A self-splicing group I intron in the nuclear pre-rRNA of the green alga, Ankistrodesmus stipitatus

Jennifer A. Dávila-Aponte¹, Volker A.R. Huss², Mitchell L. Sogin³ and Thomas R. Cech¹,⁴

¹Department of Molecular, Cellular and Developmental Biology, Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA, ²Institut für Botanik und Pharmazeutische Biologie der Universität, D-8520 Erlangen, FRG, ³Marine Biological Laboratory, Woods Hole, MA 02543 and ⁴Department of Chemistry and Biochemistry, Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA

Received May 23, 1991; Revised and Accepted July 19, 1991

ABSTRACT

The nuclear small subunit ribosomal RNA gene of the unicellular green alga Ankistrodesmus stipitatus contains a group I intron, the first of its kind to be found in the nucleus of a member of the plant kingdom. The intron RNA closely resembles the group I intron found in the large subunit rRNA precursor of Tetrahymena thermophila, differing by only eight nucleotides of 48 in the catalytic core and having the same peripheral secondary structure elements. The Ankistrodesmus RNA self-splices in vitro, yielding the typical group I intron splicing intermediates and products. Unlike the Tetrahymena intron, however, splicing is accelerated by high concentrations of monovalent cations and is rate-limited by the exon ligation step. This system provides an opportunity to understand how limited changes in intron sequence and structure alter the properties of an RNA catalytic center.

INTRODUCTION

The intron that interrupts the sequence of the large subunit (LSU) rRNA from Tetrahymena thermophila can catalyze its own excision from the rRNA precursor and concomitantly rejoin the two exons to produce the mature rRNA (1). This reaction requires Mg²⁺ and a GTP (or guanosine) cofactor. The Tetrahymena intron (Tt intron) represents the class of introns called group I. They have in common a conserved secondary structure which was determined by phylogenetic comparison of the introns (2,3). This structure consists of a number of conserved paired regions, P₁-P₁₀ (Fig. 1). The central part, which includes P₄, P₆ and P₇, has a conserved sequence as well as a conserved structure. The importance in splicing of many positions within this core region has been tested by random or site-directed mutagenesis (4–13), but the specific role of many of the conserved nucleotides remains unknown. One exception is a conserved guanosine residue in the 5' half of P₇, which contributes directly to the binding of the GTP substrate (14).

Group I introns are distributed widely across phylogenetic boundaries. They have been found in the nuclei of certain protists and fungi; in the organelles of fungi, green algae and higher plants; and in cyanobacteria and bacteriophages (15–18). The unicellular green alga Ankistrodesmus stipitatus has been shown to harbor a 394 nucleotide group I intron (As intron) between nucleotides 1262 and 1263 in the gene encoding the nuclear small subunit (SSU) rRNA (19). This represents the first example of a group I intron in the nucleus of a member of the plant kingdom.

Within the catalytic core, the As intron differs from the Tt intron at only a few nucleotide positions. The As intron is self-splicing, but its splicing characteristics differ from those of the Tt intron. Thus, the As intron offers an opportunity to determine whether any of the differences in the core have an effect on splicing.

MATERIALS AND METHODS

Nucleotides and Enzymes

Nucleoside triphosphates (NTPs) and calf intestinal phosphatase (CIP) were obtained from Pharmacia. Labelled NTPs were purchased from New England Nuclear. All restriction endonucleases except XmaIII were purchased from New England Biolabs, as was T₄ polynucleotide kinase. Restriction endonuclease XmaII and nuclease PhyM were purchased from Bethesda Research Laboratories. Dideoxynucleoside triphosphates, T₄ DNA ligase, and T₄ DNA polymerase were purchased from U.S. Biochemicals. Nucleases P₁, T₁, T₂, and U₂ were obtained from Calbiochem.

Nucleic Acids

M13 universal sequencing primers #1211 and #1201 were purchased from New England Biolabs. M13 vectors mp18 and mp19 were obtained from Pharmacia. The plasmid pTZ18U is available through Biorad.

Plasmid Construction

An M13 vector containing the gene encoding the small subunit rRNA of Ankistrodesmus stipitatus (strain 202-5 from the Sammlung von Algenkulturen, Pflanzenphysiologisches Institut der Universität Göttingen, Federal Republic of Germany) was constructed as described (19). The gene encoding the SSU rRNA of Ankistrodesmus was transferred from this M13 vector into a phagemid vector. Double-stranded replicative form of mp18A.s.
ssu was prepared according to the method of Godson and Vapnek (20). This DNA was cut with restriction endonucleases SaI and HindIII to release the fragment containing the rDNA. This fragment was inserted into the phagemid pTZ18U (previously cut with Sall and HindIII) using T4 DNA ligase. The resulting plasmid (pTZAs1) was used to construct a subclone containing only the intron (A.s. SSU1), 19 nucleotides 5' to the intron and 108 nucleotides 3' to the intron. pTZAs1 was linearized with XmaIII and the overhanging ends were filled in using the Klenow fragment of DNA polymerase I. A second digest with endonuclease Sall released a 521 bp fragment containing the intron and flanking sequences. This fragment was inserted into pTZ18U which had been cut with Sall and SmaI and treated with CIP. Bacterial strain JM83 was transformed with this construct by the CaCl₂ method (21), and the subsequent clones were checked by sequencing over the plasmid-insert boundaries using M13 universal primers that anneal to the polylinker sequence of pTZ18U. This subclone was named pTZAs2.

Transcription

Uniformly labelled precursor RNA containing a 32 nt 5' exon (19 nt natural sequence), the As intron, and a 141 nt 3' exon (108 nt natural sequence) was transcribed from pTZAs2 that had been linearized with HindIII. Template (1 μg) was incubated in a 20 μl reaction containing 40 mM Tris-HCl, pH 7.5, 12 mM MgCl₂, 10 mM DTT, 2 mM spermidine, 0.5 mM each UTP, CTP and GTP, 0.125 mM ATP, 30 U T7 RNA polymerase and α[³²P] ATP. The transcription reaction was subjected to electrophoresis on a 4% polyacrylamide-6M urea gel, and the precursor band was excised and eluted overnight in 10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA, 0.25 M NaCl. The RNA was ethanol precipitated and resuspended in a small volume of TE (10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA). Cold precursor RNA was made using the same techniques, except with 0.5 mM ATP in the transcription reaction. End-labelled precursor RNA was prepared as follows: cold precursor RNA was treated with

![Diagram of predicted secondary structure of the As intron](image)

Figure 1. Predicted secondary structure of the As intron. Lower case letters are exon sequences and capital letters are intron sequences. Bold letters are core conserved nucleotides. Boxed nucleotides differ from the same position in the Τt intron but conform to the consensus sequence for group I introns. Circled nucleotides differ from both the 71 intron and from the consensus sequence. Arrows denote the 5' and 3' splice sites. The asterisk denotes the base identified by Michel et al. (14) as the G-binding site. The shaded nucleotides are the A-G mismatch described in the text. Two areas of the As structure—the A-rich bulge in the P5 extension and the P9 helix—are drawn differently than their counterparts in some previous models of the Τt secondary structure (37). These changes are not differences between the two introns; they simply reflect new data on the phylogenetically conserved folding of a subclass of group I introns to which both the As and Τt introns belong (38). The inset at lower left shows the P1 and P10 interactions of the As intron proposed to help align the exons for splicing (3). The larger inset shows an outline of the structure of the Tetrahymena group I intron.
CIP, phenol/chloroform extracted and ethanol precipitated. The RNA was subsequently labelled using γ[32P] ATP and T4 polynucleotide kinase.

**Splicing Assays**

Precursor RNA in TE was heated to 95°C for 1.5 min, then quick-cooled on ice in the presence of 50 mM Tris·HCl, pH 7.5, and various concentrations of MgCl₂ and monovalent salts as stated for individual experiments. Splicing reactions were preincubated for 5 min at the reaction temperature and initiated by the addition of GTP to a concentration of 0.2 mM. Aliquots of the reaction were removed at given timepoints and stopped with an equal volume of loading buffer (0.02% xylene cyanole and bromphenol blue, 100 mM EDTA, 6 M urea). Reaction products were subjected to electrophoresis on a 4% polyacrylamide-6 M urea gel and quantified using a Molecular Dynamics phosphorimager system.

**Identification of Splicing Products**

5' end-labelled precursor RNA was incubated in 50 mM Tris·HCl, pH 7.5, 15 mM MgCl₂, 1 M NH₄Cl, 0.2 mM GTP for 30 min at 30°C. The splicing reaction was subjected to electrophoresis and the product bands were cut out and eluted as described above for precursor RNA. Direct sequencing of these products was performed by limited treatment with sequence-specific ribonucleases (22).

**RESULTS**

The *Ankistrodesmus stipitatus* SSU rRNA intron can be folded into the secondary structure characteristic of group I introns (Fig. 1). This intron is inserted in a conserved region of the rRNA between helices 35 and 36 (Fig. 2). The exact number of SSU rRNA genes in *Ankistrodesmus* has not been determined but is likely to be in the magnitude of hundreds (23). Following PCR-amplification of SSU rRNA genes by eukaryote-specific amplification primers (24), only a single 2.2 kb band could be detected in an agarose gel (data not shown). This band corresponds to the size of the gene including the intron. In addition, sequencing of a total of three cloned PCR-amplified SSU rRNA genes revealed no heterogeneity within the intervening sequence (data not shown). Thus, all or at least the vast majority of the genes contains the same intervening sequence.

The *Ankistrodesmus* intron self-splices

Group I introns accomplish their own excision from a precursor RNA by two transesterification reactions (Fig. 3). In the first reaction, an exogenous molecule of guanosine or a 5'-phosphorylated form of guanosine attacks the phosphorus atom

![Figure 2. Insertion site of the As intron. Section of the conserved region of As SSU rRNA showing parts of helices 35 and 36 (helix numbering is according to Dams et al.; 39). The insertion site of the intervening sequence is between nucleotides 1262 and 1263 of the mature RNA transcript at the beginning of helix 36. Bold letters adjacent to the insertion site indicate nucleotides that are almost invariant among eukaryotes.](image)

![Figure 3. Mechanism of group I splicing, illustrated with sequences of the As pre-rRNA. Precursor RNA incubated in the presence of guanosine or GTP ([italicized G] undergoes attack at the 5' splice site. The exogenous guanosine bound in the G-binding site attaches to the 5' end of the intron, releasing free 5' exon (1). In a postulated conformational change (14, 34, 35), the terminal G residue of the intron replaces the exogenous G in the G-binding site (conformational change). This activates the phosphate at the 3' splice site, which is then attacked by the 5' exon to yield ligation exons and linear intron (2). The 5' exon binding sequence is shaded throughout.](image)
at the 5' splice site and attaches to the 5' end of the intron. In the second step, the free 3' hydroxyl group of the 5' exon attacks the phosphorus atom at the 3' splice site and yields ligated exons and linear intron (L-IVS).

An artificial precursor RNA containing the entire As intron and portions of the flanking exons was synthesized and purified (Fig. 4A). When incubated in the presence of Mg++, monovalent salt and guanosine or GTP, this RNA undergoes self-splicing (Fig. 4B). The identity of the bands labelled precursor and ligated exons was confirmed by incubating a 5’ end-labelled transcript under splicing conditions (see Materials and Methods) and isolating the radio-labelled bands from a polyacrylamide gel. These molecules were sequenced directly by limited digestion with sequence-specific nuclease (data not shown). The products labelled IVS-3’ exon and IVS were identified by incubating unlabelled precursor RNA with a [32P]GTP and isolating the radiolabelled bands from a polyacrylamide gel. These RNA species were also sequenced by the enzymatic method (data not shown). These products conform to the splicing scheme described in Figure 3. The 5’ and 3’ splice sites are also shown in Figure 1.

At 0.2 mM GTP, the reaction proceeds with the accumulation of splicing intermediates (free 5’ exon and IVS-3’ exon), indicating that a significant population of molecules have completed the first step of splicing but have not yet accomplished the second step. That is, some step after the first is rate-limiting for splicing. These splicing intermediates are formed under all splicing conditions that were tested (see below) except when the concentration of GTP was very low. At lower [GTP] (0.02 mM), the first step of splicing slows sufficiently that the second step of splicing is no longer rate-limiting (Fig. 4C).
A product that migrates more slowly than precursor RNA appears at late times during the splicing reaction (Fig. 4B). A band with the same lower mobility appears when internally labelled linear IVS is incubated by itself in 50 mM Tris-HCl, pH 7.5, 50 mM MgCl₂ and 1 M NH₄Cl. Three small fragments (3–7 nucleotides) appear concomitantly with this band (data not shown). We infer that the slowly-migrating RNA consists of circular forms of the IVS produced in the manner described by Zaug et al. (26) and that the oligonucleotides are produced by the cyclization of linear IVS to three sites close to the 5' end of the linear IVS. However, the cyclization sites have not yet been directly confirmed.

Optimization of the splicing reaction

In order to accurately compare the relative abilities of the As intron and the Tt intron to splice efficiently, it was necessary to find a set of suitable conditions for splicing of the As intron. The formation of L-IVS at 2.5 min was used as the criterion for optimization. This time point was chosen because the rate of L-IVS formation is linear up to this time under all conditions tested (with the exception of high temperature). The splicing reaction was optimized in the presence of 0.2 mM GTP with regard to concentration and identity of monovalent salt, the concentration of MgCl₂ and temperature.
Splicing of the \( As \) intron is greatly accelerated by a variety of monovalent salts (Fig. 5). It prefers \( \text{NH}_4\text{Cl} \) to any other monovalent salt tested (Fig. 5B). \( \text{NaCl} \) functions nearly as well in the splicing reaction, but at higher concentrations than \( \text{NH}_4\text{Cl} \). At high concentrations, \( \text{KCl} \) is the least effective of the three salts at promoting splicing.

Figure 6 shows the progress of splicing in the presence of either no added monovalent salt (A) or 0.4 \( \text{M} \) \( \text{KCl} \) (B). While ligated exons were not quantitated in this experiment, similar experiments show accumulation of ligated exons concomitant with the appearance of IVS (data not shown). It is clear that the entire splicing reaction proceeds much faster at higher \([\text{salt}]\), as judged by accumulation of linear IVS. Both reaction profiles show a small population of molecules (\(-30\%\)) that reacts faster than the bulk of the population. In the reaction containing no \( \text{KCl} \), this population accomplishes the first step of splicing quickly, and splicing intermediate accumulates because the second step is not as fast. The reaction with 0.4 \( \text{M} \) \( \text{KCl} \) appears to have a different profile. At the earliest timepoint, two thirds of the fastest-reacting population has already completed the splicing reaction (Fig. 6B). This distribution of splicing products and intermediates seems to indicate that the second step of splicing is not significantly slower than the first. However, at 0.5 min the 0.4 \( \text{M} \) \( \text{KCl} \) reaction has already reacted to the same extent that the reaction with no salt does in 10 min; it is possible that the high salt reaction has the same rate-limiting step as the reaction with no salt, but less intermediate accumulates because the entire reaction proceeds so quickly.

The pH dependence of the \( As \) intron was tested over a range from 5.5 to 8.2 using MES (4-morpholinoethanesulfonic acid) and MOPS (3-[N-morpholino]propanesulfonic acid) buffers. Though the intron spliced two-fold better in MOPS than in MES, the splicing reaction exhibited no apparent pH dependence in the range pH 6.0–8.2 (data not shown).

At 1 \( \text{M} \) \( \text{NH}_4\text{Cl} \), the splicing reaction has a broad \( \text{MgCl}_2 \) optimum of 5–90 \( \text{mM} \) (data not shown). At 15 \( \text{mM} \) \( \text{MgCl}_2 \), the temperature optimum for splicing is 80°C, and efficient splicing persists at temperatures up to 90°C (data not shown). Only at 5 \( \text{mM} \) \( \text{Mg}^{2+} \) was it possible to show the decrease in activity at very high temperatures expected for disruption of the catalytic center. At this lower \([\text{Mg}^{2+}]\), the temperature optimum is \(-50°C\) in 0.2 \( \text{M} \) \( \text{NH}_4\text{Cl} \) and \(-55°C\) in 1 \( \text{M} \) \( \text{NH}_4\text{Cl} \) (Fig. 7).
DISCUSSION

The As intron closely resembles the Tt intron in both secondary structure and primary sequence within the core (Fig. 1). It also resembles the Tt intron in the number and size of non-conserved helices outside the core. Eight of 48 nucleotides in the core of the As intron differ from those in the Tt intron. These differences are clustered in the P4, P7, and J7/3, the joining region between P7 and P3 regions. The As intron also lacks the bulged A present in P4 of the Tt intron.

The P1 helices of these two introns also differ. The Tt intron has a purine-rich internal guide sequence (IGS) that pairs to a stretch of pyrimidines in the 5' exon. The As P1 helix is almost the exact converse of the Tt P1, having a pyrimidine-rich IGS that pairs with a purine-rich tract in the 5' exon. Base-pair changes in the P1 helix of an enzymatic form of the Tt intron have been shown to affect cleavage activity at the 5' splice site or the dependence of that activity on ionic conditions (27).

The As intron is more similar to the Tt intron than any other group I intron, including the nuclear SSU rRNA intron of *Pneumocystis carinii* (Pc) (28). The Pc and As introns differ at 10 positions within the core region; almost all of these differences are in the P7 and J4/5 region. Outside the core region, the introns differ in the number of non-conserved helices.

It is interesting to speculate how introns as similar as the As and Tt introns came to be in two different genes in organisms of two different phylogenetic kingdoms. Group I introns have been shown to behave as mobile genetic elements, as demonstrated for the rRNA intron omega in the mitochondrion of the yeast *Saccharomyces cerevisiae* (29), the LSU rRNA intron from *Physarum* (30) and the thymidylate synthase mRNA intron in bacteriophage T4 (31). In each case, mobility is mediated by a DNA endonuclease encoded by an open reading frame contained within the intron. This endonuclease creates a double-strand break within a target (intron-less) copy of the same gene. This break is thought to be repaired using the intron-containing gene as a template. The endonuclease need not recognize the same sequence in the DNA that pairs to the IGS at the RNA level during splicing. It has been suggested that mobile introns could move themselves to a new site within a non-homologous gene in a transposition event. A second possible mechanism for transposition occurs at the RNA level and requires no endonuclease to move the intron. As proposed by Woodson and Cech (32), reverse splicing could insert an intron into another RNA molecule, which could then be reverse-transcribed and reinserted into the genome. This mechanism requires the existence within the target RNA of a sequence that can pair with the IGS of the intron.

The Tt and As introns are located in different genes and their insertion site sequences are different. The Tt intron is inserted in the LSU rRNA gene at the sequence CUCUCU/AAAAU, while the As intron is located in the SSU rRNA gene at the site AUGGGGU/GGUUG, where the slash represents the insertion site. Thus, the As intron could not have acquired its present position and cleavage specificity by a simple transposition of the Tt intron as proposed by Woodson and Cech (32). Though it is possible that a progenitor of the Tt and As introns carried an open reading frame encoding an endonuclease, neither the Tt intron nor the As intron has an internal open reading frame of significant length; it is unlikely, therefore, that the current As intron inserted itself in the manner described for omega. Given these differences, it is difficult to imagine a single event that could be responsible for the existence of two very similar introns in such phylogenetically diverse locations. Thus, while it seems apparent that these introns are invasive, the mechanism of their spread remains problematic.

Standard conditions for As splicing were chosen as 50 mM Tris·HCl, pH 7.5, 1 M NH₄Cl, 15 mM MgCl₂, 0.2 mM GTP at 30°C. Although this temperature is well below the optimal temperature for splicing (see Fig. 7A), it is closer to the temperature at which the organism lives. Under these conditions (GTP) below saturation; data not shown), the splicing reaction is half-complete at -2 min (Fig. 4B). The splicing rate of the As intron is comparable to that of the Tt intron transcript from pJK43 (t₁/₂ = 1 min at 30 mM Tris·HCl, pH 7.5, 100 mM (NH₄)₂SO₄, 5 mM MgCl₂, 0.1mM GTP at 30°C; 33). Interestingly, for the As intron the second step of splicing is rate-limiting rather than the first. This is different from the splicing reaction profile of the transcript from pJK43, in which the attack of exogenous G at the 5' splice site limits the rate of reaction.

There are at least three possibilities to explain the slow second step of *Ankistrodesmus* splicing. (1) The 3' splice site is ill-defined. The terminal G of the IVS is followed by two Gs in the 3' exon. The slow second step may arise because incorrect G residues compete with the authentic terminal G for binding in the G-binding site but are unreactive, perhaps because the P9.0 and P10 interactions cannot be formed (see Fig. 3). Sequencing of the ligated exons shows no heterogeneity at the splice junction, indicating that these other G residues are not reactive. (2) The terminal G of the intron does not fit well into the G-binding site. One of the nucleotide changes within the core of the As intron is a U-G change at the 5' end of the conserved sequence forming the 5' half of P7. This change introduces an A-G mismatch in P7. This A-G mismatch in the As intron could perturb the G-binding site, elements of which are contained in the G-C pair in middle of P7. This possibility is supported by preliminary evidence that the Km for GTP is much higher for the As intron than for the Tt intron (data not shown). A similar A-G mismatch in a shortened version of the Tt intron results in a diminution of cleavage at the 5' splice site to 38% of the wild type activity (13). Several plant mitochondrial introns and several introns from cyanobacteria have the same U-G change at this position in P7, but they also have a compensatory A-C change in the 3' half of P7 (15-17). Of these introns, only those from cyanobacteria have been shown to self-splice in vitro (16). (3) The intron might be slow to undergo the conformational change that is thought to be a prerequisite for the second step of splicing (14, 34, 35). Possibilities (2) and (3) seem the most likely of the three.

The As intron displays a preference for very high [monovalent salt] and for elevated reaction temperature. The nmdB and sunV introns from bacteriophage T4 also exhibit such preferences (36). The preference for both high [salt] and high temperature seems counterintuitive if one thinks of the opposing effects of raising temperature and raising [salt] in stabilizing RNA secondary structure. It is perhaps easier to think of the effects of [salt] and temperature separately. The rate of splicing, like most chemical reactions, will increase with temperature, up to the point at which the secondary structure of the RNA is destabilized. At this point the catalytic core of the intron begins to denature and the rate decreases. The addition of monovalent salt might stabilize the secondary structure of the RNA and allow the intron to retain activity at higher temperatures. This effect can be seen in Figure 7. At 55-65°C, the intron is less active than at 50°C in 0.2 M NH₄Cl but is still highly active when the [NH₄Cl] is raised to 1 M.
In summary, the intron from *Ankistrodesmus stipitatus* is similar to that from *Tetrahymena thermophila* in structure and overall splicing rate, but is limited in rate by the second rather than the first step of splicing. This intron may be an excellent tool for studying the second step of splicing for this reason. The nucleotide changes within the core or near the 5' splice site may be responsible for the differences in behavior between the two introns. Further studies will be directed toward investigating this possibility.

ACKNOWLEDGEMENTS

We would like to thank Alexandra Gampel and Arthur Zaug for technical help and advice. This work was supported by NIH grants GM32964 (M.L.S.) and GM28039 (T.R.C.) and by an NSF predoctoral fellowship (J.D.-A.) and a stipend from the Deutscher Akademischer Austauschdienst (V.A.R.H.). We would like to thank the W.M. Keck Foundation for their support of RNA science on the Boulder campus.

REFERENCES