A Group II Intron Inserted into a Bacterial Heat-Shock Operon Shows Autocatalytic Activity and Unusual Thermostability†

Catherine Adamidi,‡ Olga Fedorova,‡,§ and Anna Marie Pyle*,‡,§

Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute, Yale University, Room 334A, Bass Building, 266 Whitney Avenue, New Haven, Connecticut 06520

Received December 9, 2002; Revised Manuscript Received January 27, 2003

ABSTRACT: Group II intron RNAs fold into catalytically active structures that catalyze their own self-splicing and subsequent transposition into DNA. Because of their remarkable enzymatic properties, it has been of interest to find new group II introns with novel properties. Here we report the cloning, sequencing, and mechanistic characterization of a new group II intron from the bacterium *Azotobacter vinelandii* (the AV intron). Although it bears the characteristics of the group IIB1 class, the AV intron is unusually G-C rich, and it has unusual insertion sequences and a minimal dependence on the EBS2−IBS2 tertiary interaction. The AV intron is the first bacterial intron that has been found to reside in a housekeeping gene which, in this case, encodes a heat-shock protein (hsp60). Consistent with a potential role in heat-shock regulation, kinetic analysis reveals that AV intron self-splicing is activated only at elevated temperatures. This suggests a novel pathway for the regulation of heat shock in prokaryotes and provides a first example of a thermally tolerant group II intron RNA.

Group II introns are large autocatalytic RNAs that can also act as retroelements. They are found in bacteria, mitochondria, and chloroplast genomes (*1*−*4*). Despite their lack of primary sequence conservation, group II intron RNAs adopt a consensus secondary structure comprised of six domains (D1−D6)[1] with specific roles in catalysis, folding, and protein expression (*1*, *5*). Comparative analysis of secondary and tertiary structural elements has defined the two major subgroups: IIA and IIB, with subdivisions A1, A2, B1, and B2 (*1*, *2*, *6*).

Group II introns catalyze a remarkable diversity of reactions during splicing and in the process of intron mobility. In addition, ribozymes derived from group II introns can cleave RNA and DNA substrates with multiple turnover (*7*−*10*). The splicing mechanism and modularity of group II intron domains suggest an evolutionary relationship between group II introns and RNA components of the eukaryotic spliceosome. The ORFs encoded by group II introns also share significant similarities with non-LTR retrotransposons and telomerases (*11*).

Although many group II introns are able to self-splice in vitro in the absence of protein cofactors, few of them have been extensively studied. The introns that have been examined exhibit a great deal of variation with respect to their ability to use hydrolysis or branching pathways, the efficiency of their self-splicing reactions, and their magnesium requirements (*12*−*14*). In particular, our knowledge of the activity of bacterial group II introns is very limited. The bacterial introns display a great diversity in sequence and secondary structure, which makes it difficult to classify them into canonical organellar groups IIA and IIB (*6*). Also, unlike their organellar counterparts, bacterial group II introns have not thus far been found to interrupt housekeeping genes. They are almost always associated with mobile genetic elements or located in intergenic regions of the bacterial genome (*3*, *15*). Therefore, the basis for the function and persistence of group II introns in bacterial genomes is unclear, and their locations suggest that they exist solely as selfish genetic elements.

In this work, we focus on a group II intron from the nitrogen-fixing bacterium *Azotobacter vinelandii* that was previously identified by Ferat and Michel (*16*). The earlier report provided a partial nucleotide sequence that spanned part of the ORF in D4−D6, from which we cloned the entire sequence of the AV group II intron, together with its flanking exons. Both intron RNA and ORF are closely related to the self-splicing mitochondrial introns from *Saccharomyces cerevisiae* (Sc.ai5γ) and *Pylaeilla littoralis* (Pl.LSU/1 and -2), and also from uncharacterized bacterial introns from *Escherichia coli* (EC.p0157), *Pseudomonas* sp., and *Xylella fastidiosa*, all of which belong to subgroup II B1. Although it is related to these other introns, the AV intron presents unusual structural features that diverge from typical subgroup IIB1 signatures. In this study, we examine the in vitro catalytic activity of the intron and probe its constituent structures by using both cis-splicing and trans-cleavage assays. Kinetic analysis reveals that the AV intron is a highly reactive RNA with an unusual level of thermostability.

† Funding provided by the Howard Hughes Medical Institute and National Institutes of Health Grant RO1 GM50313 (to A.M.P.).

* To whom correspondence should be addressed. Phone: (203) 432-5633. Fax: (203) 432-5316. E-mail: anna.pyle@yale.edu.

‡ Howard Hughes Medical Institute.

§ Department of Molecular Biophysics and Biochemistry.

[1] Abbreviations: AV, *A. vinelandii*; ai5γ, mitochondrial intron from *S. cerevisiae*; nt, nucleotide; D1−D6, domains 1−6, respectively; RT, reverse transcriptase or reverse transcription; PCR, polymerase chain reaction; ORF, open reading frame; ΔORF, open reading frame deleted; Δ292, D1 insertion deleted; hsp60, heat shock protein 60.

10.1021/bi027330b CCC: $25.00
Published on Web 03/08/2003

MATERIALS AND METHODS

Bacterial Strains and DNA Purification. A. vinelandii strain UW was obtained from D. Figurski (Columbia University, New York, NY). Cultures were grown at 37 °C in C medium (0.8 mM $MgSO_4$, 1.7 mM NaCl, 0.25% yeast extract, 0.05% casamino acids, and 1% mannitol). Genomic DNA was isolated according to a procedure that is effective with soil organisms, followed by ultracentrifugation purification through a CsCl gradient (*17*).

Ligation-Mediated PCR. We first amplified a fragment of the intron as described by Michel (*16*) using forward and reverse primers annealing to domain 6 and to RT finger 5 of the intron-encoded ORF in domain 4, respectively. This first step confirmed the identity of the intron and its presence in the UW strain.

Adaptation of the ligation-mediated PCR method (*18*) was used to recover the unknown upstream and downstream intron sequences. In this procedure, genomic DNA is first digested with a restriction enzyme. Three enzymes [*Bam*HI, *Sal*I, and *Nco*I (NEB)] were used sequentially in this study. Double-strand DNA adaptors were ligated to the each of the fragment ends (*19*). PCR was then performed using an intron specific primer and a primer that annealed to the adaptor sequence. PCR products were detected by ethidium staining agarose gels, purified using the Qiaquick gel extraction kit (Qiagen), and cloned into a TA-cloning vector (Promega) for subsequent DNA sequencing.

Consensus PCR Amplification and Cloning. The sequence of the intron and its flanking exons was obtained by standard PCR using Pfu Turbo DNA polymerase (Stratagene) and two specific primers designed to anneal to the 5′- and 3′-exons, respectively. The first construct (pAV1) used in this study contained the truncated intron DNA starting at intron position 83 and included the first 297 nucleotides of the 3′-exon. The full-length intron construct (pAV) contains the last 85 nucleotides of the 5′-exon, the 2798-nucleotide intron, and the first 297 nucleotides of the 3′-exon. Both were cloned between the *Bam*HI and *Hin*dIII sites of plasmid pBluescript (Stratagene). The plasmid constructs were used subsequently as a template for deletion mutagenesis by PCR.

Mutagenesis by PCR. Ribozyme constructs containing deletions were made using a mutagenic PCR method (*20*). The ORF encoded in domain 4 (D4, encompassing residues 920−2665) was removed in both the truncated (pAV1ΔORF) and full-length intron (pAVΔORF) constructs. The latter construct was then used as a template to delete an additional 292 nucleotides (from positions 11−302, described as the D1 292 nt insertion) that are located directly upstream of the AAG(AA) internal bulge. This third construct was labeled pAVΔ292ΔORF. For trans-cleavage assays, the ribozyme DNA (pAVΔ292ΔORFΔex) was obtained by the removal of the exons from the pAVΔ292ΔORF construct using consensus PCR amplification with a forward primer that includes a T7 promoter.

DNA Sequencing and Sequence Analysis. Sequencing was performed on an ABI 377XL automated DNA sequencer at the HHMI core facility (Columbia University). RNA secondary structure was inferred by phylogenetic comparison (*1*, *6*) and by using *mulfold* version 3 (*21*, *22*).

RNA Synthesis and Transcription. RNAs used in the self-splicing assays were obtained by in vitro transcription from *Hin*dIII-cut plasmids in the presence of [α-^{32}P]UTP, as described previously (*24*). The intron RNAs used in trans-cleavage assays (AV1ΔORF and AVΔ292ΔORFΔex) were transcribed as previously described (*27*) and purified on a 5% denaturing polyacrylamide gel. Concentrations were calculated spectrophotometrically.

Substrate RNAs were synthesized with an ABI 392 synthesizer and deprotected following the standard procedure. The oligonucleotides were then purified on 20% denaturing polyacrylamide gels. Concentrations were determined spectrophotometrically. Substrates were 5′-labeled with [γ-^{32}P]ATP and gel purified. Labeled substrate concentrations were determined from specific activity.

Self-Splicing Experiments. For each self-splicing reaction, precursor RNA (10 nM in a final volume of 20 μL) was first denatured at 95 °C for 5 min in 40 mM MOPS (pH 7.5) and 500 mM monovalent salt [$(NH_4)_2SO_4$, NH_4Cl, or KCl]. After the mixture had cooled to reaction temperature for 2 min, splicing was initiated by addition of $MgCl_2$ (100 mM). Aliquots (1 μL) were removed at specific time points, and reactions were quenched by addition of 5 μL of a formamide loading solution. Product and precursor molecules were separated by 5% denaturing gel electrophoresis and quantified with a Phosphorimager (Molecular Dynamics).

The fraction of unreacted precursor (% total RNA) was determined from the molar contribution of all intron products. Data were analyzed using Kaleidagraph (Synergy) and fit to Scheme 1 using the following equations for describing parallel kinetics (*23*, *24*):

$$\frac{\text{Precursor(t)}}{\text{Precursor}_o} = f_u + f_r \times \left(e^{-k_T t}\right) \qquad f_u + f_r = 1$$

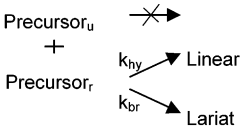
$$f_r = f_{hy} + f_{br}$$

$$\frac{\text{Linear(t)}}{\text{Precursor}_o} = f_{hy} \times \left(1 - e^{-k_T t}\right) \qquad k_T = k_{hy} + k_{br}$$

$$k_{hy} = \frac{f_{hy}}{f_{hy} + f_{br}} \times k_T$$

$$\frac{\text{Lariat(t)}}{\text{Precursor}_o} = f_{br} \times \left(1 - e^{-k_T t}\right) \qquad k_{br} = \frac{f_{br}}{f_{hy} + f_{br}} \times k_T$$

Scheme 1



where Precursor_o, Precursor_u, and Precursor_r represent the amount of precursor molecules at time 0 and of unreactive and reactive precursor molecules, respectively. The fractions of unreactive and reactive precursor molecules are labeled f_u and f_r, respectively. The abbreviations hy and br specify hydrolysis and branching reaction pathways, respectively.

Trans-Cleavage and Single-Turnover Kinetics. The oligonucleotide substrate RNAs were reacted with an excess of ribozyme in the following way. Prior to reaction, ribozyme RNA (100 nM in a final volume of 20 μL) was denatured at 95 °C for 5 min in 40 mM MOPS (pH 7.5) and 500 mM KCl. After the mixture had been cooled for 2 min to the reaction temperature, 100 mM $MgCl_2$ was added. The RNA was then preincubated for 5 min at the reaction temperature to promote proper folding of the ribozyme. Reaction then was initiated upon addition of 5′-labeled oligonucleotide substrates (1 nM). Aliquots (1 μL) were removed at specified

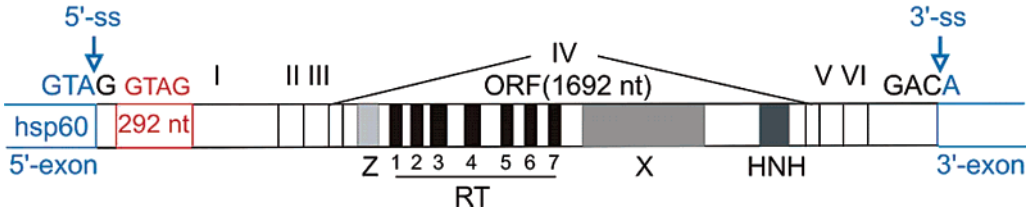


FIGURE 1: Region of DNA encoding the AVgroEL group II intron. The intron (black) is inserted between the TA$^{\downarrow}$A stop codon of the hsp60 gene. These exons are shown in blue. The intron is composed of six structural domains labeled I−VI. The first nucleotide and the last three nucleotides of the intron are shown in black. The 292 nt insertion located in D1 (I) is shown in red, which includes the putative exon−intron boundary that was initially identified. In D4 (IV), the encoded ORF is shown as a series of black and gray bars that indicate putative reverse transcriptase (RT 1−7), C- and N-terminal endonuclease (Z and HNH, respectively), and intron maturase (X) domains.

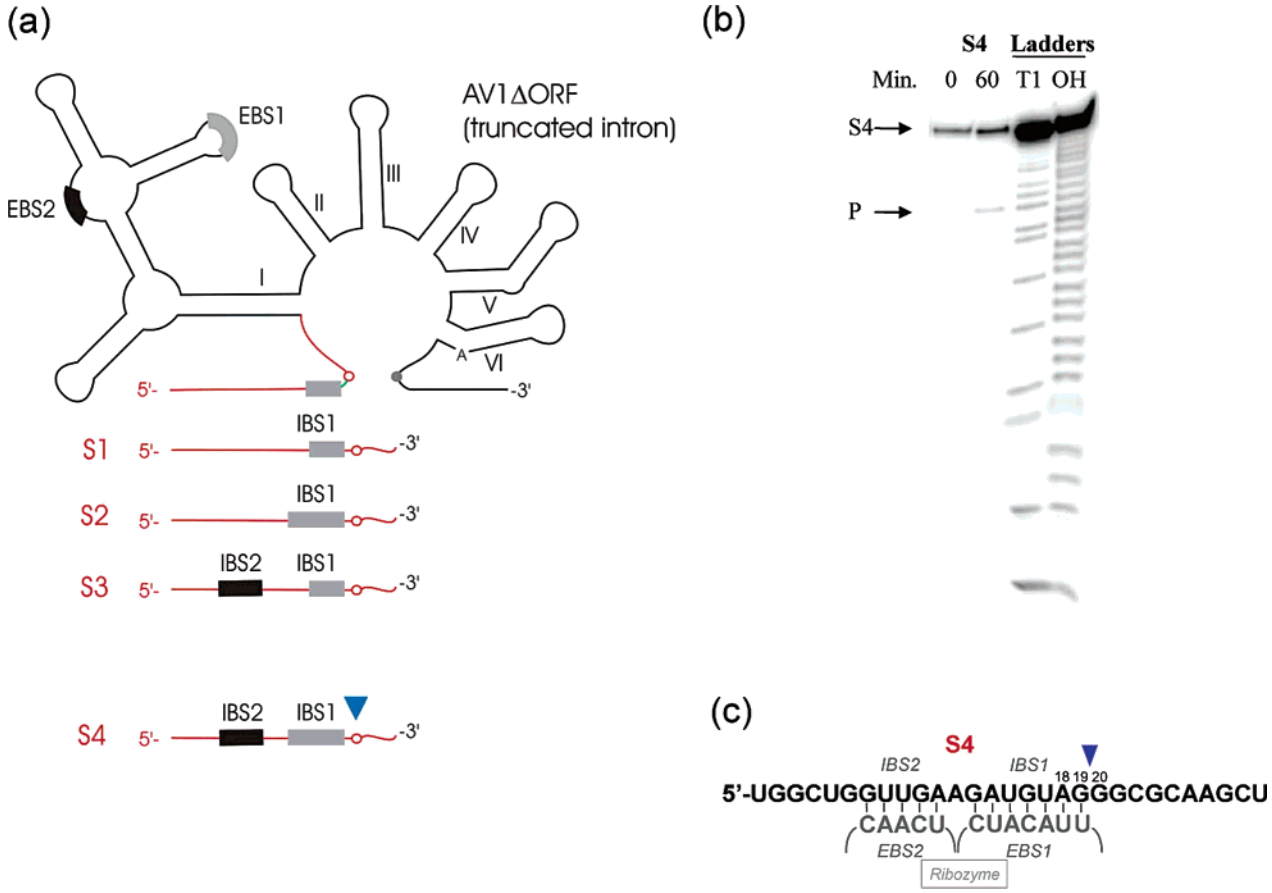


FIGURE 2: Rescuing the ribozyme activity by trans-cleavage experiments. (a) Schematic diagram of the AV1ΔORF ribozyme. The construct contains intact domains 2, 3, 5, and 6, a 5′-truncated domain 1, and domain 4 from which the ORF was removed. The putative 5′- and 3′-splice sites are labeled as empty and filled circles, respectively. This ribozyme was reacted with a set of 29 nt RNA substrates (S1−S4; see Materials and Methods) that contained 19 nt downstream and 10 nt upstream of the putative 5′-splice site. One substrate includes a putative 4 nt IBS1 sequence and no IBS2 (S1). Additional substrates contain either a 7 nt IBS1 with no IBS2 (S2), the short IBS1 sequence with a 5 nt IBS2 (S3), or both the extended IBS1 and the new IBS2 (S4). Blue arrowheads denote the cleavage site location on S4. (b) Mapping the cleavage site of substrate S4. Partial T1 nuclease and alkaline hydrolysis ladders (T1 and OH, respectively) are presented next to the corresponding ribozyme cleavage products (P) at time 0 and 60 min, respectively. (c) Scheme for cleavage site selection.

time points and reactions quenched with formamide loading dye before the mixtures were loaded onto a 20% denaturing polyacrylamide gel. Reaction products were visualized and quantified using a Phosphorimager (Molecular Dynamics). Data were fit to single-exponential equations (*10*) using Kaleidagraph (Synergy).

The sequences of substrate RNAs used in this study are shown here: S1, UGGCUGCUCCUAUCCGUAGGGCGCAAGCT; S2, UGGCUGCUCCUAGAUGUAGGGCGCAAGCT; S3, UGGCUGGUUGAAUCCGUAGGGCGCAAGCT; S4, UGGCUGGUUGAAGAUGUAGGGCGCAAGCT; S5, CAUGGGCGGCAUGAUGUAGUGCGA; S6, CAUGGUUGACAUGAUGUAGUGCGA; S7, CAUGAUGUAGUGCGA; S8, CAUGGGCGGCAUGAUGUAGAAAAA; S9, AAAAAAAAACAUGAUGUAGUGCGA.

Mapping of the Cleavage Site by Alkaline Hydrolysis, and P1 and T1 Nuclease Digests. To map the exact location of the cleavage sites, products of trans-cleavage reactions were resolved on 20% denaturing polyacrylamide gels, in parallel with partial digestions of the RNA substrates. Partial digestion with endonuclease T1 and alkaline hydrolysis were carried out under standard conditions (*44*). The method for P1 partial digests is described by Su et al. (*10*).

RT-PCR Mapping of Splice Sites. The splice sites were mapped by sequencing the ligated exons using RT-PCR. An antisense primer designed to anneal 120 nucleotides down-

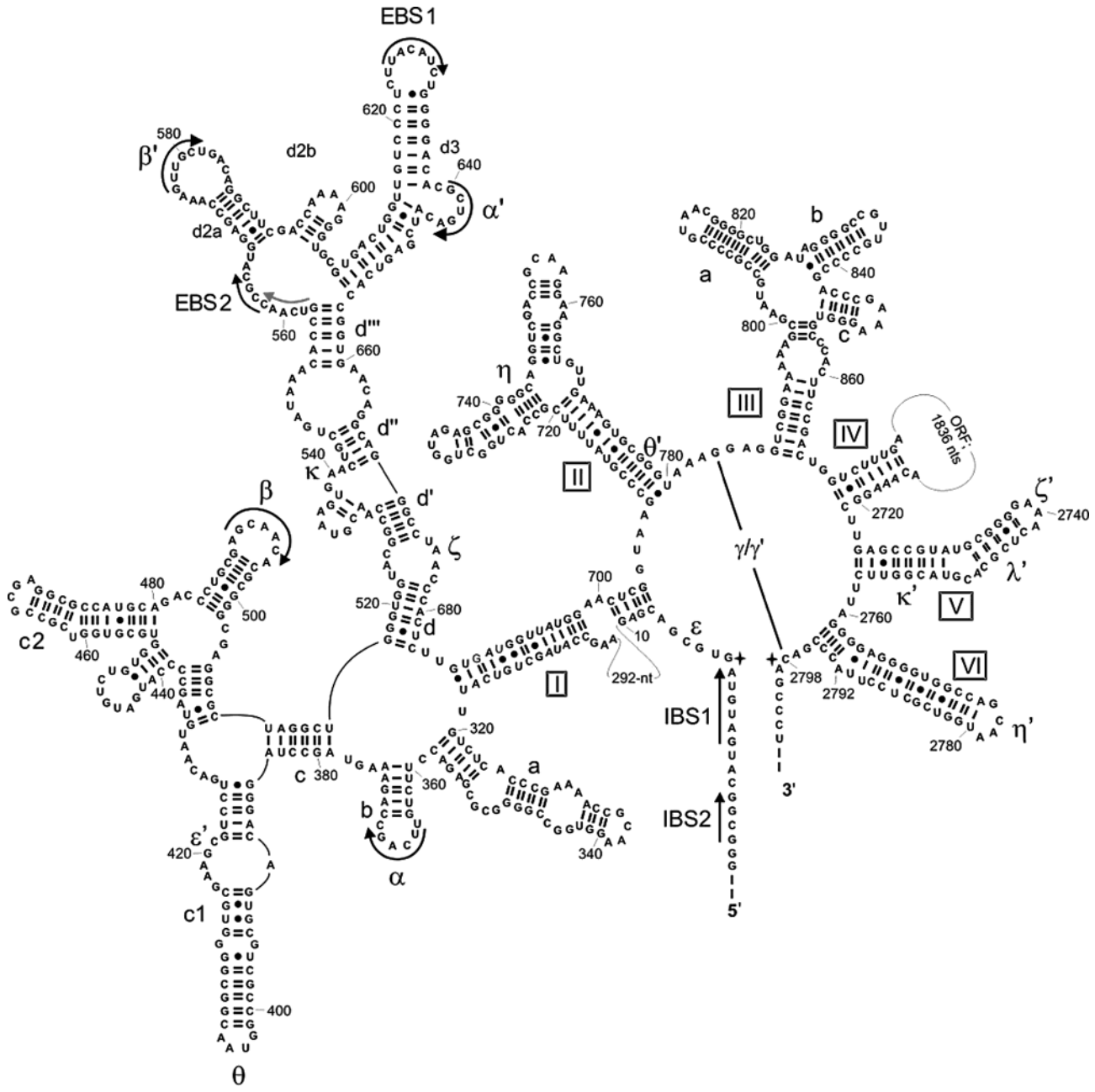


FIGURE 3: Intron sequence and secondary structure map. Secondary structural domains are labeled with boxed roman numerals (I−VI). Tertiary pairings common to most group II introns are indicated by Greek letters (i.e., $\epsilon-\epsilon'$, $\gamma-\gamma'$, $\kappa-\kappa'$, and $\lambda-\lambda'$). When the tertiary interaction consists of W-C base pairs, the arrows indicate the extent of pairing ($\alpha-\alpha'$, $\beta-\beta'$, EBS1−IBS1, and EBS2−IBS2). Part of D4 (1836 nt) that includes the encoded ORF, and the 292 nt insertion in D1, are represented as loops. In the D1d''' loop, the initially hypothesized EBS2 sequence is highlighted in gray, while the correct one is shown in black. Stars indicate the splice junctions.

stream of the 3′-splice site in the 3′-exon was used during the first RT step. In the second step, PCR was performed using the antisense RT primer and a sense primer targeting the 5′-exon complementary strand 56 nucleotides upstream of the 5′-splice site.

Self-splicing was performed at 55 °C for 1 h with 15 nM unlabeled precursor RNA in a final volume of 50 μL. Aliquots of 0.5 and 2 μL of the reaction mixture were each tested as template RNA for reverse transcription. After the RNA template had been denatured at 95 °C for 5 min, first-strand synthesis was carried out at 42 °C for 1 h using 2 pmol of RT primer and 4 units of AMV reverse transcriptase (Invitrogen), as described in the manufacturer's protocol. One-tenth of the reaction mixture was subsequently used as template DNA in a 50 μL PCR, containing 25 pmol of each RT antisense and PCR sense primers and 1 unit of Taq DNA polymerase (Roche). Preincubation at 95 °C for 5 min was followed by 25 cycles of 45 s at 95 °C, 45 s at 60 °C, and 45 s at 72 °C. Amplification products were visualized and purified on a 1.2% agarose gel stained with ethidium bromide. Products were eluted from the gel with a Qiaquick purification kit (Qiagen), and inserted into a TA-cloning vector (Promega) for DNA sequencing.

RESULTS

Sequence and Domain Structure of the AV Intron. The AV intron RNA adopts a canonical six-domain secondary structure and includes all known tertiary interactions that are characteristic of self-splicing eukaryotic group IIB1 introns

(*1*, *6*). It contains a 1692-nt ORF in D4 that encodes reverse transcriptase, maturase, and endonuclease motifs (Figure 1). The previously published 86-amino acid sequence from *A. vinelandii* strain UWR (*16*) (GenBank entry 479670) matches with 100% identity a fragment of sequence that we find in maturase domain X. Two YADD motifs are present within the intron ORF: one within the highly conserved VRYADD sequence present in finger 5 of the RT domain and the other in domain X where a YADD motif is not usually found.

Despite strong conservation of the intron domains, the 5′-exon−intron boundary was initially difficult to identify and seemed unusually degenerated. We initially presumed that the 5′-splice site was located 297 nucleotides downstream of the GroEL gene (hsp60, Figure 1). The resulting intron 5′-end did not follow the highly conserved GUGYG sequence (*1*), and it contained an EBS1−IBS1 base pairing that was truncated to a four-base pair helix; the EBS2−IBS2 interaction appeared to be missing.

When we deleted the ORF from the intron DNA and cloned the AV1ΔORF construct with flanking 5′- and 3′-exons under a T7 promoter, the resulting RNA was unable to self-splice in vitro under a variety of conditions.

Rescuing the Ribozyme Activity through Trans-Cleavage Experiments. Considering the extent of the primary, secondary, and tertiary structure conservation of the AV1 intron RNA, we predicted that its inability to self-splice in vitro was not due to misfolding or to the absence of cis- or trans-encoded factors, but rather, it was due to difficulties in forming a stable intron−exon interaction. To test this hypothesis, we decided to use the ability of group II introns to function as trans-acting ribozymes. As such, group II introns can cleave small oligonucleotides containing sequences analogous to the 5′-exon−intron boundary (*8*, *25*). Substrate binding occurs through base pairing interaction between EBS sequences on the intron and IBS sequences on the oligonucleotide substrate.

We designed a series of 29 nt substrate RNAs that were designed to restore and test the degenerated EBS−IBS interactions (Figure 2a; for sequences, see Materials and Methods). One substrate consisted of the expected 5′-exon−intron boundary (S1). The second contained a set of mutations that extended the IBS1 interaction from a 4 nt to a 7 nt helix (S2). The third substrate (S3) contained four single mutations upstream of the expected IBS1 sequence. These mutations were designed to promote base pairing with a region that normally contains the EBS2 sequence found in group IIB introns (the loop in D1d‴). Finally, a fourth substrate (S4) contained a combination of mutations that were predicted to create an extensive set of interactions between the ribozyme and the trans substrate (S4).

After chemical synthesis and 5′-end labeling of the oligonucleotide substrates, we tested the trans-cleavage reactivity of our new intron under single-turnover conditions (see Materials and Methods). Of the four substrates, only S4 was cleaved by AV1ΔORF RNA (Figure 2b). The reaction requires a high concentration of monovalent and divalent cations. Activity of the ribozyme was low (0.5 h^{-1} at 55 °C), although it is notable that reactivity increased when the incubation temperature was increased from 40 to 55 °C.

Cleavage sites were mapped using T1 and OH RNA ladders (Figure 2b) and a P1 ladder (data not shown), which showed that ribozyme cleavage occurs at the terminus of the expected EBS1−IBS1 loop−loop interaction (Figure 2c). Thus, once the apparent EBS−IBS interactions had been extended, the intron RNA acted as an active ribozyme. From these initial results, we concluded that the intron RNA folds into a catalytically active structure, even if it appears to be unable to splice.

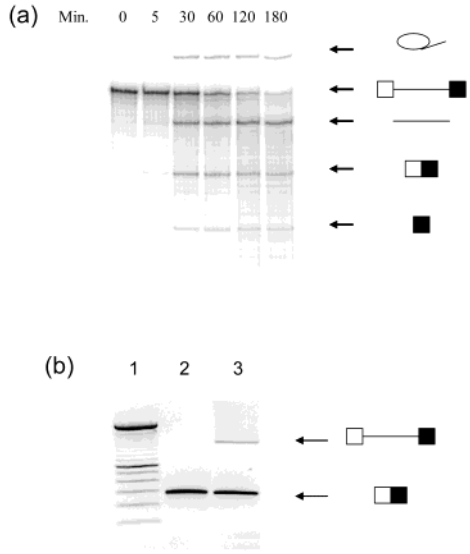


FIGURE 4: Self-splicing reaction catalyzed by the AVΔ292ΔORF ribozyme. (a) Time course of the self-splicing reaction. Products from self-splicing were separated by 5% denaturing PAGE. Time points in minutes are indicated above each lane. From top to bottom, the RNA species are the lariat intron, precursor RNA, linear intron, ligated exons, and 3′-exon. The intron is represented by a continuous line, and exon sequences are depicted as boxes, with white and black for 5′- and 3′-exons, respectively. (b) Mapping the splice sites by RT-PCR. RT-PCR products were taken from splicing reaction aliquots (2 and 0.5 μL in lanes 2 and 3, respectively) and resolved on 2% agarose gels alongside 100 bp DNA markers (lane 1).

Identification of the True Intron 5′-End. When the S4 mutations were inserted into the 5′-exon of the AV1ΔORF self-splicing construct, the self-splicing defect was not rescued. This puzzling result, together with inspection of the S4 sequence, led us to look further upstream from the putative 5′-splice site that we had first identified. In upstream regions, we identified a 6 nt sequence that is identical to the IBS1 site that we designed to create the reactive S4 trans substrate (Figure 3). The new intron 5′-terminus follows the GUGYG consensus sequence. When folded in a conventional manner, the intron did not appear to contain an IBS2 sequence that matches IBS2 in substrate S4. However, by refolding domain D1d‴ and internal loops D1d3 and D1d2, we obtained an alternative intron secondary structure that presents a reasonable IBS2−EBS2 base pairing (shown in Figure 3). The IBS2−EBS2 interaction consists of four consecutive base pairs that are G-C rich (black arrow in Figure 3).

A particularly notable feature of the revised AV secondary structure is the unusually large 292-nt insertion that is located directly upstream of the conserved AAG(AA) internal bulge that is almost always found in the upstream helical terminus of D1 (Figure 3). Insertions at this location have not been previously observed in subgroup IIB1 introns, although they have been noted in IIB2-type introns, in certain bacterial introns, and in eukaryotic subgroup IIA atpF introns.

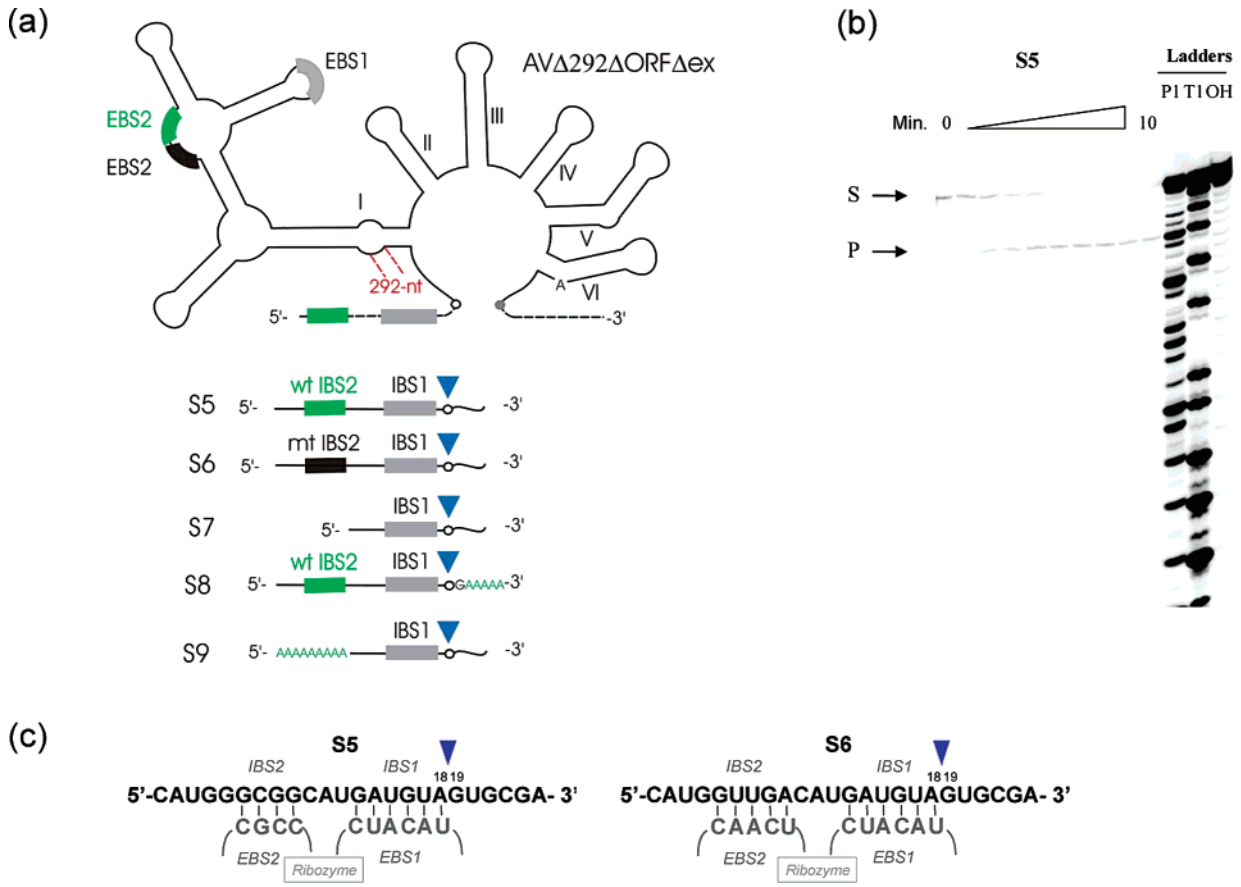


FIGURE 5: Oligonucleotide cleavage catalyzed by AV RNA. (a) Schematic diagram for the trans-cleavage assays. The ribozyme construct AVΔ292ΔORFΔex consists of intact domains 2, 3, 5, and 6, and domains 1 and 4 with deletions of the 292 nt insertion (in red) and the intron-encoded ORF. The 5′- and 3′-splice sites are labeled as empty and filled circles, respectively. The exons shown as dashed lined have been removed. Gray boxes represent IBS1 and EBS1 and green boxes IBS2 and EBS2 sequences. Black boxes represent the mutant IBS2 and the alternative EBS2 sequences. RNA substrates consist of the wild-type 5′-exon−intron boundary (S5), with mutant IBS2 (S2, mt IBS2), with deletion of the IBS2 region (S7), and with poly-A mutations downstream of the cleavage site (S8) or upstream of the IBS1 sequence (S9). Blue arrowheads denote the cleavage site. (b) Mapping the cleavage site of wild-type 5′-labeled substrate S5 using partial alkaline hydrolysis (OH) and nuclease digests (P1 and T1). The substrate (S) and product (P) were separated on 20% PAGE and run side by side with P1, T1, and OH ladders. The time course of the trans-cleavage reaction (at a T_{opt} of 55 °C) is shown here, with time points taken at 5, 15, 30, 45, 60, 90, 120, 150, 300, and 600 s (from 0 to 10 min). (c) Scheme for cleavage site selection for S5 and S6 substrates.

However, to the best of our knowledge, none of the latter insertions exceeds 100 nt (*1*, *6*).

Perhaps most importantly, the revised AV1 intron contains splice sites that interrupt the stop codon of groEL (hsp60, Figures 1 and 3). This is notable because most bacterial group II introns are located either in mobile elements or in intergenic regions of the chromosome. The AVgroEL intron DNA is the first to reside in a key gene for the function of bacteria.

In Vitro Splicing of the Revised AV Intron. To test the self-splicing activity of the full-length intron, two new constructs were cloned under a T7 promoter. AVΔORF consists of the full-length intron (containing the newly identified 5′-splice site and the 292 nt D1 insertion). AVΔ292ΔORF contains an additional deletion that eliminates the D1 insertion to reconstitute a group IIB1 canonical AAG(AA) bulge (Figure 3). The resulting in vitro-transcribed RNAs were examined for self-splicing activity, and both were proven to be active ribozymes under high-salt conditions, i.e., 500 mM $(NH_4)_2SO_4$, NH_4Cl, or KCl and 100 mM $MgCl_2$ (data not shown). The smaller construct (AVΔ292ΔORF) exhibited higher activity and was used in all subsequent assays. Splicing in the presence of ammonium sulfate led to formation of lariat, linear intron, ligated exons, as well as 5′- and 3′-exons, revealing that both splicing pathways (branching and hydrolysis) compete as observed with the ai5γ self-splicing reaction (Figure 4a) (*24*). In potassium chloride, the extent of the branching reaction is greatly reduced in favor of the hydrolysis pathway.

We mapped the 3′- and 5′-exon−intron boundaries using an RT-PCR scheme with primers that target the 3′- and 5′-exons, as adapted from Vogel et al. (*26*). Two specific RT-PCR products can be seen after separation on a 2% agarose gel, depending on the amount of spliced RNA that is used as a template for reverse transcription (Figure 4b). All sequenced clones indicated that the products arose specifically from amplification of either ligated exon or precursor RNA. The former confirmed that the canonical splice sites were accurately chosen. This result also verified that the AVgroEL intron interrupts the stop codon of the GroEL ORF. After splicing of the intron from GroEL mRNA, the UAG stop codon will be replaced by a UAA stop codon (Figure 1). Thus, the intron is inserted within an essential gene, but accurate in vivo splicing may not affect the translation reading frame.

Trans-Cleavage Experiments with the Revised AV Intron. To further investigate the ribozyme activity and to probe the respective roles of IBS1 and IBS2, we designed new trans-

cleavage experiments with the revised intron sequence (AVΔ292ΔORFΔex). The ribozyme RNA was reacted with oligonucleotide substrates that were designed to mimic the new, proper, EBS and IBS sequences (Figure 5a). The substrates consisted of 24-mer oligonucleotides that were based on substrate sequence S5. An additional substrate contained the new IBS1 sequence, but retained the original IBS2 sequence that had been used in S4 (substrate S6). Substrates S4−S6 were cleaved by the AVΔ292ΔORFΔex RNA construct with equal efficiency (2 min^{-1}).

Mapping of the cleavage sites by partial digests with P1 and T1 nucleases, and by alkaline hydrolysis, confirmed that cleavage occurs specifically at the 3′-end of the IBS1 sequence (Figure 5b,c). In contrast to previous trans-cleavage experiments with the truncated AV1ΔORF ribozyme (Figure 2), the cleavage site has now shifted by one base pair and is now located downstream of the A-U base pair in the EBS1−IBS1 helix, directly at the revised 5′-splice site. When the mutant S4 substrate is cleaved by the AVΔ292ΔORFΔex ribozyme, the cleavage site is now the same as that for substrate S5. Of particular interest is the fact that the reaction rates for oligonucleotide cleavage by the AVΔ292ΔORFΔex ribozyme are now 200-fold faster than that of the truncated construct that we studied initially.

Role of the IBS2−EBS2 Interaction. The IBS2 sequence in the S6 substrate was expected to disrupt base pairing with EBS2 sequences on the revised AV intron (Figure 5a). However, the rate constants for cleaving both the S5 and S6 substrates are equivalent. To determine if the EBS2−IBS2 interaction is actually dispensable, we deleted the IBS2 region entirely (S7 and S9). Substrates S7 and S9 were cleaved with equivalent rates, but they reacted 10-fold more slowly than S5 (Figure 6b). Cleavage site fidelity was preserved in all three mutant substrates. Thus, ablation of the IBS2−EBS2 interaction does not have the radical effect observed with another group II introns (*9*, *27*), suggesting that the IBS2−EBS2 interaction plays only a minor role in the function of the AV intron.

To understand whether sequences downstream of the cleavage site contribute to intron recognition of trans substrates, we inserted a poly-A sequence downstream of the cleavage site (S8, Figure 5a). As observed previously for ai5γ ribozymes (*10*), the poly-A sequence had no effect on reaction efficiency for oligonucleotide cleavage (Figure 6). Taken together, these data suggest that the EBS1−IBS1 interaction is predominantly important for recognition by the AV intron.

High Thermostability of the AVgroEL Group II Intron. During self-splicing at 55 °C (Figure 4a), the fraction of spliced AVΔ292ΔORF RNA reaches only 50%. To improve the splicing efficiency and the extent of reaction, we tested the effects of different temperatures, which ranged from 45 to 70 °C (Figure 7). Remarkably, the splicing reaction had a temperature optimum (T_{opt}) of 65 °C. Reactivity dropped abruptly above 67 °C, and no reaction was observed at 70 °C except for nonspecific backbone hydrolysis. On the basis of the decreasing fraction of unreactive precursor RNA, the increase in the level of spliced products as a function of temperature appears to be due to a gradual reduction in the population of unreactive (and potentially misfolded) RNA. A large unreactive RNA fraction is observed at temperatures much lower than T_{opt} (see also ref *28*). From 50 to 65 °C,

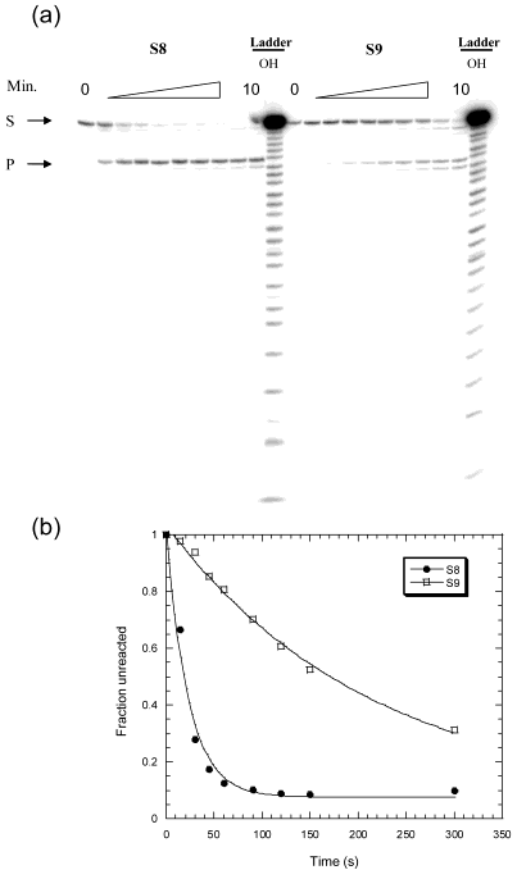


FIGURE 6: Influence of the substrate IBS2 region on the cleavage reaction. (a) Time course of cleavage for 5′-labeled substrates S8 and S9. Time points (5, 15, 30, 45, 60, 90, 120, 150, 300, and 600 s) were taken for reactions performed at 55 °C. Products of reaction were resolved by 20% PAGE, which was run with OH ladders. (b) Quantification of substrate evolution. The data were fitted to single-exponential equations from which the rates were extracted (*10*). At 55 °C, the k_{obs} values for substrates S8 and S9 are 2 and 0.2 min^{-1}, respectively. The secondary cleavage products observed here are likely to arise through the mechanism described by Su et al. (*10*).

the fraction of reactive precursor molecules increases from 30 to 85% of total RNA (Figure 7b,c). In addition to increasing the fraction of active molecules, a higher temperature also affects the apparent rate of branching. After accounting for the unreactive RNA fraction in the reaction kinetics (see methods), we observe a 6-fold increase in the branching rate (k_{br}) (from 0.002 min^{-1} at 50 °C to 0.012 min^{-1} at 65 °C). Intriguingly, the rate of the hydrolytic pathway (k_{hy}) remained constant at 0.018 min^{-1} between 50 and 65 °C. Taken together, the data indicate that thermal activation of AV RNA splicing occurs both through an increase in the population of active molecules and through increases in the rate of branching. In addition, the results indicate different thermodynamic requirements for the two reaction pathways used by group II introns. But perhaps most fundamentally, the data demonstrate the unusually high thermostability of AVgroEL group II intron RNA and a requirement for thermal activation of splicing.

To further examine the ribozyme thermostability, we tested the temperature dependence of oligonucleotide cleavage by the AVΔ292ΔORF RNA. Both substrates S5 and S6 were studied, and both reacted similarly. As observed for the cis-splicing reaction, the ribozyme is still fully active at 65 °C,

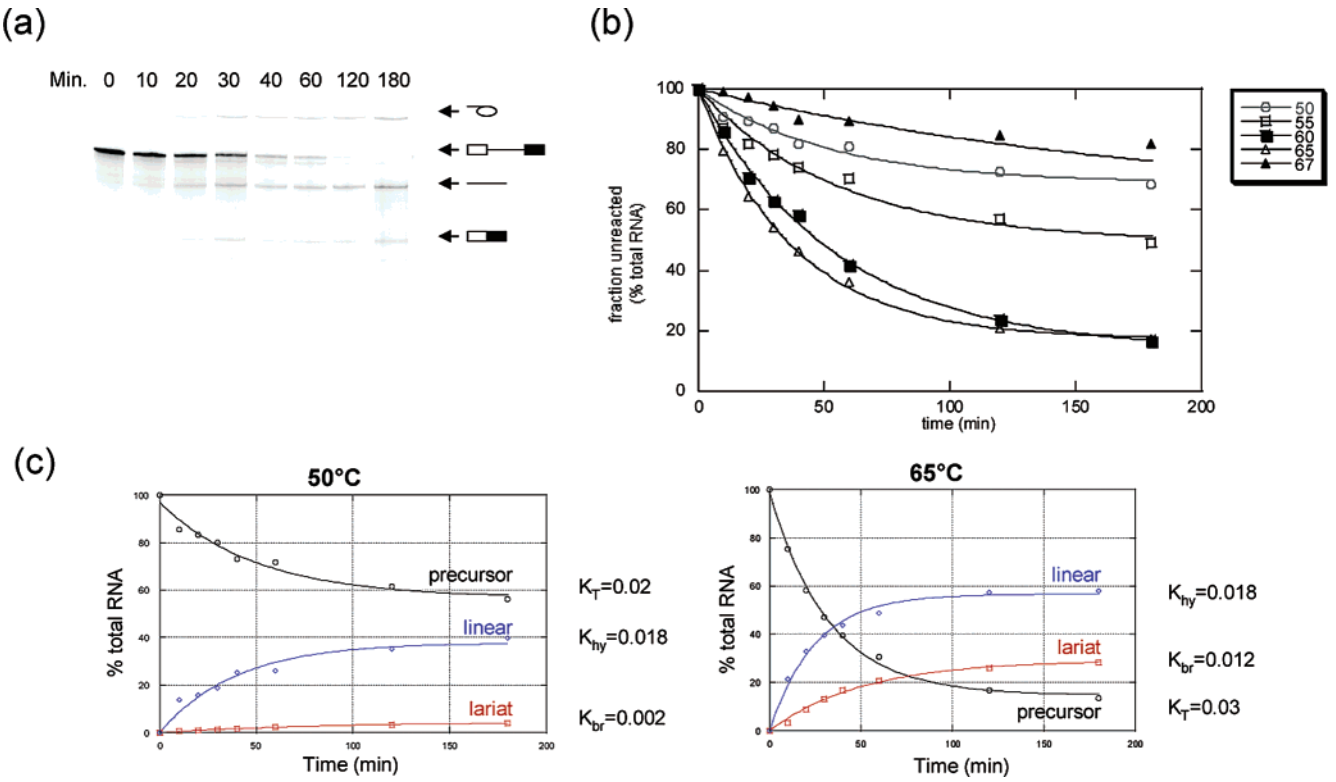


FIGURE 7: Temperature dependence of the self-splicing reaction catalyzed by the AV intron. (a) Gel electrophoresis of AVΔ292ΔORF RNA self-splicing products at 65 °C. Lariat intron, precursor RNA, linear intron, and ligated exons are marked using symbols used in Figure 4. The products were separated by 5% denaturing PAGE. (b). Temperature dependence of the self-splicing reaction. The evolution of the unreacted RNA fractions is plotted against time and fitted as described in Materials and Methods. (c) Detailed self-splicing kinetics at 50 and 65 °C. Data were fitted and rates extracted as described in Materials and Methods. Rates for hydrolysis (k_{hy}) and branching (k_{br}), as well as the total rate (k_T), are shown next to the curves for linear and lariat introns, and precursor RNA, respectively. Rate constants for branching are highly sensitive to reaction conditions, which makes it difficult to compare the kinetic behavior of different introns. However, if one compares maximal rates under optimized conditions for all introns that have been characterized, the rate constant for branching by the AV intron at 65 °C (0.012 min^{-1}) is comparable to that reported under optimal conditions for the intron from *Lactococcus lactis* (*43*) and to the slow populations of ai5γ (*24*) and *Pylaeilla litteralis* (*13*) introns. The fast populations of the latter two introns branch with rate constants that are 1 and 2 orders of magnitude faster, respectively, than that observed for AV.

although no reaction is observed at 70 °C (Figure 8). Similar to the rate constants for self-splicing through the hydrolysis pathway, the rate constants for oligonucleotide hydrolysis are relatively constant between 55 and 65 °C, although the extent of miscleavage increases at higher temperatures (Figure 8a). These results demonstrate that the tertiary fold of the ribozyme is stable at 65 °C as shown by earlier splicing assays.

DISCUSSION

Thermal Activation of in Vitro Splicing and Stability of the Catalytic Core. All autocatalytic group II introns require nonphysiological conditions for in vitro activity, i.e., high cation concentrations and temperatures that are higher than the optimal growth temperature for the host organism. However, to the best of our knowledge, none have been shown to retain self-splicing activity above 60 °C (*13*, *28*). Optimal activity of the AVgroEL self-splicing RNA is reached at 65 °C, a temperature at which other large self-catalytic RNAs undergo partial unfolding. Under the same ionic conditions, the ai5γ self-splicing ribozyme completely loses activity at temperatures above 55 °C (data not shown).

These findings raise two major questions: what is the basis for the unusual thermal stability of the AV RNA, and what function does it serve? It is conceivable that stabilization of the structure is achieved by the elevated G-C content of this intron. AVgroEL intron RNA contains 59.2% G-C nucleotides, a value which is significantly higher than those of other self-splicing group IIB introns. However, there is no obvious correlation between G-C content and RNA thermal stability, and no single mechanism for the effects of G-C content on RNA function. For RNAse P, the high stability of naturally occurring thermophilic versus mesophilic versions of the ribozyme seems to be driven by the fact that the thermostable RNA folds more homogeneously to its native state with increased cooperativity and via a less structured intermediate (*29*). In another example, improving the stability of a group I intron core through in vitro selection was shown to correlate with an increase in RNA secondary structure content (*30*). However, the effects of increasing secondary structural stability on tertiary folding are not clear-cut, and other factors, such as metal ion−RNA interactions and the coordination of water molecules, may play a critical role (*30*, *31*).

Group II introns fold into stable three-dimensional structures with a solvent inaccessible core that is formed by conserved long-range tertiary interactions (*32*, *33*). In one case (the ai5γ intron), the folding pathway appears to be devoid of kinetic traps (*34*). However, the steep temperature dependence of the AV RNA self-splicing reaction and the increase in the active RNA population as a function of temperature suggest that a misfolded, off-pathway intermedi-

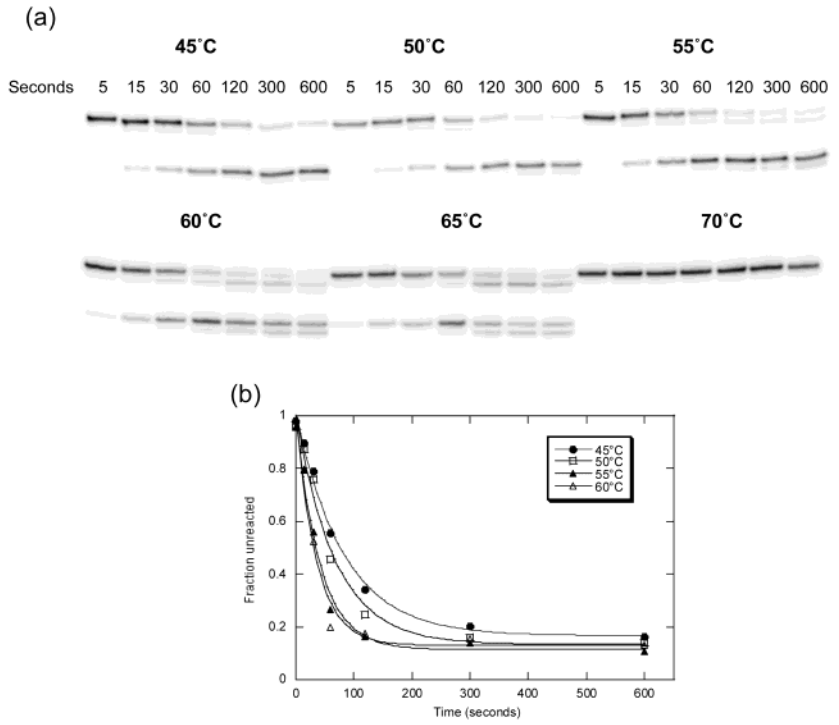


FIGURE 8: Temperature dependence of oligonucleotide cleavage. (a) Time courses for the cleavage of S5 at temperatures ranging from 45 to 70 °C. Aliquots of the reaction mixture were taken at 0, 15, 30, 60, 120, 300, and 600 s. Substrates and products were separated on 20% denaturing polyacrylamide gels. (b) Evolution of substrate RNA at different temperatures.

ate can perturb the folding pathway of the AV intron. This off-pathway intermediate may serve a function (providing a temperature-dependent switch for activation of splicing during heat shock, for example), or it may indicate a requirement for cis- or trans-acting factors that assist in vivo excision. The presence of inactive conformers and their response to temperature and encoded maturase proteins can be assessed by a variety of kinetic and structural methods that were used to elucidate the ai5γ intron folding pathway (*34*).

Unusual Location of the AVgroEL Bacterial Group II Intron. The AVgroEL group II intron interrupts the stop codon of the groEL gene within the groESL operon (*A. vinelandii* sequencing genome project, http://www.jgi.doe.gov/JGI_microbial/html/azotobacter/azoto_homepage.html). Although intron splicing does not alter the open reading frame of GroEL, the presence of the intron may have implications for the regulation of the operon. In particular, the remarkable thermostability of the intron catalytic core could contribute to post-transcriptional regulation of GroEL expression. Heat shock proteins 60, also known as GroEL, are essential proteins highly expressed in bacterial cells and eukaryotic organelles (*35*, *36*). GroEL proteins are essential for normal growth, but their catalytic activity is particularly important under conditions that lead to protein denaturation. These molecular chaperones are induced by a variety of environmental stresses, in particular heat shock. Although this induction is a universal response, the regulatory mechanisms controlling heat shock protein synthesis seem to differ widely among organisms. They can operate by transcriptional control, translational control, and proteolysis. In bacteria, the regulation of the groESL operon appears to happen mostly at the transcriptional and post-transcriptional level, by both positive and negative mechanisms (*37*). The data presented herein suggest that the thermodynamics of RNA folding and the kinetics of self-splicing may provide an alternative regulatory pathway that has yet to be explored.

To our knowledge, bacterial group II introns have not yet been found within the essential genes of host organisms, and they are mainly associated with mobile elements and intergenic regions. Thus, it would seem that the presence of AVgroEL in the groEL gene constitutes an extraordinary event. However, the presence in proteobacteria of multiple copies of the hsp60 genes has suggested their mobility within and between genomes through lateral transfer (*38*). Thus, the AVgroEL intron may in fact be associated with insertion into a putative mobile element, as with other bacterial group II introns. We also have identified very closely related group II introns in other Azotobacteriaceae strains (data not shown), suggesting the widespread nature of this interesting ribozyme.

Exon−Intron Interactions in the AV RNA. The results reported here also provide insights into the role of intron−exon pairing and the sequence requirements for exon binding in group II introns. Normally, group II introns recognize their 5′-exon via extensive base pairing between two distinct intron-binding and exon-binding sequences (IBS−EBS). Unlike the IBS1−EBS1 pairing, which is common to all group II introns, the IBS2−EBS2 pairing is not universally conserved and appears to be missing in certain introns. Disruption of the IBS2−EBS2 interaction by introducing mismatch mutations does not affect self-splicing of the ai5γ ribozyme (*39*), although the IBS2−EBS2 pairing is essential for substrate recognition in trans-cleavage experiments (*9*). It is therefore most likely that the IBS2−EBS2 pairing plays an important role in the mobility of the intron RNA, and the dispensability of the interaction could be indicative of insertion into an ectopic site (*40*). Shifting the IBS2−EBS2 interaction within the AVgroEL intron had no effect on the in vitro activity of ribozymes derived from the AVgroEL

intron. However, a 10-fold decrease in the rate constant was observed when the IBS2 region was disrupted by either deletion or mutation to eliminate all base pairs with the intron RNA. These results demonstrate the extreme stability of an enzyme−substrate complex that is based primarily on IBS1−EBS1 contacts, which may indicate the presence of a tertiary contact between the intron and 5′-exon that has to be yet identified.

A third intron−exon interaction (IBS3−EBS3) that is well-conserved among intron subgroup IIB (*41*) also seems to be absent in AVgroEL. Instead, the intron−3′-exon interaction may be guided through the $\delta-\delta'$ interaction as in ai5γ, as proposed by Jacquier and Jacquesson-Breuleux (*42*). The weak IBS2−EBS2 interaction and the absence of the IBS3−EBS3 signature interaction in the AV RNA reflect a general diversity among group II introns in the mechanism for mediating intron−exon recognition.

An Unusual Insertion within the D1 Stem. In addition to its unorthodox IBS−EBS base pairing, the AVgroEL exhibits a second feature that is atypical for subgroup IIB1 introns. Although large insertions within the conserved D1 stem have been previously reported, all were found either in organellar introns of subgroup IIB2 or IIA (*1*) or in bacterial introns of classes B and D (*6*). This is the first time that this feature has been observed in subgroup IIB1. In addition, this is the largest insertion reported yet for any group II intron. Interestingly, deletion of the insertion has no detrimental effect on catalysis. The intron RNA is able to fold into a catalytically active structure, suggesting that this insertion forms a peripheral domain that does not disrupt the catalytic core of the intron. This unusual feature could form a site of binding for a putative protein cofactor; it could help form a thermally deactivated conformational trap for regulating splicing, or it could also represent a nonfunctional remnant of the ectopic insertion of the intron into *A. vinelandii* genomic DNA. Like the thermally activated mechanism of the AV RNA core, the function of this insertion will be explored through kinetic and structural analysis in vitro and in vivo.

ACKNOWLEDGMENT

We thank Prof. David Figurski (Columbia University) and Prof. James W. Brown (North Carolina State University, Raleigh, NC) for their generous gift of *Azotobacter* strains and genomic DNA. We are also grateful to Alexandre De Lencastre and Linhui Julie Su for helpful comments and discussion.

REFERENCES

1. Michel, F., Umesono, K., and Ozeki, H. (1989) *Gene 82*, 5−30.
2. Michel, F., and Ferat, J. L. (1995) *Annu. Rev. Biochem. 64*, 435−461.
3. Martinez-Abarca, F., and Toro, N. (2000) *Mol. Microbiol. 38*, 917−926.
4. Bonen, L., and Vogel, J. (2001) *Trends Genet. 17*, 322−331.
5. Qin, P. Z., and Pyle, A. M. (1998) *Curr. Opin. Struct. Biol. 8*, 301−308.
6. Toor, N., Hausner, G., and Zimmerly, S. (2001) *RNA 7*, 1142−1152.
7. Griffin, E. A., Qin, Z.-F., Michels, W. A., and Pyle, A. M. (1995) *Chem. Biol. 2*, 761−770.
8. Michels, W. J., and Pyle, A. M. (1995) *Biochemistry 34*, 2965−2977.
9. Xiang, Q., Qin, P. Z., Michels, W. J., Freeland, K., and Pyle, A. M. (1998) *Biochemistry 37*, 3839−3849.
10. Su, L. J., Qin, P. Z., Michels, W. J., and Pyle, A. M. (2001) *J. Mol. Biol. 306*, 655−668.
11. Zimmerly, S., Moran, J. V., Perlman, P. S., and Lambowitz, A. M. (1999) *J. Mol. Biol. 289*, 473−490.
12. Pyle, A. M. (1996) in *Nucleic Acids and Molecular Biology* (Eckstein, F., and Lilley, D. M. J., Eds.) pp 75−107, Springer-Verlag, New York.
13. Costa, M., Fontaine, J. M., Goër, S. L., and Michel, F. (1997) *J. Mol. Biol. 274*, 353−364.
14. Granlund, M., Michel, F., and Norgren, M. (2001) *J. Bacteriol. 183*, 2560−2569.
15. Klein, J. R., and Dunny, G. M. (2002) *Front. Biosci. 7*, 1843−1856.
16. Ferat, J. L., and Michel, F. (1993) *Nature 364*, 358−361.
17. Wilson, K. (1992) in *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology* (Ausubel, F. M., Ed.) pp 2.4.1−2.4.5, Greene Publishing Associates and Wiley-Interscience, New York.
18. Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C., and Markham, A. F. (1990) *Nucleic Acids Res. 18*, 2887−2890.
19. Frank, D. N., Adamidi, C., Ehringer, M. A., Pitulle, C., and Pace, N. R. (2000) *RNA 6*, 1895−1904.
20. White, B. A. (1993) PCR Protocols: current methods and applications, in *Methods in Molecular Biology*, Vol. 15, pp 251−261, Humana Press, Totowa, NJ.
21. Zuker, M., Jaeger, J. A., and Turner, D. H. (1991) *Nucleic Acids Res. 19*, 2707−2714.
22. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
23. Tinoco, I., Sauer, K., and Wang, J. C. (1978) *Physical chemistry: principles and applications in biological sciences*, Prentice-Hall, Englewood Cliffs, NJ.
24. Daniels, D. L., Michels, W. J., Jr., and Pyle, A. M. (1996) *J. Mol. Biol. 256*, 31−49.
25. Jacquier, A., and Rosbash, M. (1986) *Science 234*, 1099−1104.
26. Vogel, J., Hess, W. R., and Borner, T. (1997) *Nucleic Acids Res. 25*, 2030−2031.
27. Qin, P. Z., and Pyle, A. M. (1999) *J. Mol. Biol. 291*, 15−27.
28. Franzen, J. S., Zhang, M., Chay, T. R., and Peebles, C. L. (1994) *Biochemistry 33*, 11315−11326.
29. Fang, X. W., Golden, B. L., Littrell, K., Shelton, V., Thiyagarajan, P., Pan, T., and Sosnick, T. R. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 4355−4360.
30. Juneau, K., and Cech, T. R. (1999) *RNA 5*, 1119−1129.
31. Juneau, K., Podell, E., Harrington, D. J., and Cech, T. R. (2001) *Structure 9*, 221−231.
32. Costa, M., and Michel, F. (1999) *EMBO J. 18*, 1025−1037.
33. Swisher, J., Duarte, C. M., Su, L. J., and Pyle, A. M. (2001) *EMBO J. 20*, 2051−2061.
34. Swisher, J. F., Su, L. J., Brenowitz, M., Anderson, V. E., and Pyle, A. M. (2002) *J. Mol. Biol. 315*, 297−310.
35. Bukau, B., and Horwich, A. L. (1998) *Cell 92*, 351−366.
36. Sigler, P. B., Xu, Z., Rye, H. S., Burston, S. G., Fenton, W. A., and Horwich, A. L. (1998) *Annu. Rev. Biochem. 67*, 581−608.
37. Segal, G., and Ron, E. Z. (1998) *Ann. N.Y. Acad. Sci. 851*, 147−151.
38. Karlin, S., and Brocchieri, L. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 11348−11353.
39. Jacquier, A., and Michel, F. (1987) *Cell 50*, 17−29.
40. Mohr, G., Smith, D., Belfort, M., and Lambowitz, A. M. (2000) *Genes Dev. 14*, 559−573.
41. Costa, M., Michel, F., and Westhof, E. (2000) *EMBO J. 19*, 5007−5018.
42. Jacquier, A., and Jacquesson-Breuleux, N. (1991) *J. Mol. Biol. 219*, 415−428.
43. Matsuura, M., Noah, J. W., and Lambowitz, A. M. (2001) *EMBO J. 20*, 7259−7270.
44. Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) *Nucleic Acids Res. 4*, 2527−2538.

BI027330B