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

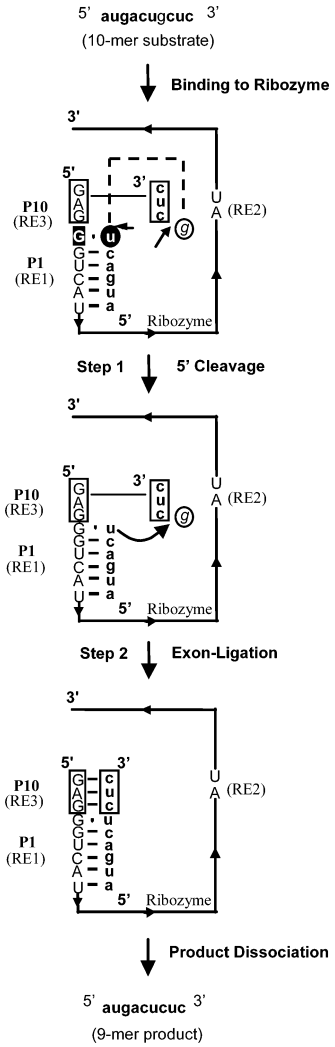

FIGURE 1: Schematic of the two-step TES reaction. The rP-8/4x ribozyme is in uppercase lettering, the 10-mer substrate is in lowercase lettering, and the single guanosine nucleotide to be excised is circled and in italics. The base to be excised corresponds to the ω position when it is part of self-splicing 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): $^{5'}$CGACTCACTATAGAGCGTCATGAAAGCGGC$^{3'}$ and $^{5'}$GCCGCTTTCATGACGCTCTATAGTGAGTCG$^{3'}$ to create P-8/4x-5′C; $^{5'}$CGACTCACTATAGAGAGTCATGAAAGCGGC$^{3'}$ and $^{5'}$GCCGCTTTCATGACTCTCTATAGTGAGTCG$^{3'}$ to create P-8/4x-5′A, and $^{5'}$CGACTCACTATAGAGTGTCATGAAAGCGGC$^{3'}$ and $^{5'}$GCCGCTTTCATGACACTCTATAGTGAGTCG$^{3'}$ to create P-8/4x-5′U. The following primer pair was used to create P-8/4x-noP10: $^{5'}$CGACTCACTATAGGTCATGAAAGCGGC$^{3'}$ and $^{5'}$GCCGCTTTCATGACCTATAGTGAGTCG$^{3'}$. 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

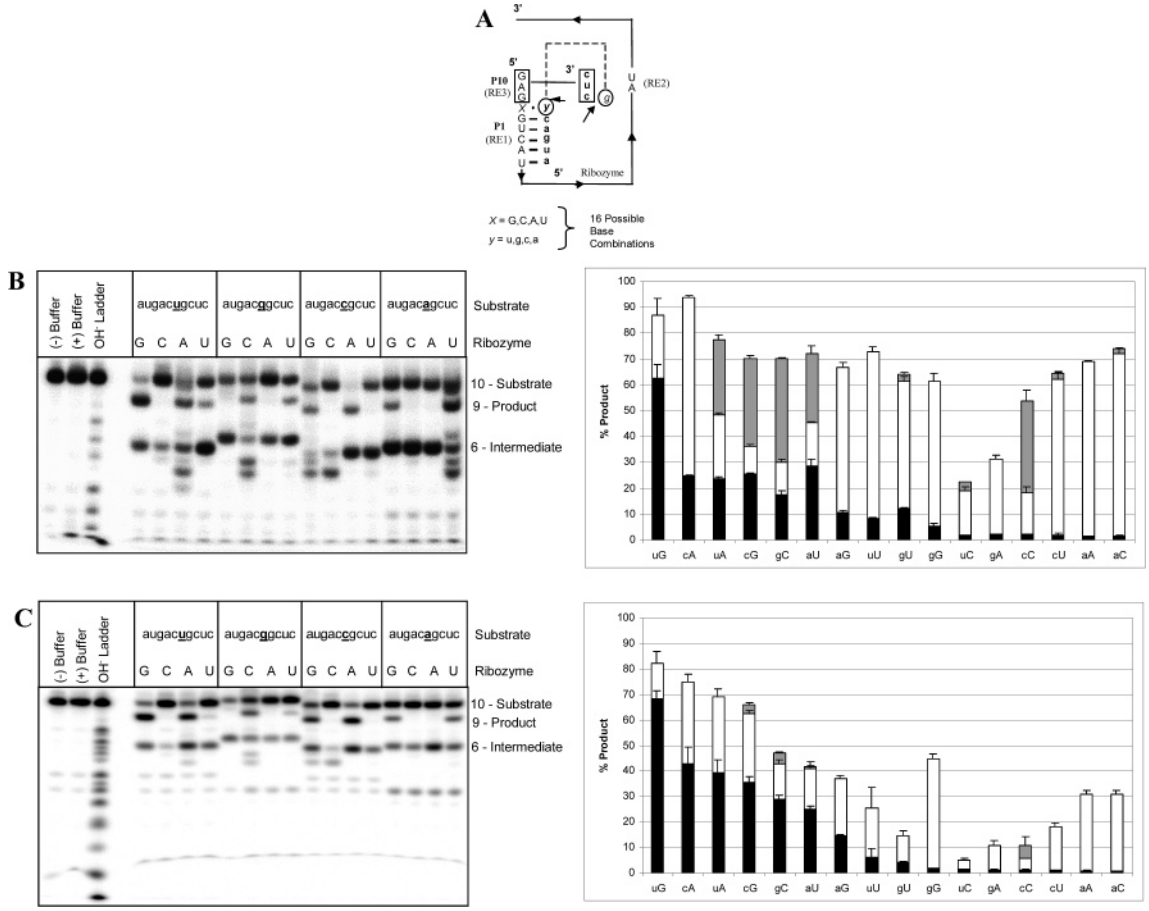

FIGURE 2: Sequence analysis of the 5′ splice site of TES reactions at 1 h and 15 min. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h (B) and 15 min (C) at 44 °C in 10 mM $MgCl_2$. (A) Diagram of the model TES reaction used. See Figure 1 for more detailed information. 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%).

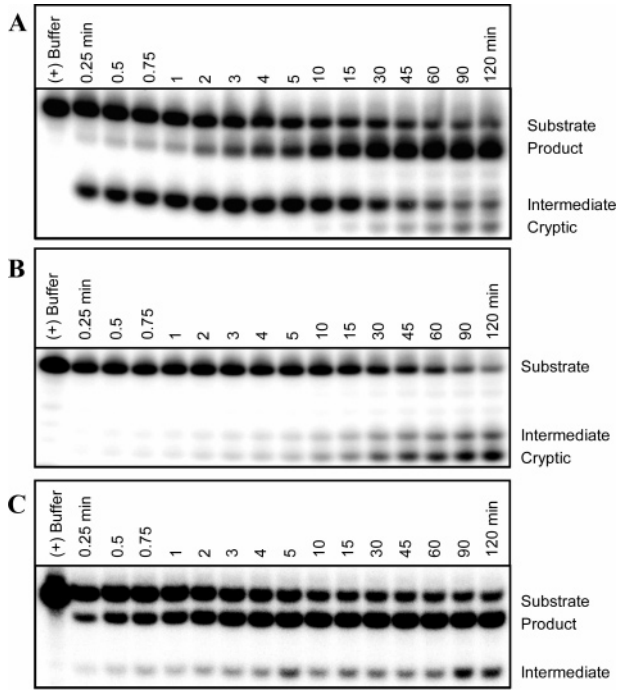

FIGURE 3: Time-dependent polyacrylamide gels of representative TES reactions. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. Aliquots were removed at the times listed above each lane. The migration position on the gel of the 10-mer substrates, 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),

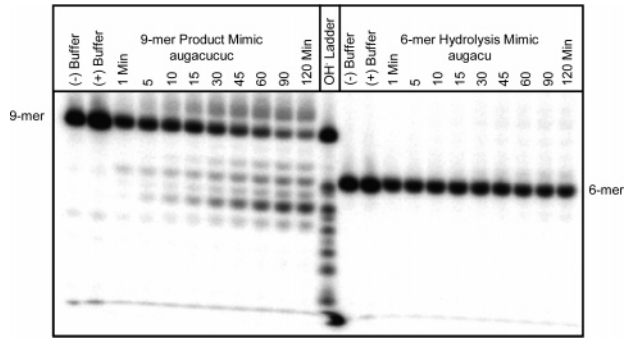

FIGURE 4: Time-dependent polyacrylamide gels of TES reactions utilizing 9-mer TES product (left) and 6-mer intermediate (right) as reaction products. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. The ribozyme used in each case was rP-8/4x-5′A, 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.

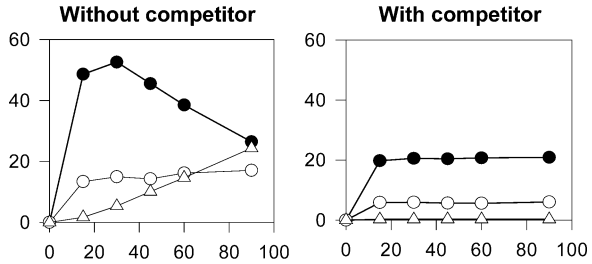

FIGURE 5: Competition TES reactions. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme at 44 °C in 10 mM $MgCl_2$. The ribozyme used in each case was rP-8/4x-5′A, which, when paired with the substrate augacugcuc, will create a u−A base pair at the 5′ splice site. After the 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

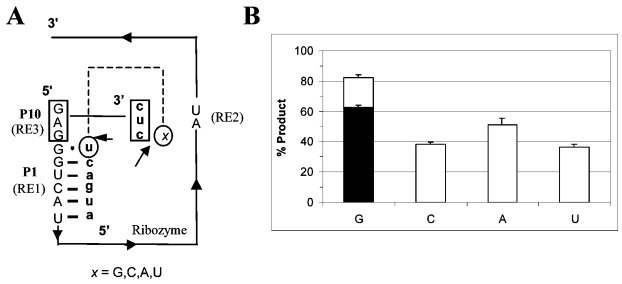

FIGURE 6: Sequence analysis of the 3′ splice site of TES reactions at 1 h. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in 10 mM $MgCl_2$. (A) Diagram of the model TES reaction used. See Figure 1 for more detailed information. The four substrates utilized 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.

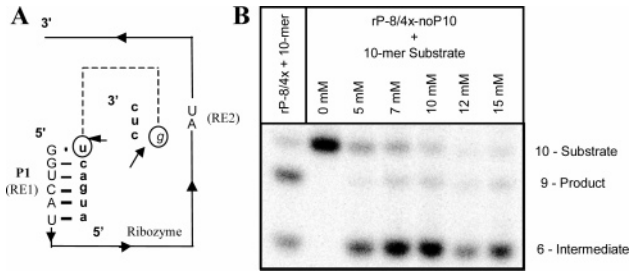

FIGURE 7: Magnesium dependence of TES reactions using the ribozyme rP-8/4x-noP10. Reactions were conducted with 1.3 nM radiolabeled substrate and 166 nM ribozyme for 1 h at 44 °C in the $MgCl_2$ concentrations listed below each gel lane. (A) Diagram of the model TES reaction used. See Figure 1 for more 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 *ω*G 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 *ω*G 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 (*ω*G, P10, and P9.0), only *ω*G 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 *ω*G 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 *ω*G 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