AD_____

Award Number: DAMD17-03-1-0329

TITLE: Repairing RNA Transcripts that Mediate Breast Cancer Susceptibility

PRINCIPAL INVESTIGATOR: Stephen M. Testa, Ph.D.

CONTRACTING ORGANIZATION: University of Kentucky
Lexington, KY 40506-0057

REPORT DATE: August 2005

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-08-2005	Final	1 May 2003 – 31 Jul 2005

4. TITLE AND SUBTITLE

Repairing RNA Transcripts that Mediate Breast Cancer Susceptibility

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0329

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Stephen M. Testa, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Kentucky
Lexington, KY 40506-0057

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

A number of genetic mutations that predispose individuals to cancer are known. For example, over 25 insertion mutations have been identified in BRCA1 and p53 transcripts that have been linked to breast cancer susceptibility. Each mutation impairs the functionality of the resultant tumor suppressor protein. Developing technology to specifically excise these mutations, thus restoring tumor suppressor activity, would be of considerable importance for the development of new molecular-based therapeutics. To this end, we have developed a novel biomolecule (a ribozyme) that can specifically excise regions from RNA transcripts. In this work, we designed a ribozyme that excises an insertion mutation that is linked to breast cancer predisposition from a short mimic of the p53 transcript in a cell-free system. We have analyzed the molecular recognition properties of this ribozyme binding substrate, which has aided our design of more effective ribozymes. Furthermore, we developed a Green Fluorescent Protein system to demonstrate that such ribozymes can excise designated insertion mutations from transcripts within bacterial cells. This ribozyme is not toxic to the cell, and appears to work with a reasonable degree of specificity. The successes in this study represent the foundation for a completely novel, yet simple molecular-based therapeutic strategy: repair disease-causing RNA transcripts by removing their insertion mutations, thereby restoring the proper function of the resultant biological products.

15. SUBJECT TERMS
Drug Development; tumor suppressor genes; development and character of model systems; complementary and alternative treatment approaches

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	30	

Table of Contents

INTRODUCTION

The development of technologies for identifying genetic mutations that lead to diseases is rapidly outpacing the development of therapeutic strategies to remediate the effects of these mutations. This work directly addresses this discrepancy by analyzing a new potential therapeutic strategy. To this end, we are analyzing the potential of a novel catalytic RNA, which we previously developed, to excise a single base insertion mutation from a transcript derived from the p53 gene. This mutation impairs the functionality of the resultant tumor suppressor protein product, leading to breast cancer susceptibility. The development of technology to specifically excise insertion mutations, which in this case would restore proper tumor suppressor activity, could be of considerable importance for the development of new, molecular based genetic therapeutics.

BODY

Task #1: Develop and analyze ribozymes that excise BRCA1 and p53 insertion mutations *in vitro*.

a. *Develop and synthesize RNAs that mimic mutant mRNAs.*
b. *Develop and synthesize ribozymes that can remove these mutations.*

We engineered two catalytic RNA ribozymes, one designed to specifically bind and excise insertion mutations from within a mutant BRCA1 transcript, and one specific for a mutation in a p53 transcript. The ribozymes were engineered via site-directed mutagenesis of a plasmid containing the *P. carinii* ribozyme by altering the recognition elements such that they base pair with the sequences flanking the desired BRCA1 and p53 mutations.

In addition, we synthesized small RNA mimics of the BRCA1 and p53 transcripts, centering on those regions flanking the mutations that are known to predispose individuals to breast cancer. In combination with one another, the ribozyme-transcript mimic combinations should change the 'Mutant' RNAs, as shown below, to the 'Fixed' RNAs.

U insertion in codon 2731 of BRCA1gene
Normal GCU CCG UUU - Ala - Pro - Phe - - -
Mutant GCU CCU **g**UU U - Ala - Pro - Gly → Stop
Fixed GCU CCU UUU - Ala - Pro - Phe - - -
End Result: Remove g (bold) and restore reading frame.

U insertion in codon 257 of p53 gene
Normal ACA CUG GAA - Thr - Leu - Glu - - -
Mutant ACA CUU **g**GA A - Thr - Leu - Gly → Stop
Fixed ACA CUU GAA - Thr - Leu - Glu - - -
End Result: Remove g (bold) and restore reading frame.

c. *Develop and perform assays in vitro to test for the excision of the designated mutations.*

Based on previous *in vitro* analysis of generic trans excision-splicing (TES) ribozymes (1,2), we developed and conducted excision reactions using the BRCA1 and p53 systems. In these cases, we radiolabeled the transcript mimics and followed the reactions by visualization on polyacrylamide gels, using migration distance as a marker for product size. Unfortunately, the BRCA1 specific ribozyme was not able to cut out the BRCA1 mutation in this *in vitro* system, even though many sequence variants of the ribozyme were analyzed. Fortunately, as can be seen in Figure 1, the 18 nucleotide p53 transcript mimic becomes the expected 17-mer product when the p53-specific ribozyme is added. In the optimized case, at 1 hour in 2 mM $MgCl_2$, approximately 50% of the transcript mimics are corrected via the excision of the breast cancer causing mutation.

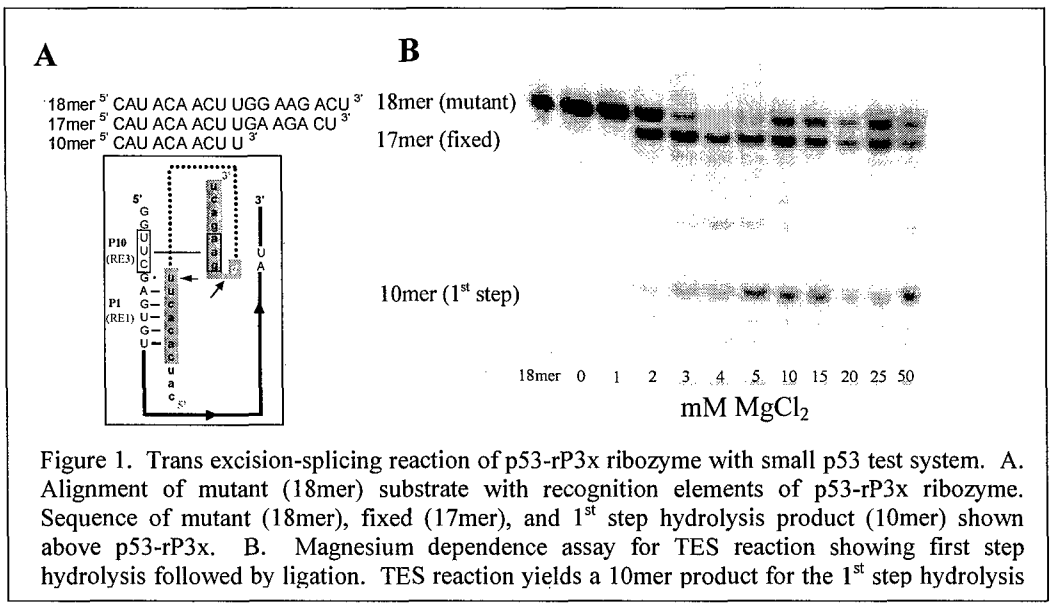

Figure 1. Trans excision-splicing reaction of p53-rP3x ribozyme with small p53 test system. A. Alignment of mutant (18mer) substrate with recognition elements of p53-rP3x ribozyme. Sequence of mutant (18mer), fixed (17mer), and 1st step hydrolysis product (10mer) shown above p53-rP3x. B. Magnesium dependence assay for TES reaction showing first step hydrolysis followed by ligation. TES reaction yields a 10mer product for the 1st step hydrolysis

We originally proposed two model systems for the *in vitro* studies (BRCA1 and p53), to enhance our chances that one system would work. This turned out to be wise, as the BRCA1 ribozyme appears to be inactive. However, the encouraging data on the p53 system lead us to continue these studies in this system (only). We exploited this data to publish a manuscript (added to the appendix) that reports the study of the how such a ribozyme sequence-specifically is able to recognize and bind its target (3). Although we attempted to analyze the excision reaction in a complete p53 transcript, the data were too difficult to interpret. The reason is that such a long mutant transcript will only differ from the fixed version by a single nucleotide, and so we could not visualize the difference on a polyacrylamide gel (as done above), and RT-PCR gave inconclusive results. We therefore decided that we needed a visual marker for transcript correction, ultimately at the protein (translated) level. Because the ultimate goal is to get the ribozyme to work *in vivo*, we designed a system whereby correction of the mutation would lead to cell fluorescence. This would therefore simultaneously address the question of *in vivo*

reactivity, ribozyme specificity for the particular mutation, and ribozyme ability to target a a full-length RNA transcript. This strategy is outlined below in Task #2.

Task #2 Develop and analyze ribozymes that excise BRCA1 and p53 insertion mutations *in vivo*.

a. *Develop and synthesize shuttle vectors for expressing mutant genes and the repair ribozymes in the bacterium E. coli.*
b. *Develop and perform assays in vivo to test for the excision of the designated mutations in this cell system.*

To develop the *in vivo* system to test excision-repair, we first analyzed a ribozyme that would fix a single base insertion mutation in green fluorescent protein (GFP) *in vivo*. In this way, we can analyze *in vivo* excision, as it is tied with fluorescence as a visual marker. The resultant manuscript (added to the appendix) outlines the first ever excision of a single base mutation from full-length transcripts (in this case GFP) *in vivo* (4), which now permits us to test the p53-specific ribozyme. Note that the same plasmids can be used to test ribozyme activity in mammalian cells, as we also have a complementary plasmid that expresses T7 RNA polymerase in mammalian cells.

In order to test whether the p53-specific ribozyme can excise out the particular cancer-causing mutation, we have linked a small portion of the p53 gene with the GFP gene, such that when the breast cancer mutation is in the p53 region of the transcript, GFP is inactivated and the cell does not fluoresces. After trying many GFP insertion sites unsuccessfully, we found that inserting the mutant p53 region between amino acids 172 and 173 of GFP still allowed GFP fluorescence. When the ribozyme is expressed and the cancer-causing mutation is excised, GFP is then re-activated, and success in the reaction can be seen and quantified via cellular fluorescence. A simple diagram of this idea is shown below in Figure 2. We have successfully synthesized such plasmids, containing both the p53-GFP 'fusion' transcript and the p53 specific ribozyme, and are about to test their *in vivo* activity.

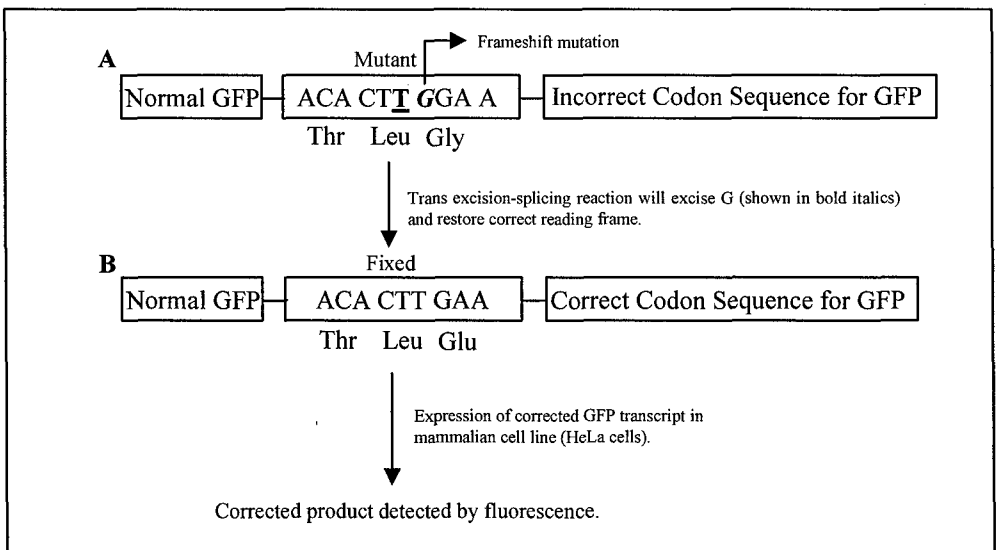

Figure 2. Trans excision-splicing reaction utilizing green fluorescent protein as a visualization marker for correction of a mutant transcript in vivo (mammalian). A. Insertion of small p53 system (ACA CTT GGA A) within coding sequence of green fluorescent protein (GFP). Insertion of T (underlined in bold) causes a frameshift mutation that yields incorrect amino acid sequence of GFP to be expressed which leads to a non-fluorescent protein. B. Trans excision-splicing reaction catalyzed by p53-rP3x ribozyme will excise out G (shown in bold italics in A) and will restore the normal reading frame of GFP. The corrected sequence will yield the correct amino acid sequence and a fluorescent protein will be observed.

A change to the original proposal is that two of the Ph.D. students have graduated and have been replaced on the project with new Ph.D. students. Another change, which does not affect the 'Statement of Work', is that GFP-p53 hybrid constructs will be used as cellular targets, because linking p53 transcript repair to cellular fluorescence allows for direct visualization of proper ribozyme activity. This has greatly facilitated our ability to conduct and analyze the *in vivo* experiments.

KEY RESEARCH ACOMPLISHMENTS

- The development, synthesis, and successful analysis of a catalytic RNA ribozyme system that can excise the U257 insertion mutation, which is a known predisposition indicator for breast cancer, from a small model mimic of the p53 transcript in a cell-free system.

- The optimization of reaction conditions for this ribozyme activity in a cell-free system, as well as a thorough molecular recognition analysis of substrate binding.

- The development, synthesis, and successful analysis of a catalytic RNA ribozyme test system that can excise an insertion mutation from Green Fluorescent Protein, which provides for a visual marker of *in vivo* transcript repair.

- The development and synthesis of a catalytic RNA ribozyme system for analyzing the potential of a p53-specific ribozyme to excise the U257 insertion mutation from a GFP-p53 hybrid transcript *in vivo*.

REPORTABLE OUTCOMES

1) Manuscripts:
    i. See Appendix for two published manuscripts.

2) Seminar Presentations:
    i. "Exploiting the Properties of Nucleic Acids for the Development of Novel Strategies in Biotechnology," Chemistry Departmental Seminar, University of Kentucky, Lexington, KY, November 7, 2003.
    ii. "Developing New Intron-Derived RNA Catalysts: Potential Therapeutic Strategies," Markey Cancer Center Weekly Tumor Board Conference, (Put on by the Health Sciences Learning Center - Presentation is viewable from their web site), University of Kentucky, Lexington, KY, February 5, 2004.
    iii. "Strategies for Targeting RNA Transcripts," Markey Cancer Center Experimental Therapeutics Program Seminar, University of Kentucky, Lexington, KY, April 23, 2004.
    iv. "Nucleic Acids, Disease, and You," Department of Chemistry, University of Dayton, Dayton, OH, October 15, 2004.

v. "Nucleic Acids, Disease, and You," Department of Chemistry, The University of The South, Sewanee, TN, October 28, 2004 (11:00 am).
vi. "Nucleic Acids, Disease, and You," Department of Chemistry, Berry College, Mt. Berry, GA, October 28, 2004 (5:00 pm).
vii. "Exploiting the Properties of Nucleic Acids for the Development of Novel Strategies in Biotechnology," Chemistry Department Seminar, Pennsylvania State University, State College, PA, November 9, 2004.
viii. "Nucleic Acids, Disease, and You", Department of Chemistry, Gonzaga University, Spokane Washington, February 15, 2005.
ix. "A Biochemical Introduction to Personalized Medicine", Department of Chemistry, East Stroudsburg University, East Stroudsburg PA, April, 21, 2005.
x. "A Biochemical Introduction to Personalized Medicine", Department of Biology, Emmanuel College, Boston, MA, April 28, 2005.
xi. "The Rational Design of Novel Catalytic RNA Reactions," Chemistry Departmental Seminar, West Virginia University, Morgantown West Virginia, June 21, 2005.

3) Meeting Abstracts and Talks
   i. Dana A. Baum, Joy Sinha, & Stephen M. Testa, "Molecular Recognition of 5' and 3' Splice-Sites in a Trans Excision-Splicing Reaction". 9th Annual Meeting of the RNA Society. University of Madison, WI. June 1 – June 6, 2004.
   ii. Dana A. Baum and Stephen M. Testa "*In Vivo* Excision of a Single Targeted Nucleotide from an mRNA by a Trans Excision-Splicing Ribozyme". 31st Annual Naff Symposium, University of Kentucky, Lexington KY 40506. April 15, 2005.
   iii. Dana A. Baum and Stephen M. Testa. "Repairing RNA Transcripts Linked to Lung Cancer". 3rd Annual Kentucky Lung Cancer Research Program Scientists' Seminar. Louisville KY 40299. May 2, 2005.
   iv. Dana A. Baum and Stephen M. Testa "*In Vivo* Excision of a Single Targeted Nucleotide from an mRNA by a Trans Excision-Splicing Ribozyme" Invited Talk: 10th Annual Meeting of the *RNA* Society. Bamf, Canada, May 28, 2005.
   v. Stephen M. Testa, Dana Baum, Patrick Doston, and Joy Sinha. "Designing a Novel Strategy for Repairing RNA Transcripts that Mediate Breast Cancer Susceptibility". Era of Hope 2005 Department of Defense Breast Cancer Research Program Meeting, Philadelphia PA, June 8-11, 2005.

4) Degrees awarded partially supported by this award:
   i. 3 doctorates obtained

5) Funding applied for based on work supported by this award:
   i. National Institutes of Health "Mechanisms of Group I Intron Ribozyme Reactions", pending.

6) Employment opportunities applied for and/or received based on training supported by this award:
    i. 3 postdoctoral fellowships obtained

CONCLUSIONS

A wide variety of heritable genetic mutations that predispose individuals to particular cancers are known. Insertion mutations, for example, cause frameshift mutations that often lead to the synthesis of inactive and truncated proteins. At least 25 known insertion mutations have been identified in the genes BRCA1 and p53, which have been linked to breast cancer susceptibility. Each mutation impairs the essential functionality of the resultant tumor suppressor protein products. The development of technology to specifically excise these mutations, thus restoring proper tumor suppressor activity, would be of considerable importance for the development of new, molecular based genetic therapeutics. Our research directly addresses this problem by developing and testing a novel therapeutic strategy for the specific removal of RNA transcript mutations that have been linked to breast cancer predisposition. Using a short transcript model of the p53 gene in a cell free system, we have been able to design a molecule that can excise an insertion mutation that is known to predispose one to breast cancer. Using a model fluorescing system, we have been able to design a molecule that can specifically excise an insertion mutation from a full-length transcript in a simple bacterium. This ribozyme is not toxic to the cell, and appears to work with some degree of specificity. Although there is much work to be done, including targeting p53 transcripts, or mimics thereof, *in vivo* and in mammalian cells, we have developed the first molecules (of any kind) that can target insertion mutations in a cell and specifically excise them, thus repairing the transcripts. Indeed, these mammalian studies are well underway. This work brings the idea of fixing insertion mutations that lead to genetic disease into the realm of possibility, and will be the basis for many translational research projects.

REFERENCES

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

2) Bell, M. A., Sinha, J., Johnson, A. K., & Testa, S. M. (2004) "Enhancing the Second Step of the Trans Excision-Splicing Reaction of a Group I Ribozyme by Exploiting P9.0 and P10 for Intermolecular Recognition" *Biochemistry* 43, 4323-4331.

3) Baum, D. A., Sinha, J., and Testa S. M. (2005) "Molecular Recognition in a Trans Excision-Splicing Ribozyme: Non-Watson-Crick Base Pairs at the 5' Splice Site and ωG at the 3' Splice Site Can Play a Role in Determining the Binding Register of Reaction Substrates" *Biochemistry* 44, 1067-1077.

4) Baum, D. A. and Testa, S. M. (2005) "*In Vivo* Excision of a Single Targeted Nucleotide from an mRNA by a Trans Excision-Splicing Ribozyme" *RNA* 11, 897-905.

APPENDIX

Molecular Recognition in a Trans Excision-Splicing Ribozyme: Non-Watson−Crick Base Pairs at the 5′ Splice Site and ωG at the 3′ Splice Site Can Play a Role in Determining the Binding Register of Reaction Substrates†

Dana A. Baum,‡ Joy Sinha,‡ and Stephen M. Testa*

Department of Chemistry, University of Kentucky, Lexington, Kentucky 40506

Received August 16, 2004; Revised Manuscript Received October 13, 2004

ABSTRACT: Trans excision-splicing (TES) ribozymes, derived from a *Pneumocystis carinii* group I intron, can catalyze the excision of targeted sequences from within RNAs. In this report, the sequence requirements of the splice sites are analyzed. These conserved sequences include a u−G wobble pair at the 5′ splice site and a guanosine in the ω position at the 3′ splice site (in the substrate). We report that 7 out of 16 base pair combinations at the 5′ splice site produce appreciable TES product. This promiscuity is in contrast to results reported for analogous self-splicing reactions using a *Tetrahymena* ribozyme. At long reaction times TES products dissociate and rebind free ribozyme, at which point product degradation occurs via the 5′ cleavage reaction. Unexpectedly, only in cases where Watson−Crick base pairs form at the 5′ splice site do we see degradation of TES products at cryptic sites, suggesting that non-Watson−Crick base pairs at the 5′ splice site are acting in concert with other factors to precisely determine the binding register of TES reaction substrates within the ribozyme. Moreover, cryptic site degradation does not occur with the corresponding reaction substrates, which additionally contain ωG, suggesting that ωG can play a similar role. We report that ωG cannot be replaced by any other base, so TES substrates require a guanosine as the last (or only) base to be excised. Additionally, we demonstrate that P9.0 and P10 are expendable for TES reactions, suggesting that ωG is sufficient as a 3′ molecular recognition element.

We have previously reported that a group I intron-derived ribozyme from *Pneumocystis carinii* can catalyze the excision of a targeted sequence from within an RNA transcript (*1*). The reaction, called the trans excision-splicing (TES)[1] reaction, consists of two steps (Figure 1): 5′ cleavage followed by exon ligation. In the 5′ cleavage reaction, a nucleophile from the aqueous solvent cleaves the phosphodiester backbone of a substrate, generating 5′ and 3′ exon intermediates. In the exon-ligation step, the newly available nucleophilic 5′ exon attacks a specific base within the 3′ exon intermediate, simultaneously ligating the exon intermediates and excising an internal segment. This reaction is analogous to the self-splicing reaction, wherein the intron itself is excised, but for which the actual chemical steps are similar. Note that one key difference is that the 5′ cleavage reaction in self-splicing uses a guanosine cofactor as the nucleophile (the G-addition reaction), whereas TES reactions appear to use a solvent water molecule (the hydrolysis reaction). Also, we have shown that TES ribozymes can excise a variety of sequences, including those where the excised segment is as small as a single nucleotide (*1*). In addition, ribozyme regions that dictate the molecular recognition between the ribozyme and its substrate can be modified, via sequence changes (*1*) or via rational ribozyme redesign [of helices P1, P9.0, and P10] (*2*), to direct the ribozyme to effectively act upon chosen targets in vitro.

Because of the similarity between the TES and self-splicing reactions, and that the body of information available regarding the self-splicing reaction is relatively large, especially concerning a ribozyme from *Tetrahymena thermophila*, we can infer many things regarding the TES reaction. For example, a u−G pair at the −1 position of helix P1 is highly conserved in group I introns (*1, 3*−*8*), and it has been shown to play a pivotal role in defining the site of 5′ cleavage. Similarly, the last base in all group I introns is a guanosine, called ωG, and it has been shown to be a critical component of the exon-ligation reaction (*9*−*15*). For the TES reaction, this means that a uridine in the target, which forms the u−G wobble pair in helix P1 upon binding the ribozyme, would have to immediately precede the excised region. Moreover, the last base of the excised region (the only base if excising a single base) would have to be a guanosine, which corresponds to ωG. Therefore, the sequences that can help define the splice sites in the TES reaction are expected to be limited by these factors. For this reason, our previous studies regarding the molecular recognition of *P. carinii* ribozymes, catalyzing both TES (*1, 2*) and suicide inhibition (*16, 17*) reactions, did not entail altering these two sequence elements. Note that throughout this report, lowercase nucle-

† This work was supported by the Muscular Dystrophy Association, the DOD Breast Cancer Research Program DAMD17-03-1-0329, and the Kentucky Research Challenge Trust Fund.

* To whom correspondence should be addressed. Tel: (859)-257-7076. Fax: (859)-323-1069. E-mail: testa@uky.edu.

‡ The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

[1] Abbreviations: TES, trans excision-splicing; IGS, internal guide sequence; RE1, recognition element #1; RE2, recognition element #2; RE3, recognition element #3; GBS, guanosine binding site.

10.1021/bi0482304 CCC: $30.25
Published on Web 12/29/2004

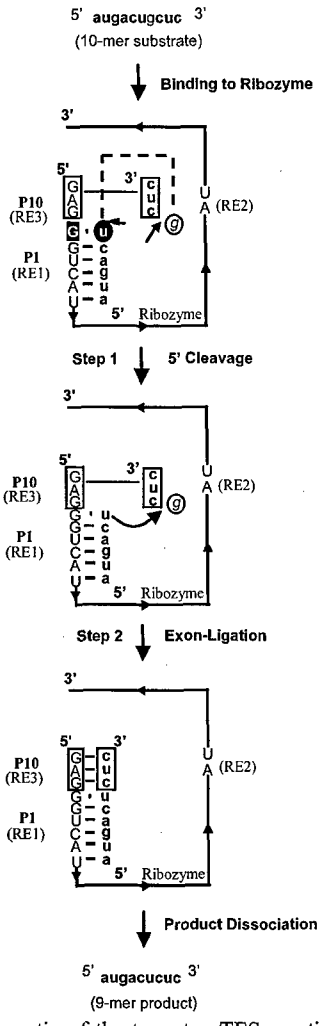

FIGURE 1: Schematic of the two-step TES reaction. The rP-8/4x ribozyme is in uppercase lettering, the 10-mer substrate is in lowercase lettering, and the single guanosine nucleotide to be excised is circled and in italics. The base to be excised corresponds to the ω position when it is part of self-splicing introns, and so will be referred to as the substrate ω position. The ribozyme recognition elements RE1, RE2, and RE3 base pair with the substrate to form the P1, P9.0, and P10 helices, respectively. Note that the P9.0 helix does not form in this system because of a lack of complementary bases. The sites of catalysis for the first step (5′ cleavage) and the second step (exon ligation) are shown with large bold arrows in the uppermost diagram. The P10 helix is boxed. The −1 position of the substrate and the 12 position of the ribozyme are shown in white lettering in the uppermost diagram and define the native 5′ splice site. Note that the diagram shows only the recognition elements of the ribozyme.

otide abbreviations refer to the substrate, uppercase refers to the ribozyme, and the 5′ splice site refers to the base pair that forms between position 12 of the ribozyme and −1 of the substrate (Figure 1).

It was previously reported, however, that c−A can be a substitute for the u−G wobble pair at the 5′ splice site in the 5′ cleavage reaction (*3*) and c−G can substitute in the self-splicing reaction (*4, 18*). It has also been reported that ωG can be changed to ωA, with either no change required to the catalytic core (*19*) or by modifying the G-binding site (GBS) of the ribozyme in a particular way to accommodate adenosine (*13, 20*). Therefore, we analyzed the TES reaction with all 16 combinations of base pairs at the 5′ splice site and all four bases at the 3′ splice site. This was undertaken to provide a more thorough understanding of what the sequence requirements are for the TES reaction. This information will be useful in terms of what sequences the ribozymes can target, what sequences the ribozymes can excise, and how specific these reactions might be (relative to each other). Note that the studies outlined here utilize RNA substrates that do not make use of P9.0 as a molecular recognition component, as it is not required (*1*). In addition, a single nucleotide, one that is analogous to the ω position in group I introns, is excised from these substrates. These design principles were employed to limit the number of reaction variables and to simplify analysis.

We now report that all 16 possible base pair combinations at the 5′ splice site undergo the first reaction step (5′ cleavage); with the u−G and c−A wobble pairs, all four Watson−Crick pairs, and the a−G pair producing appreciable amounts of the complete TES product (> 10%). This substrate promiscuity at the 5′ splice site is in contrast to results reported for analogous reactions using a *Tetrahymena* ribozyme (*3*). It was found that at long reaction times (relative to product completion), TES products dissociate and rebind the ribozyme, at which point degradation occurs via the 5′ cleavage reaction (i.e., ribozyme-mediated hydrolysis) at the new 5′ splice site. Unexpectedly, it was discovered that products that contain Watson−Crick base pairs (and no other type of base pair) at the 5′ splice site show cryptic site degradation upon rebinding free ribozyme. These results suggest that non-Watson−Crick base pairs, which suppress cryptic site degradation, can play a role in determining the binding register of TES reaction substrates. Since such cryptic site degradation does not occur with the Watson−Crick containing substrates, which additionally contain ωG, ωG is also likely playing a role in determining the binding register of TES reaction substrates. Analyses of base changes at the ω position show that the requirement for guanosine in TES reactions is absolute. Therefore, using the current incarnation of the *P. carinii* ribozyme, TES substrates require a guanosine as the last (or only) base to be excised. This requirement, however, extends only to the second step of the TES reaction (exon ligation). Last, it was found that P9.0 and P10, although beneficial, are not required for either step of the TES reaction. The implications of these results with regard to molecular recognition, as well as the potential applicability of TES ribozymes, are discussed.

MATERIALS AND METHODS

Nucleic Acid Synthesis and Preparation. DNA oligonucleotides (for site-directed mutagenesis) were purchased from Integrated DNA Technologies (Coralville, IA) and used without further purification. RNA oligonucleotides were purchased from Dharmacon Research Inc. (Boulder, CO) and deprotected following the manufacturer's recommended protocol. Designated RNA oligonucleotides were 5′-end radiolabeled with T4 polynucleotide kinase (New England Biolabs; Beverly, MA) and [γ^{32}P]-ATP (Amersham Pharmacia Biotech; Piscataway, NJ) and then purified on a 20% native polyacrylamide gel, as previously described (*2*).

The *P. carinii* ribozyme plasmid precursor, P-8/4x, was synthesized as previously described (*1*). Modifications to alter the guanosine involved in the wobble pair at the −1 position of the P1 helix and to delete the RE3 sequences that are

involved in P10 formation were made using site-directed mutagenesis (*1*). The following pairs of primers were used for altering the ribozyme at the −1 position (underlined base represents the altered bases as compared to P-8/4x): 5′CGACTCACTATAGAGCGTCATGAAAGCGGC3′ and 5′GCCGCTTTCATGACGCTCTATAGTGAGTCG3′ to create P-8/4x-5′C; 5′CGACTCACTATAGAGAGTCATGAAAGCGGC3′ and 5′GCCGCTTTCATGACTCTCTATAGTGAGTCG3′ to create P-8/4x-5′A, and 5′CGACTCACTATAGAGTGTCATGAAAGCGGC3′ and 5′GCCGCTTTCATGACACTCTATAGTGAGTCG3′ to create P-8/4x-5′U. The following primer pair was used to create P-8/4x-noP10: 5′CGACTCACTATAGGTCATGAAAGCGGC3′ and 5′GCCGCTTTCATGACCTATAGTGAGTCG3′. The site-directed mutagenesis reactions were performed as previously described (*1*), with the changes that follow. The reaction mixtures were first subjected to denaturation at 95 °C for 30 s, followed by 16 temperature cycles of 95 °C for 30 s, either 50 °C or 60 °C for 2 min, and then 68 °C for 6 min. The parental plasmids were then digested with 20 units of *Dpn*I (Invitrogen; Carlsbad, CA) in 4.2 μL of manufacturer's buffer for at least 2 h at 37 °C. A 3-μL aliquot of this mixture was then used to transform *Escherichia coli* DH5α competent cells (Invitrogen). The resultant plasmids were purified using a QIAprep Spin Miniprep kit (QIAGEN; Valencia, CA), and sequenced for confirmation (Davis Sequencing; Davis, CA).

Prior to transcription, plasmids were linearized with *Xba*I and purified from the reaction mixture using a QIAquick PCR purification kit (QIAGEN). Transcription and purification of the ribozymes were conducted under the conditions previously described (*1*).

TES Reactions. Reactions were conducted at 44 °C in H10Mg buffer, which consists of 50 mM Hepes (25 mM Na^+), 135 mM KCl, and 10 mM $MgCl_2$ (from 0 mM to 15 mM when using the rP-8/4x-noP10 ribozyme) at pH 7.5. Prior to each reaction, 166 nM ribozyme in 5.0 μL of H10Mg buffer was preannealed at 60 °C for 5 min and then slow cooled to 44 °C. The reactions were initiated by adding 1.0 μL of a H10Mg solution of 8 nM 5′-end radiolabeled or 3′-end radiolabeled substrate (*21*). The K_d of the substrate is expected to be approximately 5.2 nM, similar to the K_d of the 6-mer 5′ exon mimic (*5*). Reaction times for the TES reactions investigating the 5′ splice site were 15 min and 1 h. The 3′ splice site studies and the rP-8/4x-noP10 reactions were allowed to proceed for 1 h. Time dependence assays to determine the source of cryptic products were run from 1 to 120 min. All reactions were terminated by adding an equal volume of stop buffer (10 mM urea, 0.1X TBE, 3 mM EDTA). The urea in the stop buffer is sufficient to stop the reactions, even at the highest $MgCl_2$ concentrations used in this study, as confirmed in control reactions utilizing no EDTA in the stop buffer (data not shown). The products and reactants were denatured at 90 °C for 1 min and then run on a 12.5% polyacrylamide/8M urea gel for separation. The gel was dried on chromatography paper (Whatman 3MM CHR) under vacuum for 1 h at 70 °C. Reaction products were then visualized and quantified using a Molecular Dynamics Storm 860 PhosphorImager. All the data reported are the average of at least two independent assays.

Kinetics. The observed rate constants for the TES reactions were obtained under single-turnover, ribozyme-excess conditions. Only data from those reactions that resulted in product yields greater than 10% were found to be reliable, and so only those values are reported. These experiments were conducted under the same TES reaction conditions described above, with aliquots periodically removed and added to an equal volume of stop buffer, typically over a period of 15 min. Products and reactants were denatured at 90 °C for 1 min prior to loading on a 12.5% polyacrylamide-8M urea gel. Gels were dried under vacuum and the bands quantified using a phosphorimager. The data were fit as a single exponential (SigmaPlot, Jandel) to obtain an observed rate constant. Each observed rate constant represents the average of two independent assays. The relevant graphs are available as Supporting Information.

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

RESULTS

Molecular Recognition at the 5′ Splice Site. We synthesized four different 10-mer substrates, each containing a different nucleoside at position −1, to be utilized with four different ribozymes, each containing a different nucleoside at position 12 of the ribozyme (shown in Figure 1), allowing us to test all 16 possible base pair combinations at the 5′ splice site (Figure 2A). We chose to use our simplest TES system, in which case P9.0 is not utilized as a molecular recognition element, where a single nucleotide is excised from within the substrates. Note that this single nucleotide is analogous to ωG of the ribozyme. For these studies, we utilized reaction conditions previously optimized for the TES reaction. In addition, we used the 10-mer substrate and ribozyme that reconstitute the conserved u−G wobble pair found at the 5′ splice site (*1*). This initially included running the reactions for 1 h. Under these conditions, all 16 base pair combinations gave first-step product (5′ cleavage via hydrolysis), although the extent of reaction varies significantly as a function of sequence (Figure 2B). Note that the product of the second reaction step must necessarily have undergone the first reaction step. Therefore, the yield of the first step is the combined total of the first and second steps of the reaction. Surprisingly, multiple sequence combinations also gave an appreciable amount (>10%) of complete TES product (Figure 2B). However, in some cases (u−A, a−U, c−G, g−C, and c−C) appreciable amounts of cryptic products (that are shorter than the 9-mer TES product) were also produced. We therefore ran time dependent assays to determine when the cryptic products appear.

Time dependence assays allowed us to find a reaction time that minimized cryptic product formation while still producing appreciable TES product, and allowed us to quantify the observed rate constants for product formation (in the absence

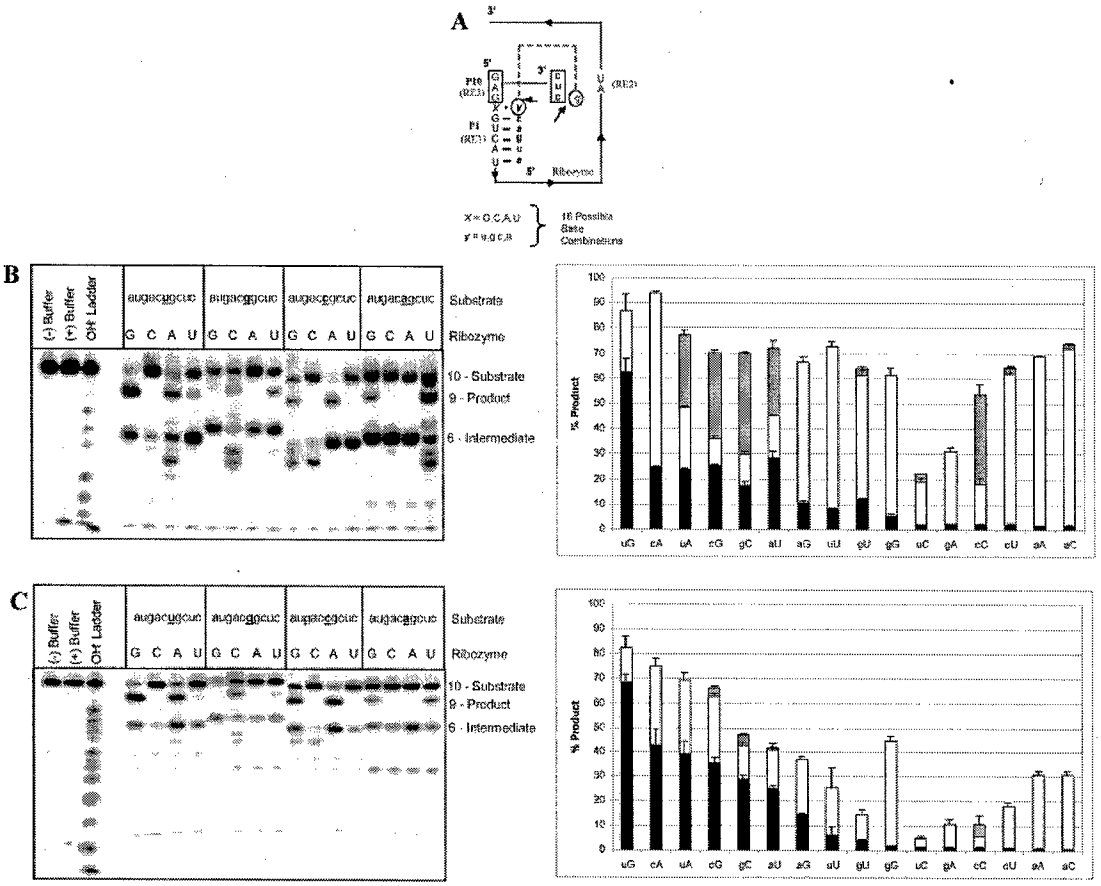

FIGURE 2: Sequence analysis of the 5′ splice site of TES reactions at 1 h and 15 min. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h (B) and 15 min (C) at 44 °C in 10 mM $MgCl_2$. (A) Diagram of the model TES reaction used. See Figure 1 for more detailed information. The substrates were 10-mers (shown in lowercase lettering) where *X* represents position 12 in the IGS of the ribozyme and *y* represents position −1 of the substrate. Every combination of the four nucleotides at *X* and *y* was analyzed. (B) A representative polyacrylamide gel using all 16 base pair combinations at the 5′ splice site (left) and graph of the percent of all products formed in 1 h in the TES reactions as a function of 5′ splice site sequence (right). Each complete substrate sequence and the base at ribozyme position 12 (in uppercase lettering) is shown above its corresponding lane. The migration position on the gel of the 10-mer substrates, the 9-mer products, and the 6-mer intermediates are labeled. All other bands represent cryptic sites. Note that there is some sequence-dependent migration variability between these lanes. The lanes marked "buffer" had substrate augacugcuc incubated as a typical reaction in the absence of ribozyme, both with (+) and without (−) added buffer. The black bars on the graph represent 9-mer TES products, the white bars represent 6-mer 5′ cleavage products, and the gray bars represent all the cryptic products formed. The results are the average of three independent assays, and the standard deviation in all cases is less than 10%. Note that the order of the data in the graph does not correspond to the loading order of the representative gel. (C) A representative polyacrylamide gel using all 16 base pair combinations at the 5′ splice site (left) and graph of the percent of all products formed in 15 min in the TES reactions as a function of 5′ splice site sequence (right). The results are the average of four independent assays and the standard deviation in all cases is less than 10%. Note that the order of the data in the graph does not correspond to the loading order of the representative gel. Data have been ordered according to percent TES product formation.

of these side products). Figure 3 shows representative gels of the three classes of results. In the first class, cryptic products appear after 15 min (Figure 3A). This class consists only of those cases where Watson−Crick pairs are present at the 5′ splice site. In the second class, consisting only of c−C, cryptic products appear much earlier (at 1 min) and to a greater extent than the other combinations (Figure 3B). In the third class, consisting of all other non-Watson−Crick combinations, appreciable cryptic products do not appear, even in cases where substantial TES products form. It seems likely, then, that the Watson−Crick pairs produce cryptic sites after TES product formation, while c−C produces them shortly after substrate binding. Therefore, based on these studies, we again investigated the 5′ splice site sequence in the TES reaction, this time using 15 min as the reaction endpoint.

The trend with regard to the yield of TES products formed was very similar at 15 min and 1 h, but cryptic sites with the Watson−Crick pairs were effectively removed at the shorter time. Again, multiple sequence combinations gave an appreciable amount (>10%) of complete TES product (Figure 2C). These include the u−G and c−A wobble pairs, all four Watson−Crick base pairs, and the a−G combination. Not surprisingly, the conserved u−G wobble pair is the most effective, producing 68% TES product, with the c−A wobble pair being the next highest (43%). It was previously known that these wobble combinations were acceptable at this position in other introns. What was surprising was that, unlike *Tetrahymena* ribozymes in 5′ cleavage and exon-ligation reactions, all four Watson−Crick base pairs produce a relatively high amount of product (>25%).

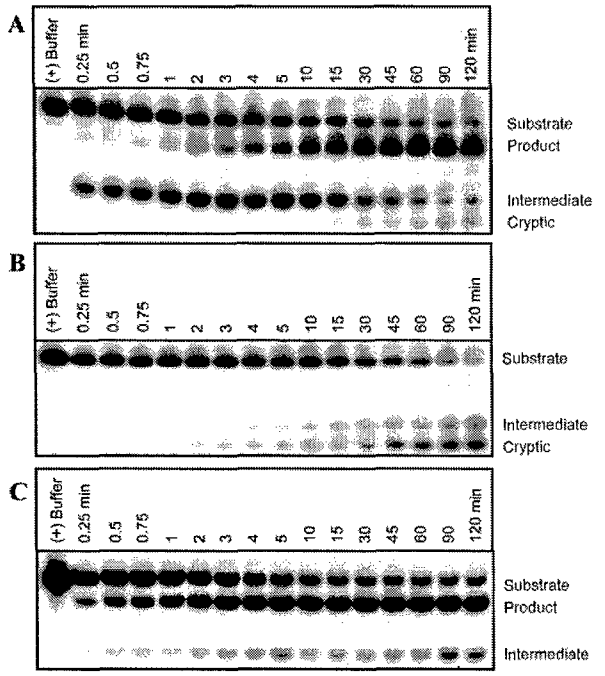

FIGURE 3: Time-dependent polyacrylamide gels of representative TES reactions. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. Aliquots were removed at the times listed above each lane. The migration position on the gel of the 10-mer substrates, the 9-mer products, the 6-mer intermediates, and the cryptic products are labeled. The lane labeled (+) Buffer contains a 120 min reaction in the absence of added ribozyme. (A) A TES reaction utilizing the substrate augaccgcuc and the ribozyme rP-8/4x, which places a c−G pair at the 5′ splice site. Note that cryptic products begin to appear at the 15 min time point. (B) A TES reaction utilizing the substrate augaccgcuc and the ribozyme rP-8/4x-5′C, which places a c−C pair at the 5′ splice site. Note that cryptic products begin to appear at 1 min. (C) A TES reaction utilizing the substrate augacugcuc and the ribozyme rP-8/4x, which places a u−G wobble pair at the 5′ splice site. Note that no cryptic products form in this case.

Apparently, the sequence of both the exon and the ribozyme (at the 5′ splice site) is important for each reaction step. Guanosine is favored in the IGS of the ribozyme, with all four possible combinations giving over 60% first step (5′ cleavage) and only the g−G combination producing less than 10% TES product during the 1 h reaction. The presence of a cytosine in either the ribozyme or the substrate is unfavorable for TES product formation, unless the cytosine is involved in a Watson−Crick base pair or a c−A wobble pair. In addition, a high level of first-step product does not necessarily lead to a high level of TES product, as is seen when adenosine is present at the −1 position in the substrate. Adenosine in the substrate leads to at least 30% first step reactivity with all four ribozymes, but only a−U and a−G lead to significant amounts of TES product (Figure 2C). Similar to that seen for the first step, then, the second step is fairly accommodating when it comes to sequence combinations that lead to product formation.

Graphs of time dependent TES assays as a function of 5′ splice site sequence are available as Supporting Information. Table 1 contains the resultant quantified observed rate constants (k_{obs}) from the graphs for which greater than 10% product forms. The conserved u−G wobble pair has the fastest rate of TES product formation. Although the c−A wobble gave the second highest amount of TES product, it had the second lowest rate constant. Apparently, a high percentage of the ribozyme molecules are properly folded for reactivity (as shown by the extent of reaction), but the rate of catalysis is nevertheless relatively slow. The c−G and g−C base pairs have similar, and the second highest, rate constants. The rate constants for the a−U and u−A base pairs were similar, but less than those for c−G and g−C. The lowest rate constant was obtained for the a−G pair. Thus, the identity of the base pair at the 5′ splice site does affect the observed rate constant of the TES reaction.

Table 1: Observed Rate Constants for the TES Reaction with Various Base Pairs at the 5′ Splice Site[a]

base pair	k_{obs} (min^{-1})
u−G	1.96
c−G	0.79
g−C	0.67
a−U	0.20
u−A	0.16
c−A	0.11
a−G	0.08

[a] TES reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. For these reactions, the base pairs at the 5′ splice site of each ribozyme−substrate complex are shown to the left, with the substrate base in lowercase and the ribozyme base in uppercase. The observed rate constants (k_{obs}) were calculated using a single-exponential curve fit of time-dependent graphs of the reactions (up to 15 min), which are available in the Supporting Information. Each k_{obs} value is the average of two independent assays and the standard deviation for each value is less than 10%.

It is important to note that the 9-mer products of the reactions result from the expected TES reactions, and not from alternative reactions. This is clear for the case with u−G at the 5′ splice site, as this product was enzymatically sequenced (*1*), and no alternative products were found. One possible alternative reaction that could lead to 9-mer products being formed could occur if the substrate binds the ribozyme in a misaligned register, resulting in the 3′ c of the substrate being cleaved off in the first step of the TES reaction. To confirm that this is not occurring in any of the sixteen 5′ splice site cases, TES reactions were run using substrates that contain a radiolabeled nucleotide added to the 3′ end of the 10-mer substrates (to give 11-mer substrates). If this alternative reaction were occurring, then the reaction product bands would now be two nucleotides in length. A representative gel of these reactions is available in Supporting Information. The absence of products two nucleotides in length shows that this alternative reaction is not occurring. In addition, that we get 10-mer product bands, and only in those cases and in the same yields as we get 9-mer product bands with the 5′ end radiolabeled material, confirms that these differentially radiolabeled substrates are undergoing the same reactions. In addition, these reaction end-products all contain both the 5′ and 3′ ends of their substrates (comparing the 5′ and 3′ end radiolabeling studies), and they also are one nucleotide shorter than their substrates. Therefore, any alternative reactions that lead to 9-mer product formation would encompass the two-step excision of a single nucleotide other than ωG. If this were the case, reaction intermediates using 5′ end radiolabeled substrates would be a size other than 6 nucleotides. That only 6-mer intermediates are visualized in all 16 cases studied, as confirmed through running size controls for each intermediate (data not shown),

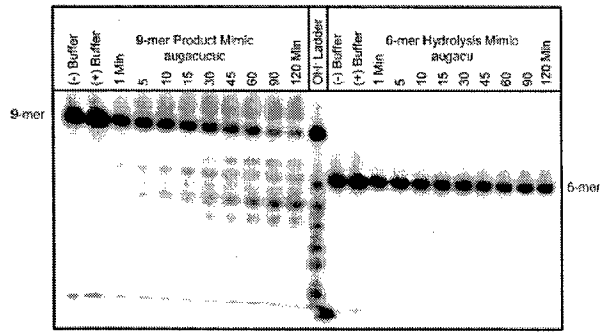

FIGURE 4: Time-dependent polyacrylamide gels of TES reactions utilizing 9-mer TES product (left) and 6-mer intermediate (right) as reaction products. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. The ribozyme used in each case was rP-8/4x-5′A, which, when paired with the substrate, will create an u−A base pair at the 5′ splice site. Aliquots were removed at the times listed above each lane. The migration positions on the gel of the 9-mer and 6-mer starting material are labeled. Unlabeled bands are cryptic products. The lanes marked "buffer" were incubated as a typical reaction for 120 min in the absence of ribozyme, both with (+) and without (−) added buffer. Note that cryptic products occur only when using the 9-mer TES product.

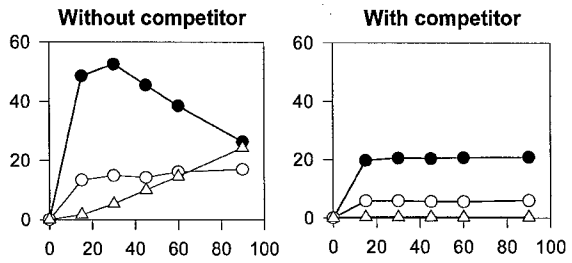

FIGURE 5: Competition TES reactions. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. The ribozyme used in each case was rP-8/4x-5′A, which, when paired with the substrate augacugcuc, will create a u−A base pair at the 5′ splice site. After the reaction proceeded for 5 min, 1.3 μM of the unlabeled TES product (1000-fold excess over substrate) was added to one of the reactions. Aliquots were removed starting at another 10 min. Shown are graphs comparing reactions in the absence (left) and presence (right) of this unlabeled competitor. TES products are represented by black circles, 6-mer hydrolysis products by white circles, and all cryptic products by white triangles. The addition of competitor at 5 min does reduce the amount of TES product formed because not all the labeled substrate has had a chance to bind the ribozyme. Of the labeled substrate that does bind, however, the addition of competitor eliminates the formation of cryptic products because the TES product cannot rebind another ribozyme once it dissociates.

then, shows that the TES reaction is always acting to excise ωG, as expected.

Determining the Source of Cryptic Products. It was intriguing that of those 5′ splice site base combinations that gave substantial TES product, only the Watson−Crick pairs produce cryptic products at extended reaction times. These cryptic sites are almost exclusively less than six bases long. To investigate this phenomenon further, we ran time dependent TES assays utilizing a 6-mer, which mimics the 5′ exon product of the 5′ cleavage reaction, and a 9-mer, which mimics the expected TES product. We already know that degradation does not occur with the 10-mer substrates, as cryptic products form only after 15 min. For these assays we utilized ribozyme−substrate combinations that produced a u−A pair (Figure 4) or a c−G pair (data not shown) at the 5′ splice site. Degradation of the 6-mer mimic would indicate that cryptic degradation is occurring at some point prior to exon ligation, while degradation of the 9-mer mimic would indicate that cryptic degradation might be occurring after the second reaction step (only after product forms). The results show that only the 9-mer, and not the 6-mer, produces cryptic site degradation (Figure 4, data not shown for c−G). Therefore, it appears that cryptic site cleavage might be originating from the TES products themselves, and not the 5′ exon intermediate of the first step of the TES reaction. This explains why cryptic products occur only at relatively long reaction times. In fact, the product is cleaved not only at the cryptic sites but also at the original 5′ splice site (producing 6-mer), showing that some amount of TES products for all base pair combinations likely recleaved at the correct 5′ splice site.

We investigated the mechanism of cryptic product formation by running competition assays to distinguish between two possible major routes. The first route involves the TES product staying bound to the ribozyme, but changing binding registers, which would put the new 5′ cleavage site at a position other than −1. This could occur through 5′ exon slippage relative to the internal guide sequence of the ribozyme (*4*) or could occur through P1 helix translocation through the catalytic core (*3, 8, 22*). Another route for cryptic product formation involves dissociation of the product from the ribozyme, followed by rebinding in alternative registers, where it is subject to the 5′ cleavage reaction in this altered state. In these assays, TES reactions were conducted and, before cryptic products started to form, a large excess of unlabeled TES product was added. If the product dissociates and rebinds, we expect the excess unlabeled product to out compete the labeled product for rebinding a free ribozyme. This competition would lead to a decrease in the amount of cryptic degradation seen in the labeled material. If the product remains bound, the excess competitor would have no affect and the amount of cryptic products formed would remain the same relative to product formation. These studies utilized ribozyme−substrate combinations that formed the u−A pair (Figure 5) and c−G pair (data not shown) at the 5′ splice site. We find that addition of the competitor leads to an almost total elimination of cryptic products, indicating that the TES products do dissociate and then rebind free ribozyme, at which point the TES products are subjected to the 5′ cleavage reaction (Figure 5, data not shown for c−G). Note that utilizing substrate−ribozyme combinations that produce cryptic sites, rather than wobble pairs (which cleave only at the −1 position) allowed us to differentiate between the 5′ exon intermediate that forms in reactions utilizing TES substrates and secondary products stemming only from TES reaction products.

Molecular Recognition at the 3′ Splice Site. To determine the sequence specificity of the 3′ splice site, we again used the simplest substrate−ribozyme system; the excision of a single nucleotide. This nucleotide, which corresponds to the ω position within self-splicing introns, was altered in four different 10-mer substrates and used in TES reactions. Typically, the ω position defines the 3′ splice site by binding to the guanosine binding site (GBS) of the catalytic core of the ribozyme (*14, 15*). Note that in systems where more than a single nucleotide are excised, this base also represents the last base of the excised region. The results (Figure 6) show

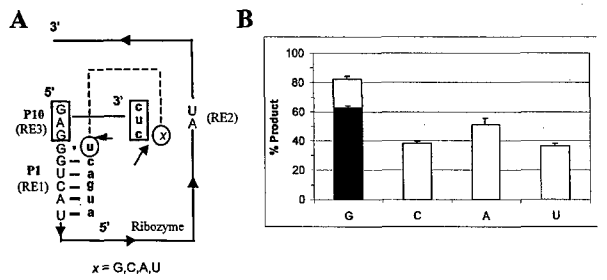

FIGURE 6: Sequence analysis of the 3′ splice site of TES reactions at 1 h. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in 10 mM $MgCl_2$. (A) Diagram of the model TES reaction used. See Figure 1 for more detailed information. The four substrates utilized were 10-mers (shown in lowercase lettering), where *x* is one of each of the four nucleotides. Note that *x* represents the substrate position analogous to the *ω* position of self-splicing introns. The recognition elements from ribozyme rP-8/4x are shown (uppercase lettering). (B) Graph of the percent of all products formed in the TES reactions as a function of 3′ splice site sequence. The black bars represent 9-mer TES products and the white bars represent 6-mer 5′ cleavage products. The results are the average of two independent assays, and the standard deviation in all cases is less than 10%. Notice that only *ω*G produces TES product.

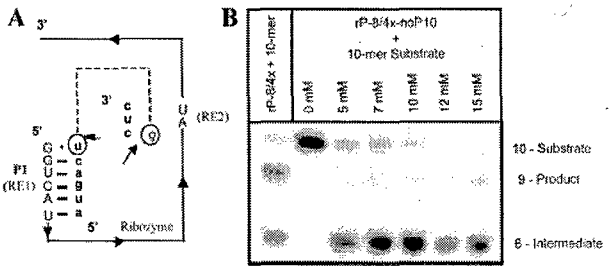

FIGURE 7: Magnesium dependence of TES reactions using the ribozyme rP-8/4x-noP10. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in the $MgCl_2$ concentrations listed below each gel lane. (A) Diagram of the model TES reaction used. See Figure 1 for more detailed information. The substrate 10-mer is shown in lowercase lettering and the recognition elements of the ribozyme are shown in uppercase lettering. In this system P10 formation is not possible. (B) Representative polyacrylamide gel of TES reactions in the absence of P9.0 and P10. The migration positions on the gel of the 10-mer substrate, the 9-mer product, and the 6-mer intermediate are labeled. The far left lane is a positive control using the 10-mer substrate and a ribozyme with P10. This reaction was run twice, and product yields were as much as 10%.

that the first step of the TES reaction (5′ cleavage) occurs in all four cases, with the maximum amount of 5′ cleavage occurring when guanosine is at the *ω* position. TES product is formed, however, only when guanosine is at the *ω* position. Multiple attempts to rationally redesign the GBS to bind adenosine, as was done with *Tetrahymena* ribozymes (*13, 20*), were unsuccessful. This indicates the mode of molecular recognition of the GBS for the *ω* position base may not be the same between the two ribozymes. Nevertheless, for the TES reaction, guanosine is required at the *ω* position.

Several elements are thought to play a role in the determination of the 3′ splice site. They include the conserved guanosine residue, *ω*G, which precedes the reactive phosphodiester bond, as well as the P9.0 and P10 helices (*9*−*15, 23*−*26*). For the TES reactions where a single nucleotide is excised, P9.0 helix formation is not possible (Figure 1) or required (*1*). Therefore, we expected that P10 formation would be critical for binding the 3′ intermediate of the 5′ cleavage reaction (first step) in preparation for the exon-ligation reaction (second step). To test this, we used the ribozyme rP-8/4x-noP10, where RE3 has been removed from the ribozyme. Figure 7 shows that while product yield is low (~10% maximum), TES product does form in the absence of P9.0 and P10. Therefore, *ω*G is sufficient as a 3′ molecular recognition element for TES reactions.

DISCUSSION

Molecular Recognition at the 5′ Splice Site. In nature, a u−G wobble pair is almost universally conserved at the 5′ splice site of self-splicing group I introns. Exceptions are rare, consisting solely of c−G pairs that exist for translational coding purposes (*18*). Therefore, one of the expected sequence limitations of the trans excision-splicing reaction is the requirement for either a u−G pair (*1, 3*−*8*) or a c−G base pair (*18*) at this position. In these cases, a guanosine is present at position 12 of the ribozyme (located in RE1, as diagrammed in Figure 1), and uracil or cytosine is at the −1 position of the substrate. For group I intron-derived ribozymes, it has also been reported that c−A pairings are active in the 5′ cleavage reaction, although they are not nearly as effective as u−G pairs (*3*). Other base pairs and mismatch combinations do not work at all or work only to a very small extent (*3, 4, 27*). The c−A substitution is most likely allowable because c−A can form a wobble base pair with a spatial orientation and accessibility of free functional groups analogous to that of u−G pairings (*3, 28*−*30*). A wobble pair may also help to define the 5′ splice site via the backbone structure of the wobble pair. X-ray crystallography studies on tRNA (*31, 32*) and an NMR structure of a model of a P1 helix (*33*) showed that a u−G wobble pair perturbs the RNA helix by positioning U closer to the helix axis and G away from it. It is believed the distorted backbone of the wobble pairs is accessible for nucleophilic attack in the 5′ cleavage reaction, which helps to determine the correct 5′ splice site. Also, the free exocyclic amine group of guanosine (at ribozyme position 12) in the minor groove is believed to be involved in tertiary interactions that help in the recognition of the correct splice site (*6, 7, 28*). It must be noted, however, that there are significant differences in molecular recognition between different introns. For example, the contribution of tertiary interactions by the 5′ splice site u−G wobble is different between introns, as the contribution to the *P. carinii* ribozyme is greater than those for *Tetrahymena* (*34*) and *Candida albicans* (*35*). Moreover, 2′-hydroxyl groups from the ribose sugar units have been shown to take part in binding of *Tetrahymena* substrates (*36*−*40*), yet these interactions do not appear to be critical for *P. carinii*-derived ribozymes (*5*). Therefore, identifying and understanding the sequence requirements of the 5′ splice site for *P. carinii* ribozymes is useful for developing effective and specific TES ribozymes. To this end, we have tested the effectiveness of each of the 16 possible base combinations at the 5′ splice site in the TES reaction.

We report that all 16 base pair combinations undergo the first reaction step (5′ cleavage); with the u−G and c−A wobble pairs, all four Watson−Crick pairs, and the a−G pair producing appreciable amounts of TES product (> 10%). In terms of the first reaction step (5′ cleavage), that all base pair combinations at the 5′ splice site are permissible suggests that the interactions involved in 5′ splice site determination

Table 2: Relative Effectiveness of Each Base in Each Position of the 5′ Splice Site in TES Reactions[a]

When the ribozyme contains G	u	42.29	When the ribozyme contains A	u	26.74	When the ribozyme contains C	u	1.51	When the ribozyme contains U	u	6.67
	c	18.29		c	30.14		c	1.27		c	1.02
	a	6.71		a	0.48		a	0.82		a	33.61
	g	0.87		g	0.71		g	44.16		g	4.23
When the substrate contains g	U	3.58	When the substrate contains a	U	21.72	When the substrate contains c	U	0.60	When the substrate contains u	U	3.56
	C	32.48		C	0.55		C	0.70		C	0.77
	A	1.13		A	0.63		A	33.86		A	27.48
	G	1.72		G	11.47		G	26.50		G	60.43

[a] Top: Effectiveness of each ribozyme construct in TES reactions as a function of base identity at position 12 using each of the four 10-mer substrates that differ at position −1. Bottom: Effectiveness of each of the four 10-mer substrates that differ at position −1 in TES reactions as a function of each of the four ribozyme constructs. These numbers were obtained by taking the extent of TES product at 15 min of a particular case (for example; u−G) and dividing that by all products formed for the other three combinations (for example; TES, 5′ cleavage, and cryptic products for u−A, u−C, and u−g) plus the nonproductive products of the particular case (for example; 5′ cleavage and cryptic products for u−G). These theoretical numbers, therefore, represent percent product formation of one particular base pair combination in relation to all other products that would be produced in a 1:1:1:1 mixture of each ribozyme (top of table) or substrate (bottom of table).

are not stringent. This substrate promiscuity at the 5′ splice site was unexpected and is in contrast to results reported for analogous reactions using a *Tetrahymena* ribozyme (*3, 4*). Apparently, for *P. carinii*-derived TES ribozymes, identification of the 5′ splice site is not entirely dependent on having specific functional groups in the ribozyme at position 12 or the substrate at −1. In terms of the second reaction step (exon ligation), that 7 out of 16 combinations give an appreciable amount of TES product, including all four Watson−Crick base pairs, suggests an unexpectedly lax sequence (and hence structural) requirement at the 5′ splice site. Some general trends to note include a benefit for having a purine in the IGS of the ribozyme, a strong benefit for the u−G pair, and the appearance of substantial amounts of cryptic sites at extended times for the Watson−Crick pairs. It is interesting that although u−G and c−A give substantial amounts of TES products, g−U and a−C do not. Apparently, it is not just the presence of a wobble pair that is beneficial for the TES reaction, but the presence of a wobble pair with a purine in the ribozyme and a pyrimidine in the substrate.

Note that effective 5′ cleavage, however, does not necessarily translate into high levels of TES product. For example, base combinations a−A, a−C, and g−G produce reasonable amounts of 5′ cleavage product, yet they each produce only about 2% TES product (Figure 2C). These base pairing combinations are expected to be thermodynamically weak, especially in relation to wobble and Watson−Crick pairs (*41*). Therefore, these results indicate that after 5′ cleavage, a thermodynamically stable base pair at the 5′ splice site, although not required, is beneficial for the exon-ligation reaction, perhaps to spatially orient the newly created 3′ hydroxyl group more favorably for this reaction step. Previous work with the *Tetrahymena* ribozyme supports this idea (*4*).

We obtained the rate constants for those reactions producing appreciable TES product in 15 min (Table 1) and determined that there is a correlation between predicted base pair strength (*41*) at the 5′ splice site and the rate of the reaction. The observed rate constants can be broken down into four classes. The native u−G wobble at the 5′ splice site produced the greatest amount of TES product and had the fastest rate constant. This result is not surprising since this is the native 5′ splice site and the ribozyme has evolved to exploit this pairing. The next fastest class is the stronger Watson−Crick base pairs, c−G and g−C. The rates for c−G and g−C are approximately 2.5-fold less than the native system, with c−G being slightly faster. This is a further indication of a preference for a purine at position 12 of the ribozyme. In addition, that cytidine can work at position 12, and almost as well as guanosine, indicates that required interactions binding to the native guanosine in the ribozyme are not at play here. The next fastest class is seen with the weaker Watson−Crick pairs u−A and a−U. The u−A and a−U Watson−Crick pairs show an approximately 11-fold reduction in the rate compared to the native system. Note that all the base pair combinations produce approximately the same amount of TES product, which indicates that the amount of ribozyme that folds properly is similar in these cases. The slowest class is the weaker wobble pairs c−A and a−G. The c−A wobble pair is unusual in that although it produces the second highest amount of TES product, it is approximately 20-fold slower than the native case. As a whole, these results suggest that there is a significant correlation between base pair strength at the 5′ splice junction and the observed rate constant of TES reactions. This is probably manifest in the second reaction step and further supports the above purported idea (looking at base pair strength versus extent of reaction) that a thermodynamically stable base pair at the 5′ splice site is beneficial for the TES reaction, apparently in terms of both the rate and extent of reaction.

Because multiple base pair combinations at the 5′ splice site give a substantial amount of TES product, and because they all give a substantial amount of first-step product, it is worth considering the relative specificity of targeting different substrate sequences with the different ribozymes. Table 2 (top) shows the relative specificity for each ribozyme construct and (bottom) the relative specificity for each base at the −1 position in the substrate. These results show that there are distinctive combinations of substrate and ribozyme that work much better in the TES reaction relative to similar combinations. In terms of the ribozyme, a cytidine is very specific for targeting a substrate guanosine, with virtually no TES product being formed with cytidine, adenosine, or uridine at the corresponding substrate position. Note that we would still get a substantial amount of hydrolysis products, however, ribozyme reconstruction methods have been developed that could be used to help overcome this problem (*2*). Uridine in the ribozyme has the next highest specificity, then guanosine, and last adenosine. Adenosine is interesting in that it will target substrates with uridine and cytidine approximately equally. In terms of the substrate position, to

target a guanosine in the substrate, cytidine is the overwhelmingly best choice for the corresponding ribozyme position. When targeting an adenosine, uridine is the best choice; when targeting a cytidine, adenosine is the best choice (even better than a guanosine); and when targeting a uridine, guanosine is the best choice. If one has the flexibility to choose any substrate base, using a cytidine in the ribozyme to target a guanosine in the substrate would appear to give the most specific TES product (although not necessarily the highest yield). Note that this analysis does not take into account the amount of product subjected to cryptic site degradation.

Molecular Recognition at the 3′ Splice Site. We looked at the molecular recognition of the 3′ splice site by altering the base at the ω position, which is the base excised from the 10-mer substrate. Any base at the ω position allowed 5′ cleavage (the first reaction step), however, the second reaction step (exon ligation) was completely inhibited when ωG was replaced with the other bases (Figure 6). Thus, the ω position must be a guanosine for complete TES reactivity, probably because of a required and specific binding of the ω position base with the GBS. This shows that, in its current incarnation, *P. carinii*-derived TES ribozymes require a guanosine as the last (or only) base to be excised in the substrate. Alteration of this specificity most likely would require an alteration of the GBS of the ribozyme. The GBS of the *Tetrahymena* ribozyme has been altered to change the specificity from guanosine to adenosine (*13, 20*); however, these same mutations did not produce the same change in specificity for the *P. carinii* ribozyme (data not shown). One possible conclusion is that there could be a difference in the recognition of the ωG by the *Tetrahymena* and *P. carinii* ribozymes.

P9.0 and P10 are not Required for TES Reactions. Of the three known elements that can aid in the molecular recognition of the 3′ splice site in self-splicing reactions (ωG, P10, and P9.0), only ωG and P10 are present in TES reactions that remove a single nucleotide (Figure 1). The P9.0 helix cannot form in these constructs, and its absence does not appear to have negative consequences on the TES reaction (*1*). Since ωG binding to the GBS is a relatively weak interaction (*42*), and because it is thought that this interaction occurs only after the first reaction step (see previous section), it was expected that the formation of P10 would play a crucial role in defining and binding the 3′ end of the substrate to the ribozyme (for example, as the first reaction step intermediate). Surprisingly, however, we found that P10 and P9.0, although beneficial, are not required for either step of the TES reaction. How is it then that the 3′ exon intermediate does not readily dissociate from the ribozyme between the two reaction steps? One possible explanation is that the 3′ exon is sterically constrained within the docked state of the catalytic core of the ribozyme after the first reaction step. In other words, the 3′ exon is physically trapped within the ribozyme in a space for which no significant molecular interactions occur between the ribozyme and the 3′ reaction intermediate. Then a conformational rearrangement takes place between the two reaction steps that places ωG in the GBS, in preparation for the second reaction step. An alternative explanation is that the interactions that hold the 3′ exon intermediate do not require base pairing. For example, it could be that the 3′ intermediate nucleotides themselves, perhaps in combination with the hydroxyl groups from the sugar−phosphate backbone, form tertiary interactions with the ribozyme.

This result was also surprising because previous studies with this same ribozyme indicated that it was unable to catalyze the complete TES reaction when excising a 20 nucleotide segment in the absence of P10 and P9.0 (*2*). It is likely that dissociation of the 3′ intermediate may have been responsible for a lack of second-step product in the previous case, as including a P9.0 interaction restored reactivity (*2*). Apparently, this problem of intermediate dissociation is not as overwhelming when excising a single nucleotide. This indicates that there are functional differences between TES reactions where a single base is excised relative to excising multiple bases.

A Mechanism for Ribozyme-Mediated TES Product Degradation. We found that at long reaction times (relative to product completion), TES products dissociate and rebind the ribozyme, at which point degradation occurs via the 5′ cleavage reaction at one or more new 5′ splice sites. We show that product cleavage does not occur through translocation of a bound P1 helix after product formation, as the product must first dissociate. Note, however, that c−C is a special case in that the substrate also appears to degrade (at cryptic sites). Nevertheless, these results show that TES ribozymes can bind the products of TES reactions and degrade them, which is essentially irreversible, as the ribozyme-mediated hydrolysis reaction is kinetically unfavorable to reverse. Of course, this degradation mechanism is probably greatly aided by the conditions under which the reactions were conducted (i.e., ribozyme excess and long reaction times).

Non-Watson−Crick Base Pairs at the 5′ Splice Site Can Play a Role in Determining the Binding Register of Reaction Substrates. The activation of cryptic sites as it relates to the first step of self-splicing (5′ cleavage) has been investigated with *Tetrahymena* ribozymes (*4, 6*−*8, 22*). It was shown that cryptic hydrolysis sites can become activated when the substrate helix shifts from the original binding register to another register without dissociation (*22*). Such a mechanism is probably occurring in our reactions where a c−C base pair forms at the 5′ splice site. For our other cases, however, whereby Watson−Crick base pairs form, cryptic sites are activated only after correct TES products form, the products dissociate from the ribozyme, and then bind a new ribozyme. Since 5′ exon translocation was not detected in these assays as a source of cryptic cleavage sites, it seems apparent that these TES products are binding to the IGS of the ribozyme, with subsequent P1 helix docking, in the wrong helical register. In contrast, TES products that do not contain Watson−Crick pairs at the 5′ splice site are not degraded at cryptic sites. Apparently, non-Watson−Crick base pairs at the 5′ splice site are acting in concert with other factors to precisely determine the binding register of TES reaction substrates with the IGS of the ribozyme. The mechanism of this role likely includes the fact that non-Watson−Crick base pairs are causing a structural perturbation of the sugar−phosphate backbone at the 5′ splice site, which would prevent the formation of a structurally uniform P1−P1ex (extended) helix. This continuous helix would contain no perturbation to structurally define the 5′ splice site. In the absence of this molecular recognition component, the 5′ exon region binds

the IGS of the ribozyme and the subsequent P1 helix docks within the catalytic core of the ribozyme in multiple registers, activating various positions within the 5′ exon for subsequent 5′ cleavage reactions. Note that the role of non-Watson−Crick base pairs in defining the correct 5′ splice site must be combined with a stable base pairing conformation (as mentioned above) in order to allow the second reaction step to proceed effectively.

Except for the case with c−C, cryptic products are created from TES products, which indicates that the 1 h yield of TES products formed for those ribozyme−substrate complexes with Watson−Crick pairs at the 5′ splice site is higher than originally calculated. The yield of TES product created is actually the yield of TES product plus the yield of cryptic products. This increases the percent TES product formed to over 50% for all Watson−Crick pairs during the 1 h reactions. Since we also see degradation of the TES products into 6-mers (which is the same size as the expected TES intermediate), some of the hydrolysis products shown in all the TES reactions could have stemmed from the degradation of TES products. Thus, the quantified TES products shown in this report at all time points represent a minimum. Note that, in terms of applicability, to prevent dead-end TES product degradation, the use of Watson−Crick base pairs at the 5′ splice site in TES ribozyme−substrate complexes should be avoided. It should also be noted that these reactions are run under ribozyme-excess conditions. Product degradation may not be as prevalent under conditions utilizing lower ribozyme concentrations.

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

Implications. These results advance our knowledge of the molecular recognition involved in both steps of the TES reaction (and the analogous steps of self-splicing) and allow us to improve our design principles for developing and applying ribozyme-mediated excision reactions. We are now able to target a wider range of TES substrates, and with more sequence specificity, particularly with regard to the 5′ splice site. This information will help us to further develop TES ribozymes as potential biochemical tools. Examples include RNA repair ribozymes, which would be a strategy complementary to trans-splicing ribozymes (*43*−*45*), and transcript cleaving ribozymes, which would exploit the first reaction step to produce transcripts ending in 3′ OH groups [which is in contrast to hammerhead and hairpin ribozymes, which leave 2′, 3′-cyclic phosphates (*46*)]. Last, exploiting the TES reaction, in contrast to self-splicing or suicide inhibition, has allowed us to observe new and unexpected molecular recognition principles exploited by group I intron-mediated catalytic reactions.

ACKNOWLEDGMENT

The authors wish to thank the members of the Testa lab for technical assistance and helpful discussions.

SUPPORTING INFORMATION AVAILABLE

Representative gel of 15 min TES reactions using 3′ radiolabeled substrates to investigate the 5′ splice site and graphs used to obtain observed rate constants. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Bell, M. A., Johnson, A. K., and Testa, S. M. (2002) Ribozyme-catalyzed excision of targeted sequences from within RNAs, *Biochemistry 41*, 15327−15333.
2. Bell, M. A., Sinha, J., Johnson, A. K., and Testa, S. M. (2004) Enhancing the second step of the trans excision-splicing reaction of a group I ribozyme by exploiting P9.0 and P10 for intermolecular recognition, *Biochemistry 43*, 4323−4331.
3. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) RNA structure, not sequence, determines the 5′ splice-site specificity of a group I intron, *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402−7406.
4. Barfod, E. T., and Cech, T. R. (1989) The conserved U−G pair in the 5′ splice site duplex of a group I intron is required in the first but not the second step of self-splicing, *Mol. Cell. Biol. 9*, 3657−3666.
5. Testa, S. M., Haidaris, C. G., Gigliotti, F., and Turner, D. H. (1997) A *Pneumocystis carinii* group I intron ribozyme that does not require 2′ OH groups on its 5′ exon mimic for binding to the catalytic core, *Biochemistry 36*, 15303−15314.
6. Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H., and Cech, T. R. (1994) Replacement of the conserved G.U with a G−C pair at the cleavage site of the Tetrahymena ribozyme decreases binding, reactivity, and fidelity, *Biochemistry 33*, 13856−13863.
7. Strobel, S. A., and Cech, T. R. (1996) Exocyclic amine of the conserved G·U pair at the cleavage site of the Tetrahymena ribozyme contributes to 5′-splice site selection and transition state stabilization, *Biochemistry 35*, 1201−1211.
8. Knitt, D. S., Narlikar, G. J., and Herschlag, D. (1994) Dissection of the role of the conserved G·U pair in group I RNA self-splicing, *Biochemistry 33*, 13864−13879.
9. Burke, J. M., Esherick, J. S., Burfeind, W. R., and King, J. L. (1990) A 3′ splice site-binding sequence in the catalytic core of a group I intron, *Nature 344*, 80−82.
10. Suh, E. R., and Waring, R. B. (1990) Base pairing between the 3′ exon and an internal guide sequence increases 3′ splice site specificity in the Tetrahymena self-splicing rRNA intron, *Mol. Cell. Biol. 10*, 2960−2965.

11. Burke, J. M. (1989) Selection of the 3′-splice site in group I introns, *FEBS Lett. 250*, 129−133.
12. van der Horst, G., and Inoue, T. (1993) Requirements of a group I intron for reactions at the 3′ splice site, *J. Mol. Biol. 229*, 685−694.
13. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) The guanosine binding site of the Tetrahymena ribozyme, *Nature 342*, 391−395.
14. Price, J. V., and Cech, T. R. (1988) Determinants of the 3′ splice site for self-splicing of the Tetrahymena pre-rRNA, *Genes Dev. 2*, 1439−1447.
15. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J., and Strobel, S. A. (2004) Crystal structure of a self-splicing group I intron with both exons, *Nature 430*, 45−50.
16. Johnson, A. K., Baum, D. A., Tye, J., Bell, M. A., and Testa, S. M. (2003) Molecular recognition properties of IGS-mediated reactions catalyzed by a *Pneumocystis carinii* group I intron, *Nucleic Acids Res. 31*, 1921−1934.
17. Testa, S. M., Gryaznov, S. M., and Turner, D. H. (1999) In vitro suicide inhibition of self-splicing of a group I intron from *Pneumocystis carinii* by an N3′→P5′ phosphoramidate hexanucleotide, *Proc. Natl. Acad. Sci. U.S.A. 96*, 2734−2739.
18. Hur, M., and Waring, R. B. (1995) Two group I introns with a C·G basepair at the 5′ splice-site instead of the very highly conserved U·G basepair: is selection post-translational?, *Nucleic Acids Res. 23*, 4466−4470.
19. Golden, B. L., and Cech, T. R. (1996) Conformational switches involved in orchestrating the successive steps of group I RNA splicing, *Biochemistry 35*, 3754−3763.
20. Been, M. D., and Perrotta, A. T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site, *Science 252*, 434−437.
21. England, T. G., and Uhlenbeck, O. C. (1978) 3′ Terminal labeling of RNA with T4 RNA ligase, *Nature 275*, 560−561.
22. Herschlag, D. (1992) Evidence for processivity and two-step binding of the RNA substrate from studies of J1/2 mutants of the Tetrahymena ribozyme, *Biochemistry 31*, 1386−1399.
23. Michel, F., Netter, P., Xu, M. Q., and Shub, D. A. (1990) Mechanism of 3′ splice site selection by the catalytic core of the sunY intron of bacteriophage T4: the role of a novel base-pairing interaction in group I introns, *Genes Dev. 4*, 777−788.
24. Russell, R., and Herschlag, D. (1999) Specificity from steric restrictions in the guanosine binding pocket of a group I ribozyme, *RNA 5*, 158−166.
25. Karbstein, K., Carroll, K. S., and Herschlag, D. (2002) Probing the Tetrahymena group I ribozyme reaction in both directions, *Biochemistry 41*, 11171−11183.
26. Karbstein, K., and Herschlag, D. (2003) Extraordinarily slow binding of guanosine to the Tetrahymena group I ribozyme: implications for RNA preorganization and function, *Proc. Natl. Acad. Sci. U.S.A. 100*, 2300−2305.
27. Green, R., Szostak, J. W., Benner, S. A., Rich, A., and Usman, N. (1991) Synthesis of RNA containing inosine: analysis of the sequence requirements for the 5′ splice site of the Tetrahymena group I intron, *Nucleic Acids Res. 19*, 4161−4166.
28. Strobel, S. A., and Cech, T. R. (1995) Minor groove recognition of the conserved G·U pair at the Tetrahymena ribozyme reaction site, *Science 267*, 675−679.
29. Masquida, B., and Westhof, E. (2000) On the wobble GoU and related pairs, *RNA 6*, 9−15.
30. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) The non-Watson−Crick base pairs and their associated isostericity matrices, *Nucleic Acids Res. 30*, 3497−3531.
31. Westhof, E., Dumas, P., and Moras, D. (1985) Crystallographic refinement of yeast aspartic acid transfer RNA, *J. Mol. Biol. 184*, 119−145.
32. Quigley, G. J., and Rich, A. (1976) Structural domains of transfer RNA molecules, *Science 194*, 796−806.
33. Allain, F. H., and Varani, G. (1995) Structure of the P1 helix from group I self-splicing introns, *J. Mol. Biol. 250*, 333−353.
34. Disney, M. D., Gryaznov, S. M., and Turner, D. H. (2000) Contributions of individual nucleotides to tertiary binding of substrate by a *Pneumocystis carinii* group I intron, *Biochemistry 39*, 14269−14278.
35. Disney, M. D., Haidaris, C. G., and Turner, D. H. (2001) Recognition elements for 5′ exon substrate binding to the *Candida albicans* group I intron, *Biochemistry 40*, 6507−6519.
36. Sugimoto, N., Tomka, M., Kierzek, R., Bevilacqua, P. C., and Turner, D. H. (1989) Effects of substrate structure on the kinetics of circle opening reactions of the self-splicing intervening sequence from *Tetrahymena thermophila*: evidence for substrate and Mg2+ binding interactions, *Nucleic Acids Res. 17*, 355−371.
37. Pyle, A. M., and Cech, T. R. (1991) Ribozyme recognition of RNA by tertiary interactions with specific ribose 2′-OH groups, *Nature 350*, 628−631.
38. Bevilacqua, P. C., and Turner, D. H. (1991) Comparison of binding of mixed ribose-deoxyribose analogues of CUCU to a ribozyme and to GGAGAA by equilibrium dialysis: evidence for ribozyme specific interactions with 2′ OH groups, *Biochemistry 30*, 10632−10640.
39. Caprara, M. G., and Waring, R. B. (1993) Important 2′-hydroxyl groups within the core of a group I intron, *Biochemistry 32*, 3604−3610.
40. Berzal-Herranz, A., Chowrira, B. M., Polsenberg, J. F., and Burke, J. M. (1993) 2′-Hydroxyl groups important for exon polymerization and reverse exon ligation reactions catalyzed by a group I ribozyme, *Biochemistry 32*, 8981−8986.
41. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson−Crick base pairs, *Biochemistry 37*, 14719−14735.
42. Bass, B. L., and Cech, T. R. (1986) Ribozyme inhibitors: deoxyguanosine and dideoxyguanosine are competitive inhibitors of self-splicing of the Tetrahymena ribosomal ribonucleic acid precursor, *Biochemistry 25*, 4473−4477.
43. Sullenger, B. A., and Cech, T. R. (1994) Ribozyme-mediated repair of defective mRNA by targeted trans-splicing, *Nature 371*, 619−622.
44. Ayre, B. G., Kohler, U., Turgeon, R., and Haseloff, J. (2002) Optimization of trans-splicing ribozyme efficiency and specificity by in vivo genetic selection, *Nucleic Acids Res. 30*, e141.
45. Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003) Efficient and specific repair of sickle beta-globin RNA by trans-splicing ribozymes, *RNA 9*, 1254−1263.
46. Doherty, E. A., and Doudna, J. A. (2000) Ribozyme structures and mechanisms, *Annu. Rev. Biochem. 69*, 597−615.

BI0482304

In vivo excision of a single targeted nucleotide from an mRNA by a *trans* excision-splicing ribozyme

DANA A. BAUM and STEPHEN M. TESTA
Department of Chemistry, University of Kentucky, Lexington, Kentucky 40506, USA

ABSTRACT

We have previously reported the development of a group I intron-derived ribozyme that can bind an exogenous RNA substrate and excise from that substrate an internal segment in vitro, which allows for sequence-specific modification of RNA molecules. In this report, the activity of this *trans* excision-splicing ribozyme in a cellular environment, specifically *Escherichia coli*, was investigated. The ribozyme was re-engineered to target for excision a single-base insertion in the transcript of a green fluorescent protein, and fluorescence was exploited as a reporter for *trans* excision-splicing. We show that the ribozyme is able to catalyze the *trans* excision-splicing reaction in vivo and can repair the mutant transcripts. On average, 12% correction is observed as measured by fluorescence and at least 0.6% correction as confirmed through sequence analysis. This represents the first report of a biomolecule (in this case a ribozyme) that can selectively excise a targeted nucleotide from within an mRNA transcript in vivo. This new class of biochemical tools makes possible a wide variety of new experimental strategies, perhaps including a new approach to molecular-based therapeutics.

Keywords: *trans* excision-splicing; ribozyme; RNA repair

INTRODUCTION

Group I introns are catalytic RNAs (Kruger et al. 1982) with the ability to splice themselves out of RNA transcripts. Eliminating the 5′ and 3′ exon sequences from these self-splicing introns produces catalytic ribozymes that are able to bind substrates in trans and perform reactions analogous to the reaction steps of self-splicing (Zaug and Cech 1986; Zaug et al. 1986; Cech 1990; Sargueil and Tanner 1993; Testa et al. 1997). Furthermore, such ribozymes have been developed with the ability to target and modify mRNA transcripts for biochemical and therapeutic purposes (Sullenger and Cech 1994; Jones et al. 1996; Jones and Sullenger 1997; Phylactou et al. 1998; Kohler et al. 1999; Rogers et al. 2002; Byun et al. 2003; Ryu et al. 2003; Kastanos et al. 2004; Shin et al. 2004). For example, it was previously reported that a ribozyme derived from a *Tetrahymena* group I intron catalyzes a *trans*-splicing reaction (Sullenger and Cech 1994), in which the ribozyme binds an RNA substrate and replaces its defective 3′ exon sequence with a corrected form (carried by the ribozyme). The in vivo activity of this ribozyme has been demonstrated with various transcript targets (Sullenger and Cech 1994; Jones et al. 1996; Jones and Sullenger 1997; Phylactou et al. 1998; Kohler et al. 1999; Rogers et al. 2002; Byun et al. 2003; Ryu et al. 2003; Kastanos et al. 2004; Shin et al. 2004). This type of RNA modification has potential therapeutic application as it can be used to repair transcripts with mutations that lead to disease.

A complementary reaction to *trans*-splicing is the *trans* excision-splicing (TES) reaction (Bell et al. 2002). In this reaction, a group I intron-derived ribozyme from *Pneumocystis carinii* binds an RNA substrate and excises a targeted segment from within that substrate (Fig. 1). In vitro, this ribozyme has been used to excise inserts from 1 to 28 nucleotides in length and can be re-engineered to target nonnative sequences, including the flanking region of the trinucleotide repeat that causes myotonic dystrophy (Bell et al. 2002, 2004). The usefulness of these ribozymes, however, will be greatly enhanced by demonstrating in vivo reactivity.

In this report, we have designed a test system to assess the ability of the *P. carinii* ribozyme to catalyze the TES reaction in vivo. A single base insertion mutation was engineered into a green fluorescent protein (GFP) gene, which consequently destroys fluorescence of the resultant protein product. The TES ribozyme was then re-engineered to target this mutation, so a successful TES reaction reestablishes the reading frame of the protein in the GFP transcript (Fig. 2A), producing a fluorescing protein and thus a fluorescing cell culture. We

Reprint requests to: Stephen M. Testa, Department of Chemistry, University of Kentucky, Lexington, KY 40506, USA; e-mail: testa@uky.edu; fax: (859) 323-1069.

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.2050505.

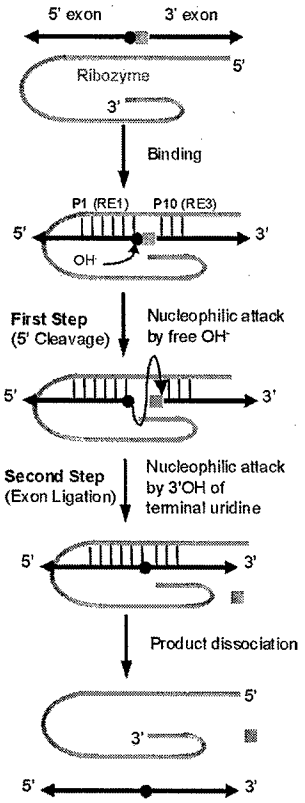

FIGURE 1. Ribozyme *trans* excision-splicing (TES) reaction. The TES ribozyme is shown as a gray line, the 5′ and 3′ exons are black lines, and the sequence targeted for excision (in this case, a single guanosine) is a gray box. The circle in the 5′ exon represents uridine. P1 and P10 are the helices that result from the ribozyme recognition elements RE1 and RE3 (respectively) base-pairing with the substrate.

now report that the ribozyme is able to catalyze the TES reaction in *Escherichia coli*. An average of 12% of the expected fluorescence was observed by using a construct with an elongated P10 helix of 5 bp, indicating an in vivo requirement for strong interactions between the ribozyme and the target transcript. Sequencing of isolated transcripts confirmed the ribozyme-mediated removal of the targeted nucleotide in at least 0.6% of the mutant transcripts. This represents the first report of a biomolecule (in this case a ribozyme) that can selectively excise a targeted nucleotide from within an mRNA transcript in vivo. This new class of biochemical tools makes possible a wide variety of new experimental strategies, perhaps including a new approach to molecular-based therapeutics.

RESULTS

Design of the in vivo test system

The design of the in vivo TES test system was based on work involving the *trans*-splicing ribozyme (Sullenger and Cech 1994; Jones et al. 1996; Jones and Sullenger 1997; Phylactou et al. 1998; Kohler et al. 1999; Rogers et al. 2002; Byun et al. 2003; Ryu et al. 2003; Kastanos et al. 2004; Shin et al. 2004), as well as design principles garnered through studies of the TES ribozyme in vitro (Bell et al. 2002, 2004). Our simplest TES reaction system involves the removal of a single nucleotide from a substrate, so we designed a test system in which we create a single base insertion mutation in the coding region of GFP (Fig. 2A). This mutation causes a deleterious shift in the reading frame of the transcribed RNA, which alters the amino acid sequence of the resultant protein product and creates a premature stop codon, resulting in a loss of fluorescence. We chose to insert a uridine 5′ to a guanosine in the RNA transcript (Fig. 2A). In the context of the TES reaction, this uridine will serve to define the 5′ splice site by forming the highly conserved u-G wobble pair with a guanosine in the ribozyme (Barfod and Cech 1989; Doudna et al. 1989; Testa

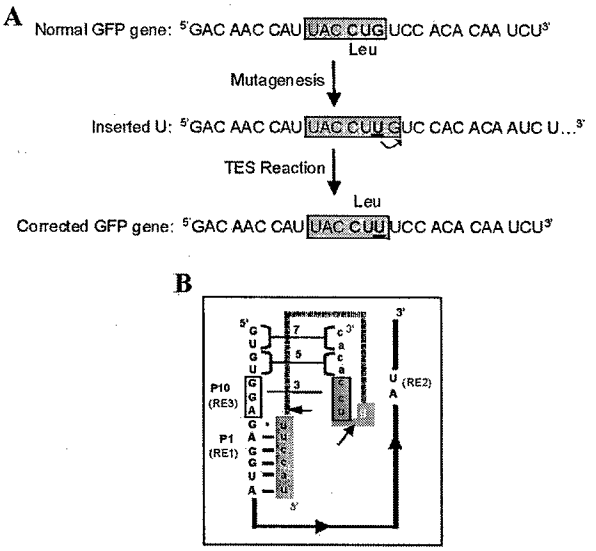

FIGURE 2. Design of the GFP target system (*A*) and schematic of the two-step TES reaction utilizing the GFP target (*B*). (*A*) The *top* line shows the sequence of interest of the GFP gene. The bold codon codes for leucine (Codon 201). The gray box is the sequence that base pairs with recognition element 1 (RE1) of the ribozyme. Site-directed mutagenesis results in the insertion of a uridine (underlined) 5′ to a guanosine in the mRNA transcript, which causes a frameshift and a premature stop codon (data not shown). The TES reaction removes the targeted guanosine and restores the correct reading frame of the GFP gene (*bottom* line). (*B*) Diagram of the recognition elements of the TES ribozyme base-pairing with the GFP target. The GFP rP3X ribozyme is in uppercase lettering, the GFP target is in lowercase lettering, and the guanosine to be excised is in white lettering. The ribozyme recognition elements RE1, RE2, and RE3 base pair with the substrate to form the P1, P9.0, and P10 helices, respectively. Note that the P9.0 helix is not utilized in this system. The sites of catalysis for the first step (5′ cleavage) and the second step (exon ligation) are shown with large bold arrows. The three-base P10 helix is boxed, with brackets indicating the extensions to five and seven bases. Note that the diagram only shows the recognition elements of the ribozyme. This ribozyme is the same as rP-8/4x (Testa et al. 1997), except for the sequences of RE1 and RE3 and the addition of a T7 terminator on the 3′ end.

et al. 1997; Bell et al. 2002). The guanosine following the inserted uridine is equivalent to the ω position in the self-splicing reaction (Price and Cech 1988; Burke 1989; Michel et al. 1989; Burke et al. 1990; van der Horst and Inoue 1993), so this ωG is the base targeted for excision (Fig. 2B). After a successful TES reaction, the inserted uridine replaces the guanosine in the coding region, resulting in a silent mutation in the restored reading frame (Fig. 2A). Thus, the site for insertion was limited to those codons where replacement of guanosine by uridine results in the same amino acid when translated into protein. Codon 201, which codes for leucine, was an attractive target for insertion and, based on RNA structure prediction estimates, was accessible in the transcript. The insertion of a uridine and subsequent removal of a guanosine allow us to determine, via the sequence of isolated transcripts from our test reactions, that the fluorescent form of GFP results from a successful TES reaction.

Changes to modify the ribozyme were kept to a minimum for these initial in vivo studies and involved altering the recognition elements of the ribozyme, which are responsible for target identification (Fig. 2B). In all the test systems, the length of recognition element 1 (RE1) was maintained at six bases. Previous work on the TES ribozyme indicated that increasing the length of RE3 can be beneficial to the reaction (Bell et al. 2004), while work with the *trans*-splicing ribozyme showed a requirement for increased 3′ exon interactions for in vivo activity (Kohler et al. 1999; Rogers et al. 2002; Byun et al. 2003; Ryu et al. 2003; Kastanos et al. 2004; Shin et al. 2004). Thus, ribozymes were tested with RE3 lengths of three bases (native length), five bases, and seven bases. A third recognition element, RE2, is not utilized in TES reactions that remove a single nucleotide, so it was not modified. As a control for ribozyme activity, we also created mutant ribozyme constructs by deleting four bases from the guanosine binding site of the ribozyme. This mutation completely inactivates the ribozyme when tested in vitro (data not shown).

The TES ribozyme is active in vivo

First, the insertion mutation engineered into the GFP target was tested to ensure that it did indeed abolish fluorescence. The expected corrected form (with the G-to-U transversion) was also tested for fluorescent activity. As seen in Figure 3, introducing the insertion mutation into the GFP gene abolishes GFP activity. The expected corrected form fluoresces, albeit at a level somewhat below that for the normal form of GFP, indicating the G-to-U transversion is not substantially disruptive to the translation of the GFP protein.

Next, the ribozymes engineered to target the mutation in GFP were tested for in vivo activity by pairing the ribozymes with the mutant form of GFP (Mut GFP). Targeted ribozymes (GFP rP3X) were tested with RE3 lengths of three, five, and seven bases. Increasing the length of RE3 increases the interaction between the ribozyme and the 3′ exon of the substrate, which helps prevent 3′ exon dissociation prior to

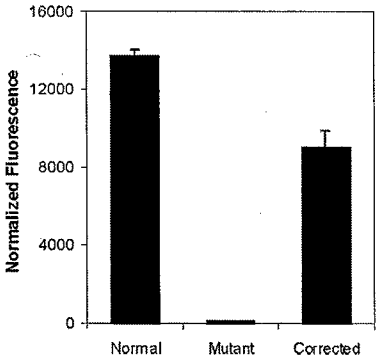

FIGURE 3. Comparison of the normalized GFP fluorescence levels for the normal, mutant, and corrected forms of GFP. All points utilize constructs with the nontargeted ribozyme and the GFP form listed on the graph. Normalized fluorescence was determined by correcting the raw fluorescence for differences in cell density (determined by the absorbance of the bulk culture at 600 nm) and normalizing to non-induced controls. Each data point represents four independent assays, and the standard deviations are <10%.

the second reaction step and can improve TES product formation in vitro (Bell et al. 2004). The native form of the ribozyme (rP3X) served as a non-GFP-targeted control for ribozyme activity, while inactive forms of the ribozymes (containing a four-base deletion in the guanosine binding site) were used as negative controls. As seen in Figure 4, increasing the length of RE3 to five bases for the targeted ribozyme (GFP rP3X RE3 = 5) significantly increases the fluorescence over the inactive ribozyme controls, indicating TES reactivity. Note the level of fluorescence for the negative controls is not zero due to a low level of inherent fluorescence in the samples. We obtained an estimate of TES reactivity by comparing the data from the constructs with that of mutant GFP (Fig. 4) and the same ribozyme constructs with the

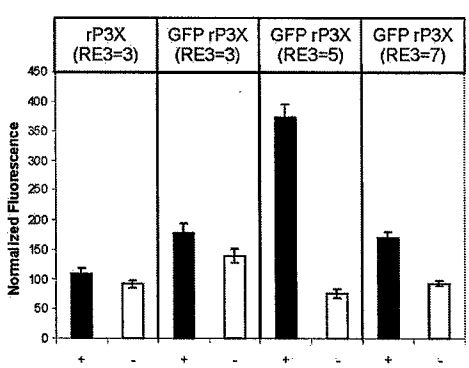

FIGURE 4. Normalized fluorescence data for constructs with mutant GFP paired with the TES ribozymes. All points utilize the mutant GFP target with the ribozyme listed above the graph. The plus symbols and solid bars indicate active ribozymes, while the minus symbols and open bars indicate the ribozymes have been inactivated by a deletion in the guanosine binding site. Normalized fluorescence was determined by correcting the raw fluorescence for differences in cell density (determined by the absorbance of the bulk culture at 600 nm) and normalizing to noninduced controls. Each data point represents four independent assays, and the standard deviations are <10%.

expected corrected form of GFP. For the corrected forms of GFP, normalized fluorescence values of 16,265.99 ± 388.75 (ribozyme with an RE3 = 3), 3105.83 ± 1940.59 (ribozyme with an RE3 = 5), and 5971.97 ± 2272.05 (ribozyme with an RE3 = 7) were obtained from at least three independent tests. Note the large standard deviations result from greater variability in reactions conducted in vivo when the RE3 length is increased from the native length of three nucleotides in constructs containing the corrected form of GFP. The estimated percentage of RNA repair was low for the targeted ribozymes with RE3 lengths of three nucleotides (1.1 ± 0.1%) and seven nucleotides (3 ± 1%), while utilizing a targeted ribozyme with an RE3 length of five nucleotides increased the estimated percentage of repair to 12 ± 8%.

Sequence confirmation of in vivo TES reaction

Total RNA was isolated from reactions involving three different constructs: the mutant GFP paired with the targeted ribozyme containing an RE3 = 5 (the repair construct that showed the greatest increase in GFP activity), the mutant GFP paired with the inactive targeted ribozyme with an RE3 = 5 (a negative control), and the expected corrected form of GFP paired with the targeted ribozyme with an RE3 = 5 (a positive control to assess the insertion efficiency). The total RNA was subjected to RT-PCR to isolate the GFP transcript and the resulting RT-PCR products were ligated into a cloning vector and were assayed for their ability to produce fluorescent protein (Fig. 5). In the screening assays, 12 fluorescent colonies (out of ~2000 screened colonies) were obtained from products from two independent in vivo tests with the active repair construct, indicating the presence of corrected transcripts. Assays involving products from three

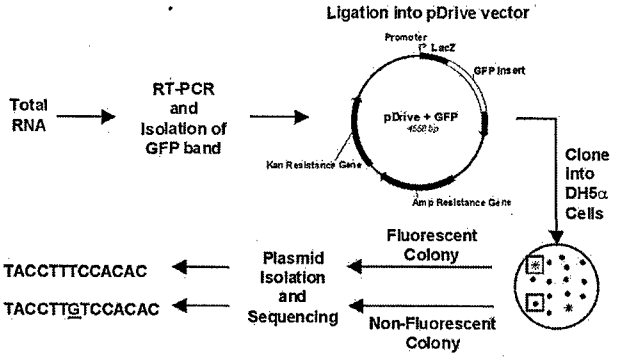

FIGURE 5. Schematic of the screening assay for isolation of corrected transcripts. Total RNA is isolated from cells that have undergone in vivo testing. The isolated RT-PCR products are ligated into the pDrive PCR cloning vector (Qiagen). Note that the RT-PCR products can be ligated into the vector in the forward direction (which puts the GFP gene in frame with the LacZ α-peptide of the plasmid) or the reverse direction, with each direction assumed to be equally possible. Only products ligated in the forward direction will produce fluorescent GFP. The resulting plasmids are isolated and sequenced to confirm the expected sequence change in the corrected transcripts.

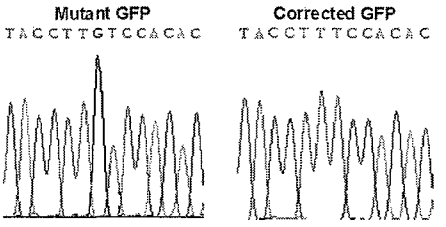

FIGURE 6. Sequencing confirmation of uncorrected GFP (*left*) and corrected GFP (*right*) transcripts. Total RNA isolated from in vivo tests involving the construct Mut GFP + GFP rP3X RE3 = 5 was utilized as the template for RT-PCR to isolate the GFP transcript.

independent in vivo tests with the mutant ribozyme construct produced no fluorescent colonies in ~2000 screened colonies, indicating that no TES reaction had occurred. The 12 fluorescent colonies from the active repair construct, one nonfluorescent colony from the active repair construct, and 18 colonies from the negative control construct were isolated, and the resulting plasmids were sequenced. All 12 sequences from the fluorescent colonies showed the removal of the targeted guanosine, resulting in the G-to-U transversion in the corrected transcript (Fig. 6). If the cell randomly fixed the transcript and produced fluorescent protein, the loss of a uridine would be anticipated to be as likely as the loss of a guanosine. As this loss of a uridine was not observed, the fluorescent protein was produced as a result of a successful TES reaction. There was also no evidence of a deletion that could serve as a compensatory mutation to restore fluorescence. The nonfluorescent colonies were confirmed to contain the unrepaired mutant GFP transcript (Fig. 6). Thus the removal of the guanosine was TES ribozyme mediated and the ribozyme is sequence specific for its target, without being deleterious to the cell. Based on this method, at least 0.6% of the mutant transcripts are corrected by the TES ribozyme.

DISCUSSION

The arsenal of molecular-based RNA tools is rapidly growing; however, the majority of these tools have focused on the destruction of the RNA message rather than the modification of the message (Sullenger and Gilboa 2002; Puerta-Fernandez et al. 2003; Scherer and Rossi 2003). Destroying the RNA message is useful for studying the effects of shutting off genes and can be useful in therapeutic applications to prevent the production of mutant proteins that lead to disease. On the other hand, modifying the message can restore the function to the transcript, thus reducing mutant protein and producing normal protein. With this in mind, we report the development of a catalytic RNA that can sequence-specifically target a single nucleotide within a cellular transcript in vivo and excise that targeted nucleotide from a transcript, with little or no apparent toxicity to the cell. Combined with the ease of adaptability, this provides for a wide variety of new in vivo experimental strategies.

As biochemical tools, TES ribozymes could be used as inducible tools to modulate the production of proteins in vivo. As demonstrated in this report, TES ribozymes can be targeted to remove sequences to restore the reading frame of a transcript, thus producing active protein. The removal of sequences to shift the sequence out of frame is also possible, so inactive or mutant proteins could be produced in an inducible manner.

The applications of TES ribozymes could also include therapeutic applications. As a therapeutic agent, TES ribozymes could be used to remove insertion mutations and premature stop codons from transcripts, thus restoring the reading frame of the transcript for protein production. We have previously demonstrated the in vitro ability of a TES ribozyme to target a small model mimic of the triplet expansion implicated in myotonic dystrophy, and to excise from that mimic a short triplet expansion (Bell et al. 2002). Thus the repair of transcripts involved in triplet expansion diseases, such as muscular dystrophy and Huntington's disease, is another potential application for TES ribozymes in vivo.

As TES ribozymes recognize their targets initially and primarily through base-pairing, engineering the ribozymes to target new transcript regions simply requires changing the sequence of the recognition elements to base pair with the desired target. Other considerations for ribozyme targeting include reconstituting a u-G wobble pair at the 5′ splice site and a guanosine at the 3′ splice site as the last (or only) base of the sequence targeted for excision (Baum et al. 2005). Also note that previous work with this ribozyme demonstrated that segments larger than a single nucleotide can be excised in vitro (Bell et al. 2002), so we anticipate that larger regions could be excised in vivo.

Comparing in vivo and in vitro TES results

In our simplest test system, which involves the excision of a single, targeted nucleotide, the TES ribozyme is able to produce 70% product in vitro under optimized conditions of ribozyme excess and 10 mM $MgCl_2$ (Bell et al. 2002). For the in vivo tests presented here, the ribozyme restores ~12% of the fluorescence. This decrease in reactivity is not unexpected and is likely due to a number of factors, including a lower concentration of magnesium, a lower concentration of active ribozyme relative to the target substrate, and target accessibility. Nevertheless, TES ribozymes are adaptable to the cellular environment and are able to produce detectable amounts of repaired RNA.

Comparison with *trans*-splicing and SMaRT

TES ribozymes offer a different approach to RNA modification compared with the *Tetrahymena*-derived *trans*-splicing ribozyme (Sullenger and Cech 1994) and spliceosome-mediated RNA *trans*-splicing, or SMaRT (Puttaraju et al. 1999; Garcia-Blanco 2003). The *trans*-splicing ribozyme replaces the 3′ exon of a targeted transcript with a new version of the exon, which is carried by the ribozyme. SMaRT utilizes the spliceosomes present in the cell to catalyze the splicing between an exon on the target transcript and an exon present in a pre-*trans*-splicing molecule (PTM). TES ribozymes are different in that they excise a targeted internal segment from the transcript.

TES ribozymes have some potential advantages over *trans*-splicing ribozymes and SMaRT. TES ribozymes are not changed in the reaction, so they have the potential to be multiple-turnover catalysts. The *trans*-splicing ribozyme splices its replacement exon into the target, making it single turnover. The PTM donates its replacement sequence during splicing, so it also is not a multiple turnover reaction. TES ribozymes can, in theory, be engineered to excise an internal segment anywhere on a transcript. In contrast, *trans*-splicing ribozymes are best suited to target mutations in the 3′ portion of a transcript, and SMaRT targets only exon/intron junctions. In addition, TES ribozymes do not carry replacement sequences, while *trans*-splicing ribozymes and PTMs do. Therefore, the size of the ribozyme-replacement exon construct or the PTM could become prohibitive for effective delivery and activity. Finally, in contrast to TES ribozymes and *trans*-splicing ribozymes, SMaRT can only be used in mammalian cells as it requires spliceosomes.

In *E. coli*, the level of activity restored by the *trans*-splicing ribozyme was ~1% (Sullenger and Cech 1994). In our studies, we observe a higher level of repair in *E. coli*, as the level of activity restored by the TES ribozyme was ~12%. SMaRT is limited to mammalian cells, so we cannot accurately compare the levels of product formation to our *E. coli* results. That *trans*-splicing ribozymes (Jones et al. 1996; Jones and Sullenger 1997; Phylactou et al. 1998; Kohler et al. 1999; Rogers et al. 2002; Byun et al. 2003; Ryu et al. 2003; Kastanos et al. 2004; Shin et al. 2004) and SMaRT (Puttaraju et al. 1999; Liu et al. 2002; Chao et al. 2003) both work fairly well in mammalian cells (upward of 50%) is encouraging, and we anticipate that TES ribozymes will also be active in mammalian systems. As is the case with *trans*-splicing (Jones and Sullenger 1997; Rogers et al. 2002; Byun et al. 2003; Shin et al. 2004) and SMaRT (Liu et al. 2002; Chao et al. 2003), we do not anticipate that TES modification will have to be 100% to have measurable effects as a biochemical tool. A low level of TES activity may produce a relevant amount of modified transcript for protein production.

Note that the challenges we face in improving the TES reaction in vivo and adapting TES ribozymes to a mammalian system are the same challenges faced by others developing ribozymes in mammalian systems (Sullenger and Gilboa 2002; Long et al. 2003; Scherer and Rossi 2003). Common issues in *E. coli* and mammalian cells include increasing the level of ribozyme activity, improving ribozyme specificity, assessing target accessibility, and effectively delivering the ribozyme. Work is ongoing to address these issues and is encouraging (Sullenger and Gilboa 2002; Long et al. 2003; Puerta-Fernandez et al. 2003; Scherer and Rossi 2003). In conclusion, this study demonstrates the potential of TES

ribozymes as new tools for the in vivo sequence-specific modification of RNA transcripts.

MATERIALS AND METHODS

Oligonucleotide synthesis and preparation

DNA oligonucleotides were purchased from Integrated DNA Technologies. Oligonucleotides for the insertion of the T7 terminator sequence were 5′ phosphorylated and PAGE purified by the company. Oligonucleotides for site-directed mutagenesis and RT-PCR were used without further purification.

Plasmid construction

The *P. carinii* ribozyme precursor plasmid was constructed as previously described (Testa et al. 1997). To stop transcription of the ribozyme in vivo, a T7 terminator sequence (Studier et al. 1990) was inserted immediately 3′ to the ribozyme sequence by using the XbaI and HindIII sites present in the P3X plasmid (Fig. 7). The ligation product was used to transform *E. coli* DH5α competent cells (Invitrogen). The resulting plasmid (P3X + T7T) was purified by using a QIAprep Spin Miniprep kit (Qiagen) and sequenced for confirmation (ACGT, Inc.).

Two unique restriction sites, BglII and SphI, were engineered into the newly created P3X + T7T plasmid to allow for the excision of the ribozyme construct (Fig. 7). The plasmid, P3X + T7TBS, was created by two successive rounds of site-directed mutagenesis. The following primer pair was used to create a BglII restriction site upstream of the T7 promoter: 5′-GGAAACAGATCTGACATGATTACGAATTT GG-3′ and 5′-CCAAATTCGTAATCATGTCAGATCTGTTTCC-3′. A SphI restriction site was created downstream of the T7 terminator with the following primer pair: 5′-GCTTACTAGTGATGCATGC TCTATAGTGTCACC-3′ and 5′-GGTGACACTATAGAGCATGC ATCACTAGTAAGC-3′. The site-directed mutagenesis reactions were conducted as previously described (Bell et al. 2002), with the changes that follow. The reaction mixtures were subjected to denaturation for 30 sec at 95°C, followed by 15 temperature cycles of 30 sec at 95°C, 2 min at 55°C, and 6 min at 68°C. The parental plasmids were digested with 20 U of DpnI (New England Biolabs) in 4.2 μL of manufacturer's buffer for at least 1 h at 37°C. A 3-μL aliquot of this digest was then used to transform *E. coli* DH5α competent cells (Invitrogen). The resultant plasmids were purified as above and were sequenced for confirmation (Davis Sequencing).

The plasmid containing the GFP target, pQBI T7-GFP, was purchased from Qbiogene. The base plasmid for testing, pQBI GFP + P3X, was created by inserting the ribozyme construct into the pQBI T7-GFP plasmid by using the BglII and the SphI restriction sites present in the pQBI T7-GFP plasmid (Fig. 7). The ribozyme construct was isolated by digesting the P3X + T7TBS plasmid with BglII and SphI. The ligation product was used to transform *E. coli* DH5α competent cells (Invitrogen). The resulting plasmid was purified and sequenced as above.

Test plasmids were generated via site-directed mutagenesis of the base plasmid pQBI GFP + P3X. The following primers were used to modify the GFP gene to contain a single base insertion mutation to destroy GFP fluorescence (Mut GFP): 5′-GCA GATTGTGTGGACAAGGTAATGGTTGT CTGG-3′ and 5′-CCAGACAACCATTACC TTGTCCACACAATCTGC-3′. The underlined base represents the insertion mutation. The corrected form of GFP expected from a successful TES reaction (Corr GFP) was created with the following primers: 5′-GCA GATTGTGTGGAAAGGTAATGGTTGTC TGG-3′ and 5′-CCAGACAACCATTACCT TTCCACACAATCTGC-3′. The underlined base represents the silent mutation expected

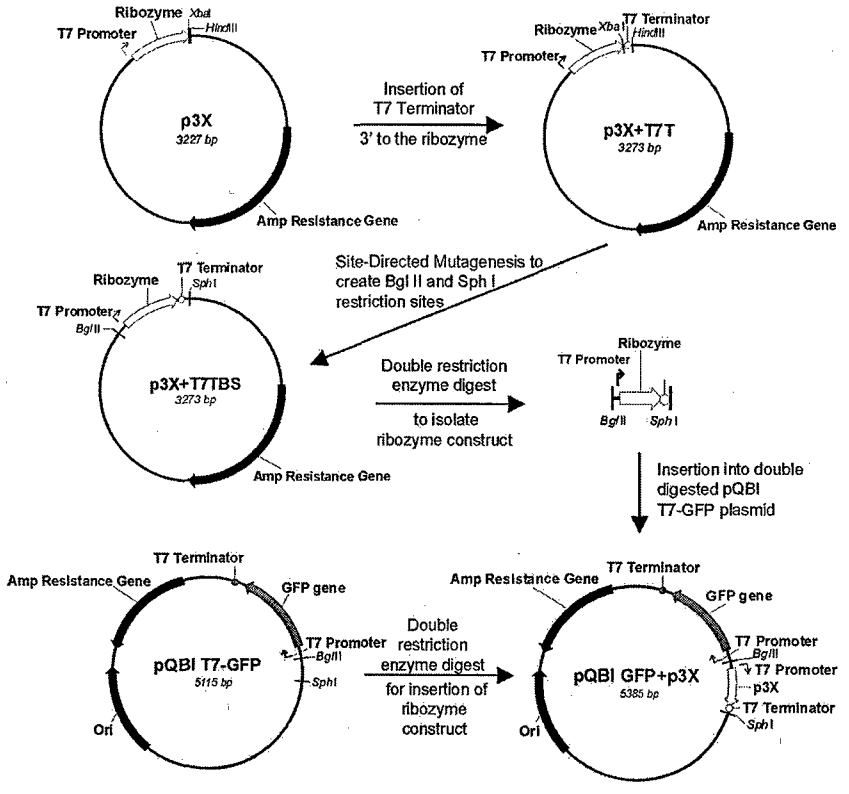

FIGURE 7. Construction of the in vivo TES test plasmid. The original TES ribozyme plasmid (P3X) was modified by the insertion of a linker containing a T7 terminator sequence 3′ to the ribozyme sequence to create P3X + T7T. Site-directed mutagenesis was used to create a unique BglII site 5′ to the T7 promoter of the ribozyme and a unique SphI site 3′ to the T7 terminator of the ribozyme (P3X + T7TBS). The ribozyme construct was isolated from the plasmid by a double restriction enzyme digest with BglII and SphI. The GFP-containing plasmid, pQBI T7-GFP (Qbiogene), was prepared for insertion of the ribozyme construct by a double enzyme digest with BglII and SphI. The ribozyme construct was ligated into the cut pQBI T7-GFP plasmid to create pQBI GFP + P3X. Modifications to the GFP gene to create the mutant and corrected forms and modifications to the ribozyme to change the recognition elements and to create the inactive forms were all created via site-directed mutagenesis.

from removal of the targeted guanosine by the TES ribozyme. These mutations alter codon 201 of the GFP gene. Site-directed mutagenesis was performed as above, with the following modifications. The reaction mixtures were subjected to denaturation for 30 sec at 95°C, followed by 15 or 18 temperature cycles of 30 sec at 95°C, 2 min at 50°C or 55°C, and 6 or 8 min at 68°C, depending on the construct. Resulting plasmids were purified and sequenced as above.

The ribozyme portion of the test plasmid was modified in the following ways by using site-directed mutagenesis. First, the recognition elements of the ribozyme (RE1 and RE3) were modified to recognize the GFP target by two successive rounds of site-directed mutagenesis. A third recognition element, RE2, is not utilized in TES reactions involving the removal of a single nucleotide, so it was not modified. The primers for the first round were as follows: 5′-CGACT CACTATAGAGGGGGTAGAAAGCGGC-3′ and 5′-GCCGCTTTC TACCCCCTCTATAGTGAGTCG-3′. The second round of changes was made with the following primers: 5′-CGACTCACTATAGGA GAGGTAGAAAGCGGC-3′ and 5′-GCCGCTTTCTACCTCTCCTA TAGTGAGTCG-3′. The length of RE3 was increased to five bases by using the following primers: 5′-CGACTCACTATAGTGGAGAGG TAGAAAGCGGC-3′ and 5′-GCCGCTTTCTACCTCTCCACTATAG TGAGTCG-3′. An RE3 containing seven bases was created by using the following primer pair: 5′-CGACTCACTATAGTGTGGAGAGGTAG AAAGCGGC-3′ and 5′-GCCGCTTTCTACCTCTCCACACTATAGT GAGTCG-3′. A mutant form of the ribozyme was created by deleting four bases (bases 250–253) from the guanosine binding site of the ribozyme (Testa et al. 1997). Site-directed mutagenesis reactions were performed as above, and the resulting plasmids were sequenced for confirmation (Davis Sequencing).

Preparing competent JM109(DE3) and transforming competent JM109(DE3)

In vivo testing of the TES ribozyme was conducted in *E. coli* strain JM109(DE3) (Promega). Competent JM109(DE3) cells were prepared by using the rubidium chloride protocol reported by Promega (Doyle 1996). Following a 1 h incubation on ice, the competent cells were transformed with the test plasmids by using a modified version of the procedure outlined by Promega (Doyle 1996). One hundred microliters of competent cells was used in each transformation reaction. Approximately 10 ng of test plasmid was added to the competent cells, and the tubes were incubated for 30 min on ice. The tubes were then heated for 50 sec at 44°C and placed immediately back on ice for 2 min; 1.5 mL of LB media was then added to each tube, and the tubes were incubated at 37°C with shaking at 225 rpm for 45 min. After incubation, the transformations were plated on LB plates containing ampicillin, and the plates were incubated overnight at 37°C.

In vivo TES reactions

The colonies resulting from transformation of the competent JM109(DE3) cells were used for in vivo testing of the TES ribozymes. Colonies were isolated and grown up in liquid culture overnight. The following day, the overnight cultures were used to inoculate fresh LB media without antibiotic to create 1:20 dilution cultures. The dilution cultures were incubated at 37°C with shaking at 225 rpm for 3 h to allow growth to reach log phase ($A_{600} > 0.4$). After 3 h, aliquots were removed from each culture for cell density analysis (500 μL) and fluorescence analysis (1 mL). The rest of the dilution culture was then subdivided. The inducer, isopropylthiogalactoside (IPTG), was added to one of the cultures to a final concentration of ~1.3 mM, while the other culture served as a noninduced control. Cultures were incubated at 37°C with shaking at 225 rpm for 5 h. After 5 h, aliquots were removed for cell density analysis (500 μL), GFP fluorescence analysis (1 mL), and, for select cultures, total RNA isolation (1 mL).

Analysis

Cell density was determined by measuring the absorbance of the culture at 600 nm (A_{600}). For GFP fluorescence analysis, the culture aliquots were centrifuged to pellet the cells. The media was removed, and the cell pellets were resuspended in 300 μL of phosphate buffered saline (PBS) at pH 7.4. The resuspended pellets were then loaded into a FluoroNunc MaxiSorp 96-well plate (Nalge Nunc International). GFP fluorescence was measured in a CytoFluor (PerSeptive Biosystems), using a 485 ± 20 nm filter for excitation and a 508 ± 20 nm filter for emission, as the excitation wavelength for this form of GFP is 474 nm and the emission wavelength is 509 nm. The raw fluorescence measurements were corrected for differences in cell culture growth by dividing the raw fluorescence values by the A_{600} readings. These corrected fluorescence values were then normalized to the noninduced values to correct for "leaky" T7 RNA polymerase activity by subtracting the corrected fluorescence for the noninduced culture from the corrected fluorescence for the induced culture. The reported normalized fluorescence values are the result of at least three independent assays. Note that omitting the corrections for differences in cell density does not significantly affect the trends obtained (data not shown). Error limits for the estimated percentage of RNA repair were calculated as for multiplicative expressions (Miller and Miller 2000).

Total RNA isolation

Total RNA was isolated by using 1 mL aliquots from designated cultures using the Ambion RiboPure-Bacteria Kit (Ambion, Inc.) and was treated with RQ1 RNase-Free DNase (Promega) to completely remove the DNA prior to RT-PCR. RNA was isolated from the DNase reaction mixture by acid phenol (pH 4.3) extraction, followed by phenol/chloroform extraction and ethanol precipitation.

RT-PCR reactions

The GFP transcripts were amplified from the total RNA by RT-PCR using the following primers: 5′-GTTGTACAGTTCATC CATGCC-3′ and 5′-GGAGAAGAACTCTTCACTGG-3′. RT-PCR reactions were performed by using the Access RT-PCR System (Promega) and consisted of 50 μL reactions containing ~1 μg total RNA, 1 mM $MgSO_4$, 45 pmol of each primer, 0.2 mM dNTPs, 5 U AMV reverse transcriptase, and 5 U Tfl DNA polymerase in the provided reaction buffer. PCR reactions were performed as a control to ensure the DNA had been removed from the samples. The reactions were subjected to 45 min at 45°C for first-strand cDNA synthesis, followed by 2 min at 94°C to inactivate the AMV reverse transcriptase. The reactions then underwent 40 temperature cycles consisting 30 sec at 94°C, 1 min at 54°C,

and 2 min at 68°C. After cycling was complete, the reactions underwent a final extension cycle for 10 min at 68°C. The RT-PCR products were separated on a 2% agarose gel, and the GFP band was excised. The band was extracted from the gel matrix by using a QIAquick Gel Extraction Kit (Qiagen). The gel-purified products were ethanol precipitated twice prior to use in ligation reactions.

Assay for corrected transcripts

The TES reaction removes a targeted sequence from within an RNA, so the 5′ and 3′ ends are the same in the mutant and corrected forms. Moreover, in our simple system, the difference between the mutant GFP and the corrected GFP is a single base. Thus, the product cannot trivially be selectively amplified by RT-PCR. Therefore, we designed an assay to analyze the pool of transcripts after the reaction for those that are corrected and thus produce fluorescent protein (Fig. 5). The GFP RT-PCR products were ligated into the pDrive cloning vector by using the QIAGEN PCR Cloning kit (Qiagen). Ligation reactions consisted of 50 ng pDrive cloning vector, 200 ng of RT-PCR product, 1 μL 50% PEG (5% final w/v), and 2× Ligation Master Mix (provided with the kit) in a final volume of 10 μL. Ligations proceeded for 1.5 h at 4°C. Following ligation, a 3-μL aliquot of the ligation reaction was used to transform *E. coli* DH5α competent cells (Invitrogen). Transformations were plated on LB media plates containing kanamycin, and the plates were incubated overnight at 37°C. The following day, the plates were placed at 4°C for at least 4 h to enhance GFP fluorescence for visualization. Those colonies containing the corrected GFP products will produce fluorescent GFP, which can be visualized on the plate. After incubation at 4°C, the plates were exposed to UV light, and colonies showing GFP activity were isolated. The resulting plasmids were isolated by using a QIAprep Spin Miniprep kit (Qiagen) and sequenced for confirmation (Davis Sequencing).

ACKNOWLEDGMENTS

We thank Rashada Alexander, Dustin Lafferty, and Joy Sinha for technical assistance and members of the Testa lab for helpful discussion. We also thank Dr. Sylvia Daunert (Department of Chemistry, University of Kentucky) and members of the Daunert laboratory for use of the CytoFluor. This work was supported by the Muscular Dystrophy Association and the DOD Breast Cancer Research Program DAMD17-03-1-0329.

Received January 10, 2005; accepted March 2, 2005.

REFERENCES

Barfod, E.T. and Cech, T.R. 1989. The conserved U.G pair in the 5′ splice site duplex of a group I intron is required in the first but not the second step of self-splicing. *Mol. Cell. Biol.* **9:** 3657–3666.

Baum, D.A., Sinha, J., and Testa, S.M. 2005. Molecular recognition in a *trans* excision-splicing ribozyme: Non-Watson-Crick base pairs at the 5′ splice site and ωG at the 3′ splice site can play a role in determining the binding register of reaction substrates. *Biochemistry* **44:** 1067–1077.

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

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

Burke, J.M. 1989. Selection of the 3′-splice site in group I introns. *FEBS Lett.* **250:** 129–133.

Burke, J.M., Esherick, J.S., Burfeind, W.R., and King, J.L. 1990. A 3′ splice site-binding sequence in the catalytic core of a group I intron. *Nature* **344:** 80–82.

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

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

Chao, H., Mansfield, S.G., Bartel, R.C., Hiriyanna, S., Mitchell, L.G., Garcia-Blanco, M.A., and Walsh, C.E. 2003. Phenotype correction of hemophilia A mice by spliceosome-mediated RNA *trans*-splicing. *Nat. Med.* **9:** 1015–1019.

Doudna, J.A., Cormack, B.P., and Szostak, J.W. 1989. RNA structure, not sequence, determines the 5′ splice-site specificity of a group I intron. *Proc. Natl. Acad. Sci.* **86:** 7402–7406.

Doyle, K. 1996. *Promega protocols and applications guide.* Promega Corporation, Madison, WI.

Garcia-Blanco, M.A. 2003. Mending the message. *Nat. Biotechnol.* **21:** 1448–1449.

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

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

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

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

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

Liu, X., Jiang, Q., Mansfield, S.G., Puttaraju, M., Zhang, Y., Zhou, W., Cohn, J.A., Garcia-Blanco, M.A., Mitchell, L.G., and Engelhardt, J.F. 2002. Partial correction of endogenous ΔF508 CFTR in human cystic fibrosis airway epithelia by spliceosome-mediated RNA *trans*-splicing. *Nat. Biotechnol.* **20:** 47–52.

Long, M.B., Jones III, J.P., Sullenger, B.A., and Byun, J. 2003. Ribozyme-mediated revision of RNA and DNA. *J. Clin. Invest.* **112:** 312–318.

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

Miller, J.N. and Miller, J.C. 2000. Statistics of repeated measurements: Propagation of random errors. In *Statistics and chemometrics for analytical chemistry*, pg. 36. Prentice Hall, Harlow, England.

Phylactou, L.A., Darrah, C., and Wood, M.J. 1998. Ribozyme-mediated *trans*-splicing of a trinucleotide repeat. *Nat. Genet.* **18:** 378–381.

Price, J.V. and Cech, T.R. 1988. Determinants of the 3′ splice site for self-splicing of the *Tetrahymena* pre-rRNA. *Genes & Dev.* **2:** 1439–1447.

Puerta-Fernandez, E., Romero-Lopez, C., Barroso-delJesus, A., and Berzal-Herranz, A. 2003. Ribozymes: Recent advances in the development of RNA tools. *FEMS Microbiol. Rev.* **27:** 75–97.

Puttaraju, M., Jamison, S.F., Mansfield, S.G., Garcia-Blanco, M.A., and Mitchell, L.G. 1999. Spliceosome-mediated RNA *trans*-splicing as a tool for gene therapy. *Nat. Biotechnol.* **17:** 246–252.

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

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

Sargueil, B. and Tanner, N.K. 1993. A shortened form of the *Tetrahymena thermophila* group I intron can catalyze the complete splicing reaction in *trans*. *J. Mol. Biol.* **233:** 629–643.

Scherer, L.J. and Rossi, J.J. 2003. Approaches for the sequence-specific knockdown of mRNA. *Nat. Biotechnol.* **21:** 1457–1465.

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

Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. 1990. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185:** 60–89.

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

Sullenger, B.A. and Gilboa, E. 2002. Emerging clinical applications of RNA. *Nature* **418:** 252–258.

Testa, S.M., Haidaris, C.G., Gigliotti, F., and Turner, D.H. 1997. A Pneumocystis carinii group I intron ribozyme that does not require 2′ OH groups on its 5′ exon mimic for binding to the catalytic core. *Biochemistry* **36:** 15303–15314.

van der Horst, G. and Inoue, T. 1993. Requirements of a group I intron for reactions at the 3′ splice site. *J. Mol. Biol.* **229:** 685–694.

Zaug, A.J. and Cech, T.R. 1986. The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* **231:** 470–475.

Zaug, A.J., Been, M.D., and Cech, T.R. 1986. The *Tetrahymena* ribozyme acts like an RNA restriction endonuclease. *Nature* **324:** 429–433.