Folding Problems of the 5′ Splice Site Containing the P1 Stem of the Group I Thymidylate Synthase Intron

SUBSTRATE BINDING INHIBITION IN VITRO AND MIS-SPLICING IN VIVO*

Received for publication, December 11, 2001, and in revised form, February 19, 2002
Published, JBC Papers in Press, February 26, 2002, DOI 10.1074/jbc.M111798200

Andrea Pichler and Renée Schroeder‡
From the Institute of Microbiology and Genetics, University of Vienna, Vienna Biocenter, Dr. Bohrgasse 9/4, A-1030 Vienna, Austria

We developed an in vitro cleaving assay for the thymidylate synthase (td) group I intron and observed that the off-rate of the substrate is faster than cleavage. From the sequence stems P1 and P2 can vary from 4 to 8 and from 6 to 10 base pairs, respectively, with folding of a long P1 stem being in competition with folding of a long P2 stem. Shorter substrates, which cannot compete with the formation of an extended P2, result in faster cleavage, suggesting that binding of the substrate indeed interferes with folding of stem P2. In vivo splicing analyses of mutants containing alterations in stems P1 and P2 indicate that the wild-type exon sequence of P1 is suboptimal for splicing. Furthermore, folding of P1 in vivo is in competition with an alternative cryptic P1 stem resulting in mis-splicing. Translation promotes splicing at the correct 5′ splice site, whereas in the absence of translation, mis-splicing is favored. The combination of the in vitro and in vivo assays clearly displays the folding problems for correct splice site selection in this group I intron.

Group I introns are found in rRNAs, tRNAs, or mRNAs (1, 2). The structure of the flanking exons influences intron folding and contributes significantly to the energy landscape for folding both in vitro and in vivo (3, 4). For example, the exon 1 structure of the Tetrahymena rRNA intron affects folding of the 5′ splice site, reducing misfolding and thereby facilitating splicing (5–7). Although formation of the secondary structure in the core of group I introns is an intrinsic property of the RNA, assembly of the tertiary structure and docking of the splice sites often require additional splicing factors (8–11).

For the Tetrahymena rRNA intron, folding of the 5′ splice site containing P1 stem has been extensively characterized in vitro using a trans-cleaving ribozyme and providing the 5′ splice site as a substrate in trans (12, 13). Substrate binding is a two-step process (14, 15). The internal guide sequence (IGS) defines the 5′ splice sites (SS) by base pairing to the exon sequences, forming a stem named P1. In the first step of binding, the substrate base pairs with the IGS, and in the second step, the P1 stem is docked into the catalytic core of the intron via several hydrogen bond interactions (14, 16–21). The structural domains after P1 vary substantially between group I introns; some introns contain no P2 stem, some contain a P2 stem alone, and others contain P2 and P2.1 stems (2, 22).

For the T4 phage-derived td intron, which is inserted into the mainly unstructured coding region of the thymidylate synthase gene, no specific splicing factor has been identified until now. We have shown that in vivo splicing of this intron requires translation (23) and that inhibition of translation results in splicing deficiency and in increased mis-splicing due to alternative base pairing of the IGS, resulting in a so-called cryptic P1 stem (24, 25). To gain a better understanding of td intron folding and substrate binding, basic kinetic parameters were established for the first step of splicing in a trans-cleaving ribozyme. A reaction overview is given in Equation 1.

\[
E + S + G = E \cdot S \cdot G = E \cdot P_1 \cdot P_2 \cdot G = E + P_1 + P_2 \quad (\text{Eq. 1})
\]

For simplicity the ribozyme td WT-12 is called ribozyme or E, the oligonucleotide substrate WT-S is referred to as substrate or S, and the guanosine cofactor is termed G. The resulting products after the cleavage reaction are referred to as P1 for the 5′ product (corresponding to exon1) and P2, the 3′ product (corresponding to the intron) to which the cofactor is added.

Kinetic analysis of the trans cleavage reaction revealed that the off-rate of the substrate is faster than the cleavage step. One possible explanation for this poor binding might be a competition between substrate binding and the formation of an extended P2 stem. This prompted us to analyze substrate binding using shorter substrates in vitro and to introduce mutations in stems P1 and P2 in vivo to better understand the folding of the td intron at the 5′ splice site in its natural context.

EXPERIMENTAL PROCEDURES

Transcription of the Ribozyme—The ribozyme was transcribed in vitro from PCR products amplified from plasmid tdAP6-2 (26) with the following primers: the coding primer with the TD T7-7 (5′-CCA AGT AAT ACG ACT CAC TAT AGG GCC TGA GTA TAA GGT GA-3′, VBC-Genomics, Austria) containing the T7 promoter and the intron sequence, starting 7 nucleotides downstream of the 5′ splice site, whereas the antisense primer TD 3′ end-5′-TGT TCA GAT AAG GTC GTT AAT C-3′, VBC-Genomics) ended 5 nucleotides upstream of the 3′ splice site. The in vitro transcription reaction mixture contained 2.5 μg of purified PCR product, 5 mM each NTP, 10 mM dithiothreitol, and 150 units of T7 RNA polymerase in 100 μl of a solution of 40 mM Tris-HCl, pH 6.9, 26 mM MgCl2, and 3 mM spermidine. After a 4-h incubation at 37 °C, the DNA template was digested with 10 units of RNase-free DNase I for 30 min at 37 °C. The reaction was stopped with the addition of EDTA to a final concentration of 50 mM. The RNA was precipitated and purified on a 5% polyacrylamide (29:1, acrylamide:bisacrylamide), 7 M urea gel. After UV shadowing the correct band was cut out of the gel. To elute the RNA the gel slice was soaked in 10 mM Tris-HCl, pH...
Starting with the addition of 20 mM EDTA in 90% formamide with 0.005% xylene cyanol, 0.01% bromphenol blue, and 1 mM Tris, pH 7.5. Substrates and products were separated by electrophoresis on 20% polyacrylamide gels as described previously (27). The labeled substrates were purified on a 20% polyacrylamide gel as described for the RNA transcription.

**Kinetics**—All reactions were single turnover, with ribozyme, td WT-12, in excess of 5′-end-labeled substrate (S′, ~0.1 nM). The ribozyme was prefolded in 50 mM NaMOPS, pH 7.0, and 10 mM MgCl₂ at 50 °C for 10 min and then shifted to 25 °C for 5 min. The reaction at 25 °C was then started with the addition of S′ and 1 mM guanosine in 50 mM NaMOPS, pH 7.0, buffer and 10 mM MgCl₂. Usually six aliquots (dependent on the kinetics of the reaction) of 2 µl were removed from 20 µl reactions at specified times and quenched by the addition of ~2 volumes of 20 mM EDTA in 90% formamide with 0.005% xylene cyanol, 0.01% bromphenol blue, and 1 mM Tris, pH 7.5. Substrates and products were separated by electrophoresis on 20% polyacrylamide, 7.5 w/v% urea gels. The ratio of products to substrates were quantitated using a Molecular Dynamics PhosphorImager and the software ImageQuant. Reactions were usually followed for ~3 half-lives. The disappearance of substrate was first order, and end points of ~97% were typically obtained in nonlinear least squares fits (KaleidaGraph, Synergy Software, Reading, PA).

**Chase Experiments**—2 µl of 1.5x ribozyme (4 µM final concentration) to be well above the KD and ensure substrate binding) were prefolded as described above in the presence of MgCl₂ and NaMOPS, pH 7.0. After shifting to 25 °C, 1 µl of labeled substrate was added. After an incubation of 1 min, 27 µl of unlabeled chase (in 10 mM MgCl₂, 50 mM NaMOPS, pH 7.0, and 1 mM guanosine) was added in a 200-fold excess to the ribozyme. Aliquots were taken after various time points to analyze the cleavage reaction in the presence of the chase. In a control the chase was omitted from the experiment to test whether the reaction goes to completion.

**Binding Assays**—For the binding assays the ribozyme was prefolded as described above. After the addition of the labeled substrate (~0.5 nM S′ in 10 mM MgCl₂ and 50 mM NaMOPS, pH 7.0) the reaction was incubated at 25 °C. After an incubation time the samples were mixed with 30% glycerol and loaded immediately on a 8% non-denaturing polyacrylamide gel running at 5 V for 1 h. The electrophoresis buffer contained 66 mM Hepes, 34 mM Tris, 10 mM MgCl₂, and 0.1 mM EDTA at a pH 7.5. The separated bands were quantified on a PhosphorImager. Binding was analyzed using the program PrismGraph (GraphPad Software) to obtain KD values for substrate binding.

**Mutated td Constructs**—All the td constructs are in the context of the vector pTU18u-ttdΔP6-2 (28). The mutants were generated by in vitro mutagenesis (29) using the following mutagenic oligonucleotides (MW-Biotech AG, Germany) with the mutagenic residues underlined: P1 stabilizing (P1 stab), 5′-GGG CCT CAA TCA GCT GAG AAA ACA TCT ACT GAG CG-3′; P1 destabilizing (P1 dest), 5′-GGG CCT CAC TAA CCC CCG TGG AAA ACA TCT ACT GAG CG-3′; P2 stabilizing (P2 stab), 5′-CCG CCT TTA AGA TTA TCG GTG AGA AGA AGA AGA AAT ACA GGC CTC AAT TAA CCC C TG-3′; P2 destabilizing (P2 dest), 5′-CCG CCT TTA AGA TTA TCG GTG AGA AGA AGA AGA AAT ACA GGC CTC AAT TAA CCC C TGA-3′; P3 destabilizing (P3 comp), 5′-CCG CCT GAT TTA AGA TTA TCG GTG AGA AGA AGA AGA AAT ACA GGC CTC AAT TAA CCC C TG-3′; P3 stabilizing (P3 stab), 5′-CCG CCT GAT TTA AGA TTA TCG GTG AGA AGA AGA AGA AAT ACA GGC CTC AAT TAA CCC C TGA-3′.

**Cis Splicing in Vitro**—The construct plasmids were linearized with EcoRV (New England Biolabs) and in vitro transcribed at non-splicing conditions in the presence of (-23)GTP as described previously (30). For the splicing reaction 500–1000 cpm/µl of labeled RNA was incubated for 1 min at 65 °C. A mixture of splicing buffer (final concentration of 500 µM Tris-HCl, pH 7.3, 0.4 mM spermidine, 5 mM MgCl₂) and the RNA was incubated for 10 min at 37 °C. Before the addition of guanosine (final concentration of 500 mM) a zero control was taken. After the guanosine cofactor addition, various time points were taken, and splicing was stopped in 1 x volume of loading buffer (1 x TBE, 7 M urea, 0.065% xylene cyanol, 0.01% bromphenol blue). The splice products were separated on 5% polyacrylamide (29:1, acrylamide:bisacrylamide), 7 w/v% urea gel.

In Vivo Splicing—The Escherichia coli strain and growth media used in this study, the in vivo RNA isolation, and poisoned primer reaction was performed as described previously (31). Quantification was done with a Molecular Dynamics PhosphorImager, and average values where calculated from three different RNA preparations.

**Results**

**Conversion of the Intron into a Trans-cleaving Ribozyme**—To analyze the first step of splicing, the td intron (Fig. 1A) was converted into a trans-cleaving ribozyme. It was shortened by the first 7 nucleotides at the 5′-end of the intron and by 5 nucleotides at the 3′-end (Fig. 1B). The resulting RNA (named td WT-12) lacks the exons with the according SS. The substrate WT-S (Fig. 1B) used for the trans assays contains the 5′ SS flanked by the last 6 nucleotides of the 5′ exon and the first 4 nucleotides of the intron.

To verify that the ribozyme td WT-12 cleaves the substrate WT-S at a single position that correlates with the natural 5′ splice site, a cleavage experiment was performed. The 32P-labeled substrate was cleaved in the presence of 400 nM ribozyme and saturating amounts of guanosine cofactor in a time-dependent manner. The product band was then compared with a labeled oligonucleotide that was designed to represent the predicted product obtained after correct 5′ SS cleavage. In addition, a partial digest with P1 endonuclease was performed to verify the length of the product. The P1 endonuclease cleavage leads to formation of a 3′-OH terminus, which is also expected for ribozyme cleavage. The td WT-12 ribozyme cleaves the WT-S substrate at the correct position, and aberrant cleavage products were not observed (data not shown).

**Measured KD Values for the Substrate Are Higher Than Predicted**—It is expected that the dissociation constant (KD) for the substrate is a function of Watson-Crick pairings and additional long range interactions. Other factors that can affect this value include alternative interactions of the IGS that must be broken before duplex formation and steric hindrances within the ribozyme (32). To get a first estimate of KD for WT-S, the Gibbs free energy for helix formation between the substrate and the ribozyme IGS was calculated as described previously (33). The Gibbs free energy at 25 °C (ΔG25) is −9.3 kcal/mol. Given that ΔG25 = RT ln KDp, the calculated KDp is 140 nM for binding of WT-S substrate (see Table I).

By using single turnover reactions at varying ribozyme concentrations and saturating guanosine, the apparent dissociation constant of substrate WT-S to the ribozyme was determined. The conversion of the substrate to the product followed pseudo-first-order kinetics (Fig. 1C) and was measured for at least three half-lives. Plots of the observed rate constants (kobs) against ribozyme concentrations revealed saturation behavior with K1/2 of 530 nM (Fig. 2A).

Because the measured K1/2 value for the substrate WT-S was higher than expected from calculation, the dissociation constant was also measured with a second method, equilibrium binding. For that, the substrate was added to the prefolded ribozyme and equilibrated for 30 min and then loaded onto a non-denaturing 8% acrylamide gel. For that, the 3′ end of the substrate was extended with an additional 8% acrylamide gel. Complex formation in the absence of guanosine cofactor gave a KD of 520 nM (Fig. 2C). This value agrees well with the K1/2 measured using saturation kinetics, suggesting that the kinetic value represents an equilibrium constant.

Both dissociation constants show a weaker binding of WT-S to the IGS of the ribozyme than expected from the calculated value. For the Tetrahymena ribozyme the measured K1/2 is 1 nM, which is much lower than the calculated value and is explained by additional tertiary interactions that assist docking of the substrate into the ribozyme core (Ref. 16 and references cited therein). One possible explanation for the weak substrate binding of the td ribozyme is that P1 is not docked into the catalytic
core and that steric restrictions hinder efficient P1 formation.

The catalytic capability of the ribozyme was determined by the second-order rate constant \( (k_{cat}/K_m)^2 \) for the reaction of free ribozyme to substrate WT-S in single-turnover experiments with increasing \( [E] \) at saturating guanosine concentration. The value of \( k_{cat} \) increases linearly with the concentration of ribozyme (2.5–200 nM \( [E] \)). The rate of the reaction of free \( td \) ribozyme with the substrate \( (k_{cat}/K_m)^{WT-S} \) is \( 2.7 \times 10^5 \text{ M}^{-1} \text{ min}^{-1} \), which is 100-fold slower than the equivalent reaction rate for the \( Tetrahymena \) ribozyme (12).

A chase experiment was performed to establish the partitioning of the ternary complex \([\text{E}\text{S}^*\text{G}]\) (Scheme 1). The complex can either react \( (k_{churn}) \), resulting in product formation, or the labeled substrate \( (S^*) \) can dissociate \( (k_{dp}) \). To ensure that after \( S^* \) has fallen off, it cannot rebind to the ribozyme, the ternary complex is chased with a 200-fold excess of unlabeled substrate \( (S) \). This complex \([\text{E}\text{S}\text{G}]\) might also react and form products, but these cannot be detected since they are not labeled. The data show that no cleavage of \( S^* \) takes place upon the addition of the unlabeled chase substrate. This suggests that the off-rate of the substrate is faster than the cleavage step (Fig. 2B).

Shorter Substrates Improve the Observed Cleavage Constants—One possible explanation for the weak binding of the substrate and the relatively slow reaction rates might be that bases A14, G13, and U12 of the IGS bind to bases U33, U34, and G35 of J2/3, forming an extended P2 stem (Fig. 3A), thereby preventing binding or displacing the substrate. This interaction would compete with substrate WT-S binding and would explain the weak binding and the fast off-rate. To test this possibility, the substrate was shortened by 1–3 nucleotides (WT-S –1 nt, WT-S –2 nt, WT-S –3 nt), and the proposed interaction of A14 with U33 was disrupted in a mutated ribozyme (\( td \) U33C), which prevents the formation of the extended P2 stem (Fig. 3A, right). Table I shows the results for the different substrates and the mutated ribozyme tested with the WT-S substrate.

Shortening of the substrates did not only result in tighter binding than predicted from base pairing (see Table I, WT-S –3 nt is binding ~470 times stronger to the IGS than calculated) but also increases the rate of cleavage by an order of magnitude (see Fig. 3B). The second order rate constant \( (k_{cat}/K_m)^2 \) for the reaction of the free ribozyme with substrate WT-S –3 nt is \( 2.9 \times 10^6 \text{ M}^{-1} \text{ min}^{-1} \), which is 10-fold faster than for the WT-S substrate (data not shown). Chase experiments with gradually shorter substrates showed that the off-rates get slower according to the substrate length and that the cleavage steps also get gradually faster than the off-rates (data not shown).

We also tested a substrate that can in principle form more Watson-Crick base pairs upon formation of P1 than the WT-S. Using substrate TDS-4 (\( 5^\prime\)-UCAGGUUCAAA-3\(^\prime\)), 2 G-U pairs were changed to an AU and a GC pair. Substrate TDS-4 has a calculated \( K_D \) of 120 pM, but the \( K_D \) as measured with equilibrium binding was 49 nM (data not shown). Again, this weak binding (~400-fold measured/calculated) suggests that substrate binding is hindered in some way. The cleavage reaction of TDS-4 with the ribozyme \( td \) WT-12 shows a biephic behavior, with 60% of the substrate cleaved in an initial fast burst (1.6 min \(^{-1}\)), which is identical to the rate observed for the short
Effects of various substrates on P1 formation tested on two ribozymes

Table I

| Ribozyme | Substrate | Calculated $K_D$ | Measured $K_D$ (SK) | Calculated/Measured $k_{cat}$ | $h_{max}$ (min⁻¹) |
|----------|-----------|-----------------|---------------------|-----------------------------|------------------|
| td WT-12 | WT-S      | 140 nM          | 530 nM (SK)         | 0.26                        | 0.16             |
|          | WT-S −1 nt| 2.16 µM         | 98 nM (SK)          | 22                          | 0.43             |
|          | WT-S −2 nt| 14 µM           | 167 nM (SK)         | 84                          | 0.61             |
|          | WT-S −3 nt| 266 µM          | 562 nM (SK)         | 473                         | 1.6              |
|          | TDS4      | 120 µM          | 49 nM (EB)          | 0.02                        | Biphasic         |
|          | WT-S      | 140 nM          | 160 nM (EB)         | 0.9                         | 0.08             |
| td U33C  | WT-S      | 11000 nM        | 84 0.6              |                             |                  |

A biphatic behavior suggests that binding of the longer substrates is sterically not favored. P2 might be longer than previously proposed (2). To test whether the interaction between IGS and J2/3 affects docking, a mutation was introduced that blocks the pairing of U33 with A14 (td U33C; Fig. 3A). Equilibrium binding of substrate WT-S to the mutated ribozyme gave a $K_D$ of 160 nM, approximately the calculated value. In this mutant the formation of the extended P2 is blocked, and the calculated value now correlates with the measured value (Table I).

Mutants with Altered P1 and P2 Stems in a Pre-mRNA Context—Because the cleavage of the substrate WT-S was so inefficient in vitro, we analyzed whether the WT sequence of P1 is also suboptimal for splicing in vivo. Because the substrate WT-S used in the in vitro assay represents the exact sequence of exon I and the first nucleotides of the intron, the question arises as to what fold the RNA acquires for in vivo splicing, a long P1 stem or an extended P2 helix. To test the in vivo splicing efficiency of mutants with different P1 and P2 stem lengths, various constructs stabilizing or destabilizing the two helices were made (Fig. 4A). First, the upstream exon sequences proposed to pair with the IGS forming P1 were mutated so that P1 was either destabilized (P1 de) or stabilized (P1 stab). Then mutations were introduced into the J2/3 region in such a way that P2 could no longer form the proposed extended helix (P2 de) or that formation of the extended helix was favored (P2 stab). A mutation in the internal guide sequence disrupts both P1 and P2 stems (P1/P2). P2 was also reinforced in mutant P2, compensatory (P2 comp) with the difference to P2 stab that in this mutant formation of extended P1 was not possible.

Mutations in P1 Slightly Increase Splicing Efficiency, and Changes in P2 Are Not Tolerated—To determine in vivo splicing of the different mutants imbedded in the full-length exon context, total RNA was isolated from E. coli cells transformed with the different constructs. The RNA was then tested using reverse transcription as described previously (31). The WT td intron (WT), a translation-deficient mutant (SH-1), a splice-deficient mutant (td C870U), and the mutant variants with different P1 and P2 stems were tested for in vivo splicing. Fig. 5 and Table II show the results of these in vivo splicing experiments. The WT intron construct splices with 57% efficiency for the mRNA. In addition 2–5% aberrant splicing takes place at a
Folding of the td P1 Stem in Vitro and in Vivo

**Fig. 3.** Cleavage rates with substrates of different lengths. A, two alternative secondary structures of stems P1 and P2 are shown. On the left, a long P1 stem with a short P2 stem is bound to the substrate WT-S. On the right a short P1 stem is shown with a long P2 stem bound to a shortened substrate, WT-S −3 nt. An elongated P2 stem allows the interactions U12-G35, G13-U34, and A14-U33 to take place. To block this interaction a mutation was introduced, resulting in ribozyme td U33C (the position of mutation is shown on the right). B, the equation used to describe the decrease of the substrate is the same as in Fig. 1. The observed rate constants are as follows: mutated ribozyme (td U33C) with substrate WT-S (cross-hatch squares) 0.08 min⁻¹, and WT ribozyme (td WT-12) with substrate WT-S (circles) 0.16 min⁻¹, with substrate WT-S −1 nt (squares) 0.43 min⁻¹, with substrate WT-S −2 nt (diamonds) 0.6 min⁻¹, and with the shortest substrate, WT-S −3 nt (triangles) 1.6 min⁻¹.

Cryptic splice site, which is located 29 nucleotides upstream of the correct 5′ SS (24).

In agreement with the higher activities of the short substrates in vitro, the mutant with a destabilized P1 (P1 de, Fig. 5B, lane 5) splices to a higher percentage (72%) than the WT (25%). Also stabilization of P1 (P1 stab) results in slightly higher splicing activity in vivo, in agreement with the in vitro data obtained with the TDS-4 substrate (Fig. 5B, lane 4). P1 is more stable in this mutant and might fold more efficiently. Reinforcing P1 improves the $K_{d}$ of the substrate without abolishing the possibility of folding an extended P2. The two mutants, P1 stab and P1 de, are both more active in splicing than the WT, suggesting that the WT P1 sequence is suboptimal for splicing.

The mutants in which the extended formation of P2 was hindered (P2 de and ΔP1/ΔP2) resulted in a strong reduction of splicing activity, where the effect was stronger when P1 was able to form, 3% splicing versus 12% splicing (P2 de, Fig. 5B, lane 7, and Table II). Disrupting P1 and P2 simultaneously (ΔP1/ΔP2, Fig. 5B, lanes 9) results in a weaker phenotype than disrupting P2 alone. Stabilization of P2 (P2 stab and P2 comp) is also not tolerated. Again, the effect is stronger when P1 cannot be formed (3% splicing), as seen for mutant P2 comp (Fig. 5B, lane 8) and less dramatic when P1 can form (10% splicing) as seen for mutant P2 stab (Fig. 5B, lane 6). These results suggest that a balance is required between stems P1 and P2, where formation of an extended P2 might be coupled to docking of P1 into the core (Table II).

**Fig. 4.** Mutations of stems P1 and P2. A, mutations were introduced into the full-sized td gene to test the effect on splicing in vivo. Boxed areas indicate the mutated regions, and changed bases are shown in red. The mutations are the following: P1 stab, stabilizing stem P1; P1 de, destabilizing stem P1; P2 stab, stabilizing stem P2; P2 de, destabilizing stem P2; P2 comp, the destabilizing mutation in P2 is compensated with the complementary sequence in the intron; and ΔP1/ΔP2, destabilizing stems P1 and P2. B, a representative cis splicing assay is shown with the WT, a double mutant containing a stop codon at position −82 (SH-1), a splicing-deficient intron mutant (C870U), and the various constructs depicted in A. The splice products are ligated exons (E1/E2), intron and exon 3 (I/E2), exon 1 (E1), circularized intron (circ. I), free exon 2 (E2), and free intron (I).

In the Absence of Translation, Cryptic Splicing Is Increased—A cryptic splice site results from an alternative folding of the IGS with sequences upstream in exon 1 forming an alternative stem P1 (Fig. 6A). This cryptic P1 stem cannot compete with the formation of an elongated P2 stem. In contrast to splicing at the 5′ splice site, splicing at this cryptic splice site is not dependent on translation. As has been shown previously, inhibition of translation by the antibiotic neomycin B results in accumulation of cryptic splice product and inhibition of splicing at the 5′ splice site (25) (Fig. 6B). In a condition where the pre-mRNA is not translated due to a stop codon upstream in exon 1, cryptic splicing is enhanced, whereas correct splicing is severely reduced. When splicing in stop codon mutants is rescued in vivo by coexpressing StpA, a protein with RNA chaperone activity, the ratio of cryptic splicing to correct
splicing is almost 1:1 (31). In the absence of translation the equilibrium between cryptic P1 and P1 is shifted in favor of cryptic P1, which cannot compete with the formation of an extended P2. We therefore suggest that the ribosome might unfold the cryptic P1 stem and also favor an extended P2.

**DISCUSSION**

An *in vitro* trans-cleaving ribozyme for the td intron was constructed to study folding of the 5′ splice site-containing P1 stem. In contrast to the *Tetrahymena* intron, which is inserted into the highly structured large ribosomal subunit RNA, the td intron disrupts the coding region of the thymidylate synthase gene, which is mainly unstructured. When group I introns transpose into novel sites, folding of the splice sites has to adapt to the new sequence and structure context. Our results point to a suboptimal sequence of stems P1 and P2 for efficient splicing. We propose that an extended P2 stem with 10 base pairs favors docking of P1 to the ribozyme since the shorter substrates, which allow formation of extended P2, resulted in a tighter binding than calculated for Watson-Crick base pairing. This is not seen when formation of extended P2 is not possible. In the absence of an extended P2, P1 cannot be efficiently docked into the catalytic center and becomes rate-limiting. The cleavage rate of the ribozyme mutant U33C with substrate WT-S is only 2-fold slower than the wild type with the same substrate. This suggests that formation of an extended P2 is not essential but that it contributes to docking and improves the cleavage rate (Table I).

P2 might stabilize P1 by stacking. In the current model by Michel and Westhof (2) this stacking is distorted with a shift of about half or one base pair. From the wild-type sequence of P1, in which length can vary from 4–8 base pairs, P2 can vary from 6 to 10 base pairs, and J2/3 can vary from 3 to 7 nucleotides.

**FIG. 5.** *In vivo* splicing activity of P1 and P2 mutants. A, scheme for the splicing assay as tested with reverse transcription of total RNA with a primer complementary to the 5′-end of exon 2 (NBS-2) was used to differentiate between pre-mRNA, mRNA, and cryptic RNA. In the presence of excess ddTTP, reverse transcription of the pre-mRNA results in the primer extended by 5 nucleotides (primer + 5), cryptic RNA (primer + 8), and mRNA (primer + 16), except for P1 stab (primer + 6) and P1 de (primer + 7) because of sequence changes. B, a representative *in vivo* splicing assay is shown with the WT and the various constructs depicted in Fig. 4. C, splicing activities of the plot shown in B were quantified using ImageQuant. The results of at least three independent RNA preparations are shown in Table II. The abbreviations are as described in Fig. 2.

**FIG. 6.** Structure of the cryptic P1 stem and splicing in the presence of neomycin. A, secondary structure of the cryptic P1 stem is shown. The exon sequence and the cryptic splice site (cryptic SS) are depicted in green with numbering as described in Fig. 1. B, splicing assay of cells incubated with 0, 55, and 100 μM neomycin; the td gene is expressed from the lysogenized vector gt11 (25).

**TABLE II**

In *vivo* splicing efficiency of various constructs stabilizing or destabilizing P1 and P2

| Mutant                  | P1 formation | P2 formation | Splicing | Total splicing |
|-------------------------|--------------|--------------|----------|----------------|
| WT                      | ++           | +            | ++       | 57 ± 5.5       |
| P1 stabilizing          | ++           | +            | ++       | 64 ± 4.8       |
| P1 destabilizing        | –            | +            | ++       | 71 ± 2.4       |
| P2 stabilizing          | +            | –            | –/−      | 3.4 ± 0.4      |
| P2 destabilizing        | +            | ++           | –        | 3.0 ± 1.5      |
| P2 compensatory         | –            | ++           | −/−      | 11 ± 2.4       |
| ΔP1/ΔP2                 | –            | –            | +/−      |                |

*In vivo* splicing activity of P1 and P2 mutants. A, scheme for the splicing assay as tested with reverse transcription of total RNA with a primer complementary to the 5′-end of exon 2 (NBS-2) was used to differentiate between pre-mRNA, mRNA, and cryptic RNA. In the presence of excess ddTTP, reverse transcription of the pre-mRNA results in the primer extended by 5 nucleotides (primer + 5), cryptic RNA (primer + 8), and mRNA (primer + 16), except for P1 stab (primer + 6) and P1 de (primer + 7) because of sequence changes. B, a representative *in vivo* splicing assay is shown with the WT and the various constructs depicted in Fig. 4. C, splicing activities of the plot shown in B were quantified using ImageQuant. The results of at least three independent RNA preparations are shown in Table II. The abbreviations are as described in Fig. 2.
The length of J2/3 is 4 bases in the current model, but a refined model would allow a length of only 3 bases, as proposed by the model with the td ribozyme and the shortest substrate, WT-S −3 nt. An important question is the role of J2/3. Is it only a linker, or is the junction involved in three-dimensional interactions? It had previously been shown that deletion of the P2 stem is not tolerated but can be bypassed by shortening the P1 stem to 5 base pairs and J2/3 to 3 nucleotides (UAA) (28). This suggests that J2/3 is not involved in essential tertiary interactions but is needed to set the distance to P3. The td intron can be split by disrupting J2/3 and by providing a substrate in trans consisting of helices P1 and P2 (34).

Phage T4 carrying the td intron proliferates in prokaryotic cells, where transcription, splicing, and translation are coupled events. Formation of stem P1 might be part of a regulatory process required to orchestrate translation and splicing. It has been demonstrated that translation is essential for efficient splicing to occur in vivo (23). On the other hand translation of pre-mRNA with termination at the 5′ splice site must be prevented to inhibit the formation of a truncated amino-terminal portion of the thymidylate synthase protein. The sequence of the extended P1 stem represents a rare leucine codon, which remains to be analyzed. We are currently studying the regulation of translation of the td gene. Preliminary results point to a down-regulation of translation of the pre-mRNA compared with the mRNA.3

In the absence of translation, splicing is not only inefficient but also aberrant, since folding of an alternative P1 stem is favored, leading to the formation of a cryptic splice site. This alternative P1 stem does not compete with the formation of the extended P2 stem. The ribosome might unfold the cryptic P1 stem and also favor an extended P2. These results point to a more complex regulation of splicing by the ribosome. Translation and splicing of the td gene are coupled. We do not yet know whether the regulation of splicing and translation are bi-directional processes where splicing is regulated by translation and vice versa.

In this study we looked only at splicing events without taking into account that translation might also be regulated by the structure of the pre-mRNA. To fully understand the role of the cryptic P1 stem as well as the overlapping P1 and P2 sequences, we need to study the translation activity and the interrelation between translation and splicing.

Acknowledgments—We thank Dan Herschlag for inviting and teaching Andrea Pichler nucleic acid kinetics in his laboratory and for constructive comments on the manuscript. We thank L. Jaeger for carefully reading the manuscript and insights on modeling the td tetractery structure. We acknowledge Christina Waldsch and Rupert Grossberger for contributions on Fig. 6. We thank Oliver Mayer, Nicolas Pignareau, Rick Russell, Herbert Wank, Eric Westhof, and Nikolai Windbichler for comments on the manuscript.

REFERENCES

1. Cech, T. R. (1990) Annu. Rev. Biochem. 59, 543–568
2. Michel, F., and Westhof, E. (1990) J. Mol. Biol. 216, 585–610
3. Pan, J., and Woodson, S. A. (1998) J. Mol. Biol. 280, 597–609
4. Treiber, D. K., and Williamson, J. R. (2001) Curr. Opin. Struct. Biol. 11, 309–314
5. Woodson, S. A., and Emerick, V. L. (1993) Mol. Cell. Biol. 13, 1137–1145
6. Nikolaeva, T., and Woodson, S. A. (1999) J. Mol. Biol. 292, 557–567
7. Cao, Y., and Woodson, S. A. (2000) RNA (N. Y.) 6, 1248–1256
8. Mehr, G., Zhang, A., Gianello, J. A., Belfort, M., and Lumbowitz, A. M. (1992) Cell 69, 483–494
9. Weeks, K. M., and Cech, T. R. (1996) Science 271, 345–348
10. Lasovská, J., Jaroslav C., and Slonínský, P. P. (1988) Cell 52, 333–348
11. Ho, Y., Kim, S. J., and Waring, R. B. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 8994–8999
12. Herschlag, D., and Cech, T. R. (1990) Biochemistry 29, 10159–10171
13. Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., and Moneuse, E. (1998) Nat. Struct. Biol. 5, 60–66
14. Herschlag, D. (1992) Biochemistry 31, 1386–1399
15. Bevilacqua, P. C., Kierzek, R., Johnson, K. A., and Turner, D. H. (1992) Science 258, 1355–1358
16. Narlikar, G. J., Khosla, M., Usman, N., and Herschlag, D. (1997) Biochemistry 36, 2465–2477
17. Narlikar, G. J., Bartley, L. E., Khosla, M., and Herschlag, D. (1999) Biochemistry 38, 14192–14204
18. Pyle, A. M., McSwiggen, J. A., and Cech, T. R. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8187–8191
19. Strobel, S. A., and Cech, T. R. (1993) Biochemistry 32, 13593–13604
20. Szewczak, A. A., Ortoleva-Donnelly, L., Zivarts, M. V., Oyelere, A. K., Kazantsev, A. V., and Strobel, S. A. (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 11183–11188
21. Herschlag, D., Eckstein, F., and Cech, T. R. (1993) Biochemistry 32, 8299–8311
22. Strausse-Soukup, J. K., and Strobel, S. A. (2000) J. Mol. Biol. 302, 339–358
23. Semrad, K., and Schroeder, R. (1998) Genes Dev. 12, 1327–1337
24. Chandry, P. S., and Belfort, M. (1987) Genes Dev. 1, 1028–1037
25. Waldisch, C., Semrad, K., and Schroeder, R. (1998) RNA (N. Y.) 4, 1653–1663
26. Schroeder, R., von Ahsen, U., and Belfort, M. (1991) Biochemistry 30, 3295–3303
27. Zaug, A. J., Grosshans, C. A., and Cech, T. R. (1988) Biochemistry 27, 8924–8931
28. Salvo, J. L., Coetze, T., and Belfort, M. (1990) J. Mol. Biol. 211, 537–549
29. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Methods Enzymol. 154, 367–382
30. von Ahsen, U., Davies, J., and Schroeder, R. (1992) J. Mol. Biol. 226, 935–941
31. Clodi, E., Semrad, K., and Schroeder, R. (1999) EMBO J. 18, 3776–3782
32. Russell, R., and Herschlag, D. (1999) RNA (N. Y.) 5, 158–166
33. Serra, M. J., and Turner, D. H. (1995) Methods Enzymol. 229, 242–261
34. Costa, M., and Michel, F. (1997) EMBO J. 16, 3289–3302
35. Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994) Nat. Struct. Biol. 1, 273–280
36. Lehner, V., Jaeger, L., Michel, F., and Westhof, E. (1996) Chem. Biol. (N. Y.) 3, 993–1009
37. Brion, P., Schroeder, R., Michel, F., and Westhof, E. (1999) RNA (N. Y.) 5, 947–958
38. Belfort, M., Chandry, P. S., and Pedersen-Lane, J. (1987) Cold Spring Harbor Symp. Quant. Biol. 52, 181–192

2 L. Jaeger, personal communication.
3 K. Semrad and R. Schroeder, unpublished results.
