>
</tbody>
</table>

$^a$H15Mg buffer consists of 50 mM Hepes (25 mM Na$^+$), 15 mM MgCl$_2$, and 135 mM KCl at pH 7.5. Nucleobases that are bold and underlined represent deviations from the native 5' exon sequence d(ATGAC)rU. $^b$It was measured by a direct band-shift electrophoresis assay with r(AUGACU) and $-\Delta G^{37, \text{BP}}$ was measured by thermal denaturation analysis. The error is the standard deviation of the measurements. $^c$Calculated from $-\Delta G^{37} = RT\ln(K_d)$ where $R = 0.001987$ kcal mol$^{-1}$ K$^{-1}$ and $T=310$ K, using more significant digits than listed in this table. $^d$BETI represents Binding Enhancement by Tertiary Interactions. $^e$Free energy increment from tertiary interactions calculated from the difference in $-\Delta G^{37}$ values [$(-\Delta G^{37, \text{TOTAL}}) - (-\Delta G^{37, \text{BP}})$]. The $-\Delta G^{37, \text{BETI}}$ error was calculated from the square root of the sum of the squares of each individual $-\Delta G^{37}$ error. $^f$The $K^2_f$ values were calculated by dividing $K_d$, BP by $K_d$, TOTAL, using $K_d$ values containing more significant digits than those listed in this table. $^g$Dissociation constants ($K_d$s) obtained through direct band-shift assays are typically 2.5 times larger than those obtained through the competitive band-shift assay (17, 73). $^h$A dash (-) indicates no measurable binding ($K_d > 1.5$ μM).
Reverse Cyclization Product

Suicide Inhibition Reaction

A Binding

Self-Splicing Reaction

B G Addition

C Exon Ligation

G Cyclization Reaction

D Nucleophilic Attack

Reverse Cyclization Reaction

E Cyclization

F Binding

Suicide Inhibition Product

G Suicide Inhibition

H Binding

Reversal Cyclization Product

Suicide Inhibition Product

Pr2 Pr1

Pr2

Pr3

Pr4

Pr2*
Figure 3.1 Schematic Diagram of Self-Splicing and Cyclization (Steps A → E), Suicide Inhibition (Steps F → G), and Reverse Cyclization (Steps H → I). Addition of an exogenous exon substrate (N represents positional randomization) leads to both suicide inhibition and reverse cyclization products. Note that steps C and G are in competition, as they occur concurrently. Uppercase lettering and black lines represent the intron. Lowercase lettering and their associated lines with arrows represent the exons. The elongated region of the exon mimics is double lined. Tertiary interactions are represented by dots. RT-PCR primers Pr1, Pr2, Pr3 and Pr4 bind to the designated regions. Pr2* indicates that Pr2 and Pr2-d(ATGAC)rU can both act as RT primers even though neither are complementary to the target region for the cyclization. Note that, at least for the *P. carinii* intron, step B is also achieved through ribozyme-mediated hydrolysis in the absence of pG (as is the case throughout this study) (6). After step C, the intron sequence is shown in more detail. Positions that can act as cyclization nucleophiles are shown with white lettering on a dark gray background and the intron sites of cyclization are shown with black lettering on a dark gray background. The internal guide sequence (IGS) is 5'GGUCAU3'.
Figure 3.2 Magnesium Dependence of the Suicide Inhibition (S.I.) and Reverse Cyclization (R.C.) Reactions. (A) A typical polyacrylamide gel of the reactions. Lanes 0→15 show reactions using 1 nM radiolabeled 5’ exon mimic d(ATGAC)rU and 500 nM of rP-h in HX Mg buffer [50 mM Hepes (25 mM Na+), 135 mM KCl, and X mM MgCl₂ (from 0 to 15) at pH 7.5]. Lanes A and B show reactions using 1 nM radiolabeled 5’ exon mimic Pr2-d(ATGAC)rU (27-mer) and 500 nM of rP-h in H4Mg (Lane A) or H15Mg (Lane B) buffer. All reactions were run for two hours, which allows maximum product formation. The sizes of the 27-mer, 52-mer, 350-mer, and 370-mer are approximate. (B) Graph of the MgCl₂ concentration dependence for the suicide inhibition (open squares) and the reverse cyclization (solid circles) reactions shown above. Note that both reactions occur simultaneously and that the graph is an average of two reactions. The standard deviations were typically below ten percent.
Figure 3.3 Competitive Binding Assay. Reactions run with d(ATGAC)rU in 50 mM Hepes (25 mM Na+), 15 mM MgCl₂, and 135 mM KCl at pH 7.5. (A) A native polyacrylamide gel showing the partitioning of bound and free radiolabeled r(AUGACU) as a function of non-radiolabeled d(ATGAC)rU concentration. Reactions utilized 30 nM rPC ribozyme, 1 nM radiolabeled 5’ exon mimic r(AUGACU), and 5 to 1500 nM of the competitor d(ATGAC)rU. (B) Plot and curve-fit of the data in panel A. Note that the results in Table 3.4 are an average of two independent assays. Refer to Table 3.4 for standard deviations.
Figure 3.4 Kinetic Analysis of Suicide Inhibition and Reverse Cyclization Reactions. Reactions run using 500 nM rP-h and 1 nM radiolabeled d(ATGAC)rU in 50 mM Hepes (25 mM Na\(^+\)), 4 mM MgCl\(_2\), and 135 mM KCl at pH 7.5. (A) A polyacrylamide gel showing a time dependence assay of a typical suicide inhibition and reverse cyclization reaction, using d(ATGAC)rU as the substrate. Note that both reactions occur simultaneously. (B) Graph showing the plot and curve-fit for the suicide inhibition (open square) and reverse cyclization reactions (solid circle) in Panel A. Note that the results in the Table 3.5 are an average of two independent assays. Refer to Table 3.5 for standard deviations.
Figure 3.5 Schematics of 5’ Exon Mimics Base Pairing with the IGS of the *P. carinii* Ribozyme. (a) The native 5’ exon, (b) the 5’ exon with a single G to C mutation at exon position –4, and (c) the 5’ exon with a G to C mutation at exon position –4 and a T to G mutation at exon position –5. The $K_d$ values given are for the respective 5’ exon mimics binding to the IGS through base pairing and tertiary interactions (in 15 mM MgCl$_2$). Exon positions –2 through –6 are deoxyribonucleotides. Note that the double mutant (c) rescues binding lost with the single mutant (b).
Figure 3.6 Schematics of the 5’ End of the Intron Base Pairing with the IGS Before Cyclization. These are three possible routes for the 5’ end of the intron binding to the G-rich region of the IGS that are consistent with the data. The cyclization reaction is similar to the self-splicing reaction with the substrate binding to the IGS, allowing the free 3’OH on the 3’ G to attack at the –1 exon position.
CHAPTER FOUR - A Tetrahymena Group I Intron-Derived
Ribozyme Catalyzes the Trans Excision-Splicing Reaction

Introduction

In the trans-excision splicing reaction (TES), a group I intron-derived ribozyme from the opportunistic pathogen *P. carinii* base pairs with an exogenous RNA substrate, excises a specific segment, and ligates the ends of the RNA substrate back together (Figure 4.1) (6). The TES reaction has two reaction steps, initiated by adding substrate to the ribozyme. The substrate forms a P1 helix with the IGS of the ribozyme. This P1 helix is positioned into the catalytic core by tertiary interactions. The first nucleophilic attack occurs at the end of the P1 helix and results in 5’-splice site cleavage of the substrate. The second nucleophilic attack occurs at the 3’-splice site between the insert and the 3’end of the substrate, resulting in TES product formation. This TES reaction has been extensively studied with the *P. carinii* ribozyme (6, 20, 21), but to make this reaction more generally applicable, we tested whether another group I intron-derived ribozyme could perform the TES reaction.

Exploration of intron self-splicing and the development of new group I intron-derived ribozyme reactions has been predominately conducted using a group I intron from the ciliate *Tetrahymena thermophila*. Therefore, we wanted to see if the *Tetrahymena* ribozyme could perform the TES reaction and if so, compare it to the *P. carinii* TES reaction. Previous results have shown that the *P. carinii* and *Tetrahymena* ribozymes utilize a different array of tertiary interactions to dock the P1 helix into the catalytic core (17-19). Since the TES reaction utilizes the P1 helix in the first nucleophilic attack, there may be differences in the ability of these two ribozymes to perform the TES reaction.

Both the 5’ and 3’-splice sites in the *Tetrahymena* self-splicing reaction are susceptible to hydrolysis (41, 74, 75), although *Tetrahymena* ribozymes do not appear to catalyze ribozyme-mediated 5’-splice site cleavage as well as *P. carinii* ribozymes (6, 76). Ribozyme mediated cleavage by a free hydroxyl (from the buffer) at the 5’-splice
site in the substrate is the first step of the TES reaction (first nucleophilic attack shown in Figure 4.1.2). Therefore, the ability of the Tetrahymena ribozyme to perform the TES reaction is suspected to be in part dependent on its ability to perform ribozyme-mediated 5’-splice site cleavage. This will be analyzed in this study.

In this work, we show that the TES reaction is not specific to the P. carinii derived-ribozyme, but also occurs in a Tetrahymena-derived ribozyme. In our model system, the Tetrahymena TES reaction was performed with a 13-nucleotide substrate, excising one nucleotide from the center. The TES reaction with the Tetrahymena ribozyme has lower optimized product yield and decreased observed rate constants for both the first and second steps of the reaction (optimized for yield), compared to the P. carinii ribozyme. For additional comparison, two hybrid ribozymes were also tested: the Tetrahymena ribozyme with the P. carinii IGS and the P. carinii ribozyme with the Tetrahymena IGS. The hybrid ribozymes allow us to separately compare the ability of the catalytic core of each ribozyme, as well as their specific native substrates, in performing the TES reaction.

The results suggest that the TES reaction is a general group I intron-derived ribozyme reaction. Both the sequence of the substrate (and hence the base paired P1 and P10 helices) and the characteristics of the individual ribozyme contribute to the accuracy and efficiency of the TES reaction. Additionally, the ability of the ribozyme and substrate to allow ribozyme-mediated 5’-splice site cleavage appears to be particularly important for efficient TES product formation.
Methods

Synthesis and Purification of Oligonucleotides

For these studies, four ribozymes were used: rT, which is transcribed from the L-21 *Tetrahymena* plasmid (kindly provided by Douglas Turner, University of Rochester; Rochester, NY); rT-PC, which is the *Tetrahymena* ribozyme with the IGS from *P. carinii*; rPC, which is rP-8/4x from previous work (6); and rPC-T, which is the *P. carinii* ribozyme with the *Tetrahymena* IGS. The ribozymes are diagramed in Figure 4.2. The hybrid plasmids, pT-PC and pPC-T, were made by changing pT and pPC via site directed mutagenesis essentially as described (77). The pT-PC plasmid was made using two rounds of site directed mutagenesis using the primer pairs CGACTCATA\text{GAGGGAGGGGAAAGTTATC} and GATAACTTTTCCCTCC\text{TCTATAGTGAGTC} for the first round and CACTAGAGGG\text{TCATAAAGTTATCAGGC} and GCCTGATAAC\text{TATGACCCCTCATAGTG} for the second round. The pPC-T plasmid was made using two rounds of site directed mutagenesis using primer pairs CGACTCATA\text{GAGGGAGGGGAAACGCG} and GCCCTTTCCC\text{CCCTTCTATAGTGAGTC} for the first round and CGACTCATA\text{GCTTTGGAGGGAAACGCG} and GCCCTTTCCC\text{CCCTAAGGTCTATAGTGAGTC} for the second round. The DNA primers were purchased from Integrated DNA Technologies, Inc. (Coralville, IA). The mutated plasmids were sequenced for confirmation (Davis Sequencing, University of California, Davis).

The ribozyme plasmids were linearized in 50 μL reactions consisting of 8 μg plasmid and either 100 U of *Sca*I restriction enzyme (New England Biolabs, Beverly, Massachusetts) for the two *Tetrahymena* plasmids (pT and pT-PC) or 50 U of *Xba*I (Invitrogen, Grand Island, New York) for the two *P. carinii* plasmids (pPC and pPC-T). Linearization was confirmed by visualization on a 1% agarose gel. The linearized DNA was purified using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA) and eluted in water. Run-off transcription was performed for two hours in 100 μL reactions consisting of 1 to 2 μg of linear DNA, 50 U of T7 RNA polymerase (New England
Biolabs), 40 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 5 mM DTT, 5 mM spermidine, 1 mM rNTP mix, and 62.5 μg/mL BSA. The ribozymes were purified using a Plasmid Midi-Kit (Qiagen) as described previously (6) or by the RNeasy purification kit (Qiagen).

RNA oligonucleotides were purchased from Dharmacon Research, Inc. (Lafayette, CO). Designated oligonucleotides were 5' end radiolabeled and purified by polyacrylamide gel electrophoresis as described previously (17). Oligonucleotide and ribozyme concentrations were calculated based on UV-absorption measurements using a Beckman DU 650 UV-Spectrophotometer (Beckman Coulter, Inc., Fullerton, CA).

**TES Reaction and Optimization**

Reactions were performed in buffer consisting of 50 mM Hepes (25 mM Na⁺), 135 mM KCl, and various concentrations of MgCl₂. The reactions were optimized for magnesium concentration (0-25 mM MCl₂), ribozyme concentration (0-500 nM), and time dependence (1-180 min). The optimized reaction conditions were: 100 nM rPC and 15 mM MgCl₂ for 30 min; 100 nM rPC-T and 4 mM MgCl₂ for 60 min; 100 nM rT-PC and 8 mM MgCl₂ for 40 min; and 200 nM rT and 10 mM MgCl₂ for 90 min (see Table 4.1). In each case, the reaction conditions were optimized to give the greatest TES product yields. The ribozyme mix (4 μL, including everything except the substrate) was preannealed at 60°C for five minutes and then allowed to slowly cool to 44°C. Reactions were initiated by adding 1 μL of 5'-end radiolabelled substrate (CUCUCUGAAAGGU for rT and rPC-T; or AUGACUGCUC for rPC and rT-PC). Reactions were run at 44°C for the optimized amount of time and stopped by adding 5 μL stop buffer (10 M urea, 3 mM EDTA, and 0.1X TBE). Reactions were denatured for 1 min at 90°C and run on a 12% polyacrylamide/8 M urea gel. The gel was dried under vacuum. The bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. All reactions were run at least twice. Reactions were also attempted in the typical *Tetrahymena* Tris buffer consisting of 50 mM Tris base, 0.1 mM EDTA, and 10 mM NaCl, and also under typical *Tetrahymena* preannealing conditions of 30 min at 50°C.
**Kinetic Studies**

The observed rate constants for the TES reaction using each ribozyme were obtained under single-turnover ‘intron excess’ conditions. In these experiments, 60 μL of ribozyme in appropriate buffer was preannealed at 60°C for five minutes and slowly cooled to 44°C. The reactions were initiated by adding a 15 μL aliquot of 5'-end radiolabelled substrate (CUCUCUGAAAGGU for rT and rPC-T or AUGACUGCUC for rPC and rT-PC), bringing the final concentrations to 1 nM substrate and 100 nM (for rPC, rPC-T, rT-PC) or 200 nM (for rT) ribozyme. A 5 μL aliquot was periodically removed and added to 5 μL stop buffer. The reactants and products were separated on a 12% polyacrylamide/8 M urea gel. The gel was dried under vacuum. The bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. The observed rate constants for the first reaction step (5’cleavage product plus TES product over time) and for the second step (TES product over time) were quantified essentially as described (17). Rate constants were calculated with data from at least two independent time studies.
Results

A new ribozyme reaction, the TES reaction (Figure 4.1), was recently reported (6). In the development of the TES reaction, a P. carinii group I intron-derived ribozyme was used. To determine if the TES reaction is specific to the P. carinii ribozyme or if it is more general to group I introns, the reaction was attempted in another group I intron derived ribozyme, the L21-Sca I Tetrahymena ribozyme (called rT in this study). The rT ribozyme was chosen because it has been extensively studied in group I intron-derived ribozyme reactions (1, 5, 29, 39, 76, 78-80). Studying the TES reaction using the Tetrahymena ribozyme also allows us to compare the two different group I intron-derived ribozymes.

The TES reaction is initiated by adding substrate (10mer or 13mer in this study) to the ribozyme. The ribozyme recognizes its substrate partially through base pairing (Figure 4.1.1). The first nucleophilic attack (Figure 4.1.2) occurs at a G-U wobble pair at the end of the P1 helix and results in 5’-cleavage of the substrate, producing a 6mer intermediate. Some of this intermediate does not perform the second nucleophilic attack and can be seen as a 6mer product in the reaction. The TES product forms when the free 3’-OH group on the 6mer intermediate performs the second nucleophilic attack at the splice site between the insert and the 3’end of the substrate (Figure 4.1.3). This second nucleophilic attack results in TES product formation (9mer or 12mer in this study).

A Tetrahymena Group I Intron-Derived Ribozyme Catalyzes the TES Reaction

As shown in Figure 4.3, the Tetrahymena ribozyme (rT) can catalyze the TES reaction, forming a 12mer product. The TES reaction with the rT ribozyme was optimized for ribozyme concentration, MgCl₂ concentration, and time dependence. A representative gel and graphs of the averaged data are seen in Figure 4.3. The optimum ribozyme (rT) concentration range is from 50-300 nM, with loss of product seen at higher concentrations (suspected to be due to additional reactions). Subsequent reactions were conducted using 200 nM rT ribozyme, although 100 nM rT appears to work just as well. The optimum MgCl₂ concentration is from 4-10 mM, and we ran subsequent reactions at 10 mM MgCl₂. The optimum reaction time for the rT TES reaction is from 45-90
minutes, and we ran the subsequent reactions for 90 minutes. The reaction was not optimized for rate, but was optimized for product yield.

The optimum TES reaction with the *P. carinii* ribozyme (rPC) was recently published (6), but reactions were repeated here for comparison (Figure 4.4). The rPC ribozyme appears to have a similar optimum ribozyme concentration range (30-200 nM) compared to that of rT. The two ribozymes appear to differ the most with their optimum MgCl₂ concentration range and time dependence. The rT reaction is optimum at MgCl₂ concentrations as low as 4 mM (Figure 4.3.B), whereas the rPC ribozyme requires at least 15 mM MgCl₂ (Figure 4.4.B). The rPC TES reaction is much faster, completed in less than 5 minutes (Figure 4.4.C). Whereas, the rT TES reaction takes up to 45 minutes to reach its optimum yield (Figure 4.3.C).

In the literature, *P. carinii* ribozyme reactions are typically performed in Hepes buffer, whereas *Tetrahymena* ribozyme reactions are typically performed in Tris buffer. Therefore, the *Tetrahymena* TES reaction was conducted in both buffers, and gave approximately the same product yields. Therefore, all the reactions for this study were conducted in the *P. carinii* Hepes buffer. In addition, the typical pre-annealing conditions for the *Tetrahymena* ribozyme are also different from the conditions for the *P. carinii* ribozyme. Studies have shown that the *Tetrahymena* ribozyme requires 30 minutes at 50°C for complete folding and maximum activity (76), whereas the *P. carinii* ribozyme only requires 5 minutes at 60°C. The *Tetrahymena* ribozyme was pre-annealed under both conditions and gave the same product yield. Therefore, all the reactions were run under the annealing conditions of 5 minutes at 60°C.

*The Tetrahymena Ribozyme does not Perform TES as well as the P. carinii Ribozyme*

The TES reaction can be evaluated based on the extent of reaction product formed. The maximum amount of TES product from optimum reactions with rPC and rT is compared in Table 4.1. The rPC ribozyme produces more TES product than rT, giving an average yield of 68.5±4.4%. The rT ribozyme produces 3-fold less TES product compared to the rPC, giving an average yield of 19.1±11.5%. The standard deviation is high because the *Tetrahymena* ribozyme exhibits marked variability in its reactivity from different syntheses. This variability has been noted in the literature (76).
Another way to analyze the *Tetrahymena* and *P. carinii* TES reactions is to compare the observed rate constants, $k_{\text{obs}}$, for each ribozyme reaction. The time studies in Figure 4.3.C and Figure 4.4.C were used to calculate the observed rate constant for TES product formation. The observed rate constant for the first step of the TES reaction is calculated as the sum of 5′-cleavage product (6mer) and TES product over time, as the TES product necessarily underwent the first step. The observed rate constant for the second step of the TES reaction is calculated as the amount of TES product over time (represented by closed circles on the graphs in Figure 4.3.C and 4.4.C).

For the *P. carinii* ribozyme, the $k_{\text{obs}}$ is 5.0 min$^{-1}$ for the first step and 3.2 min$^{-1}$ for the second step. The $k_{\text{obs}}$ values for the *Tetrahymena* ribozyme are 70-fold lower for both steps of the reaction compared to the *P. carinii* ribozyme. The $k_{\text{obs}}$ for *Tetrahymena* ribozyme is 0.07 min$^{-1}$ for the first step and 0.05 min$^{-1}$ for the second step. The first nucleophilic attack of the TES reaction, 5′-splice site cleavage (Figure 4.1.2), appears to be much slower for the rT ribozyme compared to rPC. This is not surprising since the rT ribozyme typically depends on exogenous G to perform the first nucleophilic attack (5, 39, 76, 81). Also, to keep the reactions consistent, all the reactions were performed under the optimum pre-annealing conditions for the *P. carinii* ribozyme. Both conditions were shown to give the same amount of TES product for the rT reactions.

*Hybrid Tetrahymena and P. carinii Ribozymes*

Since group I intron-derived ribozymes recognize substrates through tertiary interactions (39, 77) in addition to base pairing, we analyzed whether the native P1 and P10 sequences in either the *P. carinii* or *Tetrahymena* intron would effect the ability of the ribozyme to perform the TES reaction. Additionally, we wanted to analyze whether differences in the *P. carinii* and *Tetrahymena* reactions were due to fundamental differences in the catalytic core of the ribozymes, or whether they were due to the different IGS sequences. Therefore, two hybrid ribozymes were analyzed to observe the influence the substrate sequence and the specific ribozyme has on the TES reaction. The two hybrid ribozymes are rPC-T, which is the *P. carinii* ribozyme with the *Tetrahymena* IGS (Figure 4.2.B); and rT-PC, which is the *Tetrahymena* ribozyme with the *P. carinii* IGS (Figure 4.2.D).
Optimization of the Hybrid *P. carinii* Ribozyme with the Tetrahymena IGS and Comparison to the *P. carinii* Ribozyme

The rPC-T ribozyme is the *P. carinii* ribozyme with the *Tetrahymena* IGS (Figure 4.2.B). A representative optimization gel and graphs of averaged data are shown in Figure 4.5. The yield-optimized reaction conditions are 100 nM rPC-T and 4 mM MgCl$_2$ with a reaction time of 60 minutes. The optimum conditions for the rPC-T TES reactions differ from the rPC reaction conditions most noticeably in the narrow range of optimum ribozyme and MgCl$_2$ concentrations, as well as substantially longer reaction times to obtain maximum yields. The rPC-T requires much less MgCl$_2$, 4mM versus 15 mM for rPC. Also, the rPC reaction is complete in less than 5 minutes, whereas the rPC-T reaction takes almost an hour. Simply by changing the IGS sequence (rPC to rPC-T), the optimum reaction conditions the hybrid *P. carinii* ribozyme change substantially from the native ribozyme. In addition, the rPC-T optimum conditions are more similar to Tetrahymena ribozyme conditions. Apparently, the IGS sequence markedly influences the reaction dynamics.

The rPC-T reactions also produce substantial amounts of unidentified side products. In the rT TES reaction run with the same 13mer substrate, these products are not nearly as prevalent (Figure 4.3, all lanes). Apparently, there is something specific regarding the *Tetrahymena* ribozyme that prevents such side products when reacted with the same substrate. Indeed, it is likely that each ribozyme has evolved ways of preventing cryptic site reactions that are specific to the sequence of its own substrate.

The rPC-T TES reaction gives 25.6 ± 1.3% 12mer TES product. This is approximately a third less product than what is produced by the rPC ribozyme with its native substrate. Simply by changing the IGS of the *P. carinii* ribozyme to the IGS of the *Tetrahymena* ribozyme, the reaction yields are drastically decreased. The yields for rPC-T are more similar to those seen for rT. It appears that the P1 and P10 helices from the Tetrahymena ribozyme, even though complementary based paired, greatly reduce TES product formation with this *P. carinii* ribozyme (rPC-T). Note that the decreased product
yield, at least in part, is due to additional side reactions occurring between the rPC-T ribozyme and the 13mer substrate.

The $k_{\text{obs}}$ values for rPC-T are 0.04 min$^{-1}$ for the first step and 0.05 min$^{-1}$ for the second step. Compared to rPC, the first step of the reaction is 125-fold slower and the second step is 64-fold slower. In contrast, the $k_{\text{obs}}$ values for the TES reaction with rT and rPC-T are quite similar. Simply by changing the IGS in the *P. carinii* ribozyme, the $k_{\text{obs}}$ values for both steps of the TES reaction are greatly decreased. Presumably, changing the P1 and P10 helixes could diminish tertiary interactions at the active site, as well as the pre-organization of the substrate.

**Optimization of the Hybrid Tetrahymena Ribozyme with the *P. carinii* IGS and Comparison to the *P. carinii* Ribozyme**

The rT-PC ribozyme is the *Tetrahymena* ribozyme with the *P. carinii* IGS (Figure 4.2.D). A representative optimization gel and graphs of averaged data are shown in Figure 4.6. The yield-optimized reaction conditions are 100 nM rT-PC and 8 mM MgCl$_2$ with a reaction time of 40 minutes. The general trends for each optimized variable for rT-PC is similar to what is seen for rT (compare graphs in Figure 4.6 to Figure 4.3).

The rT-PC TES reaction gives 39.3 ± 2.4% 9mer TES product. This is approximately 60% the amount of maximum product obtained in the rPC TES reaction. Compared to the rT ribozyme, the rT-PC ribozyme produces twice as much product. Simply by changing the IGS of the *Tetrahymena* ribozyme to the IGS of the *P. carinii* ribozyme, the reaction yields are dramatically increased. Surprisingly, the specific P1 and P10 helices from the *P. carinii* ribozyme improve TES product yields with the *Tetrahymena* ribozyme. This is further proof that group I intron-derived ribozymes bind and react with complementary sequences differently. Also, this improved yield for the *Tetrahymena* with the *P. carinii* substrate may simply reflect that the *Tetrahymena* substrate is inherently problematic.

The observed rate constants for rT-PC are 0.20 min$^{-1}$ for the first step and 0.08 min$^{-1}$ for the second step. Compared to rPC, the first step of the reaction is 25-fold slower and the second step is 40-fold slower. This is a mild improvement compared to
the observed rate constants for rT. Another thing to notice about these observed rate constants, the $k_{obs}$ for the first step is improved 3-fold compared to the $k_{obs}$ for rT. Although this is a small improvement, it is possible that the particular $P. \text{carinii}$ substrate lends itself to increased rates for the first nucleophilic attack and subsequently more efficient 5’-splice site cleavage.

These results suggest that the identity of the IGS and its fully complementary substrate dramatically influence the reaction dynamics of group I intron derived ribozymes in general.
Discussion

The Tetrahymena Ribozyme Catalyze the TES Reaction

The TES reaction is a new group I intron-derived ribozyme reaction demonstrated with a P. carinii ribozyme (6). We now show that the L21-Sca I Tetrahymena ribozyme catalyzes the TES reaction as well. The Tetrahymena TES reaction is performed with a 13mer native Tetrahymena substrate (4.2.C) and produces a 12mer TES product, although not as well as the P. carinii ribozyme.

Comparison of the Four Ribozymes

Among the four ribozymes in this study, the P. carinii ribozyme gives the greatest extent of reaction with the highest observed rate constants for both the first and second TES reaction steps. This suggests that there is something inherent with the P. carinii ribozyme and its specific P1 and P10 helices that allows for efficient TES. In comparison, the observed rate constants are much lower for the Tetrahymena ribozyme as well as the two hybrid ribozymes (compare in Table 4.1).

The hybrid Tetrahymena ribozyme with the P. carinii IGS (rT-PC) has improved yields compared to the Tetrahymena ribozyme (rT), but the rates are not substantially improved. Therefore, simply changing the targeted substrate in the Tetrahymena TES reaction can improve the yield of the reaction. Also, there is no real difference between the rates or yields from the Tetrahymena ribozyme (rT) and the hybrid P. carinii ribozyme with the Tetrahymena IGS (rPC-T). Both of these comparisons suggest the specific sequence of the Tetrahymena substrate is a major determinant of low yields.

The TES reaction with the hybrid P. carinii ribozyme (rPC-T) has a 60-fold lower observed rate constant compared to the P. carinii ribozyme. In a previous study, a 10-fold decrease in the observed rate constant was observed when the P. carinii ribozyme was targeted to a different sequence (6). Clearly, just changing the complementary base paired P1 and P10 helices affects the rate of the P. carinii TES reaction. Work with other Tetrahymena reactions has shown that the Tetrahymena ribozyme recognizes the structure of the P1 and P10 helices and the specific sequence is not important (27, 67). Our results with the P. carinii ribozyme suggest the opposite. It is likely that introns
have evolved in a manner that is somewhat dependent on the sequence of their substrates, for example using tertiary interactions that help the ribozyme perform catalysis.

**The Tetrahymena Substrate is Problematic in the TES Reaction**

Unfortunately, performing the TES reaction using the native *Tetrahymena* substrate is problematic, as it results in additional and unidentified products, as well as lower TES product yields. It is likely that some of these additional products are formed because the P1 helix (formed with the CUCUCU sequence of the 13mer *Tetrahymena* substrate CUCUCUGAAAGGU) docks into the catalytic core out of register, positioning the substrate for catalysis at different positions. Although the *Tetrahymena* substrate is typically cleaved at the –1 position (CUCUCU), it has been shown to cleave at both the –2 and –3 positions (CUCUCU and CUCUCU) (69). This cleavage infidelity has been observed in other studies as well (26, 27, 38, 55). In addition to mispairing, it appears the P1 helix could also bind out of register in the catalytic core. This may explain the appearance of smaller products (approximately a 5mer in rT TES reactions and approximately a 4mer in rPC-T TES reactions).

The rPC-T hybrid TES reaction also produces bands approximately seven and eight nucleotides in length (Figure 4.5). These bands were sometimes present in rT TES reactions, although they are not prominent in the optimization gel (Figure 4.3). The identity of these products is unknown, but could be due to breakdown of the TES product. In recent work (21), TES products were shown to dissociate from the *P. carinii* ribozyme and lead to smaller breakdown products. Presumably, the TES product from the hybrid rPC-T ribozyme (12mer) could dissociate and rebind another ribozyme, leading to these unintended products (7 and 8mer).

Both the hybrid *P. carinii* ribozyme (rPC-T) and the native *Tetrahymena* ribozyme (rT) target the native *Tetrahymena* substrate (Figure 4.2 B and C). Comparing the results of the rPC-T and rT reactions (Figure 4.3 and 4.5), show that the hybrid ribozyme has more unintended products than the native ribozyme. It is suspected that the *Tetrahymena* ribozyme may be able to prevent its native substrate from slipping or the product from breaking down by the presence of specific tertiary interactions that have evolved to stabilize the native *Tetrahymena* substrate. This agrees with what we found.
for the *P. carinii* ribozyme, but disagrees with previously published results showing that changes to the *Tetrahymena* ribozyme substrate does not affect catalysis (67). Additional experiments comparing the rT and rPC-T TES reactions need to be conducted to conclude that these apparent tertiary interactions are real.

**The TES Reaction Depends on Ribozyme Mediated 5’-Site Cleavage**

The ability of a group I intron-derived ribozyme to catalyze the TES reaction depends in part on its ability to perform ribozyme-mediated 5’-splice site cleavage. The first step of the TES reaction is cleavage performed by a free hydroxyl from the buffer at the ribozyme-mediate 5’-splice site. Group I intron self-splicing utilizes a 3’-hydroxyl from an exogenous G to perform this reaction (see Figure 2.7). A study comparing the rate constants for the chemical step of ribozyme-mediated 5’-splice site cleavage with the *Tetrahymena* ribozyme reacting with the substrate GGCCCUCUAAAAA produced a $k_c$ of 350 min$^{-1}$ with saturating G and a $k_c$ of 0.7 min$^{-1}$ in the absence of G (76). Obviously, the *Tetrahymena* ribozyme does not perform the 5’-splice site cleavage very well without G. Moreover, this rate constant (0.7 min$^{-1}$) is not substantially different from the observed rate constants seen with *Tetrahymena* in the TES reaction. It appears that the low observed rate constants for the TES reaction with the *Tetrahymena* ribozyme might be linked to inefficient ribozyme-mediated 5’-splice site cleavage.

**Implications**

Showing that the *Tetrahymena* ribozyme can perform the TES reaction indicates that the TES reaction is likely a general reaction for group I introns, although the native *P. carinii* ribozyme is much more efficient at the TES reaction than the other three ribozyme-substrate combinations tested. The *Tetrahymena* TES reaction may work better, however, if run with a modified-native substrate. For example, in other *Tetrahymena* studies, the CUCUCU substrate is often changed to GGCCCUCU, apparently to help the substrate bind more tightly and prevent cleavage infidelity (54, 73, 76).
The ability of the *P. carinii* ribozyme to perform the TES reaction well appears to be due, at least in part, to its ability to perform ribozyme-mediated 5’-splice site cleavage efficiently. The slower rates seen with the *Tetrahymena* ribozymes, could be related to their decreased ability to allow 5’-splice site cleavage.

Changing the IGS of the *P. carinii* ribozyme to the *Tetrahymena* IGS substantially decreases the observed rate constants and yield. Also, the native *Tetrahymena* ribozyme with its native substrate gives fewer unintended product when compared to the *P. carinii* hybrid ribozyme targeting the same substrate. Apparently, the combination of the specific tertiary interactions with the base paired P1 and P10 helices have evolved together, and changing the IGS, even to other complementary sequences, decreases the rate of reaction. We show in this study that the specific sequence of the IGS does matter. This has been seen in another *P. carinii* study ((77) and in Chapter Three of this work) and is in contrast to other *Tetrahymena* studies (27, 67).

Finally, our results indicate that targeting TES ribozymes to particular sequences may be challenging. Ultimately, it may be difficult to overcome years of evolution that has given group I introns the ability to catalysis reactions with specific substrates. Nevertheless, the *P. carinii* ribozyme and the *Tetrahymena* ribozyme can perform the TES reaction *in vitro*, suggesting that self-spliced group I introns may have the general ability to perform the TES reaction even in the cell.
Table 4.1 Tetrahymena, P.carinii, and Hybrid Ribozyme TES Optimum Reaction Condition and Results

<table>
<thead>
<tr>
<th>Optimum Conditions</th>
<th>Result</th>
<th>Extent of TES Reaction (%)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>1st Step k&lt;sub&gt;obs&lt;/sub&gt; (min)&lt;sup&gt;d&lt;/sup&gt;</th>
<th>2nd Step k&lt;sub&gt;obs&lt;/sub&gt;(min)&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ribozyme Concentration (nM)</td>
<td>MgCl&lt;sub&gt;2&lt;/sub&gt; Concentration (mM)</td>
<td>Reaction Time (min)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rPC</td>
<td>100</td>
<td>15</td>
<td>30&lt;sup&gt;a&lt;/sup&gt;</td>
<td>68.5±4.4%</td>
</tr>
<tr>
<td>rPC-T</td>
<td>100</td>
<td>4</td>
<td>60</td>
<td>25.6±1.3%</td>
</tr>
<tr>
<td>rT</td>
<td>200</td>
<td>10</td>
<td>90</td>
<td>19.1±11.5%&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>rT-PC</td>
<td>100</td>
<td>8</td>
<td>40</td>
<td>39.3±2.4%</td>
</tr>
</tbody>
</table>

<sup>a</sup>Optimum yields are achieved in as few as 5 minutes.  
<sup>b</sup>Reactions were optimized to give maximum TES product yields. The yields were calculated with data from at least six independent reactions. Errors reported are the standard deviation.  
<sup>c</sup>The extent of reaction for the rT ribozyme is highly variable from different batches of ribozyme stock. Five independent runs from one batch gave 10.9 ± 1.1% 12mer TES product. Six independent runs from another batch gave 27.2 ± 1.8% 12mer TES product. The value of 19.1 is the average from these two batches and 11.5 is the standard deviation.  
<sup>d</sup>Observed rate constants were calculated using the average of at least two independent sets of data. The 1<sup>st</sup> step of the TES reaction is calculated as the sum of the percent of 5′cleavage product and TES product produced over time. The 2<sup>nd</sup> step of the TES reaction is calculated as the percent of TES product produced over time. Errors reported are the standard deviation.
Figure 4.1. Schematic of the Trans Excision-Splicing Reaction. The TES substrate has three parts; a 5’-exon mimic shown in pink, the insert region shown in green, and a 3’-exon mimic shown in light blue. The green segment in the TES substrate will be removed and requires a guanosine (G) as the last or only nucleotide (if cutting out just one nucleotide) in its sequence. The ribozyme is dark blue. 1) The ribozyme recognizes the TES substrate in part through base pairing, forming a P1 and P10 helix. 2) The first nucleophilic attack is performed by an exogenous hydroxyl group, presumably from the buffer. This first attack occurs at the splice site formed at the end of the P1 helix. After the first nucleophilic attack, the 5’-exon mimic (shown in pink) is no longer attached to the rest of the substrate and has a free 3’-hydroxyl (OH) group. 3) The free 3’-OH on the 5’-exon mimic performs the second nucleophilic attack at the phosphate linking the final guanosine in the insert and the 3’-exon mimic (the junction between the green and light blue segments in the TES substrate). The second attack results in removal of insert sequence (green) and the ligation of the 5’ and 3’-exon mimics (pink and light blue).
Figure 4.2. Diagrams of the *Tetrahymena*, *P. carinii*, and hybrid ribozymes base paired to their complementary TES substrate. The ribozymes are shown in black and only the IGS is detailed. The 10mer and 13mer substrates are shown in pink (5'-exon mimic), green (nucleotide to be removed), and light blue (3'-exon mimic). A) The *P. carinii* ribozyme (rPC) and the 10mer substrate produce a 9mer TES product. B) The hybrid *P. carinii* ribozyme with the *Tetrahymena* IGS (rPC-T) and the 13mer substrate produce a 12mer TES product. C) The original *Tetrahymena* ribozyme (rT) and its 13mer substrate produce a 12mer TES product. D) The hybrid *Tetrahymena* ribozyme with the *P. carinii* IGS (rT-PC) and the 10mer substrate produce a 9mer TES product.
Figure 4.3. Representative polyacrylamide gel and graphs of rT TES optimization. The TES reaction yields were optimized for ribozyme concentration, MgCl₂ concentration, and time. Reactions were run in buffer consisting of 50 mM HEPES (25 mM Na⁺) and 135 mM KCl and under optimum conditions of 200 nM rT and 10 mM MgCl₂ for 90 min, except for the changing variable. The gel shows 13mer starting material, 6mer 5’-splice site cleavage intermediate, and 12mer TES product. The size control was 13mer (CUCUCUGAAAGGU) and 6mer (AUGACU). Note that the AUGACU 6mer runs slightly higher than the CUCUCU 6mer. The graphs show the percent 12mer (TES product) with filled circles and percent 6mer (5’cleavage product) with opened circles. The graphs represent the average of two independent reactions. Standard deviations were typically below ten percent for each point.
Figure 4.4. Representative polyacrylamide gel and graphs of rPC TES optimization. The TES reaction yields were optimized for ribozyme concentration, MgCl$_2$ concentration, and time. Reactions were run in buffer consisting of 50 mM HEPES (25 mM Na$^+$) and 135 mM KCl and under optimum conditions of 100 nM rPC and 15 mM MgCl$_2$ for 30 min, except for the changing variable. The gel shows 10mer starting material, 6mer 5’-splice site cleavage intermediate, and 9mer TES product. The size control was 10mer (AUGACUGUCUC) and 6mer (AUGACU). The graphs show the percent 9mer (TES product) with filled circles and percent 6mer (5’cleavage product) with opened circles. The graphs represent the average of three independent reactions. The standard deviations were typically below ten percent for each point.
Figure 4.5. Representative polyacrylamide gel and graphs of rPC-T TES optimization. The TES reaction yields were optimized for ribozyme concentration, MgCl\(_2\) concentration, and time. Reactions were run in buffer consisting of 50 mM HEPES (25 mM Na\(^+\)) and 135 mM KCl and under optimum conditions of 100 nM rPC-T and 4 mM MgCl\(_2\) for 60 min, except for the changing variable. The gel shows 13mer starting material, 6mer 5’-splice site cleavage intermediate, and 12mer TES product. The size control was 13mer (CUCUCUGAAAGGU) and 6mer (AUGACU). Note that the AUGACU 6mer runs slightly higher than the CUCUCU 6mer. The graphs show the percent 12mer (TES product) with filled circles and percent 6mer (5’cleavage product) with opened circles. The graphs represent the average of at least three independent reactions. Standard deviations were typically below fifteen percent for each point.
Figure 4.6 Representative polyacrylamide gel and graphs of rT-PC TES optimization. The TES reaction yields were optimized for ribozyme concentration, MgCl\textsubscript{2} concentration, and time. Reactions were run in buffer consisting of 50 mM HEPES (25 mM Na\textsuperscript{+}) and 135 mM KCl and under optimum conditions of 100 nM rT-PC and 8 mM MgCl\textsubscript{2} for 40 min, except for the changing variable. The gel shows 10mer starting material, 6mer 5’-splice site cleavage intermediate, and 9mer TES product. The size control was 10mer (AUGACUGCU) and 6mer (AUGACU). The graphs show the percent 9mer (TES product) with filled circles and percent 6mer (5’cleavage product) with opened circles. The graphs represent the average of four independent reactions. Standard deviations were typically below fifteen percent for each point.
CHAPTER FIVE - Trans Insertion-Splicing: A Novel Group I Intron-Derived Ribozyme Mechanism for Inserting One RNA Sequence into Another

Introduction

We recently reported a trans excision-splicing (TES) reaction, developed with a \textit{P. carinii} group I intron-derived ribozyme, which can bind an exogenous RNA substrate, remove a section from the middle, and splice the ends back together (Figure 5.1.A) ([6, 20, 21]). We now report a new group I intron-derived ribozyme reaction called trans insertion-splicing (TIS) where one RNA substrate is directly inserted into a second RNA substrate. A simple schematic is shown in Figure 5.1.B. The ability of a group I intron-derived ribozyme to perform the TIS reaction was confirmed by enzymatic sequencing of the product. Reactions were optimized for both yield and rate, with optimum reactions ultimately run in 10 mM MgCl$_2$ for two hours. Reaction products are stable, with no visible loss at extended times. A three-step mechanism was elucidated using modified substrates to trap isolated reaction steps. Although the same \textit{P. carinii} ribozyme is used in the TES reaction, the TIS reaction is not the reverse of the TES reaction. In fact, attempts to directly reverse TES were unsuccessful, due to an inability to reverse the 5’-splice site cleavage reaction that occurs as the first step of TES.

The TIS reaction is the first example of the direct insertion of one exogenous RNA into a second exogenous RNA using a ribozyme essentially unaltered from its intronic form. Apparently, such introns are capable of more dynamic reactions than previously shown. The proposed mechanism for the TIS reaction and implications for group I intron-derived ribozyme reactions are discussed.
Materials and Methods

Synthesis and Preparation of Oligonucleotides

RNA oligonucleotides were purchased from Dharmacon Research, Inc. (Lafayette, CO) and deprotected according to the manufacture’s instructions. Oligonucleotide concentrations were calculated based on UV-absorption measurements using a Beckman DU 650 UV-Spectrophotometer (Beckman Coulter, Inc., Fullerton, CA). Designated oligonucleotides were 5’-end radiolabeled and purified by gel electrophoresis as described previously (17). Designated oligonucleotides were 3'-end radiolabeled by ligating 5’-end radiolabeled Cp to the 3’-end of the oligonucleotide. The Cp was 5’-end radiolabeled in 10 μL consisting of 5 μL of 250 μCi 32P-δATP (Amersham-Pharmacia, Piscataway, NJ), 2 μL of 50 mM cytidine 5’ monophosphate (CMP) (Sigma, St. Louis, MO), 1 μL of 10U/μL T4 poly-nucleotide kinase (New England Biolabs, Beverly, MA), and 1 μL of 10X poly-nucleotide kinase buffer (supplied by New England Biolabs). The reaction was run for 90 min at 37°C, and then the kinase was deactivated by incubation at 65°C for 15 min. 5’-end radiolabeled pCp (5’-*pCp) was ligated to the 3’-end of the oligonucleotide in a 10 μL reaction mixture consisting of 1 μL of DMSO, 2 μL of 10 μM RNA, 2 μL of 20U/μL T4 poly-nucleotide kinase (New England Biolabs), 4 μL of 5’-*pCp (approximately 20 μM), and 1 μL of 10X T4 RNA ligase buffer (New England Biolabs). The reactions were incubated for 16 h at 4°C. The 3’-radiolabeled oligonucleotides were purified by gel electrophoresis as described for the 5’-end radiolabeled oligonucleotides. The substrate names and sequences are shown in Table 5.1.

Ribozyme Preparation

The PC ribozyme plasmid (proposed secondary structure of the rPC ribozyme shown in Figure 5.2) was linearized in a 50 μL reaction mixture consisting of 8 μg of plasmid, 50 U of XbaI (Invitrogen, Grand Island, NY), and 1x React 2 buffer at 37°C for 2 h. Linearization was confirmed by visualization on a 1% agarose gel. The linearized DNA was purified using a QIAquick PCR Purification Kit (Qiagen Inc., Valencia, CA)
and eluted in water. Five other ribozymes with varying 3’-ends were made from the PCR products derived from the PC plasmid (Table 5.1). The upstream PCR primer for all five ribozymes was 5’CTCTAATACGACTCACTATAGAGGG3’. The following sequences are the downstream primers for each ribozyme (the variable region is underlined):

- 5’CACATATACTCTTTCTTCGAAAGAGG3’ for rPC-1,
- 5’TCTAGATATACTCTTTCTTCGAAAGAGG3’ for rPC-ωA,
- 5’CACGAGCTAGATATACTCTTTCTTCGAAAGAGG3’ for rPC-3,
- 5’CACGAGTCTAGATATACTCTTTTTCTTCGAAAGAGG3’ for rPC-3-ωA, and
- 5’ATGTTTCACGAGCTAGATATACTCTTTTTCTTCGAAAGAGG3’ for rPC-4.

The PCR products were gel purified using a Qiagen gel extraction kit (Qiagen Inc.). Run-off transcription was performed for 2 h in 100 μL reactions consisting of 1-2 μg of linear DNA, 50 U of T7 RNA polymerase (New England Biolabs), 40 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 5 mM DTT, 5 mM spermidine, 1 mM rNTP mix, and 62.5 μg/mL BSA. The ribozymes were purified using a Qiagen RNeasy purification kit (Qiagen Inc.).

The Trans Insertion-Splicing Reaction

The extent of the TIS reaction was optimized over a range of rPC concentration (10 to 1000 nM), 9mer insert concentration (10 to 3000 nM), MgCl₂ concentration (2 to 50 mM), time (1 to 180 min), and temperature (37°C to 50°C) for the reaction with 5’-end radiolabeled 12mer starting material AUGACUAAACAU. Ultimately, optimum reactions were found to consist of 2 h at 44°C with 200 nM rPC, 1 μM 9mer insert, and 10 mM MgCl₂. All reactions were run using approximately 1 nM 5’-end radiolabeled 12mer starting material in a buffer consisting of 50 mM Hepes (25 nM Na⁺) and 135 mM KCl at pH 7.5.

In the reactions, first 3 μL of ribozyme in appropriate buffer was pre-annealed at 60°C for 5 min. The ribozyme was then slowly cooled to 44°C. Reactions were initiated by adding 2 μL of pre-combined substrates (radiolabeled starting material and cold insert). Reactions were terminated after 2 h by adding 5 μL stop buffer (10 M urea, 3 mM EDTA, 0.1x TBE). The reactions were denatured for 1 min at 90°C and then separated on a 12% polyacrylamide/8M urea gel. The gel was transferred to
chromatography paper and dried under vacuum. The bands were visualized and quantified on a Molecular Dynamics Storm 860 Phosphorimager. The observed rate constant, $k_{obs}$, for the TIS reaction was obtained from the plot of percent TIS product formed over time (17).

**Product Isolation and Identification**

The TIS product was gel purified and sequenced by partial nuclease digestion along with the synthetic version of the expected product. The TIS reaction was scaled up 50-fold (250 μL) and run as described above using the optimized conditions, except the reactions were run in five 50 μL volumes, concentrated to 30 μL after 2 h, and terminated with 10 μL stop buffer. The product band was ultimately cut out of the gel and eluted from the gel matrix for 40 min by crushing with a stir bar in 400 μL of elution buffer (0.3 M sodium acetate, 5 mM EDTA, 10 mM Tris-HCl pH 7.5, and 0.1% SDS). The eluent was decanted and a second round of elution was performed. The product was ethanol precipitated overnight and concentrated to 30 μL. The radiolabeled TIS product and synthetic 18-mer were enzymatically sequenced using RNA nucleases T1, U2, CL-3 and, *B. cereus* (Research Unlimited; Wellington, New Zealand) essentially as described (6). *B. cereus* reactions used 0.33 units *B. cereus* in 33 mM sodium citrate (pH 5.0) and 1.7 mM EDTA.
Results

The *P. carinii* Group I Intron Catalyzes the TIS Reaction.

Work on the TIS reaction began by simply trying to reverse the TES reaction (Figure 5.1.A). Instead of just one substrate, the TIS reaction begins with two substrates (referred to as starting material and insert in Figure 5.1.B). TIS experiments initially attempted (data not shown) with a *P. carinii* ribozyme and TES products resulted only in recombination (the 3’-half of one substrate exchanged for the 3’-half of the other substrate). This recombination reaction appeared to be the first step of the TIS reaction, so the substrates were redesigned to help achieve the expected second step. Reactions with these new substrates and the ribozyme rPC (secondary structure in Figure 5.2) gave a larger product (potential insertion), but it was not the size we expected. To ensure that we were producing an insertion product, we tested whether this large product contained both the 5’ and 3’-terminal regions from the starting material. We did this by running the reaction with 5’ and 3’-end radiolabeled starting material, since the insertion product should contain both the 5’-end (Figure 5.1.B, in pink) and the 3’-end (Figure 5.1.B, in light blue). We appear to form the same product with both 5’-end radiolabeled starting material (Figure 5.3, lane A) and 3’-end radiolabeled starting material (Figure 5.3, lane B). Note that the 3’-end radiolabeled reaction product is one nucleotide larger due to the method of radiolabeling. For confirmation, the product band was excised from the polyacrylamide gel and its sequence was confirmed by enzymatic sequencing (Figure 5.4). The 18mer product AUGACUCUCUGAAACAU appears to be the result of the insertion of the underlined region of the 9mer insert, GCUCUCUG, into the middle of the 12mer starting material, AUGACUAACAU. Although the exact TIS product was unexpected, this was the first evidence that group I intron-derived ribozymes can catalyze the TIS reaction.

The reaction was optimized in terms of the yield (Figure 5.5) and the observed rate constant (Figure 5.6). Optimum reaction conditions for TIS are 200 nM ribozyme (rPC), 1 μM 9mer insert, and 10 mM MgCl₂ using 1 nM 12mer radiolabeled starting material (see Figure 5.5). The TIS reaction produces 62.9 ± 1.3% 18mer TIS product and the reaction is complete at two hours. Times studies were performed at four different
MgCl<sub>2</sub> concentrations and the observed rate constants, \( k_{\text{obs}} \), were determined (Figure 5.6). The \( k_{\text{obs}} \) value for the TIS reaction is 0.04 min<sup>-1</sup>, and is the same at 10, 14, and 18 mM MgCl<sub>2</sub>. The \( k_{\text{obs}} \) value when using 6 mM MgCl<sub>2</sub>, however, was 0.02 min<sup>-1</sup>, which is half the \( k_{\text{obs}} \) for the optimum condition and results also in half the yield. Apparently MgCl<sub>2</sub> concentrations as low as 6 mM inhibit the ability of the ribozyme to fold, as well as to catalyze the reaction.

The mechanism of the TIS reaction was elucidated by analyzing the effects of modifying functional groups in the substrates and ribozyme, and by initiating the reaction using intermediates as starting material. Key to the elucidation of the mechanism was the discovery that a large intermediate product, indicative of the substrate attached to the ribozyme, was observed only when substrates were 3’-end radiolabeled (and not when they were 5’-end radiolabeled). This suggested that the \( \omega \)G on the ribozyme was involved in catalysis and that intermediates were attached to the ribozyme. The proposed TIS mechanism is shown in Figure 5.7.

**Step 1: The \( \omega \)G on the Ribozyme Performs a Nucleophilic Attack on the Insert.**

The sequencing results show that the 9mer insert is losing its 5’-GCU (GCU-CUCGUG) during the TIS reaction (depicted in Figure 5.1.B). The loss of this GCU 3mer from the insert suggests that the insert might first be forming a P1 helix (called P1i) with the IGS of the ribozyme (Figure 5.7, 2a). This interaction would place a G-U wobble pair at the 5’-splice site where the \( \omega \)G on the ribozyme could attack the insert.

To show that the ribozyme’s \( \omega \)G is involved in the mechanism proposed in Figure 5.7, the \( \omega \)G in rPC was replaced with an \( \omega \)A (rPC-\( \omega \)A). Although an \( \omega \)A has been shown to give the second step of splicing in a group I intron from *Anabaena* (82), an adenosine was used to test the TIS reaction because the GBS in the *P. carinii* ribozyme has been shown to not functionally interact with adenosine for the second step of the TES reaction (21). We expect that if the \( \omega \)G in the ribozyme is important, its alteration would prevent the first reaction, as well as TIS product formation. In this reaction, using 5’ or 3’-end radiolabeled 12mer starting material, no appreciable 18mer TIS product formed (Figure
production of 6mer is occurring through ribozyme mediated hydrolysis at the 5’-splice site as seen in the TES reaction with the same ribozyme (6). Also, in the 3’-end radiolabeled reaction (Figure 5.3, lane D), much less of the large intermediate band forms compared to the reaction using rPC (compare to Figure 5.3, lane B), as expected if the ribozyme cannot perform the first step of the TIS reaction. Note that transcription using T7 RNA polymerase has been shown to add extra nucleotides to the 3’-end of transcripts and that a faint 18mer or intermediate band might be due to the addition of Gs to the 3’-end of the ribozyme in that process.

To test that the 9mer insert is forming a P1i helix (as depicted in Figure 5.7, 2a), the 9mer insert was shortened on either the 5’ or 3’-end. The 7mer-minus-3’UG insert (GCUCUCG _) is shortened by two nucleotides on its 3’-end. This insert can still form a P1i helix, but should give a shortened TIS product. Reactions run with 5’-end radiolabeled 12mer starting material, rPC ribozyme, and 7mer-minus-3’UG insert (Figure 5.3, lane G) show production of a shorter TIS product, as expected. Likewise, the insert was shortened by two nucleotides on its 5’-end, 7mer-minus-5’GC insert (_, UCUCGUG), which should disrupt P1i formation. Reactions run with 5’-end radiolabeled 12mer starting material, rPC ribozyme, and 7mer-minus-5’GC insert (Figure 5.3, lane F) show no TIS production and indicate that the insert is forming a P1i helix in the TIS reaction.

Since the 9mer insert loses its 5’GCU and appears to form a P1i helix, an insert with a deoxy-U at the critical position forming the 5’-splice site (9mer-dU, GCdUCUCGUG) was tested to see if this change inhibits 18mer production. If catalysis is occurring at this position in the 9mer, the substitution of a deoxy-U should either prevent or greatly inhibit TIS. The reaction run with 5’-end radiolabeled 12mer starting material, rPC ribozyme, and 9mer-dU insert gives greatly inhibited TIS production (Figure 5.3, lane H), as expected.

To directly show that the ribozyme is attacking the 9mer insert, reactions were run with 5’-end radiolabeled insert instead of radiolabeled starting material. These reactions are run with much less insert compared to the standard reaction, and therefore, are not under optimum conditions. Nevertheless, when the reaction is run with 5’-end
radiolabeled 9mer insert (GCUCUCGUG), all of the 9mer is cleaved to a smaller product (Figure 5.3, lane J), which appears to be the expected 3mer side product GCU. Running the reaction using the ribozyme rPC-ωA and 5’-end radiolabeled 9mer gives some of the 3mer product, but cleavage at this position is greatly inhibited (Figure 5.3, lane K), as expected without an ωG on the ribozyme. The 3mer produced in this reaction is presumed to be through ribozyme-mediated cleavage since the ωA in the ribozyme cannot attack at the splice site.

When the reaction is run with 3’-end radiolabeled 9mer insert (also not run with optimum TIS insert concentrations), a large intermediate band forms (Figure 5.3, lane M). This expected band is intermediate formed by the insert attaching to the 3’-end of the ribozyme (shown in Figure 5.7, 3a). The formation of this large intermediate band is greatly inhibited when 3’-end radiolabeled 9mer is run with rPC-ωA (Figure 5.3, lane O). Likewise, this inhibition also occurs when running the reaction with 3’-end radiolabeled 9mer-dU (Figure 5.3, lane N). These results indicate that the 9mer loses the 3mer (5’GCU) from its 5’-end and the 3’-portion of the insert becomes ligated to the 3’-end of the ribozyme in the process (Figure 5.7).

Taken together, these results suggest that the 9mer insert binds the IGS of the ribozyme, forming a P1i helix. The P1i helix positions the G-U wobble pair at the 5’-splice site, and the ωG on the ribozyme performs the first nucleophilic attack at this position.


Once the insert fragment is attached to the 3’-end of the ribozyme, the 12mer starting material binds the ribozyme and forms a P1 helix. This second P1 helix formation presumably requires a conformational shift. The 3’-G on the insert now acts like the ωG of the ribozyme (called ωGi) and can attack the G-U wobble pair at the new 5’-splice site in the starting material (Figure 5.7, 3a).

The importance of the ωGi on the 9mer insert was tested by changing the ωGi to an ωAi (GCUCUCGUA, called 9mer-ωAi). In the second step of TIS, if the ωGi on the insert (once its attached to the ribozyme) is acting as the ωG of the ribozyme, then
changing this position to an ωA should prevent the TIS reaction. Running the TIS reaction with 5’-end radiolabeled 12mer starting material, rPC ribozyme, and 9mer-ωA gives no 18mer TIS product (Figure 5.8, lane C). For comparison, the standard TIS reaction is shown in Figure 5.8, with the 5’-end radiolabeled reaction in lane A and the 3’-end radiolabeled reaction in lane B. The reaction run with 3’-end radiolabeled 12mer starting material, rPC ribozyme and 9mer-ωA insert also shows no TIS product (Figure 5.8, lane D). Evidently, the ωGi on the insert is critical to the TIS reaction. Since the rPC ribozyme has an ωG, the 6mer (Figure 5.8, lane C) and the large intermediate (Figure 5.8, lane D) are likely due to the competing reaction depicted in Figure 5.7, 2b.

The TIS reaction was initiated at the second step of the three-step reaction by reacting 12mer starting material with an intermediate ribozyme (rPC-3) that has the insert fragment already attached to its 3’-end (Figure 5.7, frame 3a). This TIS reaction was run exactly like the standard TIS reaction, except no 9mer insert was added. As seen in Figure 5.8, lane F, the reaction of the rPC-3 ribozyme and 5’-end radiolabeled 12mer starting material produces 18mer TIS product. Since no insert is added to the reaction, the inserted sequence must be coming from the 3’-end of the ribozyme. The lower product yields are not surprising, probably because the insert sequence on the ribozyme intermediate (rPC-3) may not be positioned properly for the second nucleophilic attack without undergoing the first reaction step.

The second step of the TIS reaction depends on a nucleophilic attack at the 5’-splice site in the P1 helix formed by the 12mer starting material. This was tested by changing the ribo-U at the 5’-splice site to a deoxy-U (12mer-dU, AUGACdUAAACAU). In the reaction with 5’-end radiolabeled 12mer-dU starting material, the ribozyme intermediate rPC-3, and no 9mer insert added), no TIS product forms (Figure 5.8, lane J) and very little 5’-cleavage (6mer production) occurs. Likewise, the same reaction run with 3’-end radiolabeled 12mer-dU shows no TIS product (Figure 5.8, lane K). Clearly, changing the ribo-U to a deoxy-U at this position in the 12mer starting material (AUGACdUAAACAU) suggests that this position forms the second 5’-splice site.
These results show that the second step of the TIS reaction depends on the 12mer starting material forming a P1 helix with the ribozyme, as well as the ωGi on the insert acting as the nucleophile in the 5’-cleavage reaction.

Step 3: The 3’-U on the 5’-half of the Starting Material Attacks the ωG of the Ribozyme, Forming TIS Product.

After the second step of the TIS reaction, the 3’-half of the starting material as well as the insert region, is attached to the 3’-end of the ribozyme (Figure 5.7, 4a). The 3’-U on the 5’-half of the starting material can then mimic the second step of self-splicing and attack at the ωG (and not at ωGi) of the ribozyme. To initiate the reaction at the third reaction step, we synthesized the ribozyme intermediate rPC-4, which has the insert and 3’-half of the starting material on the 3’-end of the ribozyme (Figure 5.7, 4a). The reaction was run with 5’-end radiolabeled 6mer (AUGACU, the 5’-half of the 12mer starting material) and rPC-4 ribozyme intermediate. This reaction produces the expected 18mer TIS product (Figure 5.8, lane M). Since the 18mer product forms with only 6mer added, the product sequence must be coming from the ribozyme, as proposed (Figure 5.7, 4a). The control reaction of 5’-end radiolabeled 6mer and rPC (the standard ribozyme) (Figure 5.8, lane O), shows no 18mer product. In addition, running the reaction with rPC-4 and the 6mer with its 3’-ribo-U changed to a deoxy-U (AUGACdU called 6mer-dU) greatly reduces TIS product (Figure 5.8, lane N), indicating that the 3’-U on the 5’-half of the starting material is the nucleophile for the third reaction step.

Finally, the third step of the reaction can be inhibited by changing the ωG in the rPC-3 intermediate ribozyme to an ωA (rPC-3ωA in Table 1). The reaction utilizing 12mer starting material and rPC-3ωA initiates at the second step of TIS (Figure 5.7, 3a), but the ωA should inhibit the TIS reaction at the third step (Figure 5.7, 4a). The results show less 18mer product forms (Figure 5.8, lane G), as expected, compared to the reaction with rPC-3 (Figure 5.8, lane F). This result suggests that the ωG in the ribozyme is not only required for the first step but is important for the third step of the TIS reaction as well.
From these results, it appears that the 3’-U on the 5’-half of the starting material attacks at the $\omega$G in the ribozyme to form the TIS product.

**P9.0 Formation Inhibits the TIS Reaction**

The TIS reaction was also run with the native *P. carinii* intron sequence (called rPC-1). In reactions with rPC-1, 18mer product yields were 6-fold lower than with rPC (11.0±0.8% compared to 62.9±1.3%). The only difference between the rPC-1 and rPC ribozymes is three nucleotides at the 3’-end of the ribozymes (Figure 5.9). This difference allows only rPC-1 to form a P9.0 helix. Apparently P9.0 formation inhibits the TIS reaction, perhaps because it inhibits the proposed conformational shift that occurs between the first and second steps, essentially halting the reaction at the second step.
Discussion

*The P. carinii Group I Intron-Derived Ribozyme Catalyzes the TIS Reaction*

We report that a *P. carinii* ribozyme can insert one RNA substrate into another (the trans insertion-splicing reaction), in a yield of up to 63%, which is surprisingly high for a three-step mechanism.

*The TIS Mechanism*

All evidence suggests there are three reaction steps in forming the 18mer TIS product (Figure 5.7). Step 1: The 9mer binds the IGS of the ribozyme forming a P1i helix, and the ωG on the ribozyme then attacks at the 5’-splice site, giving GCU + rPC-ωGCUCGUG + AUGACUAAACAU (Figure 5.7, 2a). This reaction step appears to be the same as the reverse of the second step of group I intron self-splicing (70), and has been seen in other ribozyme reactions (16, 83-85). Step 2: The 12mer starting material displaces the insert fragment and forms a second P1 helix. The ωGi from the 9mer insert (attached to the 3’-end of the ribozyme) attacks at the G-U wobble pair at the second 5’-splice site, giving AUGACU + rPC-ωGCUCGUGAAACAU (Figure 5.7, 3a). This step is similar to the first reaction step (described above). Step 3: The free 3’-OH on the 3’-U of AUGACU can then attack at the ωG of the ribozyme intermediate (rPC-ωGCUCGUGAAACAU), releasing the 18mer TIS product AUGACUCUCGUGAAACAU (Figure 5.7, 4a). This reaction step appears to be the same as the second step of group I intron self-splicing (70). A competing reaction occurs when the 12mer starting material forms the P1 helix first (before the insert forms the P1i helix). This reaction results in the formation of 6mer (AUGACU) and the 3’-end of the 12mer (AAACAU) attached to the 3’-end of the ribozyme (Figure 5.7, 2b). This ribozyme intermediate has no ωG and therefore cannot perform the TIS reaction.

The TIS reaction, along with other ribozyme reactions (16, 83-85), is dependent on the ribozyme ending in G (which is called ωG). There is evidence that the T7 RNA polymerase used to synthesize RNA frequently adds one or two template-independent nucleotides onto the 3’-end of RNA transcripts (86). There is no simple way to ensure
that extra nucleotides are not being adding to our ribozymes. Nevertheless, when we synthesize the ribozymes ending in ωA, it substantially prevents reactivity (Figure 5.3, lane C). Although template-independent nucleotides may be added to the TIS ribozymes, it appears they are not predominantly Gs. This agrees with other studies that show the extra nucleotides added by T7 RNA polymerase tend to be mostly As and Cs (86, 87).

Factors that Influence the TIS Reaction

The TIS reaction was optimized for both yield and rate. The optimum reaction conditions are 10 mM MgCl₂, 200 nM rPC ribozyme, and 1 μM 9mer insert, using 1nM radiolabeled starting material (Figure 5.5). These optimum reaction conditions give a 200:1000:1 ratio of ribozyme, insert, and starting material. The relative proportions of these reaction components are fairly specific, as decreasing or increasing the concentrations of insert or ribozyme gives decreased TIS product yields. At ribozyme concentrations higher than 200 nM, the amount of 18mer product decreases (Figure 5.5). This decrease could be due to the insert and starting material not binding the same ribozyme. The product yield might increase if the insert concentration was likewise increased. Another possibility is that higher concentrations of ribozyme may lead to TIS product breakdown, although our time studies do not support this possibility.

TIS reactions run with insert concentrations higher than the optimum (1 μM) also show a decrease in 18mer product formation. This could be due to the 12mer starting material not competing as well for binding the IGS in the presence of increased 9mer insert concentration.

The optimum MgCl₂ range is fairly narrow, with 10-14 mM MgCl₂ being optimum (Figure 5.5). Decreased TIS product at higher MgCl₂ concentrations is most likely due to the increased strength of binding of the substrates to the ribozyme, since Mg²⁺ ions allow for tighter binding. Therefore, the delicate balance between formation of the P1i and the P1 helices is likely disrupted at higher MgCl₂ concentrations.

The reactions were optimized for rates (Figure 5.6) by performing time studies at different MgCl₂ concentrations. The optimum reaction time for TIS is two hours and no degradation of the TIS product is seen at extended reaction times. This indicates that the TIS product is stable once it forms. Obtaining the same observed rate constant at 10, 14,
and 18 mM MgCl₂ concentrations indicates that the TIS reaction is optimized for rate. The yield and the observed rate constant were half the optimum values when the TIS reaction was run at 6 mM MgCl₂. This indicates that reactions run with less than the optimum MgCl₂ concentration may not allow the intron to fold properly, leading to decreased yields and rates. Obtaining an optimum rate but a decreased yield at 18 mM MgCl₂ suggests that too high MgCl₂ concentrations affect primarily the folding and not the chemistry of the reaction.

We can compare the observed rate constants for the TIS and TES reactions. The observed rate constant for the TIS reaction (0.04 min⁻¹) is 80-fold lower than for the TES reaction (3.2 min⁻¹). This is not surprising since the TIS reaction binds two exogenous substrates in succession and proceeds through three catalytic steps. In comparison, the TES reaction binds one exogenous substrate and proceeds through two catalytic steps. In addition, large conformational changes are likely necessary for the TIS reaction that are not necessary for the TES reaction.

Also, it appears that P9.0 formation inhibits the TIS reaction. The TIS reaction mechanism requires a conformational shift between the first and second nucleophilic attacks to disrupt the P1i helix and form the P1 helix. In group I introns, the P9.0 helix forms with the two nucleotides preceding ωG of the intron and can help position the ωG into the GBS (35-37). In addition, the GBS in the ribozyme binds both the exogenous G and ωG for the first and second nucleophilic attacks, respectively. Sullenger has shown that a *Tetrahymena* group I intron-derived ribozyme has different affinities for binding either the exogenous G or the ωG in the GBS between the two steps of trans-splicing (80). We suspect that a similar change in affinity for binding ωG and ωGi in the GBS is important for the TIS reaction. In our proposed mechanism, ωG on the ribozyme interacts with the GBS in the first step of the TIS reaction, although no P9.0 forms with the rPC ribozyme (Figure 5.9). After the first step of the TIS reaction, the insert is attached to the ribozyme and the formation of a P9.0 helix at this point could help position the ωGi of the insert in the GBS for the second step of the TIS reaction (see Figure 5.7, 3a). If the ribozyme forms a P9.0 in the first reaction step, this could inhibit the second step of the TIS reaction by preventing the ωGi from interacting with the GBS. The 6-fold decrease in product yields for the TIS reaction with the rPC-1 ribozyme
(compared to the rPC ribozyme with no P9.0) suggests that P9.0 formation in the ribozyme does indeed inhibit the TIS reaction.

Comparison with Previous Studies

The TIS reaction involving the direct insertion of one RNA substrate into another RNA by a group I intron-derived ribozyme has not been previously reported. The TIS reaction is the reverse of the TES reaction, where a segment is removed from the middle of an RNA substrate (6), although it is not the reverse in terms of the mechanism. Another ribozyme reaction performed with an engineered twin hairpin ribozyme results in replacement of one RNA sequence with a longer sequence, which is essentially an “insertion” product (88). The TIS reaction differs from this twin ribozyme reaction by the type of ribozyme used and also the reaction mechanism. The twin ribozyme reaction is much less efficient, taking 30 hours and producing 30% product compared to the TIS reaction that is complete in 2 hours with product yields around 60%.

Other ribozyme reactions involve the insertion of large intronic sequences into RNA transcripts. The reverse-splicing reaction is when a group I intron splices back into an RNA transcript (89). Group II introns can act as mobile genetic elements by splicing out of an RNA transcript, and then reverse-splicing into DNA (reviewed in (90)). These reactions differ from the TIS reaction because the inserted fragment is an exogenous substrate in TIS, not the intron itself.

Other group I intron-derived ribozyme reactions have taken advantage of ωG. A Tetrahymena intron was found to act as an enzyme by using its ωG to cleave and rejoin pentacytidylic acid, synthesizing a polycytidylic acid product (16). A recombination reaction performed by Lehman (85) and a polymerization reaction by Burke (83, 84) show an ωG in the ribozyme catalyzing a nucleophilic attack similar to the first step of the TIS reaction. In the recombination reaction, the substrates AB and CD form CB and AD (85), but do not continue polymerizing because there is no ωG on the resultant products. In the polymerization reaction, the substrate AB can form ABB, ABBB, and A(B)n (84) because the 3’-G on the end of B allows it to mimic the 3’-end of the ribozyme. The TIS reaction occurs in part because one substrate has a 3’-G (ωGi on the
insert) and one does not (starting material). This setup allows for insertion, but not a repeating polymerization reaction.

Implications: Mechanism

The TIS reaction is a new group I intron-derived ribozyme reaction, and it demonstrates unexpected molecular recognition interactions between the *P. carinii* group I intron-derived ribozyme and its RNA substrates. There is an added complexity in the TIS reaction compared to other group I intron-derived ribozyme reactions, since the TIS mechanism is dependent on the ribozyme binding two different substrates and forming the P1 helix twice (called P1 and P1i). Therefore, a crucial conformational shift must occur between the first and second reaction steps (Figure 5.7, between 2a and 3a) to disrupt the first P1i helix (formed with the 9mer insert) and allow formation of the second P1 helix (formed with the 12mer starting material). Each substrate can base pair with six nucleotides of the ribozyme’s IGS, suggesting that the relative binding strength of the substrates is critical for allowing the correct mechanism to occur. Another added complexity to this reaction is that the GBS appears to be used three different times in the reaction; binding $\omega G$ from the ribozyme for the first step, $\omega Gi$ from the insert for the second step, and the $\omega G$ from the ribozyme again for the third step.

The *P. carinii* ribozyme recognizes two TIS substrates; initially and primarily through base pairing. In addition, it is suspected that the recognition elements of the ribozyme can be changed to target other substrates, as seen with the *P. carinii* ribozyme in the TES reaction (6). If the P1i helix is too strong compared to the P1 helix, however, the starting material might not be able to compete off the insert from the IGS of the ribozyme. This would essentially stop the reaction after the first step. Conversely, if the P1 helix is too strong, it might preferentially bind first and inhibit formation of P1i, leading to the pathway in Figure 5.7, 2b. Apparently, the balance between formation of the two P1 helices may make targeting the TIS reaction to different substrates more complex compared to simpler group I intron-derived ribozyme reactions. Nevertheless, our mechanism does not preclude retargeting these ribozymes.
Implications: Potential Applications

The TIS reaction could potentially be useful as a biochemical tool, for example to insert a sequence, perhaps even one containing a modified nucleotide or a marker, into a large RNA transcript. Large RNA transcripts are typically made \textit{in vitro} by T7 run-off transcription, so adding site specific modifications is very arduous (91). Using the TIS reaction, a small RNA insert could be synthesized (perhaps with a desired modification), targeted to an exact location, and inserted into a large transcript. The TIS reaction could also be used as an RNA repair agent. Deletion and frame shift mutations could potentially be repaired by insertion of an RNA sequence at a specific location. The need for the ribozyme to bind two separate substrates, however, is an anticipated complication for the TIS reaction \textit{in vivo}. This can easily be overcome, fortunately, by using the TIS ribozyme with the first intermediate already attached, as seen in step 3a of Figure 5.7. This particular reaction does work (Figure 5.8, lane F), requires one substrate (instead of two), and proceeds through only two reaction steps (instead of three).

Implications: Cellular Function

The ability of group I intron-derived ribozymes to catalyze the TIS reaction also has implications for potential function of the intron once it splices out of its transcript in the cell. Since both TES and TIS ribozymes are largely unaltered from spliced introns, we speculate that at one point in evolution (and perhaps even today), introns could be shuffling RNA information by the TIS and TES reactions. Indeed, we have recently shown that the TES reaction can occur in an artificial system \textit{in vivo} by delivering a TES ribozyme into a prokaryotic cell (92).
# Table 5.1. Ribozyme, starting material, and insert sequences.

<table>
<thead>
<tr>
<th>Ribozyme</th>
<th>3’-End Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>rPC</td>
<td>~~~~CUAG</td>
</tr>
<tr>
<td>rPC-ωA</td>
<td>~~~~CUAA</td>
</tr>
<tr>
<td>rPC-1</td>
<td>~~~~UGUG</td>
</tr>
<tr>
<td>rPC-3</td>
<td>~~~~CUAGCUCGUG</td>
</tr>
<tr>
<td>rPC-3ωA</td>
<td>~~~~CUAACUCGUG</td>
</tr>
<tr>
<td>rPC-4</td>
<td>~~~~CUAGCUCGUGAAACAU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Starting Material</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>12mer</td>
<td>AUGACUUAAACAU</td>
</tr>
<tr>
<td>12mer-dU</td>
<td>AUGACUAAACAU</td>
</tr>
<tr>
<td>6mer</td>
<td>AUGACU</td>
</tr>
<tr>
<td>6mer-dU</td>
<td>AUGACU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Insert</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>9mer</td>
<td>GCUCUCGUG</td>
</tr>
<tr>
<td>9mer-dU</td>
<td>GCdUCUCGUG</td>
</tr>
<tr>
<td>9mer-ωAi</td>
<td>GCUCUCGUA</td>
</tr>
<tr>
<td>7mer-minus-3’UG</td>
<td>GCUCUCG_</td>
</tr>
<tr>
<td>7mer-minus-5’GC</td>
<td>_UCUCGUG</td>
</tr>
</tbody>
</table>

The sequences highlighted by gray boxes indicate the substrates and ribozyme used in the standard TIS reaction. The bold and underlined nucleotides indicate the positions that differ from the standard reaction components. Note that rPC-3, rPC-3ωA, and rPC-4 are the intermediate ribozymes with part of the starting material and insert attached to the 3’-end. The symbol “~~~” represents the 5’-end of the ribozyme that is the same for all the ribozymes tested. The complete ribozyme sequence is shown in Figure 5.2.
Figure 5.1. The trans excision-splicing reaction and trans insertion-splicing reaction. A) The 36mer TES starting material reacts with ribozyme rPC to give the 16mer TES product and 20mer excised region. B) The 12mer starting material and 9mer insert react with ribozyme rPC to give the 18mer TIS product.
Figure 5.2. Proposed secondary structure of a *P. carinii* group I intron-derived ribozyme (rPC).
Figure 5.3. Proof for the first nucleophilic attack in the TIS reaction. Polyacrylamide gel showing reactants and products of the TIS reaction using 200 nM ribozyme, 1 μM insert, 1 nM starting material, and 10 mM MgCl₂ run for 2 h at 44°C. Lanes E and I contain 5’-end radiolabeled 6mer, 9mer, 12mer, and 18mer (sequences in Table 5.1) run as a size control. Lane L contains 3’-end radiolabeled 9mer run as a size control. Lanes A and B contain the standard TIS reaction run with 5’-end radiolabeled 12mer (lane A) and 3’-end radiolabeled 12mer (lane B). The rest of the lanes are described in the text.
Figure 5.4. Sequencing of 5’-end radiolabeled TIS product. The 18mer TIS product was isolated from a large reaction of rPC, 9mer insert, and 5’-end radiolabeled 12mer. Both synthetic 18mer product and isolated 18mer TIS product were enzymatically sequenced next to each other using U2, CL-3, T1, and B. cer. endonucleases. The dotted line shows the position of the inserted region (CUCGUG) between the 5’ and 3’-ends of the 12mer starting material. Nuclease U2 is specific for adenosine, CL-3 for primarily cytidine, T1 for guanosine, and B. cer. for primarily cytidine and uridine. Note that the intensity of the bands in the boxed regions was enhanced relative to the rest of the gel for easier visualization.
Figure 5.5. TIS optimization graphs. Except for the changing variable, TIS reactions were run under the optimum reaction conditions of 200 nM ribozyme (rPC), 1 μM insert (9mer), 1 nM 5’-end radiolabeled starting material (12mer), and 10 mM MgCl₂ for 2 h at 44°C. The standard deviations for the data in each graph were typically under ten percent.
Figure 5.6. A representative time study gel and observed rate constants for the TIS reaction. The polyacrylamide gel shows reactants and products of the TIS reaction using 200 nM ribozyme (rPC), 1 μM insert (9mer), 1 nM 5’-end radiolabeled starting material (12mer), and 10 mM MgCl₂ run over 3 h at 44°C. The graph shows the average amount of 18mer TIS product formed over time for reactions run with 6, 10, 14, and 18 mM MgCl₂.
Figure 5.7. Proposed mechanism for the trans insertion-splicing reaction. The rPC ribozyme is represented by black lines and the IGS sequence and ωG are shown. 1) The rPC ribozyme binds the 9mer insert (pathway “a”). Alternatively, the 12mer starting material (pathway “b”) can bind the ribozyme to give dead-end products. 2a) The insert (in black and green) forms a P1i helix with the IGS of the ribozyme. The ωG of the ribozyme attacks at the 5′-splice site in the 9mer insert. 3a) The 12mer (in pink and blue) displaces the insert fragments from P1i and forms a second P1 helix. A nucleophilic attack by the ωGi on the insert (attached to the ribozyme) occurs at the 5′-splice site in the 12mer starting material. 4a) A nucleophilic attack of the 3′-U from the 5′-half of the starting material (in pink) at the ωG in the ribozyme produces the 18mer TIS product (inserted region in green between the pink and blue halves of the 12mer starting material). 2b) This alternative pathway involves the 12mer starting material forming the P1 helix first. The ωG on the ribozyme can attack at the 5′-splice site in the 12mer, leading to 6mer product (in pink) and the 3′-half of the 12mer attached to the ribozyme (in blue).
Figure 5.8. Proof for the second and third nucleophilic attacks in the TIS reaction. Polyacrylamide gel showing reactants and products of the TIS reaction using 200 nM ribozyme, 1 μM insert, 1 nM starting material, and 10 mM MgCl₂ run for 2 h at 44°C. Lanes E, I, and L contain 5’-end radiolabeled 6mer, 9mer, 12mer, and 13mer (sequences in Table 5.1) run as a size control. Note that reactions in lanes F, G, H, J, K, M, N, and O did not contain insert since the insert sequence was attached to the ribozyme intermediate. Lanes A and B contain the standard TIS reaction run with 5’-end radiolabeled 12mer (lane A) and 3’-end radiolabeled 12mer (lane B). The rest of the lanes are described in the text.
Figure 5.9. Secondary structure at the 3’-end of the rPC-1 and rPC ribozymes. The rPC-1 ribozyme is the exact sequence of the *P. carinii* group I intron, capable of forming a P9.0 helix. The P9.0 helix positions the ωG on the ribozyme into the G binding site in the ribozyme. The rPC ribozyme used in the TIS reaction differs from rPC-1 by three nucleotides and cannot form a P9.0 helix.

Copyright © Ashley Kirtley Johnson
CHAPTER SIX - Conclusions

The ability of group I introns to splice out of RNA transcripts depends on molecular recognition interactions of the 5’ and 3’-exons in the catalytic core, which helps position the exons for catalysis. Likewise, ribozymes derived from these group I introns also depend on molecular recognition of their substrates to perform catalytic reactions. These ribozymes no longer have their exons attached, so their substrates are exogenous. Presumably, the molecular recognition interactions of these exogenous substrates with the ribozyme are even more critical for correct positioning and subsequent catalysis.

Three separate studies were performed for this dissertation work. Four different ribozyme reactions were evaluated, one of which (the TIS reaction) was newly discovered and its mechanism elucidated. In general, these studies have shown that the molecular recognition interactions between the \textit{P. carinii} ribozyme and its substrates are more complex than previously thought.

\textit{Molecular Recognition in the Suicide Inhibition and Reverse Cyclization Reactions}

The combinatorial assay described in Chapter Three has proven useful for gaining a more fundamental understanding of the molecular recognition properties of \textit{P. carinii} IGS-mediated reactions. We used the combinatorial data to assess what sequences could bind the IGS and react in either the reverse cyclization or suicide inhibition reactions (Figure 3.1). As expected, the native 5’ exon sequence occurred most often for both the reverse cyclization and suicide inhibition reactions, although the fidelity for the two reactions differed (58% for the reverse cyclization reaction and 13% for the suicide inhibition reaction, each at 15 mM MgCl$_2$). Apparently, the molecular recognition interactions of the IGS and 5’ exon mimics in the reverse cyclization and suicide inhibition reactions are different, indicating a dynamic use of the IGS between the two reactions.

The combinatorial results in general show that the amount of non-pairing mismatches increases farther away from the 5’-splice site, suggesting the sequence specificity of these \textit{P. carinii} group I intron reactions is not that high. In addition, the
strength of overall binding of the exon to the ribozyme can change considerably simply by changing the specific base paired-sequence of the P1 helix. Even though the 5’ exon-IGS helices are complementary, the ribozyme likely forms tertiary interactions with only specific base pairs at certain P1 helix positions.

Molecular Recognition in the TES Reaction

The *Tetrahymena* ribozyme can perform the TES reaction (Figure 4.1), suggesting that the TES reaction might be a general group I intron-derived ribozyme reaction. The native *P. carinii* ribozyme is much more effective at the TES reaction than the *Tetrahymena* ribozyme. Since the first step of the TES reaction involves ribozyme-mediated 5’-splice site cleavage by a hydroxyl from the buffer, the *P. carinii* ribozyme’s ability to perform the TES reaction so well appears to be related to its enhanced ability to catalyze this reaction.

The hybrid *P. carinii* ribozyme with the *Tetrahymena* IGS has substantially decreased observed rate constants and TES product yields compared to the native *P. carinii* ribozyme. Apparently, changing the IGS to a different, yet complementary sequence, decreases the yield and rate of reaction. In addition, the native substrate sequence for the *P. carinii* intron appears to form critical tertiary interactions with the *P. carinii* ribozyme. This suggests that the *P. carinii* intron (and presumably other group I introns) have evolved to specifically recognize its native substrate sequence through complex molecular recognition interactions.

Molecular Recognition in the TIS Reaction

We report that a *P. carinii* ribozyme can insert one RNA substrate into another (the trans insertion-splicing reaction), and all evidence suggests that there are three reaction steps to the 18mer TIS product (Figure 5.7). The TIS reaction involves interesting molecular recognition interactions between a *P. carinii* group I intron and its RNA substrates. The TIS mechanism is dependent on the ribozyme binding two different substrates and forming the P1 helix twice (P1 and P1i). This duel use of the P1 helix gives added complexity in the TIS reaction compared to other group I intron-derived
ribozyme reactions (5, 6), and suggests that the relative binding strength of the substrates may be critical for allowing the correct TIS mechanism to occur. Since the TIS reaction is more complex, it may be more challenging to retarget the TIS reaction to a different sequence (although not impossible), compared to the TES reaction.
References


Vita

Ashley Kirtley Johnson

Ashley Kirtley Johnson was born July 12, 1975, in Owensboro, Kentucky. In 1997, she graduated from Agnes Scott College in Decatur, Georgia with a Bachelor’s of Arts Degree in Chemistry. After college, she returned to Owensboro, Kentucky and worked for three years as an assistant scientist in the analytical laboratory at a pharmaceutical manufacturing company called Large Scale Biology. In the fall of 2000, Ashley began her graduate career in the chemistry department at the University of Kentucky, working with Dr. Stephen M. Testa. She received several fellowships including: the Daniel R. Reedy Quality Achievement Fellowship, the Paul I. Murrill Fellowship, and the Research Challenge Trust Fund Fellowship. She received the One-Hundred Percent Graduate Student Award in 2002 and served one year each as both president (2002-2003) and vice-president (2001-2002) of the chemistry graduate student association.

Publications


