Structure and conformational rearrangements during splicing of the ribozyme component of group II introns

Cheng-Fang Li

UNIVERSITE PARIS-SUD 11

Ecole Doctorale Gènes, Génomes, Cellule

THESE

Présentée en vue de l'obtention du grade de Docteur de l'Université Paris-Sud 11

Discipline : Sciences du Vivant

Spécialité : Biologie Moléculaire

par

Cheng-Fang LI

Structure et réarrangements conformationnels au cours de l'épissage du composant ribozyme d'un intron de groupe II / Structure and conformational rearrangements during splicing of the ribozyme component of group II introns.

Soutenue le 27 juin 2011 devant la commission d'examen composée de :

François MICHEL	Directeur de thèse
Jenn TU	co-Directeur de thèse
Alain JACQUIER	Rapporteur
Woan-Yuh TARN	Rapporteur
Maria COSTA	Examinateur
Daniel GAUTHERET	Examinateur
Yiu-Kay LAI	Examinateur
Huey-Nan WU	Examinateur

Doctoral Thesis Manuscript

Structure et réarrangements conformationnels au cours de
l'épissage du composant ribozyme d'un intron de groupe II /
Structure and conformational rearrangements during splicing of
the ribozyme component of group II introns.

Student: Cheng-Fang Li[1,2]

Director: François Michel[1]

[1]Centre de Génétique Moléculaire du C.N.R.S., 1 Avenue de la Terrasse, 91190 Gif-sur-
Yvette, France

[2]Department of Life Science and Institute of Biotechnology, National Tsing Hua
University, Hsinchu, Taiwan.

INDEX

Acknowledgements

On the completion of my thesis, I should like to express my deepest gratitude to all the people whose help and advices have made this work possible. First, I would like to present my deepest gratitude to my director François Michel, who gave me valuable instructions on the research and generously provided me with help when I was in France. He has walked me through every step of this thesis. Without his creative genius and consistent instructions, this thesis could not have reached its present form. I would also like to appreciate Maria Costa. She has teached me most of the techniques applied in this research. Besides, I especially appreciate her generously sharing of her knowledge and advices to me. I would also like to thank my co-workers when I was in France, Boubekeur Saïfi, Jean-Luc Ferat, Gurminder S. Bassi and Jacques Daniel. Their friendship and kindness helped me to fit in the new environment when I

Abstract

Group II introns are a class of RNAs best known for their ribozyme-catalyzed, self-splicing reaction. Under certain conditions, the introns can excise themselves from precursor mRNAs and ligate together their flanking exons, without the aid of proteins. Group II introns generally excise from pre-mRNA as a lariat, like the one formed by spliceosomal introns, similarities in the splicing mechanism suggest that group II introns and nuclear spliceosomal introns may share a common evolutionary ancestor.

Despite their very diverse primary sequences, group II introns are defined by a highly conserved secondary structure. This generally consists of six domains (Domain I-Domain VI; D1-D6) radiating from a central wheel. Each of the six intronic domains has a specific role in folding, conformational rearrangements or catalysis. The native conformation of a group II intron is sustained by intra- and interdomain long-range tertiary interactions, which are critical either for folding of the intron to the native state or for its catalytic activity. In brief, Domain V interacts with Domain I to form the minimal catalytic core; Domain VI contains a highly conserved bulged adenosine serving as the branch-point nucleotide. DII and Domain III contribute to RNA folding and catalytic efficiency. Domain IV, which encodes the intron ORF, is dispensable for ribozyme activity.

Group II intron splicing proceeds through two-step transesterification reactions which yield ligated exons and an excised intron lariat. It is initiated by the 2'-hydroxyl group of the bulged adenosine within Domain 6, which serves as a branch point and attacks the phosphate at the 5'-end of the intron, thus releasing the 5'-exon while forming a lariat structure in the first step. The released 5'-exon, which is bound to the intron through

base pairing interactions, is then positioned correctly to attack the 3'-splice site with its free 3'-OH in the second step of splicing.

It is generally believed that the structure of a group II ribozyme undergoes conformational rearrangements between first step and second step and domain VI must play a central role in the process. However, despite the identification of several interdomain tertiary interactions, neither NMR nor chemical probing studies have been successful in determining the local surroundings of the branch-point adenosine and neighboring domain VI nucleotides in the ribozyme active site.

By using phylogenetic analysis and molecular modelling, we have identified several areas of the molecule which have the potential to constitute the docking site of domain VI. Mutations were introduced in putative binding sites and the resulting, mutant RNAs have been kinetically characterized. This has allowed us to identify a site within the ribozyme that appears to be specifically involved in the branching reaction. In order to further investigate the interaction between that site and domain VI, we set up a system in which the docking of domain VI into its presumed binding site is ensured by the addition of DNA/RNA oligos that position the two RNA elements in an appropriate orientation. By combining the information from such experiments, we have built an atomic-resolution model of the complex formed by domain VI, the branch site and the rest of the intron at the time at which splicing is initiated.

Keywords : group II intron, ribozyme structure, conformational rearrangements, docking site of DVI.

Introduction

The distribution of ribozymes

Molecular biology has been a rapidly growing field since James Watson and Francis Crick have discovered the double helix structure of the DNA molecule in 1953. During the 1980s, RNA molecules with the ability to catalyze chemical reactions have been found and named 'ribozymes'. The discovery of ribozymes was another most important finding which proved that protein-enzymes are not the only biomolecules that are able to catalyze chemical reactions in living cells. In 1989, the Nobel prize was awarded to two researchers, Thomas R. Cech and Sidney Altman for the first demonstration of RNA catalysis. Since then, plenty of other studies have confirmed that certain RNA molecules are capable to organize their 3-dimensional structure in order to perform catalytic functions in the presence of certain divalent cations.

Ribozymes are widespread in nature, particularly in plants, lower eukaryotes, bacteria, and viruses. Ribozymes have been categorized mainly in two groups according to their size (Table S1). The first group includes the small ribozymes, like the hammerhead, hairpin motif, the HDV RNA (hepatitis delta virus), VS ribozyme and also the more recently discovered glmS ribozyme (Winkler et al. 2004). These different catalytic RNAs were found in a size range from about 40 nt up to 154 nt. The second group includes "big" ribozymes like RNase P and Group I and Group II self-splicing introns. The molecules vary in size from as little as 100 nt up to about 1000 nt (Table S1). Besides, according to the more recently determined three-dimensional structure of the large (50S) subunit of a bacterial ribosome (Nissen et al., 2000), the formation of peptide bonds between individual amino acids must be catalyzed by the 23S RNA

molecule in the large subunit. This finding shows that the ribosome is a ribozyme as well (Cech, 2000).

Small Ribozymes

Hammerhead Ribozyme

The hammerhead is widespread in plant pathogenic viroids and virusoids, and among genomes from the *Bacteria, Chromalveolata, Plantae*, and *Metazoa* kingdoms (De la Pena and Garcia-Robles, 2010). The hammerhead is the smallest natural ribozyme that has been discovered. In the natural state, the hammerhead RNA motif exists in single strand form. Although it can act on itself in these conditions without the assistance of protein, it cannot carry out multiple turnovers. *In vitro*, hammerheads can be constructed with RNA strands and demonstrate self-cleavage in multiple turnover. In such *in vitro* experiments, the hammerhead is able to obey typical enzyme kinetics. It catalyzes the transesterification of a 3',5'-phosphodiester bond to give a 2',3'-cyclic phosphodiester and a free 5' hydroxyl as products in a Mg^{2+}-dependent reaction. The reaction is thought to involve nucleophilic attack by the 2' hydroxyl adjacent to the phosphodiester bond. The hammerhead ribozyme requires a conserved "core" of nucleotides for activity, flanked by three duplex stems. Hammerhead ribozymes having the appropriate stem-loop configuration can be thought of as an "enzyme" strand that includes the conserved nucleotide core, and a "substrate" strand that includes the site of cleavage (Pley et al., 2003).

Hairpin ribozyme

The hairpin ribozyme is found in RNA satellites of plant viruses. It was first identified in the minus strand of the tobacco ringspot virus (TRSV) satellite RNA where it catalyzes self-cleavage and ligation reaction to process the products of rolling circle replication into linear and circular satellite RNA molecules. The hairpin ribozyme is similar to the hammerhead ribozyme in that it does not strictly require a divalent metal ion for the reaction. The hairpin ribozyme-substrate complex includes two domains of secondary structure, each domain consists of two short base-paired helices separated by an internal loop. Domain A (helix 1 - loop A - helix 2) contains the substrate and the primary substrate-recognition region of the ribozyme. Domain B (helix 3 - loop B - helix 4) is larger and contains the primary catalytic determinants of the ribozyme. The two domains are covalently joined via a phosphodiester linkage that connects helix 2 to helix 3. These domains must interact with one another in order for catalysis to occur.

HDV RNA (hepatitis delta virus)

The hepatitis delta virus (HDV) ribozyme is a non-coding RNA that is considered to be the only ribozyme known to be required for viability of a human pathogen. The HDV ribozyme acts to process the RNA transcripts to unit lengths in a self-cleavage reaction. The atomic-resolution structure of this ribozyme has been solved using X-ray crystallography and shows five helical segments connected by a double pseudoknot.

Big Ribozymes

Ribonuclease P

Ribonuclease P (RNase P) was the first true RNA enzyme identified. RNase P works as an RNA–protein complex, It processes precursor tRNAs and other RNAs required for cellular metabolism. Two structures of the catalytic RNA subunit have been resolved, one from *Thermotoga maritima* at 3.85 Å resolution and the other from *Bacillus stearothermophilus* at 3.3 Å resolution. However, both structures lacked a bound substrate. Recently, however, the structure of the entire RNase P holoenzyme-tRNA complex has been solved (Reiter et al., 2010).

Group I and Group II

An intron is a nucleotide sequence within a gene that is removed by RNA splicing. During RNA transcription, introns are transcribed with exons; only during maturation of RNA ('RNA processing'), are they excised from primary RNA transcripts. This process is known as "splicing". The structures of Group I and Group II introns have been well characterized. The differences between Group I and Group II are based on their different splicing pathway: even though they both use two consecutive steps of transesterification, group I introns require a free guanine nucleoside to initiate the reaction and their secondary structure is different from that of group II introns. The secondary structure of group I introns consists of nine paired regions (P1-P9) and it folds into essentially two domains: the P4-P6 domain (composed of stacked P4, P5, P6, P6a helices) and the P3-P9 domain (formed by the P8, P3, P7 and P9 helices). During the splicing process of group I introns, the exogenous guanosine (exoG) first docks onto the active G-binding site located in P7, and its 3'-OH is aligned to

attack the phosphodiester bond at the 5' splice site located in P1, resulting in a free 3'-OH group at the end of the upstream exon and the exoG being covalently attached to the 5' end of the intron. Then, the terminal G (omega G) of the intron replaces the exoG and occupies the G-binding site to organize the second ester-transfer reaction: the 3'-OH group of the upstream exon in P1 is aligned to attack the 3' splice site in P10, leading to the ligation of the adjacent upstream and downstream exons and freeing of the catalytic intron (Brion and Westhof, 1997).

Despite being introns, both Group I and Group II self-splicing introns frequently include ORFs (open reading frames) which encode proteins, the function of which is to facilitate the mobility or splicing of the host intron *in vivo*. We will now discuss Group II introns and elaborate on the relationships between self-catalytic ribozymes and intron-encoded proteins.

Group II intron splicing mechanism and structure

Transesterification

The splicing reactions of group II introns are catalyzed by the intron RNA itself. To accomplish this, the RNA folds into conserved secondary and tertiary structures, which form an active site containing catalytically essential Mg^{2+} ions. Group II introns splice via two sequential transesterification reactions that yield ligated exons and an excised intron lariat with a 2′-5′ phosphodiester bond (Figure S4). In the first step, nucleophilic attack at the 5′-splice site by the 2′ OH of a bulged A-residue in DVI results in cleavage of the 5′-splice site coupled to formation of the lariat reaction intermediate. In the second step, nucleophilic attack at the 3′-splice site by the 3′ OH of the cleaved 5′ exon results in exon ligation and release of the intron lariat.

Hydrolytic splicing

Early in vitro studies of group II intron splicing suggested that, in addition to the lariat splicing pathway, the intron can excise via an alternative pathway: water or a hydroxyl ion is used as a nucleophile in the first splicing step (Jarrell et al., 1988; Daniels et al., 1996). The second step then proceeds as in the branching pathway of splicing, and the products of this reaction are ligated exons and a linear intron (Figure S4). The balance between branching and hydrolytic splicing is strongly influenced by the choice of monovalent cation in the reaction (Jarrell et al., 1988). As already observed for group IIB introns, the reaction mechanism shifts in favor of hydrolysis at the 5' splice site when using potassium ion as the monovalent salt. Under ammonium conditions, the situation is favorable for initiation of group IIB splicing by transesterification. It was not clear at first whether the hydrolytic pathway was just an *in vitro* artifact, but in 1998, it was shown that introns with branch-point mutations retain splicing activity *in vivo* through this pathway (Podar et al., 1998a). The discovery of introns that naturally lack a branch-point adenosine and are still active has revealed that hydrolytic splicing is an important, biologically relevant variation of group II intron splicing (Vogel and Borner, 2002).

Group II intron splicing *in vivo*

A question that may be asked is whether group II intron use the same strategy in vivo as in most in vitro experiments, which show that group II introns are spliced by two sequential transesterification reations and give rise to an intron lariat? Indeed, the most enlightening results were derived from biochemical analysis of their catalytic activity *in vitro*. In these experiments, group II introns need unphysiologically high salt

concentrations and temperatures for efficient catalytic activity, but none of the available data suggest chemical or mechanistic differences between *in vitro* and *in vivo* reactivity once the intron is folded. It is generally believed that group II introns self-excise in the form of lariat *in vivo*, but the mechanism should be carried out with the assistance of intron encoded protein (IEP).

Reverse splicing

Transesterification reactions are energetically essentially neutral and, therefore, reversible. Excised group II intron RNAs can reverse splice into ligated exons, guided by the same EBS/IBS and δ-δ' base pairing interactions between the intron and flanking exon sequences used for RNA splicing. Both steps of splicing are reversible reactions and the rate constants for the forward and reverse reactions of the first step of splicing are comparable (Chin and Pyle, 1995). Typically, the second step of forward splicing is much faster than the branching reaction, making the first step rate-limiting (Daniels et al., 1996). The reverse reaction is considerably slower, so that reverse splicing is typically an inefficient process. This reaction is not limited to RNA substrates, but also works efficiently with DNA targets. This exceptional versatility in substrate choice is biologically relevant in intron mobility, in which reverse splicing into the target DNA is a crucial step of the homing reaction (Zimmerly et al., 1995; Yang et al., 1996; Cousineau et al. 1998; Fedorova and Zingler, 2007).

Mobility of Group II (homing)

Group II intron retrohoming was demonstrated by studies of the yeast mt aI1 and aI2 and *L. lactis* Ll.LtrB introns (reviewed in Lambowitz and Zimmerly 2004). Retrohoming is mediated by the RNP formed during RNA splicing, which consists of the IEP and excised lariat RNA. RNPs initiate retrohoming by using both the IEP and intron RNA to recognize DNA target sequences. The IEP firstly helps separate the DNA strands and enable the intron RNA to base pair with target exons. Base-pairing between the intron and target DNA follows the same rules as in splicing, for which EBS/IBS and δ-δ' interactions are necessary (Mohr et al. 2000). Therefore, the insert is ensured to be excised by RNA splicing in the future. Furthermore, the intron reverse splices into the DNA strand, resulting in the insertion of linear intron RNA between the two DNA exons (top strand). Then the bottom strand is cleaved by the En (endonuclease) domain of the IEP, and the 3′ end at the cleavage site is used as a primer for reverse transcription of the inserted intron RNA. The resulting intron cDNA is integrated by cellular DNA recombination and/or repair mechanisms (Figure S5).

The lariat reverse splicing reaction is a very important feature of group II introns because it is the obligate first step of intron mobility, through which group II introns invade duplex DNA. This process may have resulted in the propagation of ancestral introns and have pushed their evolution into modern forms. The actual process of intron mobility requires more than reverse splicing by the intron RNA: it depends upon the action of an intron-encoded maturase or host proteins, which provide endonuclease and reverse transcription activities.

Linear introns have also proved their ability to undergo retrohoming by carrying out the first step of reverse splicing into complementary target molecules (Lambowitz and Zimmerly, 2004). In the case of the model ai5γ molecule, the intron catalyzes reverse splicing not only efficiently, but also with high precision. This finding raises the possibility that reverse splicing by linear group II introns may have played a significant role in certain forms of intron mobility and lateral gene transfer (Figure S5) (Roitzsch and Pyle, 2009).

Crystallography of group II introns

The first group II ribozyme to be successfully crystallized and studied using X-ray diffraction methods was a group IIC intron from the halophilic and alkaliphilic eubacterium *Oceanobacillus iheyensis* (*O. iheyensis*) (Toor et al., 2008). Group IIC introns are particularly suitable for crystallization because of their small size and apparent structural stability. The construct of O.i. intron was a 412 nt RNA in which the distal stems of Domains II, III and VI as well as the ORF of domain IV are deleted. The crystallized intron was a product of hydrolytic splicing *in vitro* and hence was not branched. Although the O.i. intron was crystallized from a full-length, self-splicing construct that contained an intact domain VI region, there was no electron density attributable to domain VI in the model. It is possible that due to its conformational instability, domain VI is degraded (Toor et al., 2010). In this model of structure, the nature of the active site of group II intron has been described in detail, but the bulged A and the surrounding residues of Domain VI are still missing. In contrast to the successful achievement of group IIC intron crystallography, the larger and IIA and IIB ribozymes have not yet been successfully crystallized.

Structure

Group II introns are very diverse in their primary sequences, there are only few short sequence stretches that share conservancy in Domain V as well as several nucleotides at the beginning of the intron. Despite their very diverse primary sequences, group II introns are defined by a highly conserved secondary structure (Figure S1) (Michel et al., 1989; Toor et al., 2001). A Group II intron generally consists of six domains radiating from a central wheel. Each of the six intronic domains has a specific role in folding, conformational rearrangements or catalysis. The native conformation of a group II intron is sustained by intra- and interdomain long-range tertiary interactions, which are critical either for folding of the intron to the native state or for its catalytic activity (Toor et al., 2010). Multiple strategies have been applied for probing the structure of group II introns. For example, powerful phylogenetic analysis, biochemical and X-ray crystallographic methods. Those works have largely expanded our knowledge of RNA folding and tertiary structure. Understanding the group II intron tertiary structure has became a hot topic since many years, because the model of group II intron is the best tool to provide researchers a better understanding of mechanisms and structure of the eukaryotic spliceosome. More broadly, research on group II introns has expanded our insight into RNA folding and RNA biochemistry, which help us explore the secrets of molecular evolution. Discoveries and accomplishments in the field will be further addressed below.

Domain I

Domain I is the largest of the six domains and contains four subdomains (A,B,C,D). Domain I is also known to be absolutely essential for catalysis

(Michel and Ferat. 1995), it binds domain V to form a catalytic core. It functions as a scaffold for other domains to assemble the catalytically active structure; for this reason, domain I is involved in several important tertiary interactions with other domains. Domain I is also responsible for exon recognition and splice site selection. Generally, there are two 5' exon substrate recognition sequences (EBS1 and EBS2), which interact with complementary regions of the 5' exon (IBS1 and IBS2) by base pairing interactions. A single mismatched mutation between EBS1-IBS1 and EBS2-IBS2 may result in a significant defect of substrate hydrolysis efficiency (Xiang et al., 1998). However, some group II introns (class IIC) do not have an EBS2 site and appear to rely solely on EBS1 for 5'- splice site recognition (Granlund et al. 2001; Toor et al., 2001). Group IIC introns are generally located downstream of transcriptional terminator motifs, some researchers proposed that the stem-loop terminator motif participates in defining the 5'-splice site to compensate the absence of an EBS2-IBS2 interaction (Fedorova and Zingler, 2007).

Besides 5'-exon recognition, Domain 1 also contributes to recognition of the 3'-exon by interacting with the first nucleotide of the 3'-exon by the EBS3-IBS3 interaction (subgroup IIB) or δ-δ' interaction (subgroup IIA), (Figure S1) (Costa et al., 2000).

Domain I is involved in many tertiary contacts that are critical for catalysis

Domain I is the largest domain and it can be further divided into various subdomains that control RNA folding, each subdomain originates from a very conserved five-way junction. RNA folding studies conducted in vitro indicate that domain I folds first and it serves as a scaffold for sequential assembly of the other domains (reviewed by Pyle et al., 2007).

Furthermore, several studies have shown that folding of Domain I is a rate-limiting step in the folding of the entire intron (Su et al., 2005). The rate-limiting step governing assembly of domain I is thought to be the introduction of a sharp bend in the ζ and κ region, which has been called a "folding control element". Phylogenetic analysis and nucleotide analog interference suppression (NAIS) studies identified three tertiary interactions between domain I and the 'catalytic center' of the intron, ζ–ζ' and κ-κ', which are important for domain V docking (Costa and Michel, 1995; Boudvillain and Pyle, 1998), and λ-λ', which positions domain V in close proximity to the 5'-splice site and is directly involved in catalysis (Boudvillain et al., 2000). The ε–ε' interaction is also critical for recognition of the 5' splice site and is directly involved in the correct positioning of the highly conserved first intron nucleotide (G1) to facilitate the nucleophilic attack at the 5' splice site (Jacquier and Michel, 1987, 1990). The ε' region and the λ region were also identified as a strong binding site for a divalent metal ion (like Mg2+) that might contribute to stabilize the intron structure at the catalytically active site.

Another phylogenetically identified long-range tertiary interaction, the highly conserved α-α' pairing that involves the terminal loop of subdomain IB, was demonstrated to be functionally important by genetic studies *in vitro*. In contrast, the β-β' interaction seems to be less important. The domain I internal base-base interaction called δ-δ' is restricted to intron members of subgroup IIB. In the IIA subgroup, the same δ nucleotide base pairs to a completely different site, the first nucleotide of the 3' exon. It has been suggested that the tetraloop-receptor interaction θ–θ' plays a role in structural stabilization of the native structure rather than being directly involved in catalysis (Costa et al., 1997). This interaction is also important for recruiting the catalytic effector domain III

and the phylogenetically conserved inter-domain joiner J2/3 into the active site (Podar et al. 1998).

Domain II

Domain II is a relatively smaller domain compared to domain I, but it harbors two essential long-range tertiary contacts with domain I (θ–θ') and domain VI (η-η') (Chanfreau and Jacquier, 1996; Costa et al., 1997). η-η' is a tetraloop-receptor interaction that is structurally conserved between different IIA and IIB introns (Costa et al., 1997). However, the locations of receptor and tetraloop are reverse in group IIA and IIB introns. In group IIA introns, the tetraloop is located in D2 and the receptor in domain VI, whereas the tetraloop is in domain VI and the receptor is in domain II in group IIB introns (Costa et al., 1997). The η-η' interaction is noteworthy, because it is believed to serve as a switch in comformational changes that occur between the two steps of transesterification. (Chanfreau and Jacquier, 1996; Costa et al., 1997).

Domain III

Domain III is generally referred to as a catalytic effector (Qin and Pyle, 1998). It is not strictly required for catalysis (Koch et al., 1992), but its presence remarkably enhances reaction rates of group II-derived ribozyme constructs (Qin and Pyle, 1998; Fedorova et al., 2003). It is believed that domain III helps to form the domain I-domain V catalytic core and stabilizes its structure. The first tertiary contact between domains III and V (μ-μ') has recently been identified by NAIS analysis (Fedorova and Pyle, 2007), but the actual function of this interaction is not clearly known.

Domain IV

Domain IV is the most variable region of the intron. In many introns, it contains an open reading frame encoding a mutlifunctional intron-encoded protein (IEP). The most common IEPs include reverse transcriptase, maturase, endonuclease. Those IEPs facilitate intron splicing (maturase) under physiological conditions and are required for intron mobility (Lambowitz and Zimmerly, 2004). Domain IV also contains the primary binding site for the maturase protein (Watanabe and Lambowitz, 2004). As it is located on the surface of the folded intron, it likely plays a more general role as a protein-binding element and may interact with various protein co-factors, facilitating intron splicing and mobility. Because domain IV is a peripheral domain that projects away from the catalytic center, this makes it ideal for encoding ORFs and it is removable in most *in vitro* experiments.

Domain V

Just as domain I, domain V is also absolutely necessary for catalysis. Domain V is composed of relatively few nucleotides (around 34 nt), but is the most conserved region of the entire intron (Michel and Ferat, 1995). Domain V has long been believed to be related to U6 snRNA in the spliceosome, domain V shares great similarity with U6 snRNA. As mentioned above, domain V interacts with domain I to form the group II catalytic center through two tetraloop-receptor interaction ζ–ζ' and κ-κ' (Costa and Michel, 1995; Boudvillain and Pyle, 1998). In two major classes of group II introns, IIA and IIB, ζ' is a canonical GNRA tetraloop (Michel and Ferat, 1995; Toor et al., 2001). However, in IIC introns it is an unusual GAAC tetraloop. Another domain I-domain V interaction, λ-λ', brings the chemical face of domain V and the 5'-splice site together (Boudvillain et al., 2000). It was also recently shown that domain V

directly interacts with the catalytic effector domain III via the μ-μ' contact, which possibly helps to anchor domain III in the catalytic core of the ribozyme (Fedorova and Pyle, 2005).

The domain V stem contains a dinucleotide bulge which is extremely important for catalysis, the bulge is also highly conserved in all functional group II introns (Schmidt et al., 1996). Terbium cleavage studies suggested that it harbors a magnesium ion-binding site that could be important for catalytic activity (Sigel et al., 2000). According to the last crystallography data of domain V, the asymmetric internal loop twists tightly upon itself and concentrates the backbone phosphates in space, thereby creating a region with extremely negative electrostatic potential allowing two divalent cations to bind tightly. The metal ion binding site is supported from below by a triple-helical structure that results from binding of a conserved junction region (J2/3) to an invariant region within the major groove of the domain V lower stem (Pyle, 2010).

The other most conserved region in domain V is the AGC triad, also frequently referred to as the 'catalytic triad' (Figure 1). This trinucleotide has also been shown to harbor a magnesium-binding site, which has been proposed to be involved in catalysis (Gordon and Piccirilli, 2001; Sigel et al., 2004). The AGC triad is a feature that group II introns share with U6 snRNA from the spliceosome.

Domain V may be also responsible for the positioning of domain VI and the joiner sequence between D5 and D6 is important as well. A shorter or deleted joiner results in the loss of the 5' transesterification reaction, while a longer joiner yields at least a reduced activity (Boulanger et al., 1996).

Domain VI

Domain VI is critical for the branching pathway of the first step of splicing, which is common to both group II introns and spliceosomal splicing. Although domain VI is a less phylogenetically conserved, short hairpin stem, a bulged adenosine, which is well conserved in group II introns that use the branching pathway, serves as a branch point. Branch-site selection by group II introns is generally very precise. Various studies have proved that the bulge adenosine and the nucleotides around the bulge are important for proper branching. For example, lariat formation is strongly reduced *in vitro* when the adenosine is trapped in a Watson-Crick pairing by adding a uracil on the opposite side of the helix. Deletion of the unpaired adenosine at the branch point completely blocks 5' transesterification and lariat formation. Substitution of the adenosine by a number of different modified nucleosides revealed that the atomic structure of the base is essential for the efficiency of the branching reaction (Chu et al., 1998).

The importance of the conformational flexibility of this region is also underscored by phylogenetic data showing a preference for a wobble or non-Watson-Crick geometry (predominantly G·U) to flank the branch point adenosine; it is believed that GU wobble base pairs promote flapping out of the bulge A (Chu et al., 1998). Further studies have shown that the exceptional accuracy of branch-site selection by group II introns rests on a combination of several partially redundant structural determinants, including the 4-bp basal stem of domain VI, the 3-nt linker between domain V and domain VI in IIB introns, and a G-U pair upstream of the branch-point adenosine (Chu et al., 2001). However, none of these features is absolutely necessary for accurate branching, except the bulged A itself. The combination of these features ensures proper branch-point

selection. The lack of characteristic features of domain VI made researchers puzzling about the position of this domain and the identity of its partners during the transesterification process.

Surprisingly, neither phylogenetic analysis nor NAIM studies have been successful in determining the local surroundings of the branch-point adenosine in the ribozyme active site. The docking site for domain VI was proposed by using UV-crosslinking to be located in the EBS3/δ'-carrying internal loop in domain I and this loop was renamed "coordination loop" (Hamill and Pyle, 2006). However, this theory has been strongly questioned (Michel et al., 2009). Although several mutations in the coordination loop confirm that this loop is important for splicing, the effects have not been shown to be specific to branching (compared to hydrolysis) and the exact nature of the proposed interaction between the coordination loop and domain VI was not clearly demonstrated. Besides, the structure of coordination loop varies between class IIA and classes IIB and IIC. It is puzzling that such a universally conserved component as the domain VI bulge and branch site should be recognised by a structure that is not universally conserved (Michel et al., 2009).

Disappointedly, although the crystal structure of the group IIC intron has provided detailed information on the structure and placement of domains I-V, domain VI lacks any attributable electron density (Toor et al., 2010). The reason why information is missing on domain VI could be the dynamic nature of domain VI. Domain VI has long been proposed to flip in and out of the intron core during the two transesterification steps of the splicing process. It has been proposed that group II introns undergo a conformational change, using the η-η' interaction, to move domain VI from first step to second step splicing conformations (Chanfreau and

Jacquier, 1996). In the crystal structure of the O.i. group IIC intron, domain VI was bited off and cleaved from the 3'-end of the intron after splicing *in vitro*, as if the dynamic behavior of domain VI rendered it vulnerable to attack by the highly reactive intron core (Toor et al., 2010). It has been proposed that domain VI engages in only a few tertiary contacts, with which it might be able to form a stable network of interactions: for example, a network of van der Waals' contacts and stacking interactions, which could be sensitive to the shape of the branch-site region. (Chin and Pyle, 1995; Chu et al., 1998; 2001). The goal of this thesis was to focus on the natural docking site of domain VI. Accordingly, this topic will be further addressed in the text.

Distribution of group II introns

While group II introns were first studied in organellar genomes, the number of known group II introns is still growing rapidly. They are wildely distributed in higher plants chloroplasts (cp), in their mitochondria (mt) and those of fungi, as well as in many bacteria, such as proteobacteria and blue algae. Group II introns are rarely found in archaea, the few that have been found there are likely to have been acquired from eubacteria by horizontal transfers (Rest and Mindell, 2003). Group II introns are widespread in eubacterial genomes and typically act as retroelements with a funtional ribozyme and RT-related enzyme. In contrast, group II introns in organelles frequently have a degenerate RNA structure and either lack ORFs or have IEPs that are no longer involved in intron mobility (Michel and Ferat, 1995). Group II introns have not been found in the nuclear genomes of eukaryotes. Instead, their hypothetical descendants, the "spliceosomal introns", are abundant in eukaryote cells.

The classification and distribution of different subclasses

Although all group II introns have similar overall secondary structures, three major subgroups, denoted IIA, IIB, and IIC, and further subdivisions (A1, A2, B1, B2) are distinguished by specific variations (Michel et al. 1989). Most group IIA introns are relatively large in size (usually > 2.5 kb), when comparing with average mitochondrial or chloroplastic group IIB introns (< 1 kb). The size difference is primarily due to the presence of a long open reading frame (ORF) in domain 4 of group IIA introns. Group IIC introns possess the simplest structure and a relatively smaller size (around 400 nt) among group II introns (Toor et al., 2008). Variation between subclasses of group II consists of rather subtle differences of structure and sequence (Figure S2). One major difference, however, is how the exons are bound and positioned into the active site by the ribozyme component of the intron.

(a) Subgroup IIA. Two terminal loops of the ribozyme secondary structure (EBS1 and EBS2) bind to the IBS2 and IBS1 segments of the 5′ exon by base pairing interactions. The 3′ exon is recognized by a δ-δ' interaction.

(b) Subgroup IIB. Binding of the 5′ exon occurs as in subgroup IIA, except that EBS2 is part of an internal, rather than terminal, loop. The first nucleotide of the 3′ exon (IBS3) is base paired to the EBS3 site. EBS3 is part of an internal loop that is tethered to the EBS1-carrying loop by the δ–δ' base pair (Costa et al., 2000).

(c) Subgroup IIC. Binding occurs as in subgroup IIB, except that (with rare exceptions) IBS2 is replaced by the stem-and-loop component of a rho-independent transcription terminator (Granlund et al., 2001; Michel et al., 2009).

There are also subtle differences in subdomains and interacting motifs between each subgroup. For example, compared to IIA introns, the structural features of D5 in IIC introns and the ε' motif of IIB and IIC are different. In IIB and IIC introns the EBS2 motif is part of an internal, not terminal loop; the "coordination loop" containing EBS3 and δ' is present in IIB and IIC but not IIA introns (Figure S2). Possibly because of such structural differences, the ability to perform *in vitro* self-splicing tends to differ also between different subgroups. For example, many group IIB introns can be spliced in low-salt buffers with low magnisum concentrations, whereas group IIA intron tend to be either slow or less reactive than group IIB introns (Lehmann and Schmidt, 2003).

The group II intron may be the evolutionary ancestor of the eukaryotic cell spliceosome

Splicing of pre-messenger RNAs to mature transcripts is a crucial and elaborate step in the expression of most eukaryotic genes. Almost all human pre-messenger RNAs undergo multiple splicing events, spliceosome-mediated splicing is the most important means of regulation of gene expression. The spliceosome, the multi-megadalton molecular machine that performs splicing, consists of five small nuclear RNAs, or snRNAs, named U1, U2, U4, U5 and U6, and over 200 different proteins. Similarities in the splicing mechanism suggest that group II introns and nuclear spliceosomal introns have an evolutionary relationship (reviewed by Michel and Ferat, 1995; Jacquier, 1996). For example, divalent metal-ion binding sites which may contribute to catalysis have been proposed to be located in domain V in the group II intron and in U6 snRNA of spliceosome. Besides, the two classes of introns share similar branch site motifs in domain VI on the one hand and the U2 snRNA-intron pairing on

the other and they do both use the 2'OH group of a bulged A residue to attack the phosphodiester bond of the 5' splice site. Group II introns generally excise from pre-mRNA as a lariat, a structure that is also adopted by spliceosomal introns. Even though group II introns share so many features with spliceosomes in common, the possible relationship between the spliceosome and group II introns remains an open question some thirty years after it was first suggested.

The endosymbiotic hypothesis of spliceosome origin

Eukaryotes are believed to have taken in group II introns from eubacteria by an endosymbiotic mechanism, which also led to the organelle structures (Koonin, 2006). Once group II introns had been absorbed into eukaryotic genomes, they became fragmented and subsequently lost the ability to self-splice. Finally, they developed into ribonucleo-protein machines such as the eukaryotic nuclear spliceosome (Wheelan et al., 2005), which processes pre-mRNA by a mechanism that is closely related to that catalyzed by group II introns.

In bacteria and some eukaryotic organelles, group II introns continue to modify the genome of their host by acting as mobile elements that bring new information when they migrate. Some organisms have successfully incorporated group II introns and developed a new RNA processing system as a gene regulatory machine. The hypothesis is also supported by many group II intron experiments showing that several domains can be separately assembled and activated *in trans* (Jarell et al., 1988; Suchy and Schmelzer, 1991; Podar et al., 1998). Fragmented parts of the intron are reminiscent of the snRNP genes distributed in the genomes of eukaryotic cells. In fact, group II introns in mitochondria (mt) and chloroplast (cp) are frequently found with structural defects that impair ribozyme activity.

These defects include mispairs, the absence of the bulged A in domain VI, and subdomains missing (Michel et al. 1989). Some introns even lack the essential catalytic domain V and contain only domain VI: these are the 'group III' introns of the Euglena chloroplast genome (Copertino and Hallick, 1993). Those highly degenerate introns presumably require trans-acting RNAs or proteins that compensate for the missing RNA structures (Lambowitz and Zimmerly, 2004). The tertiary interactions between domains make it possible for group II intron RNAs to be split readily into different trans-splicing segments (Belhocine et al. 2005). In group II intron model molecules, some domains (D1c, D3, D5, D1/2/3/4) have been demonstrated to act in trans to promote the splicing of group II intron constructs lacking them. Such fragmented group II introns and trans-acting segments underline evolutionary scenarios for the origin of snRNAs.

Group II intron-encoded proteins

Although some group II introns can perform self-splicing *in vitro*, this reaction generally requires nonphysiological conditions, for example high concentration of bivalent or monovalent salt. *In vivo*, most people believe that proteins are required to help the intron RNA fold into a catalytically active structure. In many cases, proteins required for splicing of a group II intron are encoded within the intron (they are called IEP, intron-encoded protein), in the loop of domain IV in most cases. The IEP most frequently comprises four domains that are the RT domain, X domain, D domain and En domain. The best characterized IEP is the LtrA protein from the *Lactococcus lactis* Ll.LtrB intron. The RT domain is defined by seven conserved sequence blocks (RT1-7), its sequence is similar to that of the reverse transcriptases found in non-long-terminal-repeat (LTR)

retrotransposons. In particular, RT5 contains the highly conserved sequence YADD that is part of the RT active site. Domain X is sometimes referred to as the "maturase domain" because it was identified as a site of mutations affecting RNA splicing activity; maturases bind specifically to the intron RNA to stabilize the active structure (Lambowitz and Zimmerly, 2004). Domain D contributes to DNA binding, whereas the En domain encodes a magnesium-dependent DNA endonuclease that cleaves a target DNA strand to generate the primer for reverse transcription (San Filippo and Lambowitz 2002). The carboxy-terminal D and En domains interact with the target DNA during intron movement.

LAGLIDADG family of homing endonucleases in group II introns

Finally, a small subset of fungal mtDNA group II introns stands apart in encoding proteins of the LAGLIDADG family, to which group I intron homing endonucleases also belong. The LAGLIDADG proteins promote homing of group I introns by cleaving recipient alleles to initiate double-strand break repair (DSBR) recombination, and some have also adapted to function in RNA splicing by stabilizing the active RNA structure. However, in a group II intron from the fungus *Leptographium* (Mullineux et al., 2010), the encoded LAGLIDADG homing endonuclease cleaves the target DNA to generate an intron insertion site, but does not enhance intron splicing. Furthermore, this LAGLIDADG protein does not appear to bind to the intron RNA precursor transcript (Mullineux et al., 2010). It will be of great interest to further explore the role of LAGLIDADG proteins involved in Group II intron homing.

Group II intron IEP Lineages

Group II intron ribozymes and IEPs function together as RNPs, with each IEP binding specifically to the intron RNA that encodes it. As a result, the intron RNAs and IEPs have co-evolved over long times to form phylogenetic lineages of mobile introns (Toor et al. 2001). Based on phylogenetic analysis, the IEPs can be divided into eight major lineages denoted ML (mitochondrial-like), CL (chloroplast-like 1 and 2), and bacterial A-E (Zimmerly et al. 1999), Each IEP lineage is associated with a specific RNA subgroup (Figure S3).

Applications of group II introns

Some group II introns are mobile genomic elements, which can rencognize DNA target sites largely by base pairing of the intron RNA to the DNA target sequence in double-stranded or single-stranded DNA (reviewed in Lambowitz and Zimmerly 2004). Moreover, the linear form of a group II intron has been reported to catalyze an autocatalytic reverse reaction of the second step of splicing with high efficiency and precision (Roitzsch and Pyle 2009).

It is possible to retarget group II introns to insert them into desired DNA sites simply by modifying the sequence of the base pairing segments in the intron RNA (Karberg et al. 2001). This feature, combined with the high efficiency and specificity of the retrohoming reaction, enabled the development of gene targeting vectors ("targetrons") for genetic manipulation in biotechnology and molecular therapy (Lambowitz and Zimmerly 2004).

Part I: Mitochondrial ribosomal introns that carry 5'-terminal inserts and splice by hydrolysis

Introns with 5'-terminal inserts among mitochondrial subgroup IIB1 introns

As we discussed in the Introduction, group II introns comprise six secondary-structure domains radiating from a central wheel. Although the secondary structure of group II introns is very conserved, most of the sequence is not conserved, except domain V. Therefore, by using the characteristics of sequence conservancy of domain V, once we detect a candidate domain V, it is feasible to use comparative sequence analysis to build the potential secondary structure of the entire ribozyme step-by-step. Moreover, the first five nucleotides of the intron tend to obey a characteristic consensus sequence, GUGYG, which is conserved in some 85 percent of known group II members, so that this consensus sequence helps to define the boundary of the 5' extremity of the intron. By proceeding in this manner, we have established a database of candidate group II introns in the organelle DNA sequences that have been published. During the multiple alignment and analysis of group II intron sequences, our attention was brought to a small subset of introns that diverged somewhat from normal introns. Ten group II members were selected on the following criterion: strikingly, in those ten introns, the end of the 5' exon – as inferred by comparison with uninterrupted versions of the host gene – and the GUGYG consensus sequence are separated from one another by 1 to as many as 33 intervening nucleotides (Table 1 of manuscript #1 and Figure 1). Importantly, the position of the 5' splice site can be verified by checking that the sequence of EBS1 in domain I always base pairs with the last nucleotides of the 5' exon (IBS1).

Moreover, at the other intron end, the potential secondary structure of domain VI was found to lack a bulging A at the expected location for the branch site (Figure

2). These introns, which happen to belong to the same ribozyme structural subgroup (IIB1; Michel et al., 1989) and are inserted in ribosomal RNA precursor transcripts, show additional remarkable features (Table 1): they all lack the second, EBS2-IBS2 pairing, between the ribozyme and 5' exon, which is potentially present in most group II introns, and several of them appear to code for a homing endonuclease, rather than a reverse transcriptase. A list of published sequences of mitochondrial subgroup IIB1 members, which comprises the ten organelle group II introns we found to possess an insert at their very 5' extremity, is provided in Table 1.

When we looked carefully at the introns with a 5'-terminal insert, we found that the sequence and secondary structure of domain VI was more variable than in evolutionary closely related introns: not only the branch-point adenine is missing at its expected location, but the well-conserved 3-bp helix and (GAA:CUA) internal loop immediately distal of it are unrecognizable. However, some traits of domain VI are well conserved. For example, domain VI of all introns start with G:C pairs and most of the introns with 5' insertions still keep the GNRA tetraloop at the tip of dVI. That tetraloop is generally believed to participate in the η-η' tertiary interaction during the second step transesterification (Chanfreau and Jacquier, 1996). (Figure. 3)

Sequence analysis of the *Grifola frondosa* and *Pycnoporellus fulgens* SSU788 introns

Those introns with 5' terminal insertions intrigued us and we wondered what might be their performance in *in vitro* self-splicing experiments. To further investigate this point, we chose to clone three SSU788 introns among the introns with a 5' terminal insert. One was the *Pycnoporellus fulgens* SSU788 intron (GenBank entry AF518690), the sequence of which was incomplete in the database. The other ones were insertion-lacking, related SSU788 introns that

came from the basidiomycete fungi *Grifola frondosa* and *Aleurodiscus botryosus* and for which only partial sequences (accession numbers AF334880 and AF026646) were available as well.

As shown in Figure 3, the predicted secondary structure models of the *Grifola* and *Pycnoporellus* ribozymes are very similar. The nucleotides known to participate in intra- or inter-domain, long-range tertiary interactions (Toor et al., 2006; Michel et al., 2009) are especially well conserved. However, there is a striking feature that has drawn our attention, both introns are missing EBS2-IBS2. The EBS2-IBS2 pairing is an extended canonical pairing that involves nucleotides upstream of IBS1 on the one hand, and a single-stranded loop in the distal section of subdomain ID on the other. The EBS2-IBS2 pairing is potentially present in a majority of group II introns, to the exception of members of subgroup IIC, whose 5' exon displays a hairpin structure at the expected location for the IBS2 sequence (Granlund et al., 2001). Moreover, the entire subdomain containing the EBS2 and β' nucleotides is also missing. This subdomain carries an additional sequence that, in many introns, potentially participates in the β-β' long range interaction with subdomain IC2 (Michel et al., 1989).

When we superimposed the secondary structures of the *Grifola* and *Pycnoporellus* introns, the domain VI sequences were seen to be dramatically variable. Only the first three base pairs of this domain are conserved between the *Grifola* and *Pycnoporellus* sequences. The ORFs encoded in domain IV are rather similar, both ORFs consist of closely related (91 identical amino-acids) members of the LAGLIDADG family of DNA double-stranded homing endonucleases (Stoddard, 2005). Comparing with other SSU788 introns, four out of seven published sequences of these introns contain coding sequences for additional double strand LAGLIDADG homing endonucleases (Stoddard, 2005).

This fact implies that the homing endonuclease associated with SSU788 introns is quite conserved.

Deletion of the ORF from the *Grifola* and *Pycnoporellus* introns and adjustment of conditions for transcription

The *Grifola frondosa* and *Pycnoporellus fulgens* SSU788 introns were PCR amplified and cloned in the pUC19 plasmid by G. Bassi and M. Costa. During several preliminary tests of the two introns, it was found that although the *Grifola frondosa* and *Pycnoporellus fulgens* introns are able to perform well transesterification and hydrolysis *in vitro* (data not shown), the resolution on gel is quite poor. On the one hand, the huge lariat and lariat-intermediate of *Grifola* (>1500 nts) cause a smear and are impossible to analyse properly on 4% acrylamide gels. On the other hand, the large hydrolysis products of the *Pycnoporellus* intron are difficult to distinguish from precursor (data not shown). In order to ensure better observation and quantification, both the ORFs of the *Grifola frondosa* and *Pycnoporellus fulgens* introns in domain IV were deleted. For deletion of ORF sequences, standard PCR and molecular cloning techniques were applied with primers GRXHOFWD, GRXHOREV, PYXHOFWD and PYXHOREV (Table 4). In the final constructs, 1098 bp and 898 bp of domain IV were deleted from the *Grifola* and *Pycnoporellus* introns, respectively. Transcription of digested GR1ΔORF plasmid (see below) results in a precursor that is 1057 nt in length, with 266 nt of 5' exon and 156 nt 3' exon; while the PY1ΔORF precursor is 1073 nt in length with 274 nt of 5' exon and 156 nt of 3' exon.

Plasmids comprising the group II introns were purified on CsCl gradients and linearized by SmaI, after which transcription was performed as described in Materials and Methods. For *in vitro* transcription of template DNA, we used a

relatively low concentration of MgCl$_2$ (<25mM) with extra 10% DMSO. The purpose of low MgCl$_2$ is to prevent premature splicing during transcription. The reason we used extra 10% DMSO in the transcription buffer is that several T7 transcription stops were found in the sequence; 5-15 % DMSO have been proved able to overcome the secondary structure of DNA and achieve a better transcription. Actually, without DMSO, transcription yields always a mixture, with several minor transcription products on gel. Those premature transcription stops dramatically decrease the yield of transcription. Therefore, we concluded that lower concentrations of MgCl$_2$ and 10% DMSO are favourable conditions for transcription of the *Grifola* and *Pycnoporellus* introns (data not shown).

Contrasting self-splicing products of the *Grifola* and *Pycnoporellus* SSU788 introns

The lack of the group II branch-point bulged A in domain VI of the *Pycnoporellus* SSU788 intron suggested that splicing would be initiated by hydrolysis at the 5' splice junction, rather than by transesterification (Jacquier and Jacquesson-Breuleux, 1991; Daniels et al, 1996. Vogel and Borner, 2002). This was confirmed by incubating precursor transcripts containing the *Grifola* and *Pycnoporellus* SSU788 introns under conditions that allow *in vitro* self-splicing. Several different parameters were varied for finding the optimal conditions; those include temperature, monovalent salt and MgCl$_2$ concentration. Finally, *in vitro* self-splicing of the *Grifola* SSU788 intron (Figure 6) was found to be a reasonably efficient process at 42°C in 1 M NH$_4$Cl and at 20 mM magnesium. Reaction is a kinetically complex process, the reaction of precursor is divided into two phases, part of precursor molecules reacted fast in the first two min, while the second population of precursor transcripts reacted relatively slowly (Figure 6B). Splicing products are dominated by the lariat intron and ligated exons as for other typical group II introns. Only very small amounts of a

linear intron form were observed: even when the ammonium ions were replaced by potassium ions, there was still less than 15% of linear intron products. Because there was so little linear intron found in NH_4Cl conditions, the possibility exists that the linear intron was generated from hydrolysis.

The *Pycnoporellus* SSU788 intron showed a rather rapid reaction process when incubated in the same condition as for the *Grifola* intron (Figure 6C.): 80% of the precursor was converted to products in about 10 min. The population of molecules was also divided into two populations. However, there was no branched product to be seen, all products seem to be generated through the hydrolysis pathway. At late time points of reaction, the *Pycnoporellus* intron shows unknown additional fragments, which may have been generated from the linear intron. The size of the main unknown fragment is around 70 nt shorter than the linear intron, which is 642 nt in length. We assume that this unknown fragment is derived from the linear intron by a cryptic cut, although the precise cutting site cannot be predicted from these data alone. Most importantly, varying the concentration of magnesium did not make it possible to observe lariat molecules among self-splicing products of the *Pycnoporellus* intron, a higher concentration of magnesium only slightly increases the extent of reaction for both the *Grifola* and *Pycnoporellus* precursors. However, over 50 mM magnesium also caused ligated exon reopening. Finally, we confirmed that the optimal magnesium concentration in terms of reaction rate and final extent of reaction is around 10 to 20 mM.

Reverse transcription of splicing products

After testing the splicing ability of both the *Grifola* and *Pycnoporellus* introns, there were several questions we needed to further clarify.

1. Does the *Grifola* intron use the 'correct' (predicted) branching point during lariat formation?
2. Where is the cleavage site of *Pycnoporellus* during the hydrolysis reaction? Is cleavage as precise as for the *Grifola* intron?
3. During the hydrolysis reaction of *Pycnoporellus*, an unknown fragment was seen at late time points. We assume that is a side product from hydrolysis. However, what is the precise cutting site?

In order to clarify the questions listed, reverse transcription mapping experiments were performed. The lariat and ligated exons of GR1ΔORF and the linear intron, unknown fragment with additional cut (AC) and ligated exons of PY1ΔORF were collected from gels. The lariat-3' exon intermediate product of GR1ΔORF was isolated separately with 20 mM calcium. However, reaction with calcium is much more inefficient than with MgCl$_2$. Only a small proportion of precursors was active and the yield of lariat-3'exon intermediate was about 25%. To determine whether the *Grifola* and *Pycnoporellus* introns recognise the 3' terminus of the 5' exon and ligate the two exons correctly, primer extension mapping experiments were applied. Oligonucleotides BMS103B and Gr-R2 were kinased and ^{32}P-labelled. BMS103B was used to sequence ligated exons and determine the branch-point of the *Grifola* intron-3'exon lariat, while Gr-R2 was used to determine the 5' splice of the *Grifola* intron lariat.

Results of reverse transcription are revealed in Figure 5. The predicted bulged A in domain VI of the *Grifola* intron correctly connected to the first nucleotide of the intron and the 5' exon and 3' exon were also correctly ligated. The same

procedure was applied for the ligated exons and linear intron of *Pycnoporellus* as shown in Figure 5. One thing worth noticing is that the purified linear intron of *Pycnoporellus* is a mixture: beneath the major linear intron, a light band can be observed. We assume that part of the linear intron may bear an extra cleavage. In order to determine the cleavage site, the two bands were purified together and reverse transcribed. If the cutting site was located at the 5' end of the intron, the site should be seen during primer extension. However, there was also a probability that the mysterious cutting site was located at the 3' end of intron. Therefore, linear introns were labelled by RNA ligase with primer 18873 at their 3' extremity (see Materials and Methods). Using PYCXHOFOR and 18873 as primers, the intron and its flanking sequence was amplified by RT-PCR and subcloned in the pGEM TA cloning vector (Promega). However, sequence analysis of the tagged linear intron revealed a PCR artifact in which primer 18873 had hybridized in domain VI (right after GTTCTTAT).

As previously mentioned, we had assumed that the unknown fragment (AC) corresponded to the linear intron with a cryptic cut. In an attempt to identify the site of cleavage, we applied the same tagging method used for the *Pycnoporellus* linear intron. After subcloning and sequence analysis, we unexpectedly found a candidate cryptic cleavage site located in the terminal loop of domain IV of the ORF-less construct, right after the introduced CTCGAGCTT sequence. This result is also controversial, because our best guess based on sequence analysis was that ribozyme-catalysed, hydrolytic cleavage of the linear intron would occur at position 110, 3' of the sequence UAGGAC, which offers a better match to EBS1 (GUCCUU) than the IBS1 sequence (UAGGAU) at the 3' end of the 5' exon.

Disscusion

Analysis of self-splicing products reveals the inability of an intron with additional inserted nucleotides at the 5' end to initiate splicing by transesterification

In our experiments, we have tested the self-splicing ability of both the *Pycnoporellus fulgens* and *Grifola frondosa* introns *in vitro*. Strikingly, the *Pycnoporellus fulgens* intron generates only linear intron despite being a close relative of the *Grifola frondosa* intron which produces lariat. In fact, when comparing the sequences of the two introns, they might seem to differ from one another only in minor areas. However, when the *Pycnoporellus* intron is examined in detail, one's attention is drawn to the lack of a bulging A at the expected location for the branch site and the presence of an insert at the 5' end.

It was shown long ago in multiple investigations that deletion or base-pairing of the bulging A at the branch site inhibits branching of the *S. cerevisiae* cox1/5γ intron (Chu et al., 1998 and references therein). Indeed, all of the splicing products from *Pycnoporellus* intron are generated by hydrolysis in our experiments. Moreover, constructs with additional nucleotides inserted between the IBS1 sequence in the 5' exon and the GUGCG consensus sequence at the intron 5' end were also reported long ago by Jacquier and Jacquesson-Breuleux (1991) to have undergone loss of the cox1/5γ branching reaction *in vitro*: when 17 nucleotides were added, splicing was found to proceed exclusively by hydrolysis at the 5' splice site, and the latter was shown to coincide with the 3' end of the IBS1 sequence, rather than with the 5' end of the GUGCG consensus sequence. However, this is the the first time that group II introns with additional nucleotides at the 5' end are reported in a natural environment.

Comparison of features between 'degenerated' and normal introns

The most striking feature of introns with a 5' insertion is the diversity of their domain VI sequences and structures. Such variability strongly suggests rapid, unconstrained divergent evolution. Introns with a 5' insertion miss the branchpoint bulging A, while it is well reported that the deletion or base pairing of this bulged A inhibits branching (Chu et al., 1998). Besides, they also lack the two G:U pairs flanking the branchpoint: their replacement by G:C pairs has been reported to specifically decrease the rate of branching (Chu et al., 1998). Moreover, it is difficult to recognize the characteristic AAA:CUA internal loop, which exists normally in mitochondrial subgroup IIB1 introns. However, despite introns with a 5' insertion having lost such essential features of normal group II introns in domain VI, we found that seven of the ten intron sequences with additional nucleotides at the 5' extremity have retained a 4-nucleotide terminal loop of the GNRA family, which also exists in all mitochondrial and bacterial members of subgroup IIB1. The GUAA loop that caps domain VI of the *S. cerevisiae* cox1/5γ intron was shown by Chanfreau and Jacquier (1996) to interact with a specific receptor in ribozyme domain II: mutations that disrupt the interaction inhibit exon ligation and increase the rate of first-step transesterification.

This η-η' interaction is widely conserved in group II ribozymes (Costa et al., 1997) and a potential η receptor exists indeed in domain II of each of the intron sequences in Figure 2 that share a GNRA loop at the tip of domain VI (data not shown). Interaction of the η' GNRA loop with its η receptor appears to mediate a structural shift from a ribozyme conformation that allows the branching reaction to another one which favors exon ligation and, also, hydrolysis at the 5' splice site. Specifically,

binding of domain VI to domain II after branch formation was proposed to drag the first-step product – the 2'-5' ligated A-G dinucleotide – out of the catalytic site, so as to make way for the 3' splice site (Chanfreau and Jacquier (1996). Persistence of the η-η' interaction in introns that have lost the branchpoint implies that formation of this interaction contributes also to the specific positioning of the 3' splice site for exon ligation. Formation of η-η' may help correct exon ligation by reducing the complexity of the conformational space to be explored in order to bring the 3' splice site into the catalytic center of the ribozyme.

Homing endonuclease gene habored in introns with 5' terminal insertions

Group II introns that possess 5'-terminal inserts constitute a quite small population compared with the thousands of other group II introns that have been published. In our study, those particular introns are found only in mitochondrial ribosomal RNA genes and they only belong to the IIB1 subgroup of ribozyme structures. Another coincident fact is that all known group II introns encoding proteins unrelated to reverse transcriptases also belong to subgroup IIB1 and come from mitochondrial genes encoding ribosomal RNA precursor transcripts. Actually, 4 out of the 10 members of our data pool of group II introns with a 5'-terminal insert happen to encode a protein of the the LAGLIDADG family of endonucleases (some of the ORFs in our database are defective or missing). It is a fact that a majority of the introns containing non-RT ORFs are lacking 5' inserts and have a normal domain VI, while six out of ten introns with 5' inserts lack any significant protein-coding potential. Still, such a coincidence raises the possibility that some causal relationship exists between the acquisition of a non-RT ORF, with putative homing endonuclease activity, and that of

a 5'-terminal insert. The special features of those introns may have the potential to develop into a new class of group II introns.

LAGLIDADG family endonucleases promote homing by generating a double-stranded cut with 4 nucleotide (nt) 3' OH overhangs in DNA; the break is repaired by the host's double-stranded break repair processes using the intron containing allele as a template (Stoddard, 2005). Homing endonuclease genes invade DNA sequences encoding self-splicing intron. The homing endonuclease genes spread in the population through the cleavage activity of the endonuclease and once all possible homing sites have been invaded, the homing endonuclease gene is no longer under selection pressure and soon begins to accumulate mutations and degenerate. Unless the homing endonuclease becomes essential to its host by acquiring 'maturase' activity: maturases help the ribozyme to fold into an active structure and participate in the splicing process, as is the case indeed for the LAGLIDADG proteins encoded by some group I introns, which have been shown to function as maturases and promote the splicing of their host group I intron (Bassi et al. 2002; Bassi and Weeks 2003).

Based on that, we can hypothesize that the invasion by an endonuclease gene is not permanent; that gene that once invaded may become degenerate beyond recognition and it is not strange therefore to see some of those groups II introns with 5'-terminal insert having ambiguous features of ORFs. We assume that introns with 5'-terminal inserts spread with the help of encoded homing endonucleases. Since the homing process mediated by DNA endonucleases rests on resealing of the double-strand break by homologous recombination, the intron RNA should not be necessary in this mechanism, like in the model provided by group I self-splicing introns and archaeal introns. In the group II model of

transposition, the group II lariat structure undergoes retrohoming by an inverse splicing mechanism. However, once a group II retrotransposon has been converted into a DNA transposon (class I mobile element; Wicker et al., 2007) by the loss of its reverse transcriptase and the acquisition of the coding sequence of a homing endonuclease, a 2'-5' phosphodiester bond should no longer be required for optimal mobility and the ability to generate this bond may become lost by mutations of the branch site or else, the insertion of nucleotides at the 5' splice site.

Part II: Discovery of a docking site for domain VI

A novel phylogenetic analysis provides clues to the potential docking site of domain VI during the first step of splicing

Group II introns are self-catalytic RNAs that are quite divergent in sequence, but their secondary structure is well conserved. Multiple phylogenetic analyses have been carried out with the hope of discovering inter and intra-domain interactions that contribute to the splicing function of group II intron. In 2008, after more than fifteen years of analyses and researches, a fair number of interactions and tertiary contacts had been published, and a consensus, overall three-dimensional model of the group II intron structure was emerging (Costa et al., 2000; F. Michel and E. Westhof, unpublished data; Dai et al., 2008). This was the year when crystallographic analysis of the *Oceanobacillus iheyensis* group IIC intron was published by Toor et al. (2008). Thanks to this breakthrough, many of the predicted interactions could be further confirmed and the catalysis mechanism has been better elucidated. These achievements reach beyond the area of group II intron researches in that they have potential implications for the structure and function of the spliceosome, which many believe is related to group II introns, but also contribute to pushing our knowledge of RNA to an upper level.

Even though more and more data are becoming available concerning the structure and functions of group II intron, intriguing questions are still waiting for explanations. Probably foremost among questions that needed to be answered after the publication of the atomic-resolution structure of the *Oceanobacillus* intron was the location of the docking site of domain VI during the first step transesterification. Although this topic has been much discussed and debated for a long time and multiple strategies have been applied, none could successfully answer this question.

The failure of standard phylogenetic/covariation analysis implies that docking of domain VI rests on non-canonical base pairing: no obvious co-evolutionary patterns can be discerned. Nevertheless, during our phylogenetic alignment of group IIB1 intron sequences, our attention was brought to a small subset of introns that appeared to diverge somewhat from normal introns. By comparing variations in sequence and structure between standard group II introns and those introns, which all have 5' terminal inserts, we have discovered several distinctive features that might contribute to the partly defective splicing of the introns with inserts. In the following results, we describe our attempts to find the docking site of domain VI during the first step of splicing by examining the possible functional implications of structural variations with the help of mutagenesis and kinetics analyses. For our experimental system, a well established model, the subgroup IIB1 *Pylaiella* LSU1787 intron (Pl.LSU/2; Costa et al., 1997; Figure 8) from the mitochondrial genome of the brown alga *Pylaiella littoralis*, was selected. The Pl.LSU/2 intron carries out accurate splicing at an optimal magnesium concentration of less than 10 mM and can undergo steady reaction by both the transesterification and hydrolysis pathways. Moreover, it generates more than 90% of lariat product and ligated exons in optimal concentrations of ammonium and magnesium.

Deletion of EBS2

As shown by our phylogenetic analysis of Group IIB1 introns in fungal mitochondrial ribosomal RNA (article #1), the absence of one of the exon-binding sites (EBS2) and the β- β' interaction (Figure 8) in introns with 5'-terminal insertions is particularly noteworthy. In order to explore the contributions of EBS2 and β- β' to the group II first step transesterification, we have deleted domain ID2 from the molecule (ΔID2

or PL2-54 mutant; Figure 9). Because deletion of the entire domain ID2 may also change the geometry of domain ID3 and such an effect is hard to predict, additional versions of the mutagenesis were applied. In constructs PL2-55.1 (ΔID2 + U) and PL2-55.2 (ΔID2 + A), an extra U and A nucleotide, respectively, was added between domains ID and ID3 in order to increase the flexibity of stem ID3 (Figure 9).

Three ID2 deletion mutants were analyzed kinetically under conditions that favor hydrolysis, that is, with 50 mM $MgCl_2$ and 1M KCl (see Materials and Methods). The three ID2 deletion mutants reacted as well as the wild type intron in terms of total reaction (Figure 9-1). Not only did 90% of precursors react within 120 mins, but they also showed similar total rates of reaction. When reacted fractions are further examined, it is seen that all four introns can splice by both the transesterification and hydrolysis pathways. However, for the ID2 deletion mutants, reaction through the hydrolysis pathway (which yields linear intron) is significantly increased, when compared with the wild type. Meanwhile, when rates of transesterification and hydrolysis are compared, the k_{br}/k_{hy} ratio gives the opposite result: for the wild type this ratio is around 0.75, while for mutants it is around 2.5. That means that ID2-deleted introns react significantly faster by the transesterification pathway. To sum up, ID2 deletion is reflected in the increase of products from the hydrolytic pathway, but the branching pathway is accelerated instead. The precursor molecules can be divided into two populations in the reaction, one is prone to producing lariat by transesterification, while the other is prone to generating linear intron by hydrolysis. The deletion of ID2 causes a partial switch of the population from transesterification to hydrolysis, while the total reacted fraction remains constant.

Judging from our results, we can conclude that deletion of ID2 influences to some extent the choice of precursor molecule between the transesterification and hydrolysis pathways. Although deletion of ID2 shows influence as well over the branching reaction, it must be recalled that the ID2 domain includes an important tertiary interaction, EBS2, that contributes to structural stability and 5' exon recognition. Our data cannot distinguish between defects in catalysis and structural instability.

Mutagenesis of domain VI

We have shown in our previous studies (article #1) that group II introns with an extra insertion at the 5' terminus, like the *Pycnoporellus fulgens* SSU788 intron, are only able to perform splicing through the hydrolysis pathway, while normal group II introns undergo preferentially splicing by the transesterification pathway. Showing typical sequence variation of their respective subgroups, the *Grifola frondosa* and *Pycnoporellus fulgens* SSU788 introns, which belong to the same lineage, were picked out as model molecules (Figure 3). It is noticeable that most variations are located in domain VI (Figure 4). Not only is the branch-point adenine missing, but the well-conserved 3-bp helix and (GAA:CUA) internal loop are unrecognizable in the *Pycnoporellus* intron. However, G:C pairs at the base of the DVI stem and the η' GNRA loop (Chanfreau and Jacquier, 1996) at its tip are well conserved. This fact implies that while the bulge-lacking intron fails to generate a lariat molecule, it still forms the η-η' interaction after the first step of splicing.

We speculated from these data that the internal loop of domain VI may participate in the branching reaction. In order to understand the contributions of the stem of domain VI to first-step transesterification, we manipulated those nucleotides by truncation and mutagenesis. Two versions of a modified domain VI were introduced: firstly, the stem was

truncated down to two nucleotides after the bulge to generate a short version of domain VI (PL2-57, or DVI-2bp in article #2; Figure 10-1); secondly, the internal loop of domain VI was zipped up and replaced by a continous domain VI helix from *Pseudomonas spp.* (PL2-58 or DVI stem). The GAA: CUA internal loop has long been known not to be strictly necessary for a proper branching reaction, as was shown by a series of sequence manipulations (Chu et al., 1998, 2001). Moreover, our continuously helical domain VI was borrowed from a naturally existing domain VI from a bacterial group II intron that must perform a typical branching reaction. The advantage of replacing an internal loop with a continous helical stem is that the latter can be further manipulated, taking advantage of its geometric stability; in contrast, the geometry of RNA loops is always difficult to predict due to their flexibility. Especially for a struture like domain VI, randomly changing any of its nucleotides would possibly lead to unpredicable changes of geometry.

When domain VI with a continuous stem was compared to the wild type intron, the former was found to perform splicing by both transesterification and hydrolysis, despite the fact that it seems to favour somewhat hydrolysis when placed under KCl conditions. PL2-wt, PL2-57 (DVI-2bp) and PL2-58 (DVI stem) were then kinetically analyzed and compared under both transesterification-favouring conditions, with 1 M NH$_4$Cl, and hydrolysis-favouring conditions with 1 M KCl (Figure 10-2). In our results, although the mutant with a truncated domain VI (PL2-57; DVI-2bp) keeps a bulged A flanked by two G:U wobble pairs, it nevertheless loses the ability to perform splicing through transesterification. This fact seems to imply that competent docking of domain VI needs not only the bulged A but also surrounding nucleotides in domain VI to contact with. It is reasonable to believe that the stem part

of domain VI plays an important role in its docking, although the internal loop of domain VI tolerates certain modifications.

Among further questions we wished to answer was which part of domain VI is used for docking during the first step transesterification? To address that, the continuous-stem domain VI was further truncated. PL2-70 (DVI-4bp) and PL2-71 (DVI-7bp) are mutants with 4- and 7-base-pair distal helices, respectively (Figure 11-1). Firstly, mutant RNAs were tested under transesterification-favouring conditions, in a solution containing 1 M NH_4Cl and 10 mM $MgCl_2$. The products from the transesterification and hydrolysis pathway were both analyzed kinetically (Figure 11-3). For the wild type, the transesterification pathway dominated the whole reaction, over 80 % of reaction occured by transesterification, while only a small proportion linear intron could be observed. Because some of the linear intron seen on gel also probably came from broken lariat generated by the transesterification reaction, which migrates at the same location as linear intron, the fraction resulting from transesterification can only be higher than what was observed. For the mutant with a 7 bp helix (DVI-7bp), the products from both pathways were quite similar in abundance to the wild type, the transesterification pathway was very active. One thing worth noticing is that the mutant with a 7-nt helix showed more lariat-intermediate (lariat with 3' exon), which must reflect the fact that the second step of transesterification became slower. This can be explained by the fact that the shortening of domain VI also affects the η-η' interaction between domain II and domain VI. For the mutant which is further deleted with only a 4-nt helical stem (DVI-4bp) it shows slightly fewer amounts of branched products and also a slower reaction rate. To summarize, modifications of the domain VI stem have only a minor influence on the transesterification pathway in the presence of ammonium. Probably, the

monovalent salt helps stabilize the structure of the group II intron, while that was not initially seen in our experiments.

Secondly, we examined the same mutants under hydrolysis-favoring (1 M potassium) conditions. The Pl.LSU/2 wild type precursor generated three times more lariat-related products than hydrolysis-related products during a 120 min reaction, while when reaction rates were compared, the transesterification pathway was also seen to be definitely faster than the hydrolysis pathway (k_{br}/k_{hy} is around 2.5). When DVI-4bp was reacted under the same conditions, we found that the direction of reaction was opposite. The hydrolysis pathway was more active, hydrolysis-related products were three times more abundant than transesterification-related products. At the same time, the rate of hydrolysis was two times higher than the rate of transesterification. However, when the other domain VI truncated mutant with a 7-nt helix was compared under the same conditions, we found that the transesterification-related products were dramatically increased both in terms of quantity and rate of production (Figure 11-2), even though the products of hydrolysis still dominated over transesterification products. Based on this result, we have reasons to suppose that the area between the 4th and 7th base pairs of domain VI contributes some significant function during first step transesterification.

Although the D6 7 nt-helix mutant does not perform as well as the wild type, it should be recalled that modification of domain VI is also altering the η-η' interaction between domains II and VI, which is an interdomain interaction that contributes to the second step of the transesterification pathway.

IC1 mutagenesis

Another feature that was made noticeable by our phylogenetic analysis of Group IIB1 introns was the presence of two consecutive G:U wobble pairs in domain IC1 of introns that lack a 5'-terminal insertion. When the sequence of the group II intron found in the *Grifola frondosa* mit SSU gene is compared with that of the homologous intron in *Pycnoporellus* (Figure 4) and the secondary structure models are superimposed with one another, it is seen that a majority of nucleotides and pairings is well conserved. Except for significant variation in domain VI, there were rather few differences between the two introns, so that the two G:U wobble pairs in domain IC1 of *Grifola* truly stand out. More generally presence/absence of these two G:Us co-varies with the presence of a canonical domain VI structure, which led us to speculate that the two G:Us actually play a role in the docking of domain VI during first-step transesterification.

Accordingly, mutagenesis was applied again to explore the function of these wobble parings in IC1. The following set of IC1 mutants was created. Firstly, the two G:U wobble pairs were changed to two A:U Watson-Crick pairs (PL2-56, IC1 UA:UA Figure 12-1). However, this change made us concern that the geometry of domain IC1 could also be altered, which might affect the position of the tetraloop at the tip of domain IC1. As already known, that tetraloop forms an inter-domain interaction with domain II. This interaction, which is also known as the θ–θ' interaction, is a canonical tetraloop-receptor interaction that joins the tip of the IC helix with the basal stem of DII (Costa et al., 1997). This important interaction serves to brace the IC helix and to govern the ultimate orientation of the DII and DIII stems.

To clarify this concern, two other mutants were created in addition to PL2-56 (IC1 UA:UA), in which the two consecutive G:U wobble pairs of domain IC1 are replaced by two U:A's. In mutant PL2-63 (IC1 $\Delta\theta$), the tetraloop of domain IC1 was replaced with a UUCG loop which is designed as a non-functional component, since it cannot react with the receptor of GNRA loops (Costa and Michel, 1997). Finally, in mutant PL2-64 (IC1 $\Delta\theta$ UA:UA), the two features – deletion of the GNRA terminal loop and substitution of the two GU pairs – were combined.

The wild type Pl.LSU/2 and the IC1 modified mutants PL2-56, PL2-63 and PL2-64 were tested in 1 M KCl with 10 mM $MgCl_2$ at 42°C. The products from both pathways (transesterification and hydrolysis) were quantitated and kinetic parameters were calculated. PL2-63 (IC1 $\Delta\theta$) showed features similar to the wild type, in which the branching reaction dominates over the hydrolysis reaction. Only less than 20% products were from the hydrolysis pathway. The reaction rates of both pathways were affected, but hydrolysis was more severely affected than branching (Figure 12-2). The θ-θ' interaction is seen to have only minor influence on the branching reaction and this influence seems to be more related to structural stability, instead of catalysis. Strikingly, the reactions of the two UA:UA mutants (PL2-56 and PL2-64) were strongly driven towards hydrolysis pathway. Over 80% of products came from hydrolysis, and only little lariat-related products could be observed. The kinetic parameters reflected the same tendency, with the branching rate (k_{br}) becoming five times slower than for the wild type, while the rate of hydrolysis is only slightly affected and its reduction can be incriminated upon the direct (or indirect) alteration of the θ tetraloop. Specifically, the ratio of reaction rates of the two UA:UA mutants were 0.66 and 0.82, respectively. These numbers indicate that the transesterification pathway was significantly impaired and hydrolysis became dominating (Figure 12-

2). It is remarkable that such a slight change in domain IC1 should have such a strong influence. Because these two G:U pairs are co-varying with the loss of the bulged A in domain VI in our phylogentic analysis, it is reasonable indeed to hypothesize that the two G:U pairs form the docking site of domain VI during first step transesterification.

To further adress the role of these two G:U pairs, the A:U substitution mutants were tested under multiple conditions with different combinations of monovalent salts. No matter what was the combination of NH_4Cl and KCl, the G:U pair mutants show the same tendency to the hydrolysis pathway, even in conditions favoring branching (1 M ammonium monovalent cation). It should be noted, however, that conditions of higher ammonium concentration help restore the function of branching and the branching pathway cannot be totally blocked no matter the precise conditions.

Bimolecular reaction system *in trans* (two separate pieces)
Although we had shown that the two G:U wobble pairs have an important function in the transesterification pathway, we still did not know how their substitution affects transesterification. In order to understand whether the two G:U pairings in IC1 are really a docking partner for domain VI, we designed a bimolecular/'in trans' reaction system. The IC1 subdomian has been shown to be an important subdomain, there are several inter-domain interactions that have been reported to involve IC1 and the surrounding structure appears to be essential for maintaining the catalytically reactive conformation of DV (Toor et al., 2008). The latter function is achieved through the λ-λ' interaction between domain V and the internal loop region of domain IC1 (Boudvillain et al., 2000). Besides, the tetraloop-receptor interaction, θ-θ' also plays an important role in

stabilizing the group II ribozyme core, especially for efficient self-splicing at elevated temperatures (Costa et al., 1997).

In relation with these facts, domain IC1 can also be used to set up a bimolecular reaction system. A defective θ-θ' interaction does not affect catalysis, but the lack of the λ-λ' interaction has devastating effects for group II intron catalysis. In our experimental design, domain IC1 was truncated and replaced with the AATT sequence to seal the internal loop (see also Costa et al., 1997). In order to diminish the effect of the η-η' interaction between domains II and VI, domain II was also deleted and replaced with a UUCG loop (PL2-67, Figure 14). Meanwhile, domain IC1 was reconstructed and subcloned in a pUC19 plasmid (reviewed in detail in Materials and Methods). In this construct, domain IC1 can be transcribed separately from the rest of the intron. In this system, the fragment consisting of domain IC1 acts as an "enzyme" participating in the reaction, The group II intron without domain IC1 should not be able to perform catalysis. By manipulating the sequence of the domain IC1 fragment, we hoped to further understand the role of each nucleotide participating in transesterification. Furthermore, individual RNA nucleotides may be modified by exploiting alternatives to RNA synthesis in which each atom may be altered and replaced to test its function. By combining the information from such experiments, it might be possible to build an atomic-resolution model of the complex formed by domain VI, the branch site and the rest of the intron at the time at which splicing is initiated.

Unfortunately, our first experiments conducted under conditions of 1 M ammonium choride, 20 mM magnesium, 45°C were not encouraging. Although the intron-containing precursor transcript did not react in the absence of domain IC1, it also reacted very poorly even in the presence of

a large amount of the domain IC1 piece (IC1 ranging from 10 μM to 20 μM in this experiment). The reaction was very fast initially, but progressed very slowly after 2 min and only about 20% of precursors were able to convert into spliced products. This fact may result from misfolding of the precursor at the beginning, the well folded fraction of molecules reacted fast, but the remaining fraction reacted poorly due to a lack of proper folding or quick dropping off of domain IC1 (low affinity reflecting a high k_{off} of domain IC1).

To improve the performance of the system, we tried to either lower the temperature or increase the concentration of magnesium used in the reaction: lowering the temperature to 30°C should increase the affinity between domain IC1 and the precursor, while a higher concentration of $MgCl_2$ could also help stabilize the folding of precursor. However, low temperatures also decrease the ability to perform catalysis, only less than 10% of the precursor can be reacted within 90 min. On the other hand, the precursor became too active at high concentrations of $MgCl_2$, precursor molecules were able to react even in the absence of the domain IC1 'enzyme'. Even though the same kind of experiment was carried out with similar methodology with the ai5γ group IIB1 intron (Costa et al., 1997), our own bimolecular reaction system did not work as well as we expected.

Demonstration of the identity of the first-step receptor of domain VI by the use of DNA oligonucleotides as bridging linkers

In order to gather more convincing evidence in favor of our hypothesis that IC1 is the actual docking site of domain VI in first-step transesterification, we designed another system that uses an oligonucleotide as a chain to anchor domains IC1 and VI. The general concept of this set-up rests on the fact that should domain VI contact

domain IC1 during first-step transesterification, parts of these two domains should be very close in space. If the section of domain VI used for docking is mutated, the transesterification pathway is expected to be hindered and most of the splicing reaction will be redirected to the hydrolysis pathway. However, the mutated domain VI should still be floating around in the vicinity of its docking site. Therefore, we created a special version of Pl.LSU/2 in which the sequences of domain IC1 and VI were both changed so as to adapt to a specifically designed oligonucleotide. Once the designed oligos are added in the reaction, the Pl.LSU/2 precursor is expected to see its function restored and to initiate again the branching reaction. The oligonucleotides function as a chain to constrain the movement of domain VI and force it to dock correctly. If domain VI actually docks on domain IC1, the constrained domain VI should ensure restoration of the transesterification pathway to a certain level. By combining the information from such experiments, we hoped to be able to build an atomic-resolution model of the complex formed by domain VI, the branch site and the rest of the intron at the time at which splicing is initiated.

We started by constructing a group II intron in which domain IC1 was truncated after its GG bulge and its distal section was replaced (Figure 15; PL2-72) with a sequence of which six nucleotides are complementary to a bridging oligo called PLI55 (Table 5.). Domain VI was similarly truncated two nucleotides after the bulged A and its distal section replaced (Figure 15 construct A; PL2-73) with a sequence that pairs with another part of the PLI55 sequence. The two mutants were then combined to generate a new mutant whose domains VI and IC1 are both able to pair with oligo PLI55 at the same time (Figure 15 construct A; PL2-74). In this system, PL2-72 and PL2-73 were intended as controls.

The three mutants and the wild-type were tested under transesterification-favouring conditions of 1 M ammonium chloride, 10 mM magnesium at 37 °C with or without the addition of varied concentrations of oligo PLI55 (Table 5; generally, the concentration is calculated so as to have presumably at least 95% of oligonucleotide molecules pairing with precursor). For the wild-type and IC1 mutant PL2-72, no difference in reactivity was observed with or without PLI55. These negative results show that there do not appear to exist non-specific interactions between the sequences of the precursor intron and oligonucleotide. PL2-74 behaved as we expected: transesterification could be restored by using bridging oligo PLI55 to pull domain VI closer to domain IC1. However, with the help of bridging oligo PLI55, domain VI mutant PL2-73 is also able to perform first-step transesterification to some extent. Although the PL2-74 mutant precursor performed transesterification better than PL2-73, it was not clear to what extent restoration resulted truly from the docking of domain VI or from interaction between domain VI and the bridging oligo. Because the proposed docking face of domain VI was removed in these constructs, we speculate that the phenomenon we observed in PL2-73 resulted from reconstruction of domain VI by oligonucleotide PLI55. As also expected, transesterification can be improved by increasing the concentration of oligo PLI-55, and the effect is much more pronounced with PL2-74 than with PL2-73. These data suggest that two phenomena are superimposed in our system: one is the reconstruction of domain VI, the other one is the docking of domain VI. However, both reactions remain far away from complete compensation, the maximal practical concentration of the bridging oligo is neither sufficient to fully compensate the structure of domain VI nor enough to bring domain VI to

its docking site. These data implied that a better construct was necessary so as to compare quantitatively the two reactions

In order to reach a better understanding of the interaction between domains IC1 and VI, a new set of mutants was designed. PL2-83B was designed to replace PL2-72, while PL2-84B (Figure 15, construct B) replaced PL2-74. The newly designed mutants have higher affinity to the bridging oligo and a more stable structure. The higher affinity between oligo and precursor should help decrease the Km between the oligo and precursor. In order to ensure sufficient flexibility of domain IC1, two versions of this domain were designed: the first nucleotide (a C) to pair with the bridging oligo is part either of a C-U mismatch or a C:G Watson-Crick pair in (the isolated) domain VI. After preliminary testing of both mutants, we decided to forsake the C:G version due to its poor performance.

In this system, reaction products from oligo:precursor paired molecules are mixed with those generated from the unpaired precursor alone. The ratio of paired and unpaired molecules depends on the k_{on} and k_{off} of the oligo, the branching rate and the extent of reaction. In order to improve quantification, we have modified the analysis of data, as explained in Materials and Methods: we found that observed rates of branching and hydrolysis reactions, which can easily be estimated at t=0, are more reliable than rate constants or extents of reactions, if only because their ratio depends little of the state of RNA transcripts, which tend to become less and less reactive with time (from one experiment to the next). This type of analysis in which the relative rate of branching is plotted over the concentration of oligonucleotide yields a saturation curve which is

specific to each mutant under a given set of conditions. Typical curves are shown in Figure 16.

As seen in Figure 16, the ability of precursor PL2-84B to initiate splicing by transesterification is efficiently restored by increasing concentrations of oligonucleotide PLI68. PLI68 acts as a linker that pairs with both domain IC1 and domain VI and helps domain VI dock into domain IC1 correctly, it strongly enhances transesterification that reaches up to 60% of the total reaction at saturating concentrations of the oligonucleotide. The K_m is also much lower, down to around 5.4 μM, compared with a Km of 58 μM with previous mutant PL2-74. Thus, the affinity of the oligo for the intron was hugely improved.

A mismatched oligo (PLI69; IC1 mismatched, 15 mer) was used as a control: its sequence matches that of domain VI, but not that of domain IC1 in the PL2-84B precursor. The products generated from transesterification dropped from 60% to about 30%, and more importantly, the Km was abruptly increased, from 5.4 μM to 278 μM, just because of the mismatched sequence with domain IC1. These data imply that only a well-matched oligo can efficiently act as a linker to restore the branching reaction by pulling domain VI closer to domain IC1. Nevertheless, the mismatched oligo still contributes to some restoration of transesterification in our experiment. However, we must not forget that we had observed the same phenomenon with our first set-up PL2-73: oligos that base pair with domain VI are also able to reconstruct domain VI and restore functional transesterification to a certain level.

To verify this possibility (Table 2), PLI71 (anti-D6, 7 mer) was tested with PL2-84B. PLI71 is a 7 mer with a sequence complementary to

domain VI (anti-D6), it should contribute the same function as PLI69, but should not be able to further enhance transesterification by bringing together domains VI and IC1. In our experiments, it only helped to restore transesterification to a similar level as for PLI69. The control experiment with both oligo PLI71 (anti-D6, 7-mer) and PLI72 (anti-IC1, 7-mer) showed a similar result, as also previously observed for the combination of PL2-73 (D6) and PLI58 (anti-D6), which was used as a control as well. Additional oligonucleotides (listed in Table 5), used at the same 100 μM reference concentration, were unable to provide significant restoration of branching. Combining the results from all different oligos, we conclude that bridging oligonucleotides do restore transesterification by helping domain VI to dock with domain IC1.

The length of an oligo poly-linker largely affects its ability to restore branching

Although an oligonucleotide with a matched sequence can restore the branching reaction of a functionally defective precursor, the paired molecular complex still does not perform the branching reaction as well as the wild type. We believe that this is because the position of domain VI is somewhat ill-defined in space, since the bridging oligos connect domain VI and domain IC1 as a flexible chain. The branching reaction only occurs when the bulged A of domain VI occasionally docks in a correct position. Based on this assumption, we designed a set of oligos that have the same pairing sequences as PLI68, but in which the length of the linker varies from zero to 4 nucleotides (Table 5.). These oligonucleotides were tested at a concentration of 5 μM, which is close to the K_m of the PL2-84B:PLI68 combination. To our surprise, the length of the linker was found to hugely influence the efficiency of branching. Compared with our initial, standard 3T-linker oligonucleotide (oligo PLI68), the fractional

rate of branching is much improved with 1T and 2T linkers, while no-linker and 4-T linker oligos showed relatively poor performances (no T=0.24, 1T=0.78, 2T=0.46, 3T=0.30, 4T=0.20). These results strongly suggest that oligos with shorter linkers (1T and 2T) enhance the branching reaction by constraining the movement of domain VI in three-dimensional space, thus providing domain VI with higher likelihood of docking correctly on domain IC1. Likewise, a linker with 4T gives too much space flexibility to domain VI and no linker between domain VI and domain IC1 makes it difficult to have a proper docking.

Since the oligo with a 1-T linker has proved optimal for the docking of domain VI, we chose to present experiments with this oligo in a more detailed way (Figure 17). We have found that at a saturating concentration of the 1T-linker oligo, the reaction originated mainly from transesterification, and the ratio of the branching rate over total reaction rate (80%) is again largely improved compared with the 3T-linker oligo PLI-68, for which the maximum ratio of transesterification over total reaction rate was only 60% (Figure 17). The affinity of the oligo was also markedly improved: the K_m of the 1T-linker oligo PLI-74 is no more than 73 nM, compared with 5.4 μM for the Km of the 3T-linker oligo PLI-68. In conclusion, our results show that the best length for the linker is just 1 T, which, combined with what we have learnt from our three-dimensional modeling of the group II ribozyme, makes it possible to locate domain VI in space more precisely.

Mismatched oligos and compensatory mutations

By using mutagenesis and domain IC1-VI tethering experiments, we have reached confidence that domain IC1 is associated with the docking of

domain VI. However, inevitable questions still linger. Could the phenomenon of restoration be specific of certain sequences? Can we exclude the possibility that oligos affect branching by interacting with other sequences within the intron? To clarify this question, we created an additional mutant (PL2-86; Figure 15, construct C). The difference with the previous setups was that domain IC1 of the new mutants was changed to pair with oligo PLI-69. PLI69 (a 15-mer, anti-DVI and anti-IC1 to PL2-84B) was used as a negative control with PL2-84B, its sequence is mismatched with that of PL2-84 in domain IC1 so that the corresponding part of PLI69 does not significantly contribute to the restoration of branching. Moreover, since we already had learnt that oligos with a 1-T linker performed best, we replaced in the following experiment oligo PLI69 by another oligo PLI77 (a 13-mer) with the same pairing sequence but a 1T linker. The analytic methodology used is the same as presented in Figure 17. The results (Figure. 18) showed that PL2-86, with a IC1 sequence matched to that of oligo PLI77 is still able to undergo the same restoration phenomenon displayed by the PL2-84B and PLI-74 combination, although transesterification was not restored quite as well as in the experiment involving PL2-84B and PLI-74. This result proved that the phenomenon we have discovered is robust, i.e. not sequence-specific.

Domain IC1-VI docking by pairing with an RNA bridging molecule (PLI79)

Our experiments in which domains IC1 and VI are brought together by pairing with an oligonucleotide splint were interpreted by proposing that domain VI performs the branch-generating transesterification by docking onto domain IC1. According to our modelling, domain VI most probably docks through a ribose-zipper contact between the stem of domain VI and the two consecutive G:U wobble pairs in domain IC1. While the latter are

still present in our pairing experiment, the use of DNA oligos as chains to pull domain VI closer to its natural receptor in IC1 may be interfering with the postulated RNA-RNA interaction. Also, a DNA-RNA interaction is expected to be weaker than an RNA-RNA interaction, could it be possible to enhance the contact between the two domains by using an RNA oligo?

To address these questions, we attempted additional pairing experiments with the RNA oligo PLI79. PLI79 was designed following PLI74, and is a 13-mer oligo with a 1-T linker and sequence pairing with both domains IC1 and VI. An experiment with PL2-84B and the PLI79 RNA bridging oligonucleotide showed that this RNA oligo indeed shows improved binding between the oligo and precursor RNA: the K_m for the RNA oligo is much lower than for the DNA oligo, it was estimated to be below the 1 nM level. Ironically, such high affinity of the RNA oligo makes it difficult to measure the real K_m if only since successive dilutions of the RNA oligo tend to lead to large errors (data not shown).

Although the K_m for restoration of transesterification is very low, the fraction of branched over total reaction products (Br/Rxn) is lower than with DNA oligo PLI74: only 60% of the reaction is initiated by branching with the RNA oligo in saturating concentration. Furthermore, when a control experiment was performed with mutant PL2-73, in which only domain VI is mutated, and a complementary, 7-mer RNA oligo (PLI82), we found that that RNA oligo was able to reconstruct domain VI much better than the corresponding DNA oligo – almost as well in fact as oligo PLI79. However, the K_m is much higher than for oligo PLI79, which we calculated to be around 57 nM. This phenomenon can be explained indeed by that RNA oligo having better affinity for the RNA precursor and the paired oligo being able to restore the structure of domain VI more

efficiently. To summarize our results, RNA oligos help to restore the structure of domain VI and the ability to perform transesterification by providing higher RNA-RNA affinity. Basically, these results are not in conflict with our previous hypotheses. However, it was a bit of a disappointment, that an RNA oligo should not seem to provide better restoration.

Verification of the branchpoint location and splice junctions

Group II introns perform self-splicing through two-step transesterifications. During the first step, the 2' hydroxyl of a bulged adenosine in domain VI attacks the 5' splice site, followed by nucleophilic attack on the 3' splice site by the 3' OH of the upstream exon. In our experiments, the sequences of domains IC1 and VI of the precursor were modified to pair with an additional oligonucleotide. Although we had every reason to believe that the branching reaction of our precursor was achieved through transesterification by using the stem of domain VI to interact with domain IC1, the question remained whether the branching reaction was truly authentic or an artefact. Specifically, we needed proof that the bulged A was contacting the 5' extremity of the intron correctly as in the wild type.

In order to confirm that the group II intron precursor performs the transesterification reaction with the same mechanism as the wild type, an RT-PCR and sequencing method was applied to the isolated putative lariat. This method has been used previously to verify the splicing mechanism *in vivo* (Vogel and Börner, 2002). During the first step of splicing, the bulged A of domain VI gets connected to the first nucleotide of intron to generate a lariat. By using a primer located downstream of the intron 5' extremity, the reverse transcriptase is given the opportunity to walk

occasionally past the connecting site between the bulged A and the 5' intron extremity. These rare events can be selected by PCR with a set of primers specific to the sequence of the lariat – the reverse primer is close to domain IC1, the forward primer is near to domain IIIC (Table 5). The RT-PCR products were further purified and subcloned into a pGEM cloning vector (Promega), and the precise connecting sites of individual clones were then checked by DNA sequencing. The detailed procedure is described in Materials and Methods.

An experiment in which lariat was generated by mixing the PL2-84B precursor with a saturating concentration of the 1-T linker oligo PLI74 was selected to verify that the mutant still uses the bulged A to attack the bond at the 5' extremity of intron. Primer 22299, located close to domain IIIC was used as forward primer and primer 7118 near domain IC1 was used as reverse primer. After the procedure described here above, 5 clones were picked and sequenced: In 3 out of 5 clones, the bulged A appeared to be replaced by T and to be connected with the 5'-terminal sequence of the intron (GTGCG, Table 3). This result is fully convincing, when compared with the reference method first published in 2002 (Vogel and Börner, 2002): the bulged A is read most frequently as a T indeed, but only 60%-90% of the clones show the correct connection, presumably because of errors during bypassing of the branch by the RT. Although we did not test additional clones, we believe that this result is solid and clear. Meanwhile, ligated exons of PL2-84B were also purified and analysed by sequencing: the two exons were correctly ligated in all 5 clones, which provides proof that a reconstituted domain VI can both ensure 3' exon recognition and cleave the 5' intron extremity precisely.

Part III: General Discussion

EBS2 does not appear to be involved in domain VI docking

In our phylogenetic analyses, introns with a 5' terminal insert were carefully compared with their close relatives without an insert. The first surprise came from the fact that not only all introns with a 5' terminal insert but several of those without one lack both EBS2 and β' at the same time. As we discussed in the previous section, the presence of EBS2 does not seem obligatory for an intron to recognize the 5' exon. In fact, in group IIC introns and in some group IIB introns, lack of EBS2 is well reported. For example, in the SSU778 intron clade (Figure 1), introns without an insert like the ones in *Cryphonectria parasitica, Coccidioides sp.* and *Ascosphaera apis* also lack EBS2 and β', but their domain VI structure is just as intact as in the rest of normal introns (not only do they have a bulged A but also the AAA:CUA internal loop). These facts imply that their splicing reaction is mainly through the transesterification pathway. Interestingly, those introns that lack EBS2 and β' also comprise coding sequences for endonucleases of either the LAGLIDADG or GIY..YIG families. Since introns carrying coding sequences for a DNA endonuclease rather than a reverse transcriptase are hypothesized to undergo homing through a DNA-transposon-like mechanism, such a coincidence makes us wonder the possibility that the EBS2 region may contribute some function to reverse transcription/integration of the intron. In our EBS2 deletion experiments, we found that removal of the EBS2 region only slight affects the fraction of branched over hydrolysis products and even increases the ratio of k_{br}/k_{hy}. Therefore, we tend to exclude the possibility that subdomain ID2 could be a receptor for domain VI; the observed shift from branching to hydrolysis seems more likely to

reflect some involvement in structural stability and exon recognition. One indication in favor of this idea is that the EBS2-lacking precursor also generates an additional fragment, which may come from the linear intron (this molecule is slightly shorter than the linear intron and has been verified to result from intramolecular cleavage; data not shown).

The investigation of domain VI

As we verified by assaying the *in vitro* self-splicing potential of the *Pycnoporellus fulgens* SSU788 intron, introns with a 5' terminal insert have lost the ability to carry out branching. As a consequence, sequences and structures specifically related to the branching process should get lost in the absence of natural selection. By comparing the secondary structure of introns without and with 5'-terminal insertions, it is made apparent that the 3-bp helix and well-conserved 6-nt internal loop of introns of the former set are missing or unrecognizable in members of the latter set, whereas the basal and distal sections of domain VI remain well conserved. This indicates that not only the bulged A and its two flanking base pairs are necessary for proper branching, but the entire middle part of domain VI must also be involved in branching. This stands in partial contradiction to previous studies that have been published on domain VI. In a paper reporting mutagenesis of domain VI, it was concluded that only the bulged A and the two wobble pairs that flank it are necessary for branching (Chu et al., 1998). This conclusion was somewhat hasty, for it was based on sealing of the internal loop by pairing of its nucleotides through canonical base pairs. In fact, most bacterial members of subclass IIB1 lack an internal loop in their distal dVI stem, even though they share tandem IC1 G:U pairs with their mitochondrial counterparts. Unsurprisingly, in our study as well, replacement of the internal loop with a continuous helix shows that intron Pl.LSU/2 is still able to carry out

branching with rather high efficiency. We went further and proved the importance of the middle part of domain VI by manipulating the length of the stem distal to the branchpoint. While only when domain VI was trimmed down to two base pair, is the branching reaction obviously blocked with 1 M-ammonium salt (stems trimmed to 7 bp and 4 bp were only slightly hampered), loss of the ability to perform branching is gradual instead in 1 M KCl (see Table I of Article #2).

The probable IC1 receptor of domain VI

The Group II intron ribozyme has been assumed to exist in two conformations ever since Chanfreau and Jacquier reported the existence of the η-η' tertiary interaction between domain VI and domain II. The η-η' tertiary interaction is believed to be responsible for removing domain VI from the catalytic site after the first step of splicing, thus liberating the space for the 3' exon in the second-step transesterification. In contrast to the identification of η-η', the search for interactions specific for first-step transesterification proved disappointing. Only in 2006 did Hamill and Pyle propose a candidate receptor for the branchpoint and neighboring nucleotides by using UV crosslinking: this receptor was proposed to be located in a subdomain ID internal loop, which was designated as the "coordination loop". However, the site is not conserved in subgroup IIA introns, which triggered an unsolved debate.

In our studies, another possible receptor, located in subdomain IC1 (Figure 4), is proposed for the first time. Evidence in favor of this site comes not only from our phylogenetic alignment and analysis of introns with and without 5' terminal inserts, but also from our nucleotide substitution experiments. Mutagenesis of the two G:U pairs in the IC1 distal helix (positions 78, 79, 100 and 101 of the Pl.LSU/2 intron) hugely

shifts the preferred pathway from transesterification to hydrolysis (Figure 12-2). Interestingly, this position happened to be among those whose potential importance for branching was uncovered in a NAIM (Nucleotide Analog Interference Mapping; Strobel, 1999) experiment on the subgroup IIB1 ai5γ ribozyme (Boudvillain and Pyle.,1998). Removal of the NH_2 at position 2 of G79 (Pl.LSU/2 numbering) and also of the 2'OH groups of U78 and U100 was reported to interfere with activity.

Additional evidence in favor of our proposed receptor was obtained by creating a system in which DNA oligonucleotides are used as bridging molecules that restore the branching ability of an intron which can only perform a hydrolysis default reaction by itself. This experiment successfully proved that only when domain VI interacts with domain IC1, is it possible for the branching reaction to be initiated. Our newly identified receptor also provides long-sought evidence that domain VI truly undergoes a major translocation before the ligation of the two exons. Furthermore, by manipulating the number of linker elements between the pairing sections of the bridging oligonucleotide, we were able to narrowly constrain the movement of domain VI in the space, which helped us in turn to define more precisely the position of domain VI during the docking event.

Bimolecular reaction system

To prove the connection between domains VI and IC1, the first method that came to our mind was using a separate domain IC1 to generate a two-piece bimolecular reaction *in trans*. A similar setup had been successfully used to explore the tertiary interaction θ-θ': the separate small piece formed by domain IC1 can act as an enzyme to activate catalysis by complementary, pre-folded intron Sc.cox1/5. Unfortunately, we failed to

reproduce the phenomenon with Pl.LSU/2: catalysis is not completely blocked in the absence of domain IC1 and this difference may reflect structural differences between the two introns. However, we believe that such a setup may still be feasible with the Pl.LSU/2 intron. For example, an excess of 5' exon could be added to the reaction to increase the stability of the intron at low magnesium concentrations through the IBS1-EBS1 interaction between the 5' exon and domain I. Another way to achieve this goal could be to introduce an extra subdomain in domain ID, providing Pl.LSU/2 with the missing β-β' interaction. The β-β' interaction is believed to help stabilize the structure of the intron when the precursor starts to fold by using domain I as a scaffold. Unfortunately, both methods are not guaranteed and testing them is time consuming; we decided to give up this project because of the lack of time.

Otherwise, another possibility is to use domains V and VI as a separate molecule in order to perform a bimolecular reaction, something which has long been known to work in many group II introns, including Pl.LSU/2. The only drawback is that domain V-VI (domain VI alone is not recognized in a productive manner by the rest of the molecule) is much larger than domain IC1 alone; first, interactions between domain V and domain I are likely to dominate the system; and second, the cost for RNA synthesis and atom substitution is also largely increased. Of course, a major advantage of bimolecular systems is that single RNA nucleotides can be manipulated in detail; here single atoms on domains IC1 or DVI could be altered and tested, which would greatly help building an atomic-resolution model of the first-step interaction of these two domains.

Three-dimensional modelling of the position of domain VI

A great step forward in the RNA splicing field was the publication (Toor et al., 2008) of the first crystallographic model of a group II intron, the

group IIC intron of *Oceanobacillus iheyensis* (*O. iheyensis*). However, this major achievement left people puzzling about the potential first-step receptor of domain VI, because the structure of domain VI, probably due to its instability, could not be visualized in the crystal structure. More recently, it has been speculated that domain VI would probably lie within a crevice formed between the IC helix and the coordination loop substructure (Pyle. 2010). The latter hypothesis is still different from our newly proposed receptor, which stands also in contradiction to previous UV cross-linking results from the same laboratory (Hamill and Pyle, 2006). Current models of the first-step configuration of domain VI are all based on these UV cross-linking data, which are questionable, as we discussed before. Although our proposition seems to be standing against mainstream thinking, our model has been tested by mutagenesis and the brigding oligo experiments support our point of view with particularly strong evidence.

The brigding oligo system was refined by changing the numbers of T's in the linker between the sequences complementary to the engineered IC1 and dVI stems. Because restoration of the branching reaction results from pulling domain VI closer to its natural docking site by using the brigding oligo as a string, the shorter the length of the chain (numbers of links) the more restricted will be the flexibility of domain VI in space. We found that shortening the linker does increase the probability of domain VI docking into its receptor in domain IC1. However, there is a limit to this, and the branching ratio abruptly falls down again when going from a 1-T linker to a no-linker oligo. The best explanation for this is that attempting to pair a precursor RNA with a no linker oligo does not leave sufficient space between domain VI and IC1, which would clash unless the pairing between the precursor and oligo is disrupted by at least one base pair; loss

of one or more base pairs results in turn in the K_d of the oligo and precursor becoming much higher (see Materials and Methods and Fig. 4B of Article #2). In fact, in preliminary tests of the same setup with a five-nucleotide pairing between the oligo and precursor, we found that the Km of the oligo was some 100-fold higher than for a six-nucleotide pairing, which readily explains the poor branching reaction with the no-linker bridging oligo. That the setup with a 1-T linker should work best is important, because this is the optimal distance for domain VI to interact with domain IC1. This finding helped us to place domain VI during the first step of splicing by computer modelling.

By compiling all available data, we have attempted to model the missing domain VI into the latest atomic-resolution models (Toor et al., 2010) of the *Oceanobacillus* group II ribozyme. In our model (Figure 21), we chose to have domain VI as a continuous helix, despite the presence of a very well conserved internal loop, which is closed in our experimental setup. As shown in the Figure, the 5' backbone of this continuous domain VI distal helix fits neatly into the shallow groove of the IC1 stem, and domain VI specifically contacts the section of IC1 encompassing the G79:U100 base pair. The interaction between domains VI and IC1 is probably a kind of ribose-zipper interaction: this consists in a series of hydrogen bonds involving the riboses and the shallow/minor groove edges of several consecutive nucleotides. Compared to most other interactions that exist in group II, it is relatively weak and unstable. Another ribose-zipper interaction recently reported in a group IIC intron is the ω−ω' interaction discovered by X-ray crystallographic analysis (Toor et al., 2008). The flexibility of the ribose-zipper interaction explains why this type of contact is so difficult to bring to light.

The bridging oligonucleotides anchoring model is providing new insights for group II research

By designing specific oligonucleotides that anchor domain VI to its novel binding site, we are now able to activate the branching process over a default hydrolysis reaction of the group II intron. This is truly a breakthrough for exploring the group II intron structure during first-step transesterification, because our results imply the possibility to lock the group II intron into a stable first-step comformation by stabilizing the connection between precursor and oligo. The locked molecule can then be further probed with biochemical and biophysical methods. Although we have yet failed to completely lock the molecule into a first-step ground state (The K_d values of our DNA bridging oligo are still too high), it should be possible to achieve this goal by using higher affinity RNA oligonucleotides. In fact, a preliminary test with an RNA oligo confirmed that RNA-RNA interactions provide higher affinity, and a better configuration of the molecule should still improve the performance of the system. This novel approach to RNA engineering might even make it possible to obtain crystals and visualize at last the ribozyme branchpoint and its molecular context at atomic resolution.

MATERIALS AND METHODS

Sequence analyses of mitochondrial subgroup IIB1 introns

Published sequences of mitochondrial introns that possessed characteristic sequence and secondary structure features of subgroup IIB1 (Michel et al., 1989) were collected (Table 1). Computation of the phylogenetic tree was done by PAUP* 4.0b10 (Swofford 2002), the EBS1, EBS2 and EBS3 sites were removed from the alignment in order to avoid biasing the tree-building procedure in favor of subsets constituted by introns that share homologous insertion sites.

Sequencing and cloning of fungal introns

The group II introns of interest were amplified from DNA extracted from *Grifola frondosa* and *Pycnoporellus fulgens* by PCR and cloned in pUC19 by G. Bassi and M. Costa. The procedure was as follows: PCR amplifications of the SSU788 intron and surrounding exons were performed in 50 µl with 1 µM primers BMS65MOD and BMS103E (Table 4.) using 1 unit of high-fidelity Phusion polymerase in HF buffer (Finnzymes) and 33 cycles (10 s at 98°C, 45 s at 60°C, 90 s at 72°C). Sequencing of amplification products was carried out on both strands by GATC Biotech using the same primers as well as species-specific primers listed in Table 4. Accession numbers for assembled sequences are FR773978, FR773979 and FR773980. For cloning into *E. coli*, amplification products were reamplified with primers BMS65MODT7 and BMS103EZ, digested with BamHI and XmaI and ligated into the pUC19 vector plasmid.

For deletion of ORF sequences from ribozyme domain IV of the *G. frondosa* and *P. fulgens* introns, primers GRXHOREV (or PYXHOREV) and GRXHOFWD (or PYXHOFWD) (Table 4) were used in combination with vector-specific primers ANT7 and 24mer, respectively, in order to generate PCR products. These products were digested with XhoI and either BamHI or XmaI, and cloned back into pUC19. The resulting constructs, pUC19-GR1ΔORF and pUC19-PY1ΔORF, in which most of domain IV has been replaced by a XhoI site (Fig. 3), were verified by sequencing.

DNA construct of Pl.LSU/2 used in this study

The Pl.LSU/2 constructs used in this study originate from Costa et al. (1997) with certain modifications. In brief, intron 2, the last 50 nt of its 5' exon, and the first 71 nt of its 3' exon in the mitochondrial LSU rRNA gene of *P. littoralis* were subcloned into the HindIII site of plasmid pBluescript II KS (-) (Stratagene) in the right orientation for transcription of the intron from the T7 promoter of the vector. Domain IV is largely removed, the section extending from gene positions 4554 to 6361 was replaced by CCTAGGATCT: the resulting domain IV terminal loop is 56 nt long. A series of further modified precursors were generated by PCR induced mutagenesis. The oligodeoxyribonucleotides used in this work were chemically synthesized on an Applied Biosystems 392 DNA/RNA and are listed in Table 5.

Deletion of EBS2

Deletion was carried out by a PCR which used PLI12 in the antisense orientation and PLI13 in the sense direction. The PLI13 sequence extends

across the EBS2-containing loop and is able to pair with PLI12. PLI12 and PLI13 were amplified separately with other primers corresponding to sequences at the 5' and 3' ends of the insert. Amplified fragments were digested and subcloned into pBluescript II KS (-) by a 3-piece ligation. In the resulting molecule, domain ID2 was eliminated (PL2_54). To maintain the flexibility of the ID2-deleted stem, another construct was created in this experiement by using oligo PLI14 instead of PLI13. The extra W (A or U) residue between domains ID and ID3 should give flexibilty to domain ID3 in constructs PL2_55-1 and PL2_55-2 (see Figure 9-1).

Mutation of domain IC1

Two G:U wobble pairs in domain IC1 were changed to two A:U pair in construct PL2_56, Nucleotide substitutions were introduced in the IC1 loop by using mutagenic oligonucleotides as deletion of domain IC1. The DNA amplicons for synthesis of the mutated IC1 transcripts were obtained by PCR, with oligonucleotides PLI19 (sense) and CGM8350 on the one hand, 24mer and PLI20 (reverse) on the other. Both amplicons contain a BsaI site at their 5' end. After restriction enzyme digestion, fragments were able to pair with each other and were subcloned into pBluescript II KS (-) by a 3-piece ligation.

Mutation of D6

The PL2_57 (DVI-2bp) construct was generated by PCR amplification by oligonucleotides PLI21, containing a BsaI site, and 24mer, at the insert 5' end. The PLI21-containing sequence deletes part of DVI and replaces the deleted part with a UUCG loop. Domain VI was replaced by a continous

helix designed after domain VI of *Pseudomonas spp* (PL2_58 ; DVI-stem). The method was similar to that used to create PL2_56, with oligonucleotides PLI22 (antisense) and PLI23 (sense) as primers. Generated single mutants were then used to create double mutants PL2_59 and PL2_60. The same method was applied to generate other mutants and combination mutants, the corresponding oligos used for mutagenesis are listed in Table 5.

DNA construct for bimolecular experiment

Deletion of the IC1 subdomain was carried out by using oligonucleotides PLI_34 and PLI_35 (in the antisense orientation). In the resulting molecule, intron positions 78 to 123 are replaced by the sequence TTAA (Pl2-67). The DNA matrices for synthesis of the IC1 "enzyme" transcripts were obtained by PCR, with oligonucleotides PLI_36, which contains the promoter sequence of the T7 promoter sequence and a XbaI restriction site, and PLI_37, which is used as a reverse primer containing two restriction sites, EcoRI and BsaI, fused at the 5' end. The PCR product was then digested and cloned into pUC19 with EcoRI and XbaI sites (PUCIC1-1; PUCIC1-2). For synthesis of the IC1 enzyme transcripts, plasmids were digested with BsaI to generate a homogenous 3' extremity.

In vitro transcription and purification of Pl.LSU/2 precursor RNA

Transcription of RNA is performed with the T7 RNA polymerase, Pl.LSU/2 precursor RNAs were obtained using plasmid DNA as the template in a transcription reaction. Templates were generated from the corresponding plasmids by linearization with Acc65 I prior to the transcription reaction. After the restriction digestion, the linearized

plasmid was extracted with phenol:chloroform:isoamyl alcohol, ethanol precipitated and suspended in pure water before the DNA was used for in vitro transcription reactions.

RNA synthesis and purification were carried out as described by Costa & Michel (1995). Basically, Pl.LSU/2 constructs were transcribed under the following conditions: 40 mM Tris-HCl pH 7.9, 3 mM spermidine, 0.1% Triton X-100, 50 mM DTT, 5 mM rNTPs (except 2.5 mM rUTP   and addition of 2.5 mM α-^{32}P radioactively labeled rUTP); 26 mM MgCl$_2$, ensuring a 0.66 molar concentration ratio of free magnesium over nucleotides was used during transcription in order to prevent intron splicing. DNA template is removed by the addition of 25 U RNase-free DNase I and incubation for 30 min at 37°C. The reaction was stopped by adding 1/10 volume of 0.5 M EDTA. The newly synthesized RNA may be effectively separated from unincorporated nucleotides by size-exclusion chromatography through a small Sephadex G-25 Prepacked column (Amersham Biosciences) in water. Samples were then mixed with RNA loading buffer (formamide, containing 40 mM EDTA) and purified in denaturing 4% (w/v) polyacrylamide gels: intact linear intron was separated form precursor products and intron-3' exon lariat intermediates. Samples were transferred to MES Buffer (pH = 6.2) by over night passive difussion, and finally adjusted to the desired volume by ethanol preciptation and resuspension.

In vitro transcription for *Grifola frondosa* and *Pycnoporellus fulgens* constructs

Transcription of RNA was performed with the T7 RNA polymerase, templates for synthesis of precursor RNA from pUC19-GR1ΔORF and pUC19-PY1ΔORF precursors were obtained by digestion with SmaI.

After restriction digestion, the linearized plasmid was extracted with phenol:chloroform:isoamyl alcohol, ethanol precipitation and suspended in pure water before being used for in vitro transcription reactions. RNA synthesis and purification were performed as described in Costa et al. (1995), except that transcription was carried out under a 0.55 molar concentration ratio of magnesium over nucleotides and 10% DMSO was added in order to prevent intron splicing and secondary structure stopping of RNA transcription.

Self-splicing reactions

For self-splicing experiments, all precursor transcripts were internally labelled by transcription in presence of ^{32}P-UTP. Concentrations of precursor RNA were routinely set at 20 nM. Reactions were initiated by adding 2X-concentration splicing buffer to RNA samples that had been incubated in water at the reaction temperature. In addition to $MgCl_2$ and monovalent cation salts (condition varied by requirement), all splicing buffers used in this work contained 40 mM Tris-HCl (pH 7.5 at 25°C) and 0.02% (w/v) SDS. Reactions were stopped by addition of an equal volume of formamide loading solution with 120 mM Na_2EDTA added. Samples were heated at 40°C for 10 min before being loaded onto 8 M urea/4% polyacrylamide gels. Products were quantified on fixed and dried gels with a PhosphorImager (Molecular Dynamics). The fraction of unreacted presursor molecules was determined from the molar contribution of all intron-containing forms.

Quantification applied in the presence of oligonucleotides (bridging oligo)

Accumulation of branched and linear intron products was fitted to simple exponentials: $[Lar] = [Lar]_\infty (1 - \exp(-k_{br}.t))$ and $[Lin] = [Lin]_\infty (1 - \exp(-k_{hy}.t))$, where $[Lar]$ and $[Lin]$ are the molar fractions of branched and linear molecules at time t, $[Lar]_\infty$ and $[Lin]_\infty$, the corresponding, estimated final values, and k_{br} and k_{hy}, the observed rate constants for branching and hydrolysis.

For reactions in the presence of an oligonucleotide (Sigma), the latter was added to concentrated splicing buffer (final concentrations: 40 mM Tris-HCl pH 7.5 at 25°C, 1M NH_4Cl, 10 mM $MgCl_2$, 0.02% sodium dodecyl sulfate) prior to mixing with the solution of purified precursor molecules (final molar concentration 20 nM) at reaction temperature (37°C). Reaction time courses were modelled according to the following scheme, in which pre:oligo is the unreacted complex between a precursor and an oligonucleotide molecule (whereas hydrolysis at the 5' splice site is irreversible, transesterification is expected to be reversible; however, the intron-3'exon lariat intermediate was either absent or barely detectable, even at short reaction times, for all construct and oligonucleotide combinations we tested, so that in this experimental system, branching may be regarded as irreversible for all practical purposes).

Provided k_{off} and k_{on} are much larger than the rate constants for reactions, the rates of formation of lariat and linear intron products become:

$$d[Lar]/dt = [Pre] (k_{br,U} + k_{br,B}.[OLI]/K_d) \qquad (1)$$

$$d[Lin]/dt = [Pre] (k_{hy,U} + k_{hy,B}.[Oli]/K_d) \qquad (2)$$

where [Pre] is the molar fraction of unbound precursor molecules at time t; $k_{br,U}$, $k_{hy,U}$, $k_{br,B}$ and $k_{hy,B}$ are rate constants for branching and hydrolysis in the absence and presence of the oligonucleotide, respectively; $K_d = k_{off}/k_{on}$; and [OLI] is the molar concentration of oligonucleotide. Let f be the fractional (relative) rate of formation of lariat intron:

$$f = (d[Lar]/dt)/(d[Lar]/dt + d[Lin]/dt)$$

$$= f_0 + (f_{max} - f_0) / (1 + K_m/[OLI]) \quad (3) \qquad \text{with}$$

$$f_0 = k_{br,U}/(k_{br,U} + k_{hy,U}) \quad (4)$$

$$f_{max} = k_{br,B}/(k_{br,B} + k_{hy,B}) \quad (5)$$

$$K_m = K_d (k_{hy,U}/ k_{br,B}) (f_{max}/(1 - f_0)) \quad (6)$$

In practice, (i) the accumulation of lariat and linear intron forms for a given oligonucleotide concentration was fitted to a simple exponential or, exceptionally, when reaction was both slow and limited, to a linear function; (ii) initial rates at t = 0 were obtained from these fits, f was calculated and plotted as a function of oligonucleotide concentration (the relative error of f was estimated by adding the relative errors of branching and total reaction rates, which were calculated from standard errors associated with initial rates); (iii) the resulting plot was fitted with equation (3) to determine f_0, f_{max} and K_m; (iv) K_d was extracted from equation (6) after $k_{hy,U}$ and $k_{br,B}$ had been obtained from initial reaction rates in the absence and at saturating concentrations of the oligonucleotide, respectively.

Reverse transcription and identification of reacted products (fungal introns)

The lariat and ligated exons of GR1ΔORF and the linear intron, unknown fragment with additional cut (AC) and ligated exons of PY1ΔORF were generated under 1 M NH$_4$Cl, 20 mM Mg2+, 42°C, 40 min incubation. The lariat-3' exon intermediate molecule of GR1ΔORF was isolated additionally from a splicing reaction including 20 mM CaCl$_2$. Molecules destined for reverse transcription and primer extension were prepared under denaturing condition by loading reacted samples onto 1.5 mm, 8 M urea/4% (w/v) polyacrylamide gel. Purification of splicing products from preparative denaturing polyacrylamide gels and their reverse transcription with ^{32}P-labelled oligonucleotides were performed essentially as described before. Oligouncleotides used for primer extension – BMS103B, Grifo-Rev2 and Pynco-Rev2 (see Table 4) – were kinased and 5'-end labelled with γ^{32}P-ATP. BMS103B was used to sequence ligated exons and determine the branch-point of the *G. frondosa* intron-3'exon lariat; Gr-R2, to determine the 5' splice of the *G. frondosa* intron lariat; Py-R2, to determine the 5' extremity of *P. fulgens* linear intron molecules. After gel purification from 1.5 mm 20% acrylamide-urea denaturing gel, the labelled oligonucleotides were ready to use for primer extension. For the linear intron (I) and fragment with additional cut (AC) of PY1ΔORF, those purified molecules were tagged by T4 RNA ligase with a gel-purified RNA transcript (5'-GGGAAAGCUUUUAUCUUUGACACAAAAUCGGGGGUGGCGAC UGUUUAUUAAAAAGUGCGACAAGAAGUU; this transcript had been dephosphorylated and then kinased so that it carried a single phosphate at its 5' end, see Ferat et al., 2003). The tagged AC and I fragments were reverse transcribed with oligo 18873 (5'-ACCAGATCTAGATTTTTAATAAACAGTCGCCAC) and PCR

amplified with 18873 and PYCHOFOR. The PCR products were gel purified and subcloned into pGEM-TA cloning vector (Promega). Plasmids including the PCR fragment were purified and sequenced by the GATC sequencing center.

Verification of splice junctions and the branchpoint (Pl.LSU/2 constructs)

The lariat molecule and ligated exons to be used for reverse transcription were generated from PL2-84B in the presence of oligonucletide PLI74. Samples were purified by gel purification under denaturing conditions and dissoved in water. For the lariat, the annealing reaction (10 μl) contained 25 μM of RNA and 10 μM primer 7118 (Rev) 5'-GCAGGTACATTGTCTCCAGA (complementary to intron positions 58-77) in 50 mM Tris-HCl (pH 8.3) and 50 mM NaCl. The reactions were heated at 80°C for two minutes prior to slow cooling to 30°C. For elongation reactions (in a volume of 50 μl), portions (10 μl) of the annealing reactions were mixed at room temperature with 5 μl of a solution containing 400 mM Tris-HCl (pH 8.3), 500 mM NaCl, 100 mM DTT, 30 mM MgCl2, 5 μl of dNTP mix (5 mM each) and 2 μl SuperScriptP RNase H reverse transcriptase (Gibco-BRL) and incubated at 50°C for 45 min. The reacted product was taken for PCR amplification with DNA primer 7118 (reverse) and primer 22299 (forward: 5'-GAAAGGCTGCAGACTTATTA, corresponding to part of ribozyme domain III). The 290 nt PCR product was gel purified from 2% NuSieve agarose and cloned into pGEM TA cloning vector (Promega). 5 clones with insert were sent for DNA sequencing. Ligated exons were reverse transcribed with primer 5'-GAGGTCGACGGTATCGATAA (which matches positions 70-89 of the 3' exon). PCR amplification was carried out with the same primer and 5'-

AGCTTTTATCTTTGACACAAAATCGGGGGTG (positions -19 to -49 of the 5' exon) and products cloned with the pGEM TA cloning vector (Promega): all clones examined had the expected sequence for the ligated exons.

REFERENCES

Bassi, G. S., D. M. de Oliveira, et al. (2002). "Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA." Proc Natl Acad Sci U S A 99(1): 128-33.

Bassi, G. S. and K. M. Weeks (2003). "Kinetic and thermodynamic framework for assembly of the six-component bI3 group I intron ribonucleoprotein catalyst." Biochemistry 42(33): 9980-8.

Belhocine K, Yam KK, Cousineau B (2005) Conjugative transfer of the Lactococcus lactis chromosomal sex factor promotes dissemination of the Ll.LtrB group II intron. J Bacteriol. 187(3):930-9.

Boudvillain, M. and A. M. Pyle (1998). "Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis." EMBO J 17(23): 7091-104.

Boudvillain M, de Lencastre A, Pyle AM (2000) A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. Nature. 406(6793):315-8.

Boulanger, S. C., P. H. Faix, et al. (1996). "Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3' splice site selection." Mol Cell Biol 16(10): 5896-904.

Brion P, Westhof E (1997). "Hierarchy and dynamics of RNA folding". Annu Rev Biophys Biomol Struct 26: 113–37.

Cech TR.(2000). "Structural biology. The ribosome is a ribozyme". Scienc 289(5481):878-9.

Chanfreau G, Jacquier A (1994) Catalytic site components common to both splicing steps of a group II intron. Science. 266(5189):1383-7.

Chanfreau G, Jacquier A (1996) An RNA conformational change between the two chemical steps of group II self-splicing. EMBO J. 15(13):3466-76.

Chin K, Pyle AM (1995) Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. RNA. 1(4):391-406.

Chu, V. T., Q. Liu, et al. (1998). "More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns." RNA 4(10): 1186-202.

Chu, V. T., C. Adamidi, et al. (2001). "Control of branch-site choice by a group II intron." EMBO J 20(23): 6866-76.

Copertino DW, Hallick RB (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. Trends Biochem Sci. 18(12):467-71.

Costa, M. and F. Michel (1995). "Frequent use of the same tertiary motif by self-folding RNAs." EMBO J 14(6): 1276-85.

Costa M, Deme E, Jacquier A, Michel F (1997) Multiple tertiary interactions involving domain II of group II self-splicing introns. J Mol Biol. 267(3):520-36.

Costa M, Fontaine JM, Loiseaux-de Goer S, Michel F (1997) A group II self-splicing intron from the brown alga Pylaiella littoralis is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. J Mol Biol. 274(3):353-64.

Costa M, Michel F, Westhof E (2000) A three-dimensional perspective on exon binding by a group II self-splicing intron. EMBO J. 19(18):5007-18.

Costa M, Michel F, Toro N (2006) Potential for alternative intron-exon pairings in group II intron RmInt1 from Sinorhizobium meliloti and its relatives. RNA. 12(3):338-41.

Cousineau, B., D. Smith, et al. (1998). "Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination." Cell 94(4): 451-62.

Dai, L., D. Chai, et al. (2008). "A three-dimensional model of a group II intron RNA and its interaction with the intron-encoded reverse transcriptase." Mol Cell 30(4): 472-85.

Daniels DL, Michels WJ Jr, Pyle AM (1996) Two competing pathways for self-splicing by group II introns: a quantitative analysis of in vitro reaction rates and products. J Mol Biol. 256(1):31-49.

De la Pena, M. and I. Garcia-Robles (2010) "Ubiquitous presence of the hammerhead ribozyme motif along the tree of life." RNA 16(10): 1943-50.

Edgell, D. R., M. Belfort, et al. (2000). "Barriers to intron promiscuity in bacteria." J Bacteriol 182(19): 5281-9.

Fedorova, O., T. Mitros, et al. (2003). "Domains 2 and 3 interact to form critical elements of the group II intron active site." J Mol Biol 330(2): 197-209.

Fedorova, O. and A. M. Pyle (2005). "Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3." EMBO J 24(22): 3906-16.

Fedorova O, Zingler N. (2007) Group II introns: structure, folding and splicing mechanism. Biol Chem. 388(7):665-78.

Gordon PM, Piccirilli JA (2001) Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. Nat Struct Biol. 8(10):893-8.

Granlund, M., F. Michel, et al. (2001). "Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci." J Bacteriol 183(8): 2560-9.

Hamill S, Pyle AM (2006) The Receptor for Branch-Site Docking within a Group II Intron Active Site. Mol Cell. 23(6):831-40.

Jacquier A, Jacquesson-Breuleux N (1991) Splice site selection and role of the lariat in a group II intron. J Mol Biol. 219(3):415-28.

Jacquier A, Michel F (1987) Multiple exon-binding sites in class II self-splicing introns. Cell. 50(1):17-29.

Jacquier A, Michel F (1990) Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. J Mol Biol. 213(3):437-47.

Jacquier A (1996) Group II introns: elaborate ribozymes. Biochimie. 78(6):474-87.

Jarrell KA, Peebles CL, Dietrich RC, Romiti SL, Perlman PS. （1988）Group II intron self-splicing. Alternative reaction conditions yield novel products.

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM (2001) Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat Biotechnol. 19(12):1162-7.

Koonin EV (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? Biol Direct. 1:22.

Koch JL, Boulanger SC, Dib-Hajj SD, Hebbar SK, Perlman PS (1992) Group II introns deleted for multiple substructures retain self-splicing activity. Mol Cell Biol. 12(5):1950-8.

Lambowitz AM, Zimmerly S. (2004) Mobile group II introns. Annu Rev Genet. 38:1-35.

Lehmann K, Schmidt U (2003) Group II introns: structure and catalytic versatility of large natural ribozymes. Crit Rev Biochem Mol Biol. 38(3):249-303.

Michel F, Umesono K, Ozeki H (1989). Comparative and functional anatomy of group II catalytic introns--a review. Gene. 82(1):5-30.

Michel F, Ferat JL (1995) Structure and activities of group II introns. Annu Rev Biochem. 64:435-61.

Michel, F., M. Costa, et al. (2009). "The ribozyme core of group II introns: a structure in want of partners." Trends Biochem Sci 34(4): 189-99.

Mohr, G., D. Smith, et al. (2000). "Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences." Genes Dev 14(5): 559-73.

Mullineux, S. T., M. Costa, et al. "A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions." RNA 16(9): 1818-31.

Nissen, P., Hansen, J., Ban, N., Moore, P.B., Steitz, T.A. (2000). "The structural basis of ribosome activity in peptide bond synthesis." Science 289(5481):920-30.

Pley, H. W., D. S. Lindes, et al. (1994). "Crystals of a hammerhead ribozyme." J Biol Chem 269(6): 4692.

Podar M, Perlman PS, Padgett RA. (1998) The two steps of group II intron self-splicing are mechanistically distinguishable. RNA. Aug;4(8):890-900.

Podar, M., J. Zhuo, et al. (1998). "Domain 5 binds near a highly conserved dinucleotide in the joiner linking domains 2 and 3 of a group II intron." RNA 4(2): 151-66.

Pyle, A. M., O. Fedorova, et al. (2007). "Folding of group II introns: a model system for large, multidomain RNAs?" Trends Biochem Sci 32(3): 138-45.

Pyle, A. M. "The tertiary structure of group II introns: implications for biological function and evolution." Crit Rev Biochem Mol Biol 45(3): 215-32.

Qin, P. Z. and A. M. Pyle (1998). "The architectural organization and mechanistic function of group II intron structural elements." Curr Opin Struct Biol 8(3): 301-8.

Reiter NJ, Osterman A, Torres-Larios A, Swinger KK, Pan T, Mondragón A. (2010). The structure of the entire RNAseP holoenzyme -tRNA complex has been solved: Nature. 9;468(7325):784-9.

Rest JS, Mindell DP (2003) Retroids in archaea: phylogeny and lateral origins. Mol Biol Evol. 20(7):1134-42.

Roitzsch M, Pyle AM. (2009) The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. RNA. 15(3):473-82. Epub 2009 Jan 23.

San Filippo J, Lambowitz AM (2002) Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. J Mol Biol. 324(5):933-51.

Schmidt, U., M. Podar, et al. (1996). "Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing in vitro and in vivo: phenotypes and suppressor mutations." RNA 2(11): 1161-72.

Sigel, R. K., A. Vaidya, et al. (2000). "Metal ion binding sites in a group II intron core." Nat Struct Biol 7(12): 1111-6.

Sigel RK, Sashital DG, Abramovitz DL, Palmer AG, Butcher SE, Pyle AM (2004) Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif. Nat Struct Mol Biol. 11(2):187-92.

Stoddard BL (2005) Homing endonuclease structure and function. Q Rev Biophys. 38(1):49-95.

Su, L. J., C. Waldsich, et al. (2005). "An obligate intermediate along the slow folding pathway of a group II intron ribozyme." Nucleic Acids Res 33(21): 6674-87.

Suchy M, Schmelzer C (1991) Restoration of the self-splicing activity of a defective group II intron by a small trans-acting RNA. J Mol Biol. 222(2):179-87.

Toor, N., G. Hausner, et al. (2001). "Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases." RNA 7(8): 1142-52.

Toor N, Zimmerly S (2002) Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. RNA. 8(11):1373-7.

Toor N, Robart AR, Christianson J, Zimmerly S (2006) Self-splicing of a group IIC intron: 5' exon recognition and alternative 5' splicing events implicate the stem-loop motif of a transcriptional terminator. Nucleic Acids Res. 34(22):6461-71.

Toor, N., K. S. Keating, et al. (2008). "Crystal structure of a self-spliced group II intron." Science 320(5872): 77-82.

Toor, N., K. S. Keating, et al. (2010). "Tertiary architecture of the Oceanobacillus iheyensis group II intron." RNA 16(1): 57-69.

Vogel J, Borner T (2002) Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. EMBO J. 21(14):3794-3803.

Watanabe, K. and A. M. Lambowitz (2004). "High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA." RNA 10(9): 1433-43.

Wheelan, S. J., Y. Aizawa, et al. (2005). "Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution." Genome Res 15(8): 1073-8.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). "A unified classification system for eukaryotic transposable elements". Nat Rev Genet 8: 973-982.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428: 281–286.).

Xiang, Q., P. Z. Qin, et al. (1998). "Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets." Biochemistry 37(11): 3839-49.

Yang, J., Zimmerly, S., Perlman, P.S., Lambowitz, A.M. (1996). "Efficient integration of an intron RNA into double-stranded DNA by reverse splicing". Nature 381(6580):332-5.

Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., Lambowitz, A.M. (1995). "A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility". Cell 83(4):529-38

Zimmerly S, Moran JV, Perlman PS, Lambowitz AM (1999) Group II intron reverse transcriptase in yeast mitochondria. Stabilization and regulation of reverse transcriptase activity by the intron RNA. J Mol Biol. 289(3):473-90.

Manuscrit de thèse

(section en français)

Structure et réarrangements conformationnels au cours de l'épissage du composant ribozyme d'un intron de groupe II

Etudiant: Cheng-Fang Li[1,2]

Codirecteurs de thèse: François Michel[1]

Jenn Tu[3]

[1]Centre de Génétique Moléculaire du C.N.R.S., 1 Avenue de la Terrasse, 91190 Gif-sur- Yvette, France

[2]Department of Life Science and Institute of Biotechnology, National Tsing Hua University, Hsinchu, Taiwan

[3]Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan

Résumé

Les introns de groupe II forment une classe d'ARN connus avant tout pour leur activité ribozymique, qui leur permet de catalyser leur propre réaction d'épissage. Sous certaines conditions, ces introns peuvent s'exciser des ARN précurseurs dont ils font partie et assurer la ligation des exons qui les bordent sans l'aide d'aucune protéine. Les introns de groupe II sont généralement excisés sous forme d'un lariat, semblable à celui formé par les introns des prémessagers nucléaires, dont l'épissage est assurée par le spliceosome. De telles similarités dans le mécanisme d'épissage suggèrent que les introns de groupe II et les introns des prémessagers nucléaires pourraient avoir un ancêtre évolutif commun.

Malgré leurs séquences très diverses, les introns de groupe II peuvent être définis par une structure secondaire commune, hautement conservée. Celle-ci est formée de six domaines (domaine I à domaine VI ; D1-D6), émergeant d'une roue centrale. L'épissage des introns de groupe II comprend deux étapes, et autant de réactions de transestérification, qui produisent les exons liés et l'intron excisé sous forme lariat. Il est généralement admis que la structure du ribozyme subit des changements conformationnels entre les deux étapes de l'épissage et que le domaine VI est un acteur clé dans ce phénomène. Cependant, malgré l'identification d'un certain nombre d'interactions tertiaires entre domaines, ni la RMN, ni les études faisant appel à des modifications chimiques ne sont parvenues à déterminer l'environnement immédiat, au niveau du site actif du ribozyme, de

l'adénosine qui sert de point de branchement de la structure en lariat, ainsi que des nucléotides qui entourent cette adénosine au sein du domaine VI.

A l'aide d'analyses phylogénétiques et d'une modélisation moléculaire tridimensionnelle, nous avons identifié plusieurs sections du ribozyme susceptibles de constituer le site de fixation du domaine VI au cours de l'étape de branchement. Des mutations ont été introduites dans ces sites de fixation potentiels et la cinétique de réaction des ARN mutants résultants a été déterminée. Afin de démontrer formellement l'interaction du domaine VI avec le site récepteur le plus probable, une molécule de ribozyme dont la réaction de branchement est assurée par l'addition d'oligonucléotides ADN ou ARN qui positionnent correctement le domaine VI vis-à-vis de son partenaire a été construite. En combinant l'information apportée par différentes expériences de ce type, nous avons pu générer un modèle à résolution atomique du complexe formé par le domaine VI, son site de branchement et le reste de l'intron au moment où l'épissage est initié.

Mots-clé: intron de groupe II, structure d'ARN, ribozyme, réarrangements conformationnels d'ARN, point de branchement d'intron.

Introduction

La distribution des ribozymes naturels

La biologie moléculaire s'est développée rapidement depuis la découverte de la structure en double hélice de l'ADN en 1953 par James Watson et Francis Crick. Au cours des années 1980, des molécules d'ARN naturelles capables de catalyser des réactions chimiques ont été trouvées et baptisées 'ribozymes' : cette découverte, qui montrait que les enzymes protéiques n'étaient pas les seules macromolécules biologiques susceptibles de catalyser des réactions chimiques dans les cellules vivantes, a été récompensée en 1989 par l'attribution du Prix Nobel de Chimie à deux chercheurs, Thomas R. Cech et Sidney Altman. Depuis, de nombreuses études ont confirmé que certaines molécules d'ARN sont capables de se structurer en trois dimensions afin d'assurer des fonctions catalytiques en présence de certains cations divalents.

Les ribozymes sont répandus dans la nature, particulièrement chez les plantes, les eucaryotes dits primitifs, les bactéries, et leurs virus. Les ribozymes peuvent être rangés en deux groupes principaux selon leur taille (Tableau S1). Le premier groupe comprend les petits ribozymes, tels que les molécules dites 'en tête en marteau' et en 'épingle à cheveux', l'ARN satellite du virus de l'hépatite delta (HDV, hepatitis delta virus), le ribozyme VS et aussi le ribozyme glmS (Winkler et coll. 2004), découvert plus récemment. Le second groupe inclut de "grands" ribozymes, tels celui de la RNase P, les introns autoépissables de groupe I et de

groupe II et aussi, comme on le sait maintenant, l'ARN de la grande sous-unité du ribosome.

ribozyme	source	size	function	reaction products
self-cleaving RNAs				
hammerhead motif	plant viroids and satellite RNAs, salamander	≈ 40 nt	replication	5'-OH; 2', 3'-cyclic phosphate
hairpin motif	plant satellite RNAs	≈ 60 nt	replication	5'-OH; 2', 3'-cyclic phosphate
HDV	hepatitis delta virus (human)	≈ 80 nt	replication	5'-OH; 2', 3'-cyclic phosphate
VS ribozyme	Neurospora crassa mitochondria	154 nt	replication	5'-OH; 2', 3'-cyclic phosphate
RNase P RNAs	eukaryotes (nucleus, organelles), prokaryotes	140-490 nt	tRNA processing	products with 5'-phosphate and 3'OH
self-splicing RNAs				
group I introns	eukaryotes (nucleus, organelles), prokaryotes, bacteriophages	200-1500 nt	splicing	intron with 5'-guanosin and 3'-OH; 5'/3' ligated exons
group II introns	eukaryotes (organelles), prokaryotes	300-3000 nt	splicing	intron with 2'-5' lariat and 3'-OH; 5'/3' ligated exons

Tableau S1. Liste des différent types de ribozymes.

Données rassemblées par Karola Lehmann et Udo Schmidt (2003)

Structure des introns de groupe II

Les introns de groupe II ont des séquences très diverses; seuls sont bien conservés quelques courts segments de séquence dans le domaine V ainsi que plusieurs nucléotides au début de l'intron. Malgré cette diversité, les introns de groupe II peuvent être définis par une structure secondaire hautement conservée (Figure S1) (Michel et al., 1989; Toor et al., 2001). Un intron de groupe II consiste généralement en six domaines émergeant d'une roue centrale. Chacun de ces six domaines introniques a un rôle spécifique dans le repliement, les réarrangements conformationnels ou la catalyse. De multiples stratégies ont été appliquées pour explorer la structure des introns de groupe II et elles ont largement contribué à étendre notre compréhension du repliement de l'ARN, de sa structure tertiaire, de sa biochimie et de son évolution. Comprendre la structure tridimensionnelle des introns de groupe II reste un problème d'actualité depuis bien des années, car les introns de groupe II constituent le meilleur modèle avec lequel comparer et mieux appréhender les mécanismes et la structure du spliceosome eucaryote.
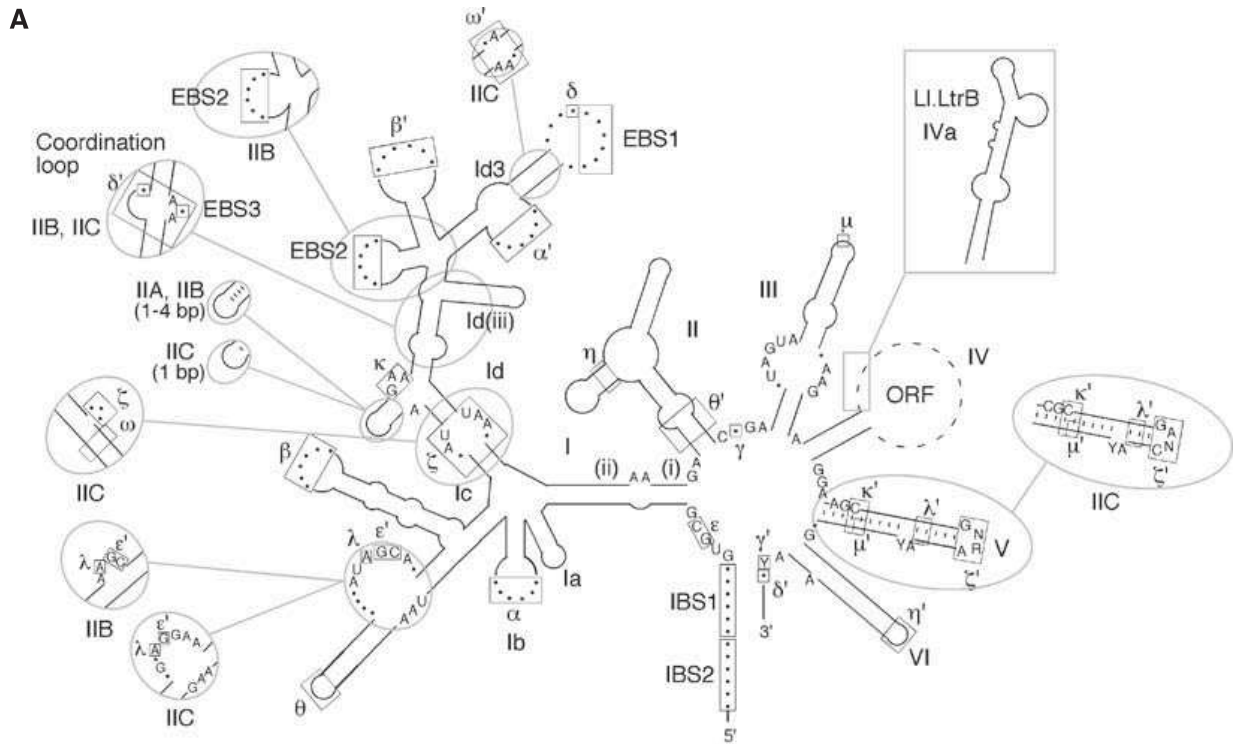
Figure S1. Structure d'un intron bactérien représentatif du sous-groupe IIA1.

Les zones cerclées indiquent les différences avec les sous-groupes IIB et IIC. La "boucle" de DIV, qui comprend la séquence codante de la protéine spécifiée par l'intron (IEP), est représentée par un cercle tireté ; un site de fixation à haute affinité pour l'IEP est montré en insert. Les domaines et sous-domaines étiquetés comprennent ceux discutés dans le texte (Figure tirée de Lambowitz et Zimmerly, 2010)

Deux voies d'épissage pour les introns de groupe II

Initiation de l'épissage par transestérification

Les réactions d'épissage des introns de groupe II sont catalysées par l'ARN intronique lui-même. Pour accomplir cette tâche, l'ARN se replie selon une

structure secondaire et tertiaire conservée et forme un site actif contenant des ions Mg^{2+} essentiels à la catalyse. Les introns de groupe II épissent via deux réactions séquentielles de transestérification qui génèrent les exons liés et l'intron excisé sous forme lariat avec une liaison phosphodiester 2′-5′ (Figure S2). Dans une première étape, l'attaque nucléophile du site d'épissage 5′ par le groupement 2′ OH d'une adénosine protubérante dans DVI a pour conséquence la coupure de la jonction 5′ couplée à la formation de l'intermédiaire de réaction, en forme de lariat. Dans une deuxième étape, l'attaque nucléophile de la jonction 3′ par l'extrémité 3′ OH de l'exon 5′ clivé a pour conséquence la ligation des exons et la libération de l'intron sous forme lariat.

Initiation de l'épissage par hydrolyse

Les premières études *in vitro* de l'épissage des introns de groupe II suggéraient qu'en plus de la voie d'épissage par formation de lariat, l'intron pouvait être excisé par une voie alternative, dans laquelle l'eau ou l'ion hydroxyle est utilisé comme nucléophile lors de la première étape de l'épissage (Jarrell et coll., 1988); Daniels et coll., 1996). La deuxième étape est ensuite la même que dans la voie initiée par transestérification/formation de lariat et les produits de la réaction sont les exons liés et l'intron linéaire (Figure S2). L'équilibre entre épissage par branchement et épissage par hydrolyse est fortement déterminé par le choix du cation monovalent dans le milieu de réaction (Jarrell et coll., 1988).
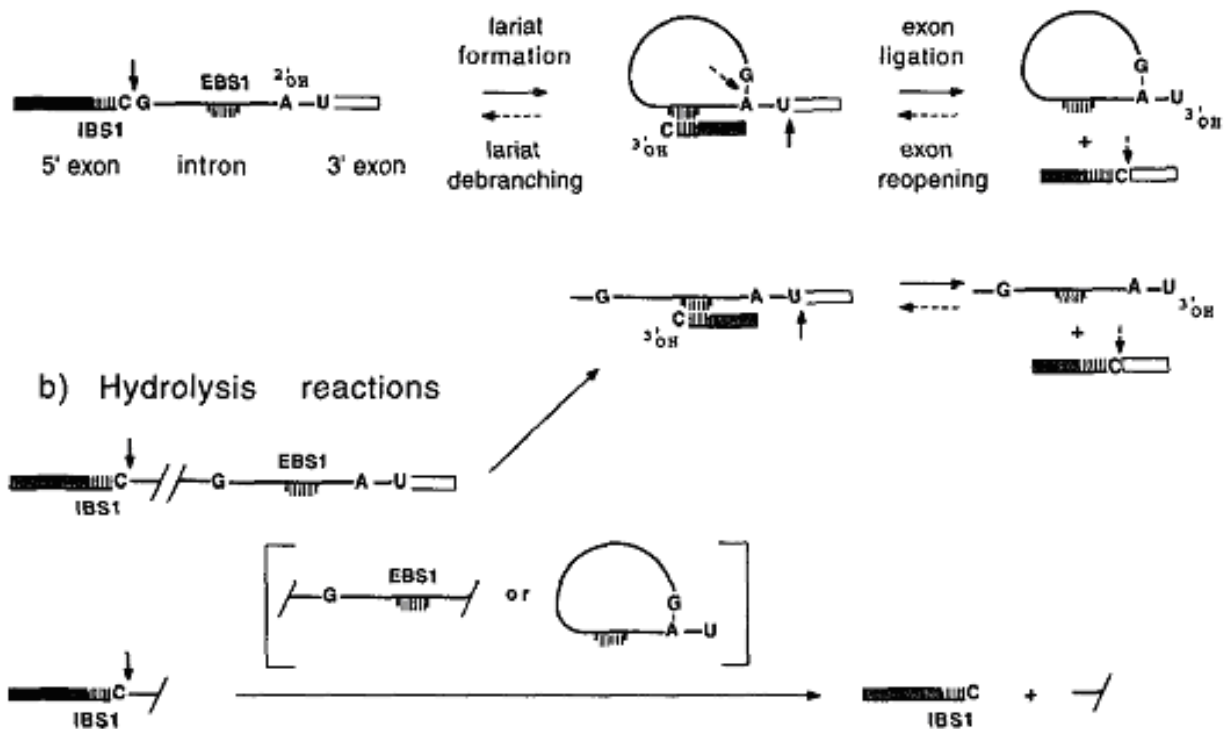
Figure S2. Deux voies d'épissage principales pour les introns de groupe II..

Dans la voie initiée par transestérification, les introns de groupe II épissent via deux réactions séquentielles de transestérification qui génèrent les exons liés et l'intron excisé sous forme lariat, comportant une liaison phosphodiester 2'-5'. Dans la voie initiée par hydrolyse, un ion hydroxyle est utilisé comme nucléophile dans la première étape de l'épissage et l'intron ext excisé sous forme linéaire. Figure tirée de François Michel et Jean-Luc Ferat (1995).

Résultats et Discussion

Introns avec insertions 5'-terminales parmi les introns mitochondriaux de sous-groupe IIB1.

A l'occasion de l'alignement et de l'analyse de séquences d'introns de groupe II d'organelles, notre attention fut attirée par un petit sous-ensemble d'introns qui divergeaient quelque peu de la norme. Un total de 10 introns avaient en effet en commun que l'extrémité de leur exon 5' (définie par comparaison avec des versions ininterrompues du gène-hôte) et la séquence consensus GUGYG qui marque normalement le début d'un intron de groupe II se trouvaient séparées par une insertion, pouvant compter de 1 à 33 nucléotides (Figure 1).

De plus, à l'autre extrémité de l'intron, la structure secondaire potentielle du domaine VI ne comportait pas de A protubérant à l'emplacement attendu du site de branchement (Figure 2). Ces introns, dont le ribozyme se trouve appartenir à un même sous-groupe structural (IIB1; Michel et coll., 1989), ont d'autres attributs remarquables en commun (Tableau 1): le deuxième appariement entre le ribozyme et l'exon 5' (EBS2-IBS2), qui est potentiellement présent chez la plupart des introns de groupe II, paraît manquer; de plus, plusieurs de ces introns spécifient une endonucléase de 'homing', plutôt qu'une transcriptase inverse (Tableau 1 et Figure 1). Un examen plus attentif de ces introns à insertion 5'-terminale, révèle que la séquence et la structure secondaire du domaine VI y est plus variable que

dans leurs proches parents évolutifs: non seulement l'adénine du point de branchement est absente à l'emplacement attendu, mais il n'y a pas de trace de l'hélice de trois paires de bases et de la boucle interne bien conservée (GAA:CUA) qui devraient jouxter cette adénine distalement.

Figure 1. Arbre phylogénétique des introns mitochondriaux du sous-groupe IIB1.
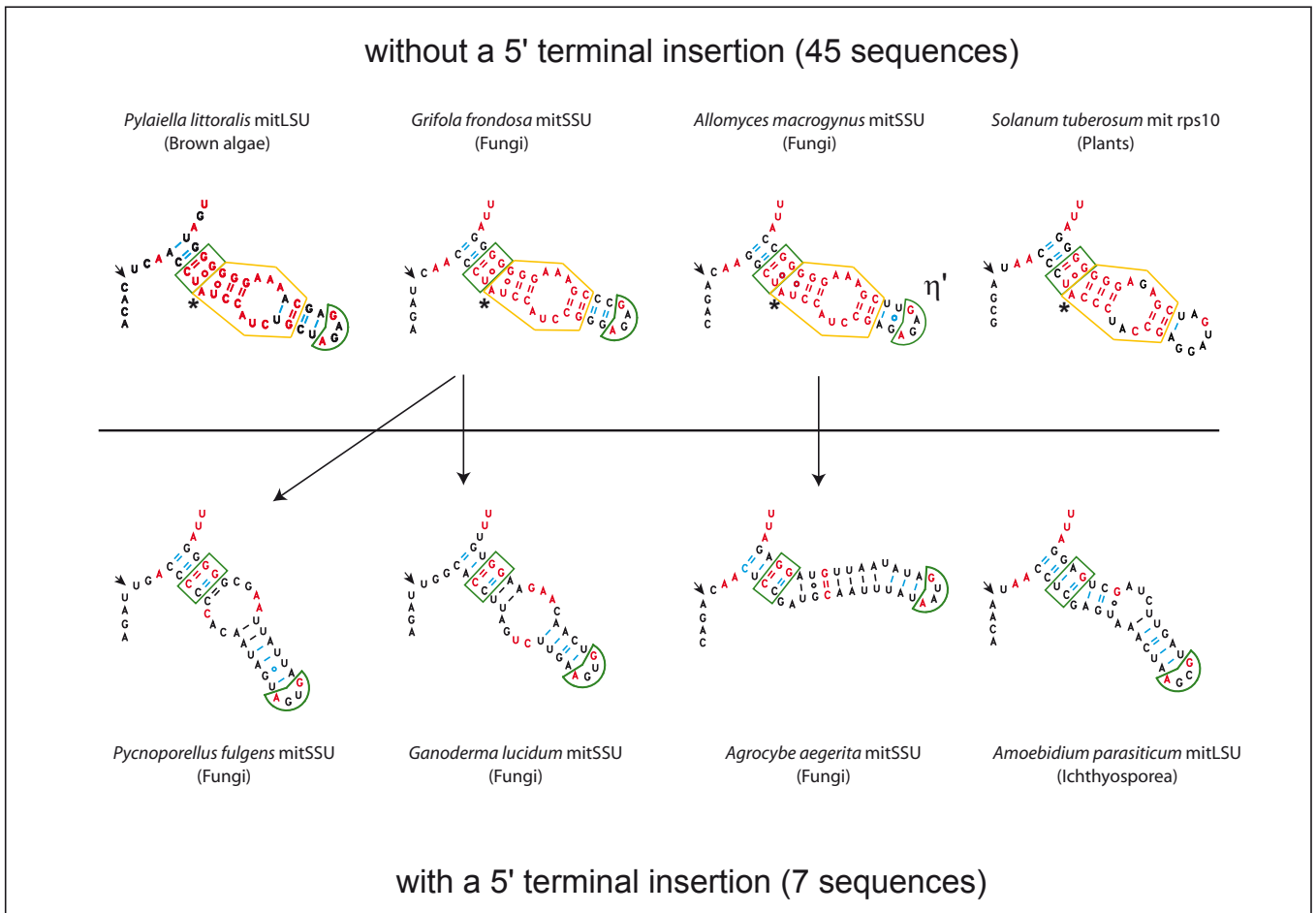
Figure 2. Variations du domaine VI chez les introns à insertion 5' terminale.

Tableau 1. Liste des introns mitochondriaux de sous-groupe IIB1 (informations collectées par François Michel)

Organism, gene and intron	Accession number	Intron coordinates[2]	5' insert[3]	EBS2	ORF product[4] and location[10]
Fungi					
Paracoccidioides brasiliensis cox1/2[5]	AY955840	41071-43890			RT (IV)
Candida parapsilosis cox1[5]	X74411	12690-15605			RT (IV)
Saccharomyces cerevisiae cox1/5?	V00694	8746-9632[*]			
Saccharomyces pastorianus cox1/5	EU852811	53565-54476			
Saccharomyces cerevisiae cob/1	EU004203	38472-39239			

Candida zemplinina LSU2584[6]	AY445918	2516-324[*]				RT (IV)
Candida ipomoeae SSU531	AY393889	176-801				
Glomus intraradices C16g1_2 LSU1787	AM950209	2541-3896				LAGLIDADG (IV)
Uncultured Glomus W9/1 LSU1787	FN377588	1827-3215				LAGLIDADG (IV)
Allomyces macrogynus LSU2059	U41288	2416-3192				
Rhizophydium sp. 136 LSU2059	NC_003053	3880-4564	+6	GUGCGAC	no	
Pichia angusta LSU2059	AL432964 AL434946 AL433470	625-1,469-1,294-879[*]			no	LAGLIDADG (IV)
Ustilago maydis SRX2 LSU2059	EU921807	3413-5372				LAGLIDADG (IV)
Agrocybe aegerita LSU2059	AF087656	9088-10871[*]	+23	UUGCGAC	no	LAGLIDADG[11] (IV)
Suillus luteus LSU2059	L47586	2675-3341	+25	UAGCGAC	no	
Cryphonectria parasitica SSU952	AF029891	7168-9235				LAGLIDADG (III)
Leptographium truncatum 1435 SSU952	TM				no	LAGLIDADG (III)
Cordyceps konnoana SSU952	AB031194	897-2724[*]				LAGLIDADG (III)
Paracoccidioides brasiliensis SSU952	AY955840	25574-27362				Unidentified (III)
Aleurodiscus botryosus SSU788	FM[7]				no	LAGLIDADG (IV)
Ceriporiopsis subvermispora SSU788	EU546103	345-907[*]			no	
Grifola frondosa SSU788	FM[7]				no	LAGLIDADG (IV)
Pycnoporellus fulgens SSU788	FM[7]		+6	UUGCGAC	no	LAGLIDADG (IV)
Ganoderma lucidum SSU788	AF214475[8]	1056-2562	+6	UUGCGAC	no	LAGLIDADG (IV)
Trametes cingulata SSU788	GU723273	39037-40442[*]	+6	AUGCGAC	no	LAGLIDADG[11] (IV)
Usnea antarctica SSU788	DQ990920	397-1473[*]				
Cryphonectria parasitica SSU788	AF029891	2415-4596[*]			no	LAGLIDADG (IV)
Ichthyosporea						
Amoebidium parasiticum SSU788	AF538044	855-2198			no	GIY-YIG (IV)
Amoebidium parasiticum LSU2449	AF538042	5337-5909	+33	GAGCGAC	no	

Analyse des séquences des introns SSU788 de *Grifola frondosa* et *Pycnoporellus fulgens*

Afin de poursuivre l'étude de ces introns atypiques, nous avons choisi de cloner trois introns insérés à la position 788 de l'ARN de la petite sous-unité ribosomique mitochondriale (SSU788 ; la numérotation est celle de *E. coli*). L'un de ces introns est pourvu d'une insertion 5' et provient de *Pycnoporellus fulgens* : une séquence partielle de cette intron avait été déposée dans les bases de données (code d'accès GenBank : AF518690). Les deux autres introns sont également insérés en SSU788 et assez étroitement apparentés à celui de *P. fulgens*, mais dépourvus d'insertion 5'; ils proviennent des champignons basidiomycètes *Grifola frondosa* et *Aleurodiscus botryosus*, et là aussi, des séquences partielles étaient disponibles (codes d'accès AF334880 et AF026646).

Comme on peut le voir Figure 3, les structures secondaires prédites pour les ribozymes des introns de *Grifola* et *Pycnoporellus* sont très semblables. Cependant, la superposition de ces structures montre que le domaine VI diffère considérablement entre les deux introns : seules les trois premières paires de bases et l'extrémité distale de ce domaine sont conservées entre les séquences de *Grifola* et *Pycnoporellus*.

Grifola frondosa
mit SSU
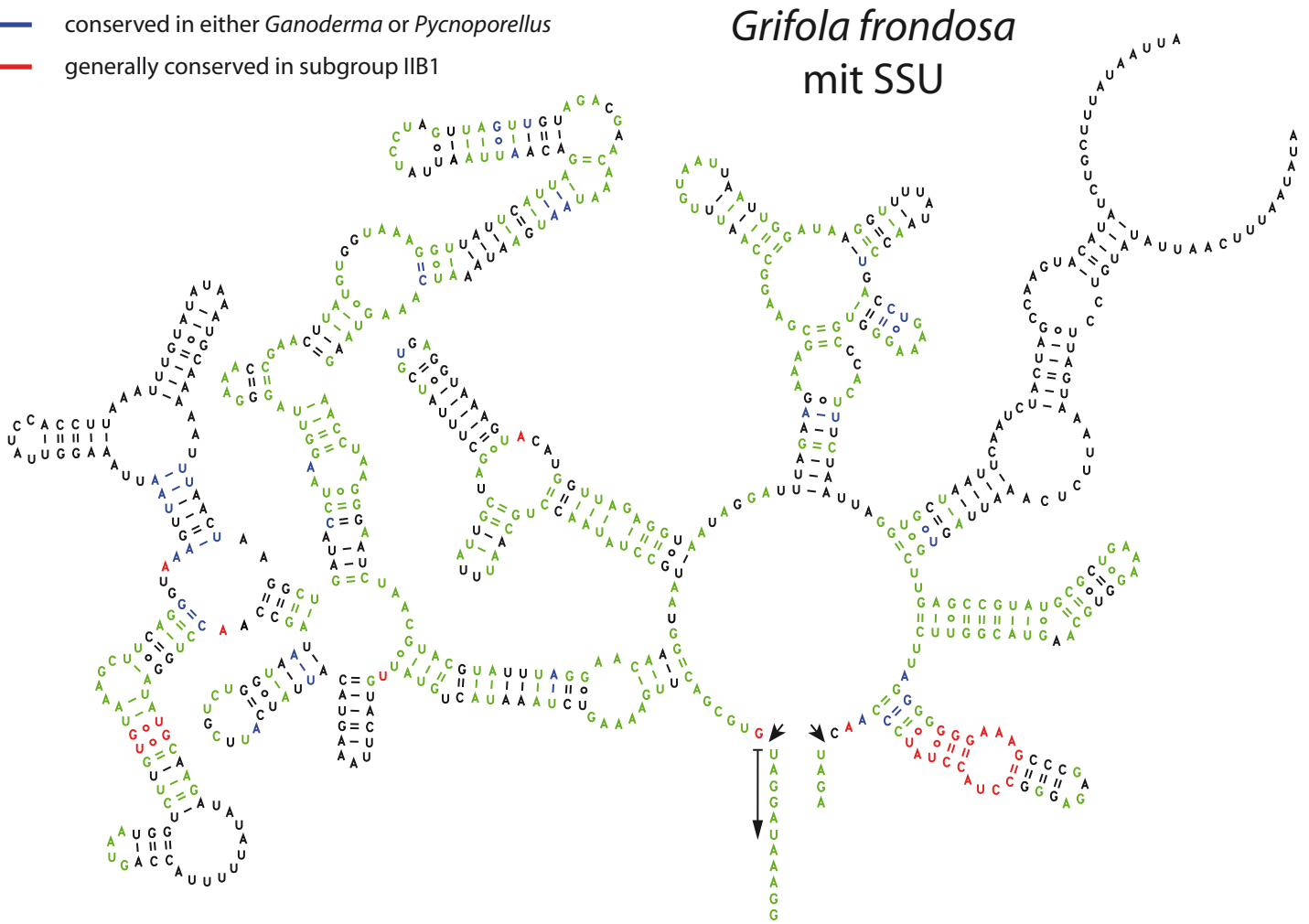
Figure 3. Superposition du modèle de structure secondaire de l'intron SSU788 de *Pycnoporellus fulgens* à celui de l'intron SSU788 de *Grifola frondosa*.

L'auto-épissage *in vitro* des introns SSU788 de *Grifola* and *Pycnoporellus* donne naissance à des produits bien distincts

Afin de décrire l'auto-épissage *in vitro* des introns de *Grifola* and *Pycnoporellus*, nous avons suivi cinétiquement les réactions d'épissage et caractérisé leurs produits. La réaction d'auto-épissage *in vitro* de l'intron SSU788 de *Grifola* à 42°C dans 1 M NH₄Cl et 20 mM magnésium est montrée Figure 4. La réaction du précurseur se divise en deux phases, une partie des molécules de précurseur

réagissent rapidement dans les deux premières minutes, tandis qu'une deuxième population de transcrits précurseurs réagit relativement lentement (Figure 4B). Outre les exons liés, les produits de réaction sont dominés par l'intron lariat, comme pour les autres introns de group II typiques. Une forme linéaire de l'intron n'est observée qu'en très petite quantité, même quand les ions ammonium sont remplacés par des ions potassium.

Quand maintenant l'intron SSU788 de *Pycnoporellus* SSU788 est incubé dans les mêmes conditions que celui de *Grifola*, la réaction reste assez rapide (Figure 6C.): 80% du précurseur est converti en produits en 10 min environ. Et la population de molécules réagies contenant l'intron se répartit de nouveau en deux formes. Cependant, il n'y a pas de traces d'un produit branché, tous les produits de réaction paraissent avoir été générés par voie hydrolytique.
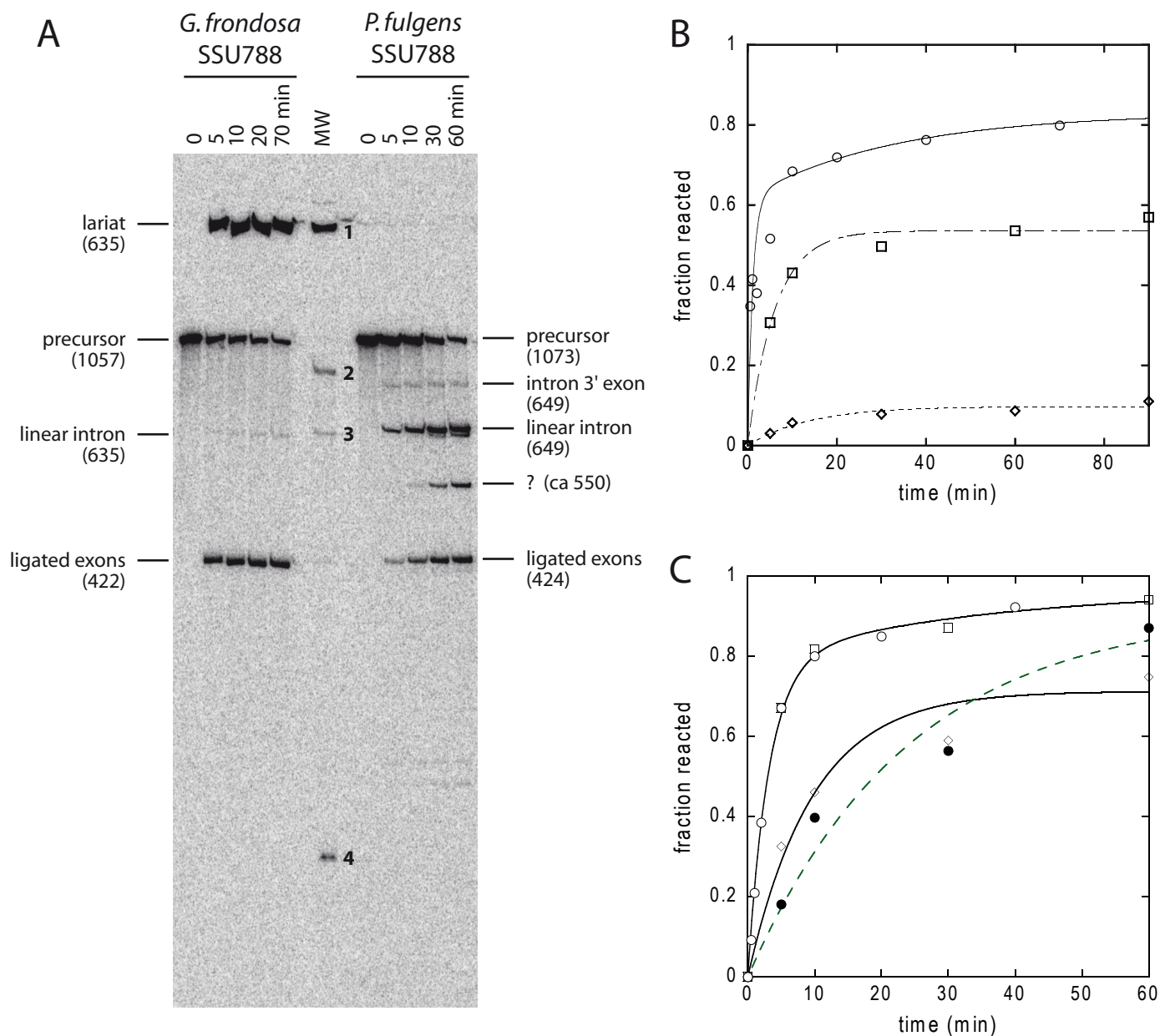
Figure 4. Auto-épissage des introns SSU788 de *Grifola* et *Pycnoporellus*.

A. Cinétiques d'auto-épissage d'ARN précurseurs contenant les introns SSU788

de *Grifola* et *Pycnoporellus* à 42 °C dans 1 M NH$_4$Cl, 20 mM MgCl$_2$, 40 mM Na-

MES (pH 6.2). Les mobilités électrophorétiques sont comparées à celles des

produits d'épissage connus d'un ARN précursuer contenant l'intron LSU1787 de

Pylaiella littoralis (piste MW: bande 1, 640 nt, lariat; bande 2, 872 nt, précursor;

bande 3, 640 nt, intron linéaire; bande 4, 232 nt, exons liés).

B. Cinétique d'auto-épissage d'un ARN précurseur contenant l'intron SSU788 de *Grifola* à 42 °C dans 40 mM Tris-Cl pH 7.5, 20 mM MgCl$_2$ et soit 1 M NH$_4$Cl (cercles et courbe pleine, générée par ajustement à une équation exponentielle biphasique avec k$_1$ = 0.9±0.2 min^{-1} et k$_2$ = 0.03 min^{-1}), soit 1M KCl (carrés et courbe tiretée pour l'intron lariat; losanges et courbe pointillée pour l'intron linéaire; les deux courbes ont été générées par ajustement à une équation exponentielle unique).

C. Cinétique d'auto-épissage d'un ARN précurseur contenant l'intron SSU788 de *Pycnoporellus* à 40 mM Tris-Cl pH 7.5, 1 M NH$_4$Cl et 10 mM MgCl$_2$ (carrés vides), 20 mM MgCl$_2$ (cercles vides), 50 mM MgCl$_2$ (losanges vides), ou 40 mM Na-MES pH 6.2 et 20 mM MgCl$_2$ (cercles pleins et courbe tiretée). Les réactions à 10 et 20 mM Mg, pH 7.5, ont été ajustées à un processus biphasique (k$_1$ = 0.32±0.03 min^{-1}, k$_2$ =0.030±0.016 min^{-1}), les autres à des exponentielles simples.

Partie II

Une analyse phylogénétique inédite suggère une piste pour explorer le site potentiel de fixation du domaine VI lors de la première étape de l'épissage

En comparant la variabilité des séquences site à site chez les introns de groupe II normaux et ceux pourvus d'une insertion à l'extrémité 5', nous avons découvert des variations de séquence et structure qui pourraient contribuer à expliquer l'absence de formation de la structure lariat chez les introns à insertion. Chez ces derniers, en effet, non seulement l'adénosine du point de branchement manque, mais la section moyenne du domaine VI, qui comprend normalement (chez les introns mitochondriaux de sous-groupe IIB1) une boucle interne bien conservée de 6 nt et une hélice de trois paires de bases, qui relie cette boucle au point de branchement, est méconnaissable. Ceci suggère que non seulement le point de branchement et les deux paires G :U qui l'encadrent (Chu et al, 1998), mais toute la section moyenne du domaine VI pourrait être impliquée dans la réaction de branchement. Outre les nucléotides appartenant au domaine VI, nous avons découvert que plusieurs sites dans l'hélice distale du domaine IC1 montrent également une variabilité différente selon qu'il existe ou non une insertion à la jonction 5' et pourraient donc contribuer aussi à la réaction de branchement : il s'agit – chez l'intron ribosomique Pl.LSU/2 de Pylaiella – des positions 78, 79 et 100.

Substitutions de nucléotides dans le domaine VI et dans son site récepteur potentiel IC1.

Nos analyses de séquences suggèrent qu'aussi bien la section moyenne du domaine VI que les paires G :U 79:100 et 78:101 de IC1 sont spécifiquement impliquées dans la réaction de branchement. Par conséquent, il devrait être possible d'observer un déplacement de l'équilibre entre transestérification et hydrolyse en réponse à l'introduction de substitutions nucléotidiques à ces sites. Au cours de notre étude, nous avons introduit diverses substitutions aussi bien dans le sous-domaine IC1 que dans le domaine VI, afin de montrer que ces composants jouent un rôle clé dans la réaction de branchement. (Figure 5).
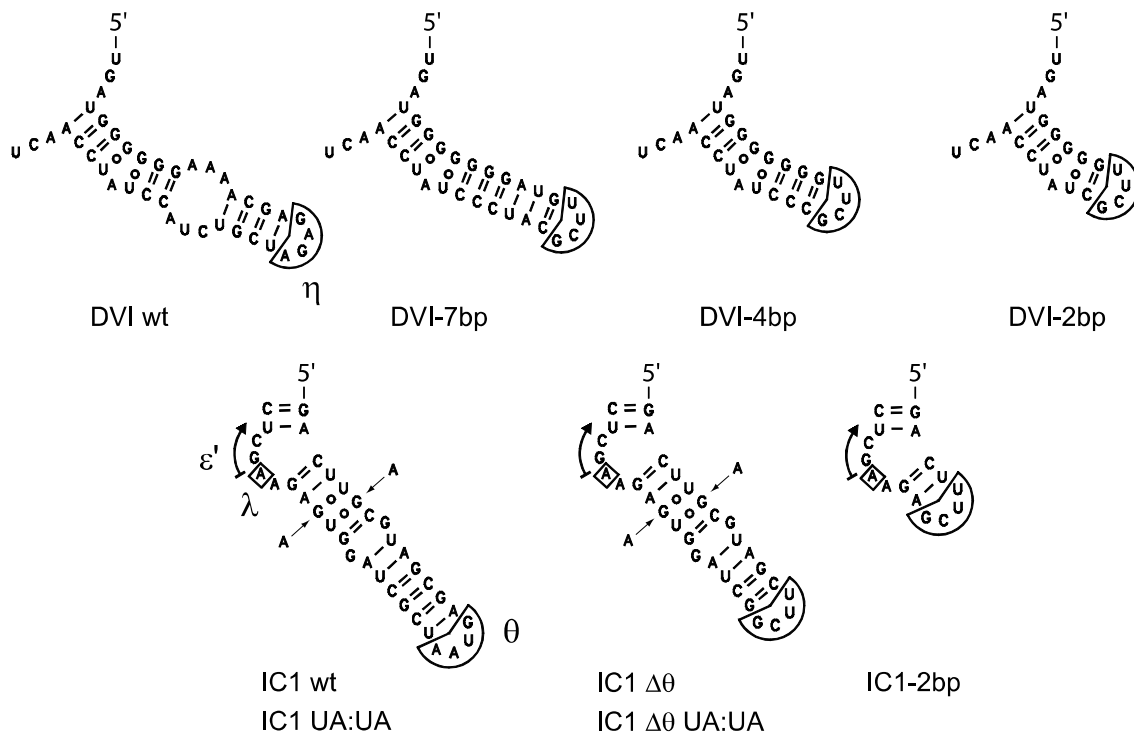
Figure 5. Ribozyme mutants comportant des substitutions dans les domaines VI et IC1.

Mutagénèse du domaine VI

Premièrement, la tige du domaine VI a été tronquée de manière à ne laisser que deux paires de bases au-delà du site de branchement (version courte, DVI-2bp dans l'article #2; Figure 5); puis la boucle interne du domaine VI a été refermée et remplacée par une hélice de domaine VI continue, tirée d'un intron de groupe II de *Pseudomonas spp.* (mutant 'DVI stem'). Les cinétiques de réaction des molécules mutantes PL2-wt, PL2-57 (DVI-2bp) et PL2-58 (DVI stem) ont alors été analysées et comparées aussi bien dans des conditions favorisant l'initiation de l'épissage par transestérification (1 M NH$_4$Cl), que dans des conditions privilégiant l'hydrolyse (1 M KCl). Il en est ressorti que même si le mutant avec un domaine VI tronqué (PL2-57; DVI-2bp) paraît conserver un A protubérant bordé de deux paires G :U, il a perdu néanmoins l'aptitude à initier l'épissage par transestérification. On en conclut que la fixation du domaine VI, pour être productive, requiert la présence non seulement du A protubérant, mais aussi des nucléotides qui entourent ce A : il est donc raisonnable de penser que la tige du domaine VI joue également un rôle important dans la reconnaissance du domaine VI.

Quelle partie exacte du domaine VI est utilisée pour assurer sa fixation au cours de la réaction de transestérification de première étape ? Afin de répondre à cette question, l'hélice continue du domaine VI du mutant PL2-58 a été soumise à des troncations successives. PL2-70 (DVI-4bp) et PL2-71 (DVI-7bp) sont des mutants avec des hélices distales de 4 et 7 paires de bases, respectivement (Figure 5). Les

ARN mutants ont d'abord été examinés et caractérisés cinétiquement dans des conditions favorisant la transestérification : il ressort de ces expériences que modifier la longueur de la tige du domaine VI n'a qu'une influence mineure sur l'épissage par transestérification en présence d'ammonium. On peut penser que le sel monovalent monovalent contribue à stabiliser la structure de l'intron de groupe II intron, ce qui n'apparaissait pas initialement dans nos expériences.

Nous avons ensuite examiné les mêmes mutants dans des conditions favorisant l'hydrolyse (1 M potassium). La vitesse d'hydrolyse du mutant DVI-4bp est grandement accrue et les produits de branchement réduits en proportion, par comparaison à la molécule de séquence 'sauvage'. Toujours dans les mêmes conditions expérimentales, chez l'autre mutant, avec une hélice DVI distale de 7 paires de bases, les produits de transestérification à la jonction 5' sont par contre beaucoup plus importants, aussi bien en termes de quantité que de vitesse de production (Tableau 1 de l'article #2).

Table 1. Kinetic parameters of dVI and IC1 mutants.

Construct	Fraction of products branched	$k_{branching}$ (min^{-1})	$k_{hydrolysis}$ (min^{-1})	k_{br}/k_{hy}
ammonium				
wt[1]	0.90 ± 0.07	0.136 ± 0.019	0.024 ± 0.010	5.5
	0.88 ± 0.11	0.166 ± 0.032	0.023 ± 0.008	7.2
dVI -7 bp	0.89 ± 0.04	0.092 ± 0.006	0.019 ± 0.002	5.0
dVI -4 bp	0.84 ± 0.06	0.058 ± 0.006	0.014 ± 0.002	4.2
dVI -2 bp	$0.02^{(2)}$	$<0.008 \pm 0.002^{(3)}$	0.013 ± 0.002	<0.62
IC1 $\Delta\theta$	n.d.	n.d.	n.d.	n.d.
IC1 UA:UA	0.89 ± 0.09	0.028 ± 0.003	0.024 ± 0.004	1.3
IC1 $\Delta\theta$ / UA:UA	n.d.	n.d.	n.d.	n.d.
IC1-2bp	0.90 ± 0.11	0.016 ± 0.004	0.024 ± 0.004	0.69
potassium				
wt[1]	0.76 ± 0.08	0.160 ± 0.030	0.064 ± 0.023	2.5
	0.77 ± 0.06	0.149 ± 0.020	0.065 ± 0.009	2.3
dVI -7 bp	0.41 ± 0.04	0.132 ± 0.021	0.057 ± 0.012	2.3
dVI -4 bp	0.15 ± 0.01	0.045 ± 0.006	0.072 ± 0.008	0.63
dVI -2 bp	0	0	$[0.135 \pm 0.011]^{(4)}$	0
IC1 $\Delta\theta$	0.69 ± 0.05	0.097 ± 0.006	0.019 ± 0.004	5.1
IC1 UA:UA	0.10 ± 0.007	0.028 ± 0.002	0.042 ± 0.005	0.67
IC1 $\Delta\theta$ / UA:UA	0.067 ± 0.005	0.025 ± 0.003	0.029 ± 0.002	0.84
IC1-2bp	0.063 ± 0.025	0.026 ± 0.013	0.031 ± 0.011	0.85

n.d. : not determined

[1] determinations from different RNA preparations

[2] observed value at 180 min

[3] estimated from the fraction branched at 180 min

[4] determined at 50 mM Mg

Mutagénèse du sous-domaine IC1

Nous avons recouru de nouveau à une mutagénèse dirigée pour explorer la fonction des paires G:U du domaine IC1. Premièrement, les deux paires 'wobble' G:U furent changées en deux paires Watson-Crick A:U (mutant PL2-56, IC1 UA:UA, Figure 5). Cette substitution était cependant susceptible d'altérer la

géométrie du sous-domaine IC1. Pour remédier à ce probblème, deux autres mutants ont alors été construits dans lesquels l'élément θ manquait. Il s'agit de PL2-63 (IC1 Δθ) et PL2-64 (IC1 Δθ UA:UA), le premier servant de référence pour le second.

Les molécules mutantes furent examinées dans 1 M KCl avec 10 mM MgCl$_2$ à 42°C. La délétion de θ s'est révélée être sans conséquences sévères pour l'aptitude des molécules précurseurs à réagir dans les conditions expérimentales que nous avions choisies (Tableau 1 de l'article #2). De manière frappante, les réactions des deux mutants UA:UA (PL2-56 and PL2-64) sont fortement déplacées vers la voie d'initiation par hydrolye : la transestérification est significativement inhibée et l'hydrolyse devient dominante (Figure 6). Il est remarquable qu'un changement apparemment si minime dans le sous-domaine IC1 ait une influence aussi forte sur le mode de réaction. Après notre analyse phylogénétique qui avait montré que la perte de ces deux paires G:U coïncide avec la perte du A protubérant dans le domaine VI, ces données expérimentales viennent étayer l'hypothèse que ces deux G:U forment tout ou partie du site de fixation du domaine VI durant la première étape de transestérification de l'épissage.

PL2-WT

PL2-63 IC1 ΔΘ

PL2-56 (IC1 UA:UA)

PL2-64 IC1 ΔΘ UA:UA
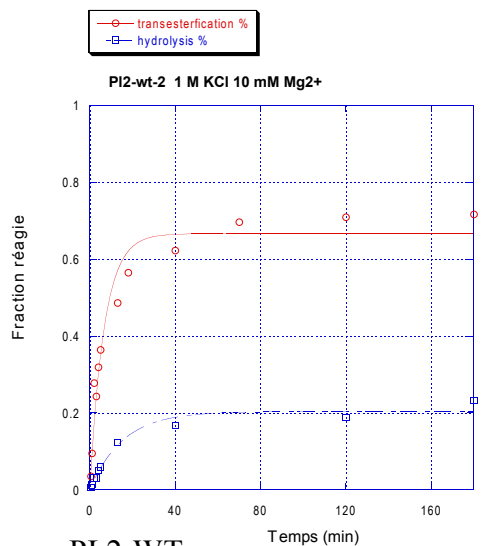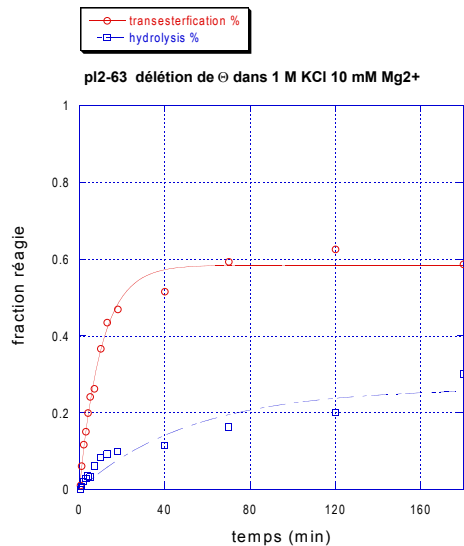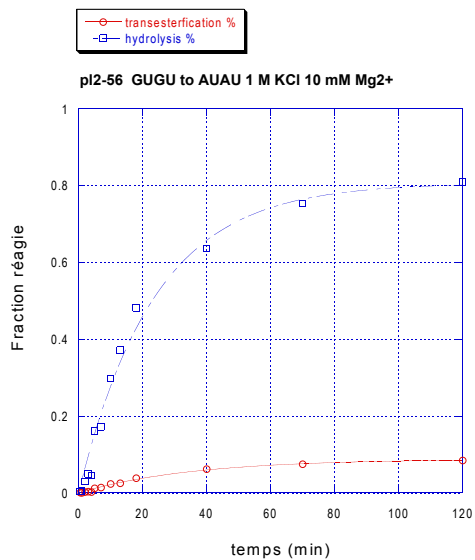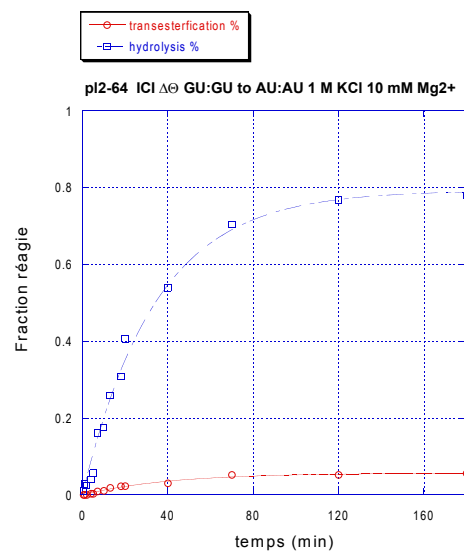
	k_{br} （1/min）	k_{hy} （1/min）	k_{br}/k_{hy}
wt	0.149	0.064	2.3
IC1 ??	0.097	0.019	5.1
IC1 UA:UA	0.028	0.042	0.66
IC1 ?? UA:UA	0.024	0.029	0.82

Figure 6. Fractions réagies d'intron branché et linéaire à partir de l'ARN précurseur PL2-wt et des mutants du sous-domaine IC1 à 42 °C dans 1 M KCl, 20 mM MgCl$_2$. Le Tableau indique la constante de vitesse de branchement (k_{br}) et celle d'hydrolyse (k_{hy}), calculées à partir des fractions réagies en function du temps.

Démonstration de l'identité du récepteur de première étape du domaine VI à l'aide oligonucléotides ADN utilisés comme connecteurs pontants

Afin de rassembler des données plus convaincantes en faveur de notre hypothèse de travail – que le sous-domaine IC1 est le véritable site de fixation du domaine VI au cours de la première étape de transestérification de l'épissage – nous avons construit un système expérimental qui se sert d'un oligonucléotide comme d'une chaîne pour ancrer le domaine VI au sous-domaine IC1. Le concept sous-jacent est que si le domaine VI contacte effectivement le sous-domaine IC1 during la réaction de branchement, alors ces deux composants doivent être situés très près l'un de l'autre dans l'espace. Et si la section du domaine VI impliquée dans la reconnaissance de ce domaine par le reste du ribozyme est mutée, on s'attend à ce que la réaction soit redirigée de la voie de transestérification à la voie d'hydrolyse. Nous avons donc créé une version spéciale du ribozyme Pl.LSU/2 dans laquelle les séquences des domaines IC1 et VI sont toutes les deux changées pour s'adapter à un oligonucléotide spécifique. Quand cet oligonucléotide est ajouté à

la réaction, on s'attend à ce que la fonction du précurseur Pl.LSU/2 soit restaurée et qu'il initie de nouveau l'épissage par une réaction de branchement.

Pour commencer, nous avons construit un ensemble d'introns mutants (Figure 7) dans lesquels deux séquences de six nucléotides, l'une dans le domaine VI, l'autre dans le sous-domaine IC1, sont complémentaire d'un oligonucléotide pontant. Ces mutants PL2-72, PL2-73 and PL2-74 ont été testés séparément avec ou sans addition de concentrations variables d'un oligonucléotide pontant appelé PLI55 (Tableau 5 de la version anglaise). Dans des conditions favorisant la transestérification (1 M ammonium), 10 mM magnésium à 37 °C, l'addition de PLI55 est sans effet sur la réactivité de la molécule sauvage ou du mutant PL2-72. Ces résultats négatifs montrent qu'il ne paraît pas exister d'interactions non-specifiques entre les séquences de l'ARN précurseur et de l'oligonucléotide. Quant au mutant PL2-74, il s'est comporté comme attendu : il y a restauration de l'aptitude à initier l'épissage par transestérification quand l'oligonucléotide pontant PLI55 amène le domaine VI à proximité du sous-domaine IC1. Cependant, le mutant PL2-73, avec un domaine VI modifié comme dans PL2-74, se révèle aussi capable d'initier dans une certaine mesure l'épissage par branchement en présence de l'oligonucléotide PLI55.

Afin de poursuivre notre analyse de l'interaction entre les domaines IC1 et VI, nous avons alors construit un nouveau lot de mutants (Figure 7, panneau B). Nous attendions de ces mutants qu'ils aient une affinité plus élevée pour l'oligo pontant,

ce qui devrait diminuer le K_m des couples précurseur-oligonucléotide, et aussi une structure plus stable.
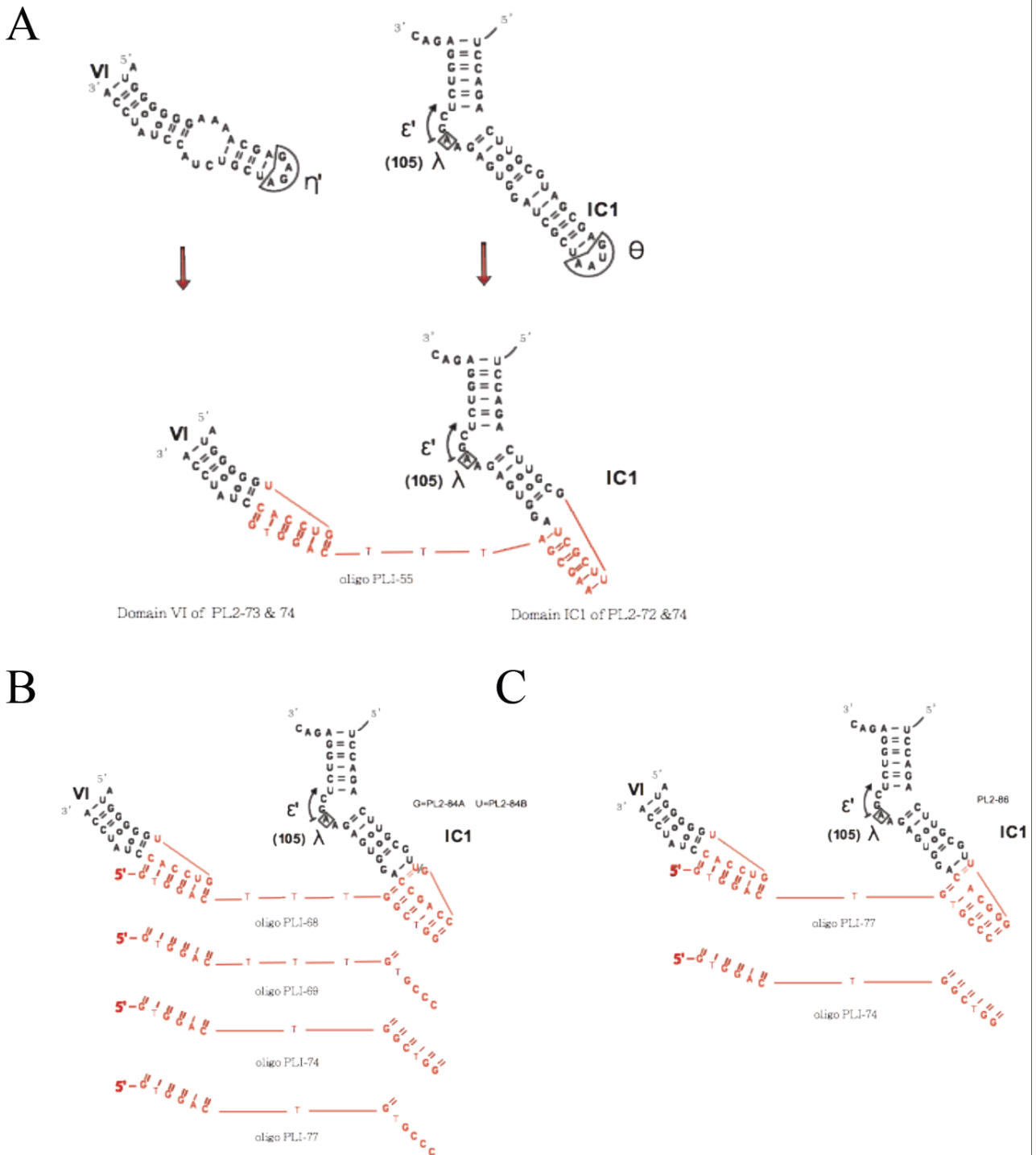
Figure 7. Principe des expériences d'ancrage VI-IC1 par un oligonucléotide

(A). Les séquences des domaines VI and IC1 sont modifiées pour s'adapter à un oligonucléotide utilisé comme connecteur. (B). Une construction prévue pour permettre aux deux poignées oligonucléotidiques de s'apparier présente une efficacité de branchement grandement accrue. Cette construction a été aussi utilisée pour tester des oligonucléotides avec des segments connecteurs de longueurs différentes et une séquence IC1 mésappariée. (C). La séquence du sous-domaine IC1 dans cette construction est modifiée de manière à s'apparier avec celle de l'oligonucléotide PLI77, précédemment mésappariée vis-à-vis du mutant PL2-84B (c'est maintenant l'oligonucléotide PLI-74 dont la séquence n'est plus complémentaire de celle du sous-domaine IC1).

Comme on le voit Figure 8, l'aptitude du précurseur PL2-84B à initier l'épissage par transestérification est efficacement restaurée par des concentrations croissantes de l'oligonucléotide PLI68. PLI68 se comporte comme un connecteur qui s'apparie à la fois avec le domaine IC1 et le domaine VI et aide ce dernier à se fixer correctement au domaine IC1. PLI68 accroît fortement la transestérification, qui atteint jusqu'à 60% de la réaction totale à une concentration saturante de l'oligonucléotide. Le K_m est aussi beaucoup plus petit, d'environ 5.4 μM, comparé à un Km de 58 μM pour le mutant précédent PL2-74. Ainsi, l'affinité de l'oligonucléotide pour l'intron a été considérablement augmentée, comme nous l'espérions.
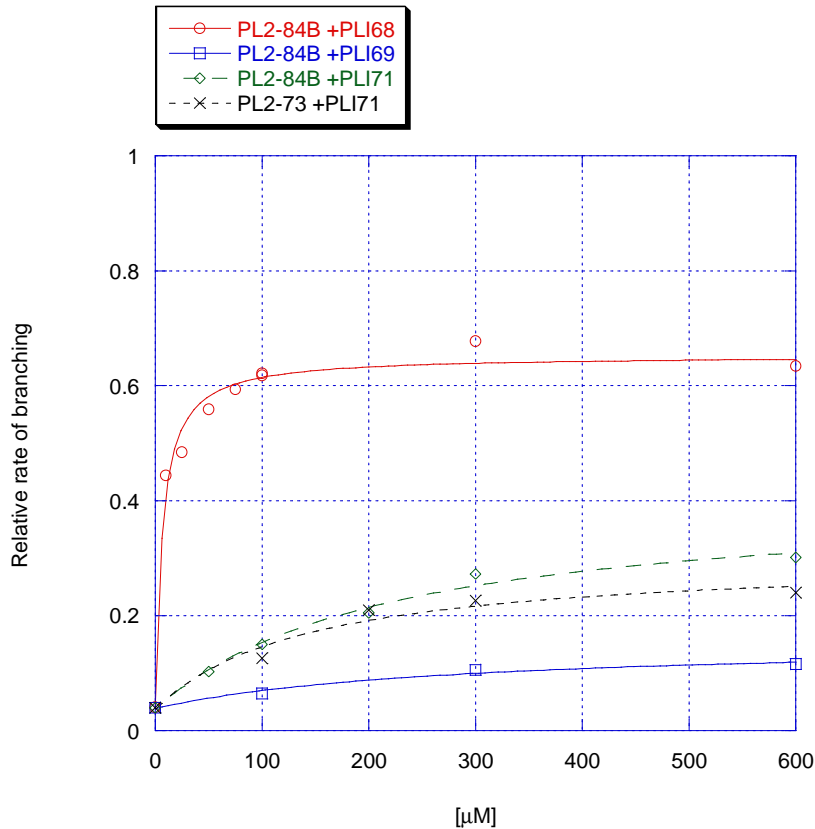
Figure 8. Vitesse relative de branchement en fonction de la concentration d'un oligonucléotide pontant pour différents couples mutant-oligonucléotide. La méthode de calcul est indiquée dans Materials and Methods. La vitesse relative de branchement est maximale pour le couple formé par le mutant PL2-84B et l'oligonucléotide complémentaire PLI68. Le remplacement de cet oligonucléotide par l'oligonucléotide mésapparié PLI-69 a pour effet de réduire et la vitesse de branchement et la fraction branchée. La combinaison (PL2-73 + PLI71) constitue un contrôle experimental, dans lequel il n'y a appariement qu'avec DVI (PL2-73 a la même séquence DVI que PL2-84B, mais un sous-domaine IC1 sauvage; PLI-71 est un 7-mère dont la séquence est complémentaire du domaine VI de PL2-84B; PL2-84B + PLI71 fournit le contrôle correspondant, avec IC1 mutant).

Un oligonucléotide mésapparié (PLI69; 15-mère avec poignée IC1 mésappariée) a été utilisé comme contrôle : sa séquence est complémentaire de celle du domaine VI, mais pas de celle du domaine IC1 dans le précurseur PL2-84B. On observe une réduction de 60% à environ 30% des produits générés par transestérification et surtout, une augmentation abrupte du K_m, de 5.4 µM à 278 µM, juste à cause du mésappariement avec le sous-domaine IC1. Ces données montrent que seul un oligonucléotide bien apparié est capable d'agir efficacement comme connecteur et de restaurer la réaction de branchement en amenant le domaine VI à proximité du domaine IC1.

PLI71, un 7-mère avec une séquence complémentaire de celle du domaine VI (anti-D6), a été testé avec PL2-73 et PL2-84B afin d'estimer dans quelle mesure la reconstruction du domaine VI participe à la restauration de l'activité de branchement : comme nous nous y attendions, l'effet reste modeste.

La longueur du segment connecteur a des effets importants sur l'aptitude d'un oligonucléotide à restaurer l'activité de branchement

Même si un oligonucléotide avec une séquence bien appariable peut restaurer la réaction de branchement d'un précurseur fonctionnellement déficient, le complexe entre cet oligonucléotide et l'intron ne réagit pas aussi bien que la molécule sauvage d'origine. On peut penser que la raison de cette situation est que la position dans l'espace du domaine VI est insuffisamment bien définie. A partir de cette hypothèse, nous avons conçu un lot d'oligonucléotides qui ont les mêmes

poignées d'appariement que PLI68, mais dans lesquels la longueur du segment connecteur poly-T varie de 0 à 4 nucléotides (Tableau 5 de la version anglaise). Ces oligonucléotides ont été utilisés à une concentration de 5 µM, proche du K_m de la combinaison PL2-84B:PLI68. A notre surprise, la longueur du segment connecteur a des effets très marqués sur l'efficacité de la réaction de branchement. Comparée à celle de l'oligonucléotide d'origine, avec un connecteur standard de 3 T (oligo PLI68), la vitesse relative de branchement est considérablement améliorée avec des connecteurs composés de un ou deux T, tandis que l'absence de connecteur ou un connecteur de 4T conduisent à des performances relativement médiocres (vitesses relatives : sans T=0.24, 1T=0.78, 2T=0.46, 3T=0.30, 4T=0.20 ; voir Figures 9 et 10). Ces résultats suggèrent fortement que les oligonucléotides avec un connecteur court (1T et 2T) améliorent la réaction de branchement en contraignant le mouvement du domaine VI dans l'espace. Toujours dans le même ordre d'idées, un connecteur composé de 4T confère trop de flexibilité à la position du domaine VI, tandis que l'absence de connecteur antre les domaines VI et IC1 rend difficile le positionnement correct de ces domaines l'un par rapport à l'autre.
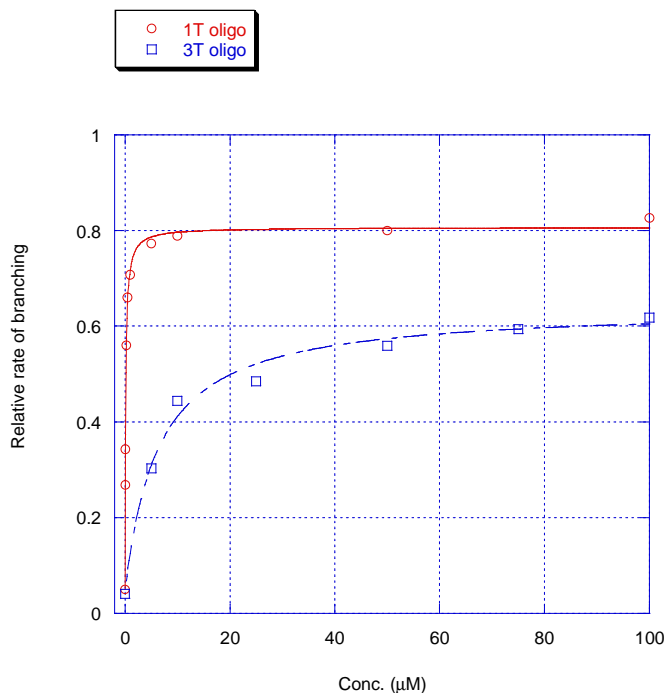
Figure 9. Vitesses relatives de branchement en fonction de la concentration d'oligonucléotides complémentaires différant par la longueur de leur segment connecteur. La réaction de branchement d'un précurseur PL2-84B est progressivement restaurée par des concentrations croissantes d'oligonucléotides avec un connecteur composé d'un ou 3 T, mais le premier oligonucléotide est beaucoup plus efficace – son K_m est beaucoup plus petit – que le second.
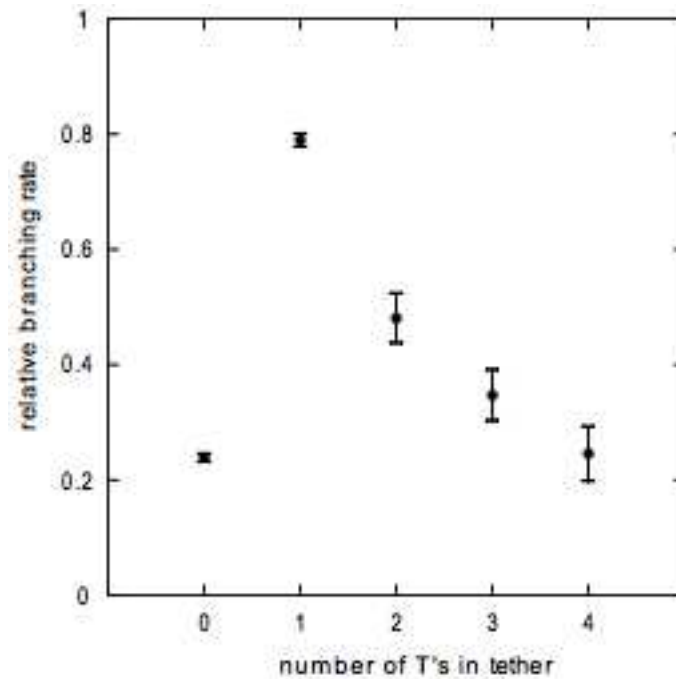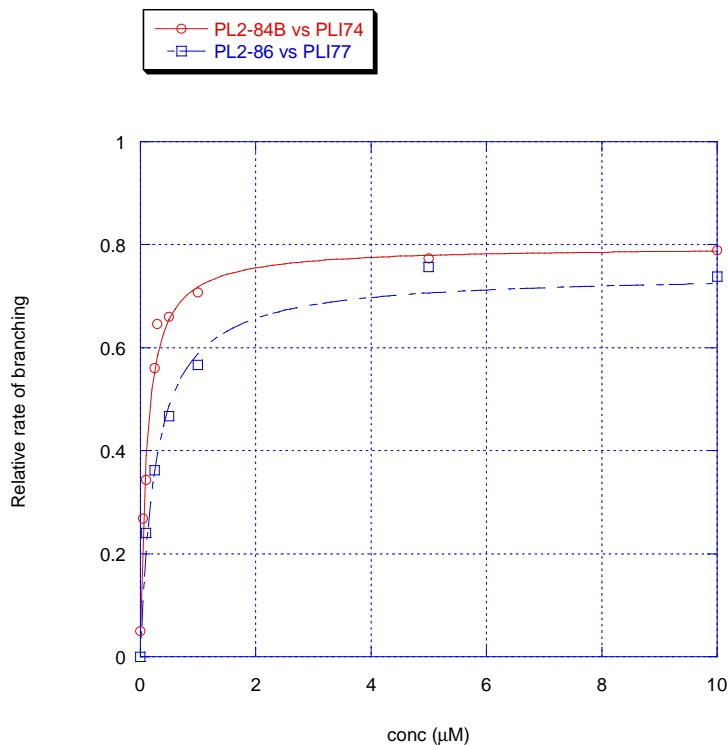
Figure 10. Vitesse relative de branchement du précurseur PL2-84B en fonction du nombre de T dans le segment connecteur de l'oligonucléotide 5'-GTGGAC[T]$_n$GGCTGG. La concentration des oligonucléotides était 5.0 μM, valeur proche du K_m mesuré pour l'oligonucléotide pontant à 3 T, PLI-68.

Oligonucléotides mésappariés et mutations compensatoires

Se pourrait-il que le phénomène de restauration de l'aptitude à effectuer la réaction de branchement soit spécifique de certaines séquences? Pour répondre à cette question, nous avons créé un mutant supplémentaire, PL2-86 (Figure 7; construction C). La différence avec les combinaisons précédentes est que le sous-domaine IC1 du nouveau mutant est complémentaire de l'oligonucléotide PLI-69. PLI69 (un 15-mère) avait été utilisé comme témoin négatif pour les expériences

avec PL2-84B : sa séquence est complémentaire de celle du domaine VI, mais mésappariée vis-à-vis de celle de PL2-84B dans le sous-domaine IC1, de telle sorte que la section correspondante de PLI69 ne contribue pas significativement à la restauration de la réaction de branchement. Les résultats (Figure 11) montrent que PL2-86, dont la séquence IC1 est complémentaire de celle de l'oligonucléotide PLI77, voit lui aussi sa réaction de branchement restaurée par la présence de cet oligonucléotide, même si cette restauration est un peu moins efficace que pour la combinaison PL2-84B:PLI-74. Ce résultat prouve que le phénomène que nous avons découvert est robuste vis-à-vis des changements de séquence.

Figure 11. Vitesse relative de branchement de deux constructions DVI-IC1 en fonction de la concentration d'un oligonucléotide complémentaire (voir partie

C de la Figure 7). L'oligonucléotide mésapparié PLI77 ne peut rétablir l'aptitude du PL2-84B à effectuer la réaction de branchement (Figure 4 de l'article #2), mais ce même oligonucléotide restaure la réaction de branchement de la construction appariable PL2-86 presque aussi bien que pour le couple PL2-84B + PLI74.

## Modélisation de l'interaction entre le domaine VI et son récepteur présumé dans le sous-domaine IC1

La première structure à résolution atomique d'un ribozyme de groupe II a été établie par Toor et coll. en 2008. Malheureusement, il manque au modèle publié les coordonnées atomiques du domaine VI, peut-être, comme l'ont proposé ces auteurs, parce que la flexibilité de ce domaine conduit à sa dégradation. Notre tentative de modélisation tri-dimensionnelle (Fig. 12) a cherché à concilier le modèle cristallographique à haute résolution de l'intron d'*Oceanobacillus* par Toor et coll. (plus précisément, la dernière version de ce modèle, publiée en 2010) et l'interaction, que nous pensons avoir découverte entre le domaine VI et un récepteur spécifique de la première étape de l'épissage. Dans notre modèle, la partie distale du domaine VI est traitée comme une hélice continue en dépit de la présence d'une boucle interne bien conservée dans les introns mitochondriaux de sous-groupe IIB1. Ce modèle est en excellent accord avec notre analyse comparée de séquences introniques et nos expériences de substitution de nucléotides, puisque la section de IC1 qui y est spécifiquement contactée par le domaine VI comprend la paire de base G79:U100 (G81:U101 dans l'intron d'*Oceanobacillus*).
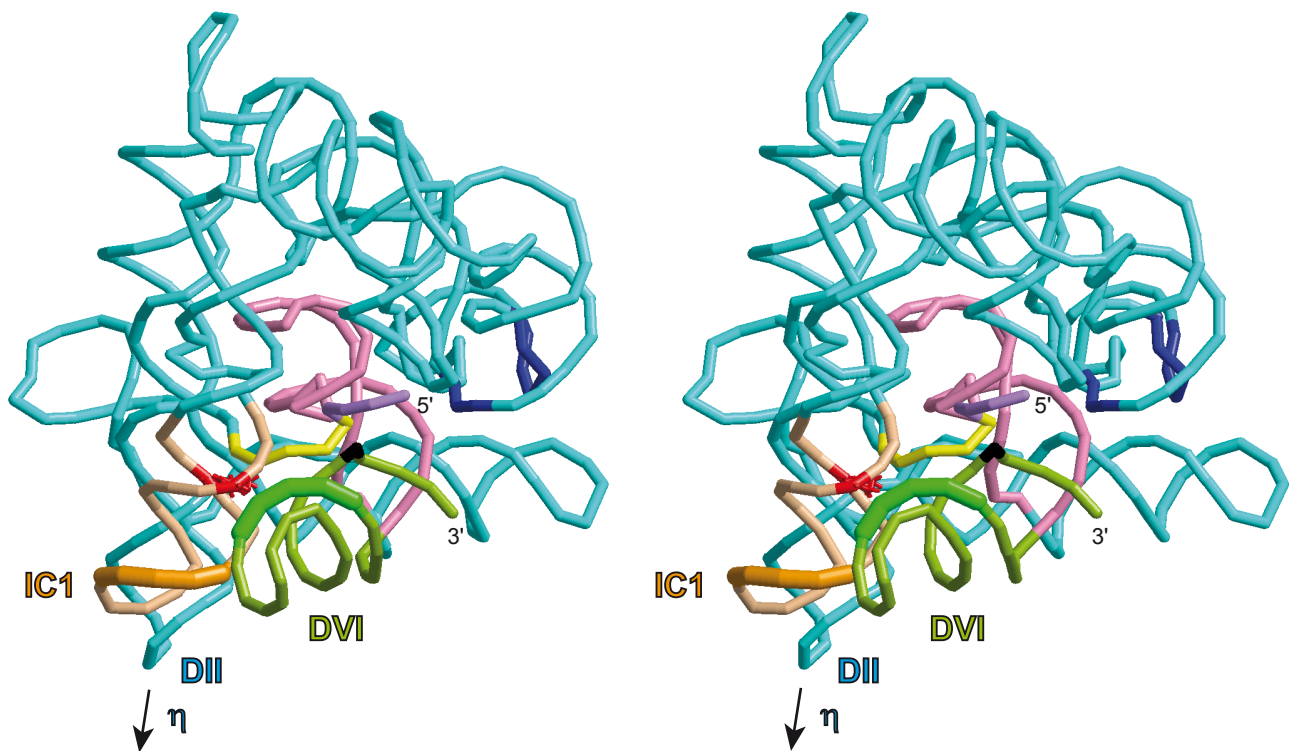
Figure 12. Modèle tridimensionnel de l'interaction entre les domaines VI et IC1 du ribozyme durant l'étape de branchement.

Ces vues stéréo ont été générées à partir d'un jeu de coordonnées du ribozyme de sous-groupe IIC d'*Oceanobacillus iheyensis* (Toor et coll., 2010), à l'exception du domaine VI, du segment de 3 nt qui relie DV et DVI et des deux premiers résidus de l'intron, qui ont été modélisés de novo. Couleurs: noir, adénosine du point de branchement; vert, domaine VI; rose, domaine V; violet, exon 5'; jaune, nt 1-5 de l'intron; ocre, sous-domaine IC1; rouge, paire de bases 79:100 (81:101 chez le ribozyme d'*Oceanobacillus ribozyme*); bleu foncé, 'coordination loop'.

Figures, Supplementary Figures and Table
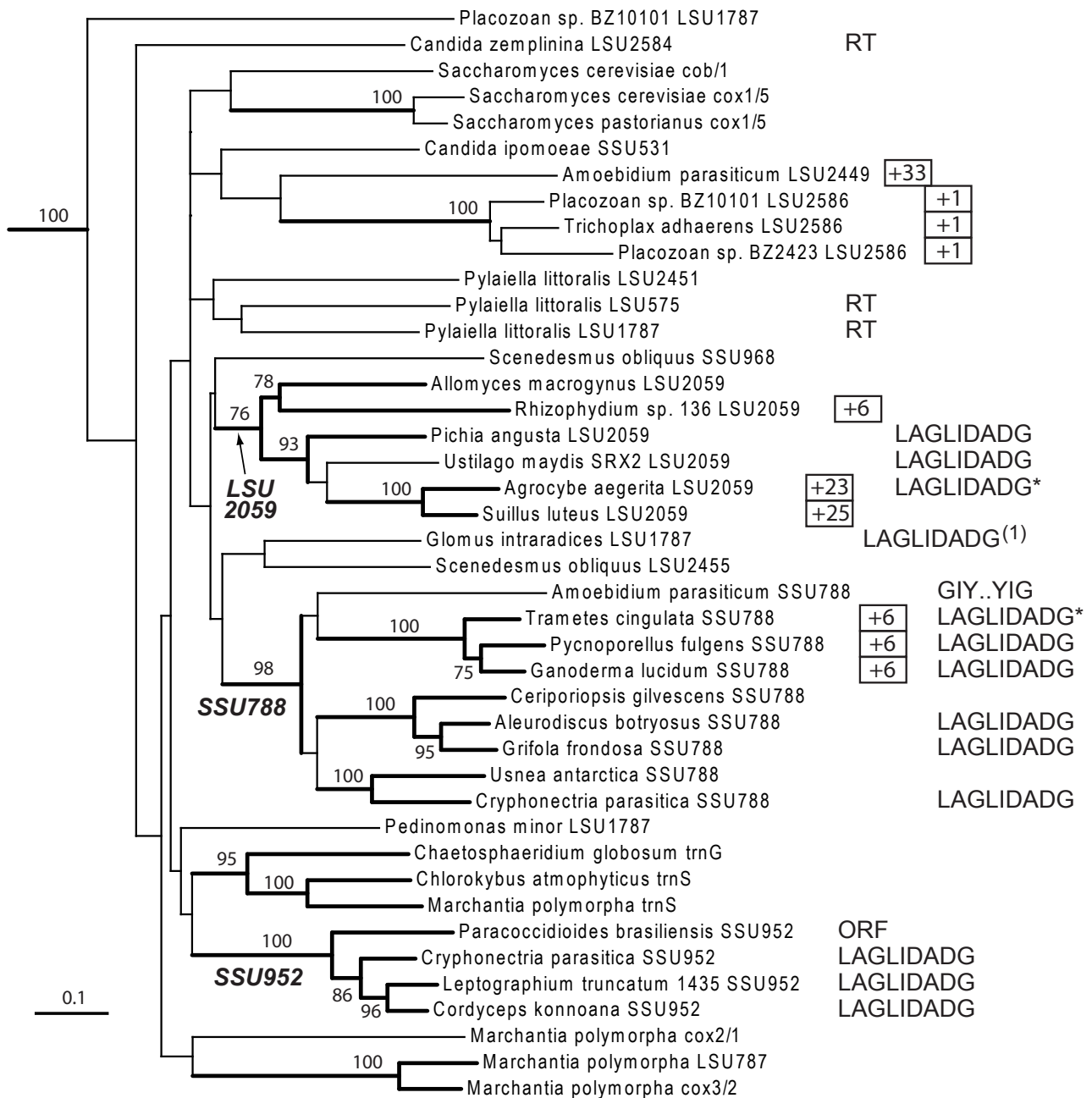
Figure 1. Phylogenetic tree of mitochondrial subgroup IIB1 introns.
Numbers next to nodes are bootstrap proportions (200 replicates) equal to or higher than 75 percent (corresponding branches are thickened). The roots of well-supported, major clades of ribosomal introns are indicated. The length of the 5' terminal insertion, when present, is provided at right of an intron name (boxed numbers preceded by + sign). RT, LAGLIDADG, GIY-YIG and 'Unknown' designate proteins potentially encoded by the introns. The cox1 introns from *P. brasiliensis* and *C. parapsilosis* are used as outgroups.
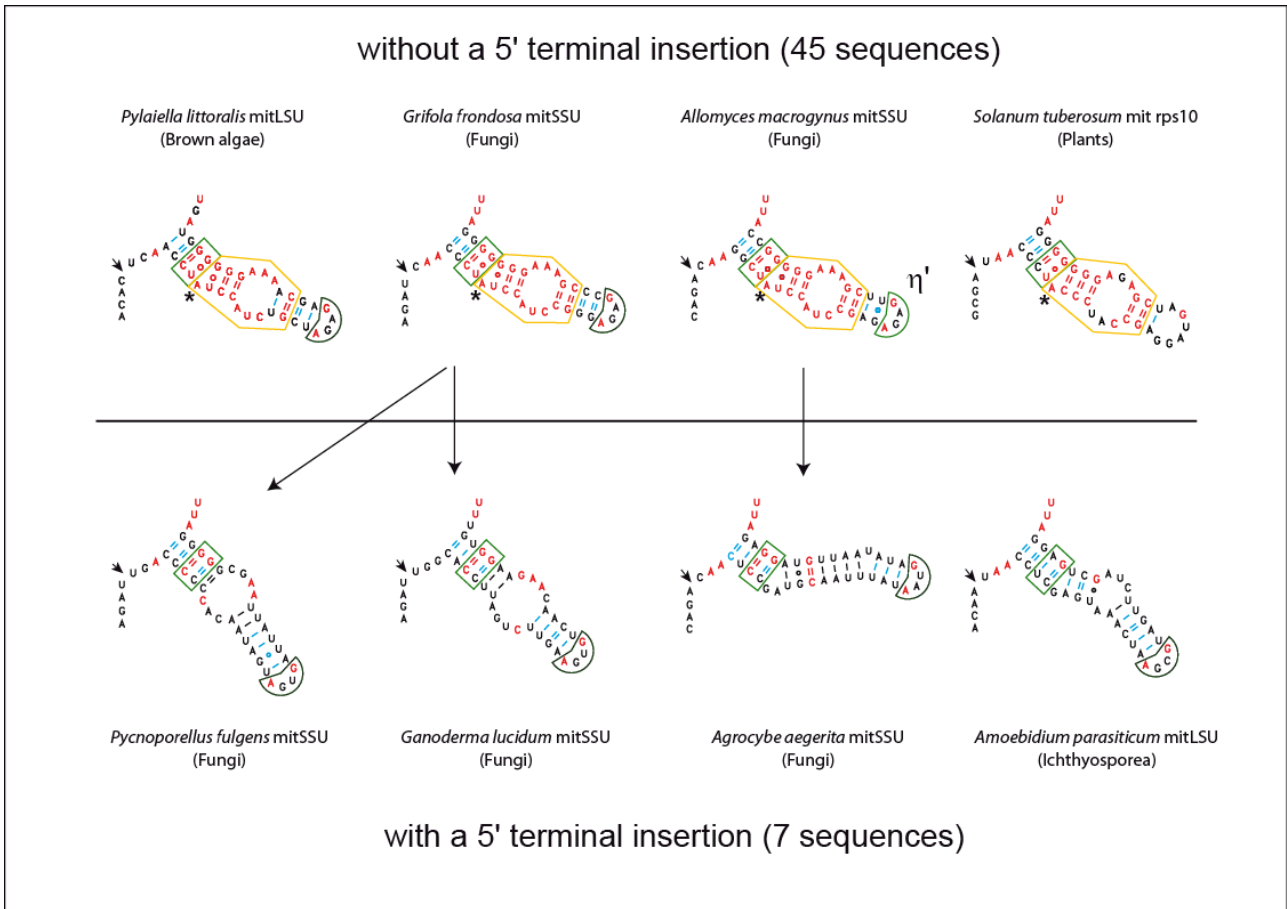
Sequence analyses

Figure 2. Variations in domain VI of introns with a 5' terminal insert.
The domain VI sequences of various introns with a 5' terminal insert were selected and compared with those of related introns without a 5' terminal insert. Nucleotides conserved in the subset IIB1 are in red; conserved hydrogen bonds are labelled in blue. The structure conserved in the middle part of the dVI stem in introns with no insert is framed by a yellow line and features conserved in all introns are boxed in green. The terminal loop structure, which follows the GNRA consensus, is also marked.
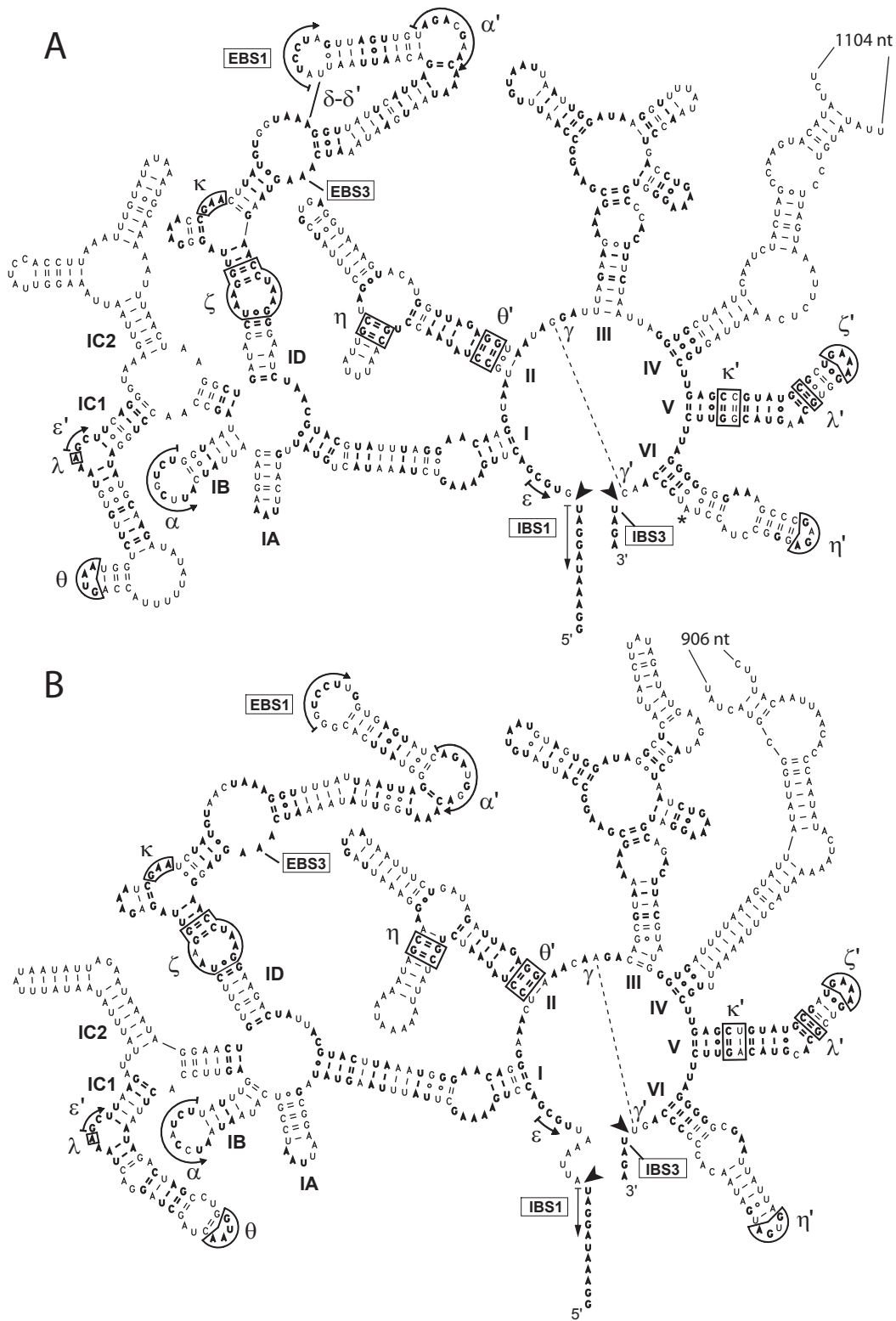
Figure 3. Secondary structure of (A) the *Grifola frondosa* SSU788 intron, which is used as the model of an intron without a 5' terminal insert. (B) the *Pycnoporellus fulgens* SSU788 intron, which stands as a model of an intron with a 5' terminal insert. The 5' terminal insert is defined by using the GUGYG consensus sequence and the IBS1 sequence as boundaries.
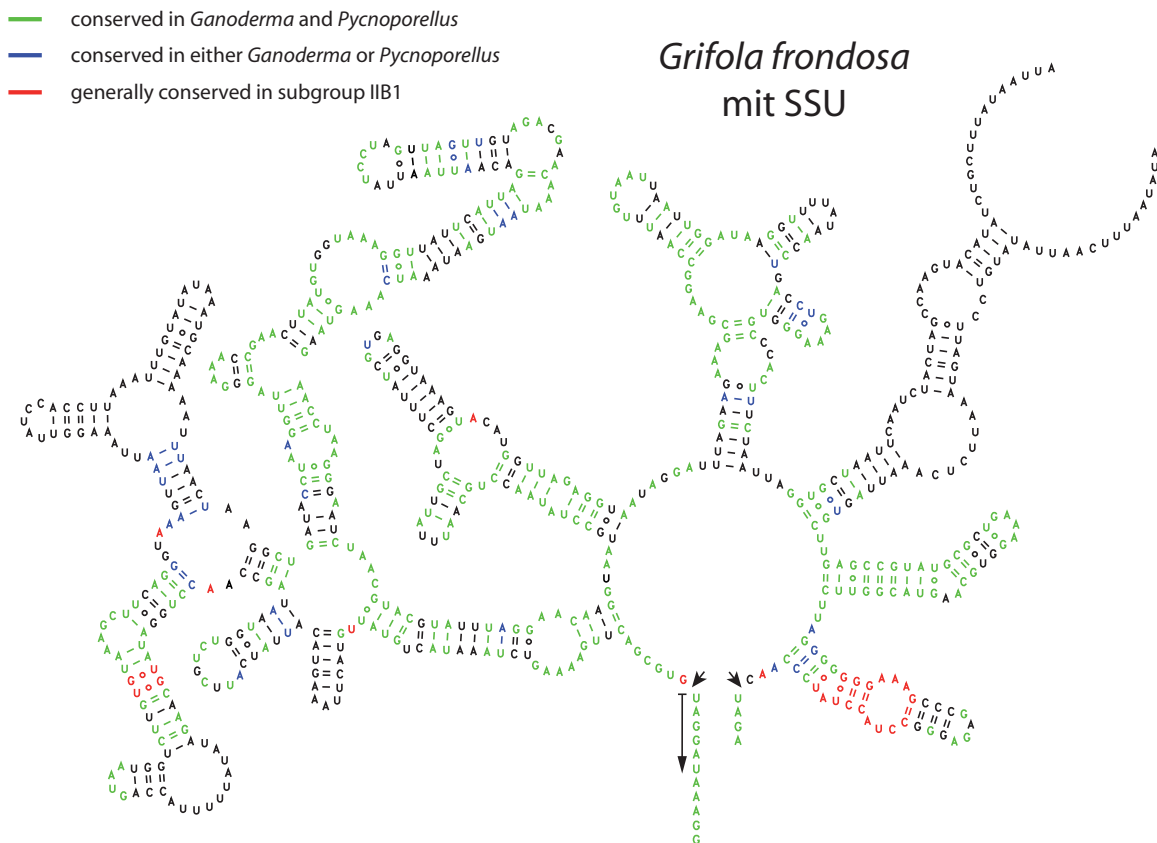
Figure 4. Superimposition of the secondary structure model of the *Pycnoporellus fulgens* SSU788 intron over the *Grifola frondosa* SSU788 intron. Light green nucleotides are conserved in both introns and other introns from the same subset. Red nucleotides are conserved as well in the IIB1 mitochondrial subset, but not in the *Pycnoporellus fulgens* SSU788 intron.
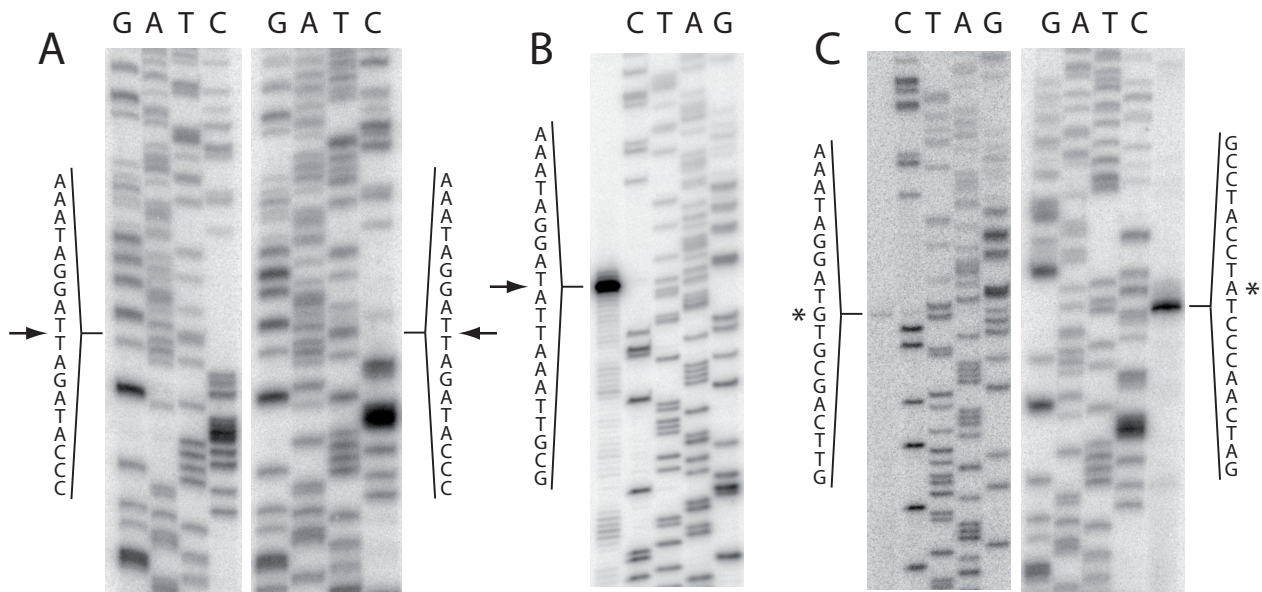
Figure 5. Mapping of intron-exon junctions and the branch site. Sequencing lanes are labelled by the base complementary to the dideoxynucleotide added.

A. Sequencing by reverse transcription of gel-extracted ligated exons; left panel, *Pycnoporellus*; right panel, *Grifola*. Arrows indicate splicing junctions.

B. Mapping of the 5' extremity of gel-extracted linear intron molecules generated by *in vitro* self-splicing of a *Pycnoporellus* precursor transcript; the latter was used as a template to generate the sequencing lanes at right with a primer located downstream of the intron 5' extremity. Elongation from the same primer using the excised intron molecules as template generated the strong stop in the lane at left; the arrow marks the 5' splice site.

C. Mapping of the branchpoint of gel-extracted lariat intron molecules generated by in vitro self-splicing of a *Grifola* precursor transcript. Left panel: elongation from a primer located downstream of the intron 5' extremity, the stop (marked by an asterisk) corresponds to the first intron nucleotide; sequencing lanes (at right) were generated by the same primer on a precursor RNA template. Right panel: elongation from a primer located in the 3' exon (intron-3'exon branched molecules were used as template), the asterisk marks the branch site (elongation stops on the nucleotide immediately 3' of the branch site); sequencing lanes were generated by the same primer on a precursor RNA template.
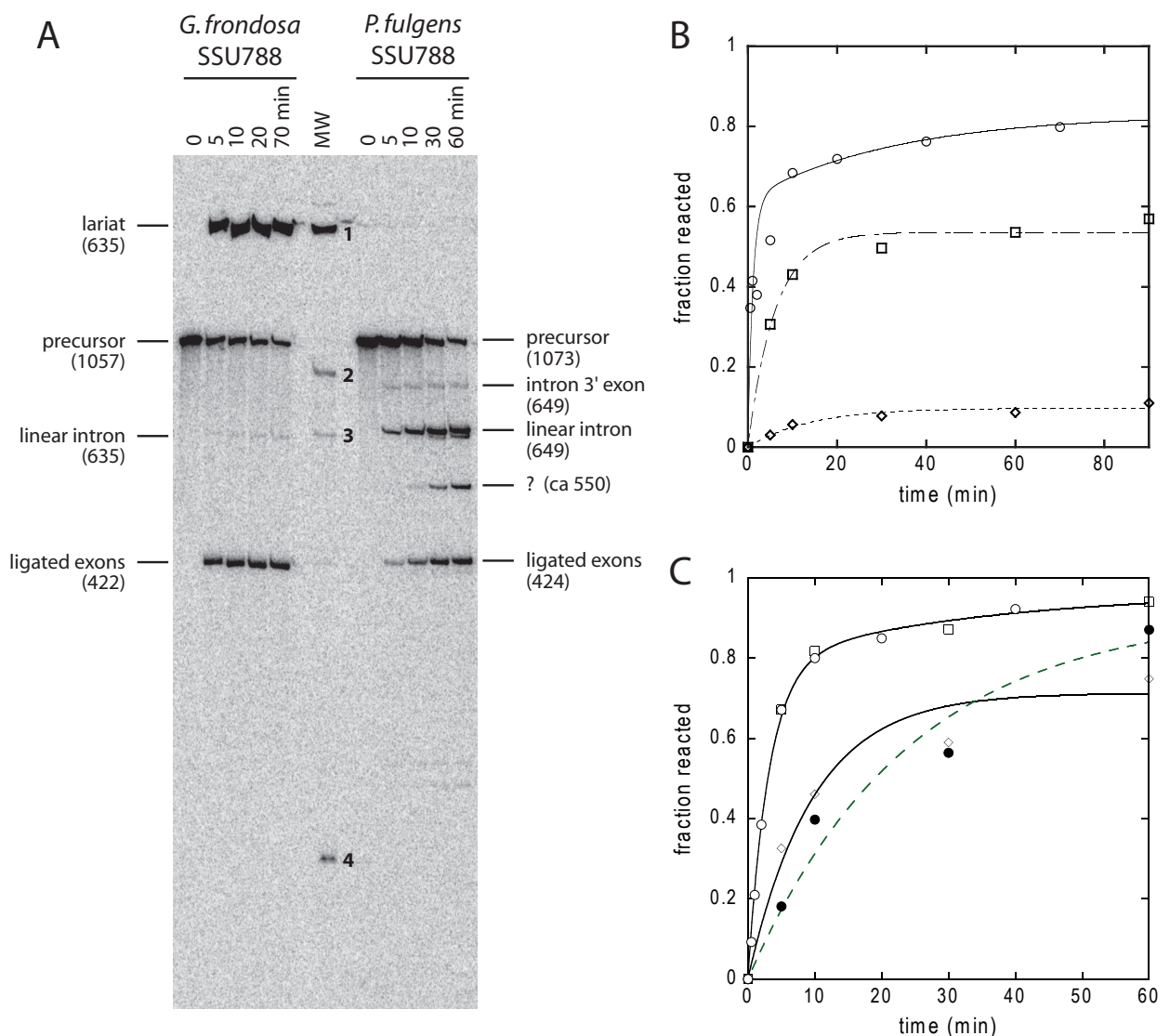
Figure 6. Self-splicing of the *Grifola* and *Pycnoporellus* SSU788 introns.
A. Time course of self-splicing reactions at 42 °C in 1M NH$_4$Cl, 20 mM MgCl$_2$, 40 mM Na-MES (pH 6.2). Electrophoretic mobilities are compared to those of known splicing products of a *Pylaiella littoralis* LSU1787 precursor transcript (MW lane: band 1, 640 nt, lariat; band 2, 872 nt, precursor; band 3, 640 nt, linear intron; band 4, 232 nt, ligated exons).
B. Time course of self-splicing reactions of a *Grifola* SSU788 precursor RNA at 42 °C in 40 mM Tris-Cl pH 7.5, 20 mM MgCl$_2$ and 1 M NH$_4$Cl (circles and solid curve, generated by a biphasic exponential fit with $k_1 = 0.9\pm0.2$ min^{-1} and $k_2 = 0.03$ min^{-1}) or 1 M KCl (squares and dashed curve, lariat intron; lozenges and dotted curve, linear intron; both from single exponential fits).
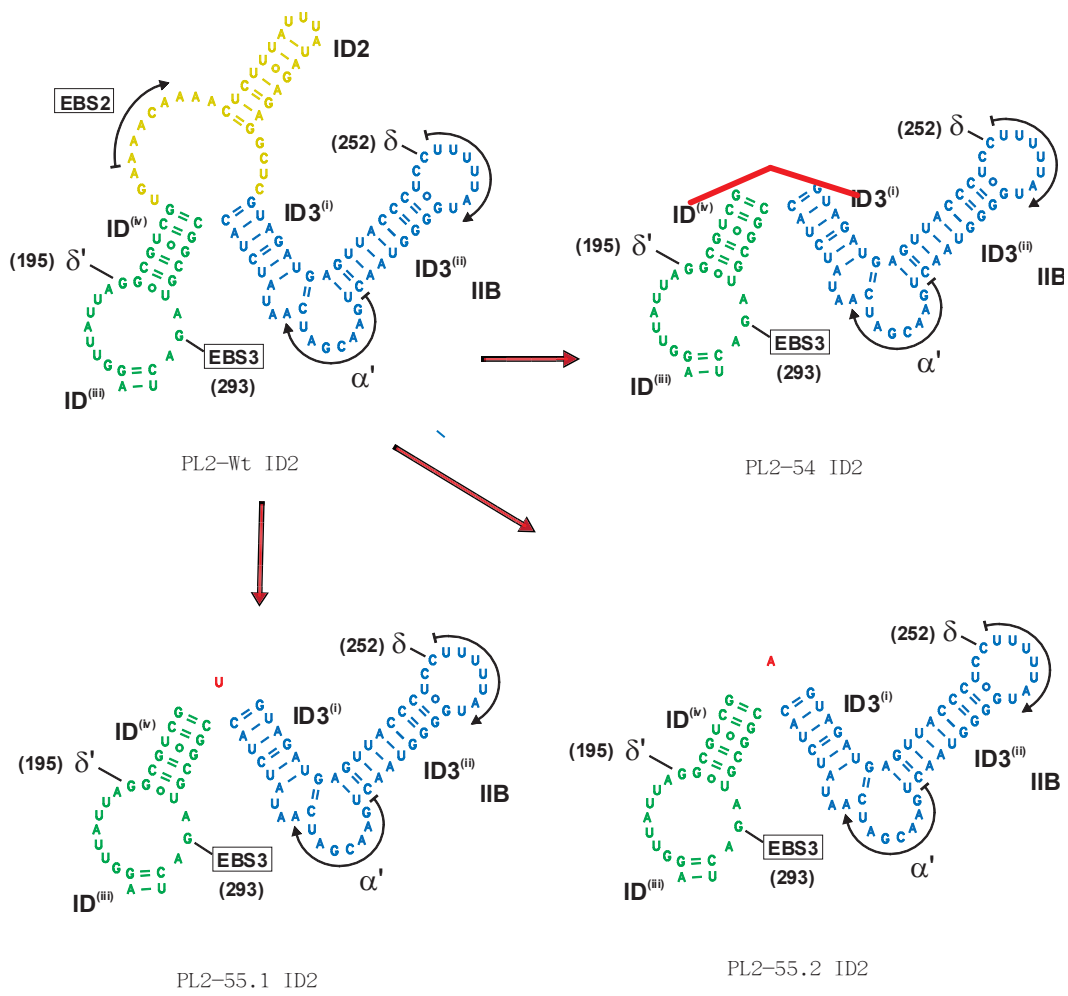C. Time course of self-splicing reactions of a *Pycnoporellus* SSU788 precursor RNA in 40 mM Tris-Cl pH 7.5, 1M NH$_4$Cl and 10 mM MgCl$_2$ (empty squares), 20 mM MgCl$_2$ (empty circles), 50 mM MgCl$_2$ (empty lozenges), or in 40 mM Na-MES pH 6.2 and 20 mM MgCl$_2$ (filled circles and dashed curve). Reactions at 10 and 20 mM Mg, pH 7.5, were fitted to a biphasic process ($k_1 = 0.32\pm0.03$ min^{-1}, $k_2 = 0.030\pm0.016$ min^{-1}), the other ones to simple exponentials.
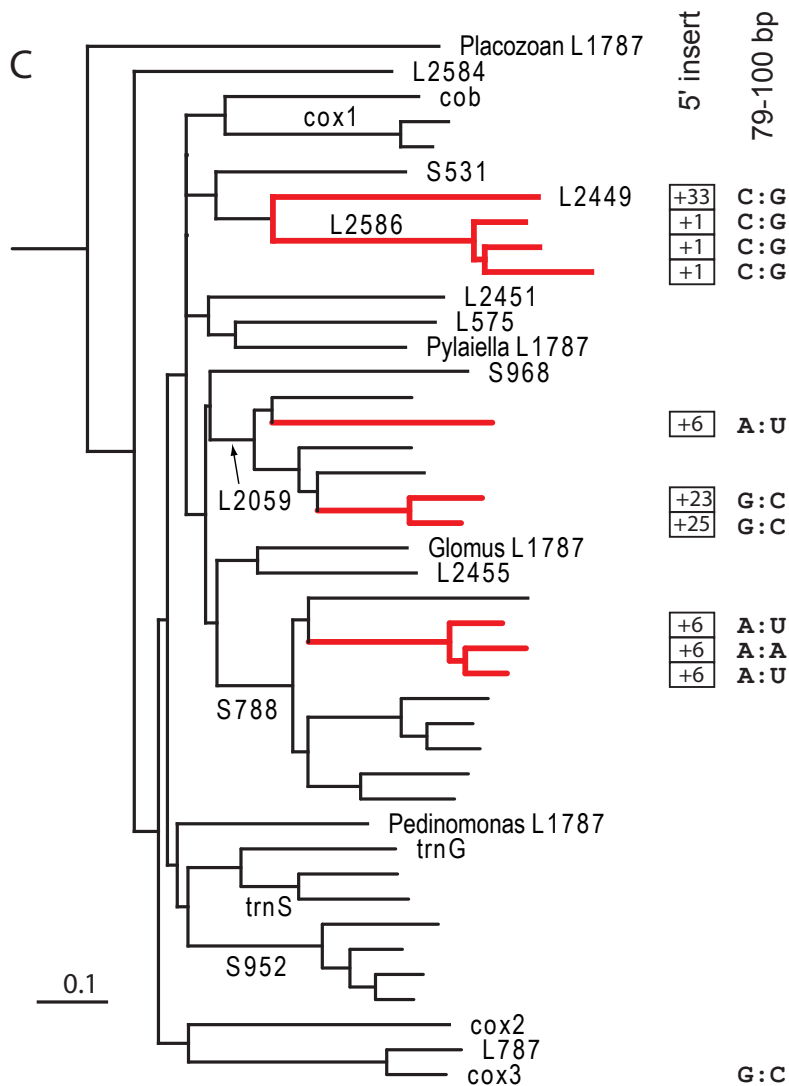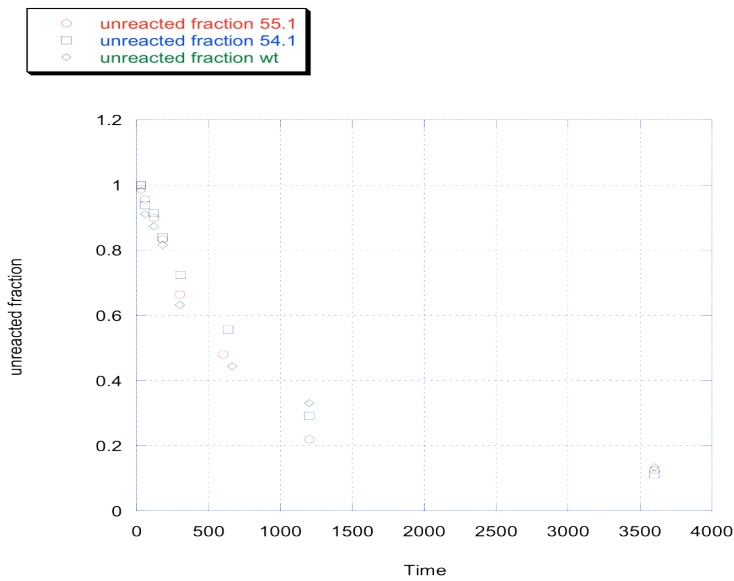
Figure 7. The phylogenetic tree of 42 subgroup IIB1 mitochondrial intron sequences. Thick red branches lead to the10 introns with 5'-terminal inserts. The colum marked '5' insert' indicates the number of nucleotides inserted between the 5' splice site and the GUGCGAC consensus sequence normally present at the intron 5' end. In the rightmost column, variants of the G79:U100 wobble pair in domain IC1 of some introns —mostly those with a 5' terminal insert – are also indicated. L and S designate the large and small subunit rRNA genes, respectively, and the following number corresponds to the site of insertion, according to *E. coli* numbering.
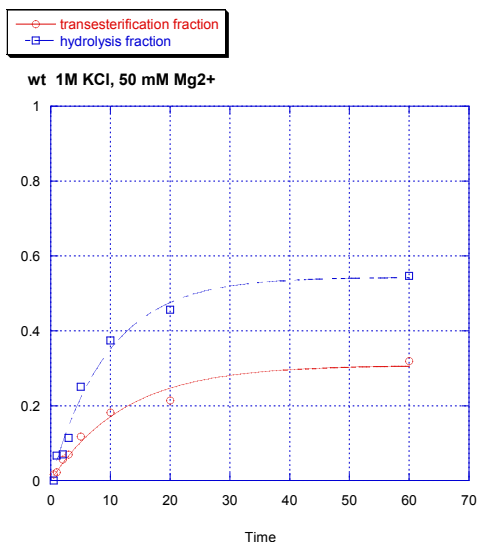
Figure 8. Secondary structure model of the Pl.LSU/2_ribozyme (Costa and Michel, 1999).
Intra- and interdomain tertiary interactions are labelled with greek symbols.
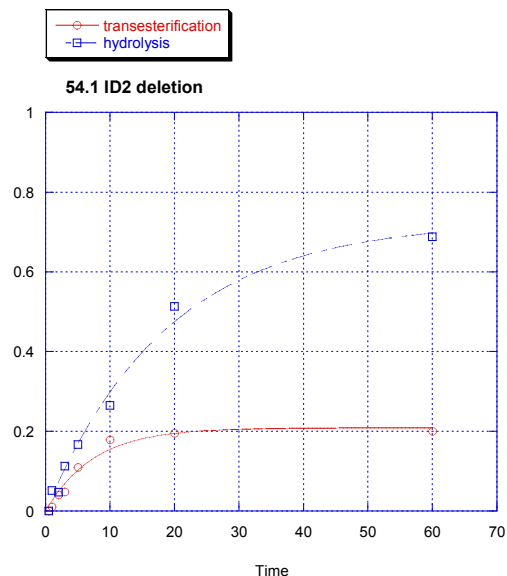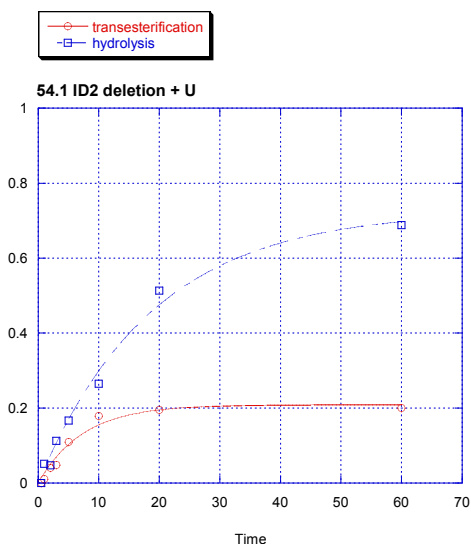
A

B

Figure 9-1. A. Ribozyme constructs of domain ID2 deletion mutants. B. Total unreacted fraction of wt and mutants as a function of time (in seconds).

Wt

ΔID2

ΔID2 +U

	K_{br} (1/min)	K_{hy} (1/min)	K_{br}/K_{hy}
wt	0.080	0.104	0.75
ΔID2	0.135	0.052	2.55
ΔID2 +U	0.135	0.053	2.84

Figure 9-2. Fraction reacted by branching and hydrolysis for the wild type and ID2 mutants as a function of time (in min). Rates were measured at 42 °C in 1 M KCl, 50 mM $MgCl_2$ and calculated as in Materials and Methods. Since PL2 55.1 and 55.2 show no difference, only one curve has been shown.
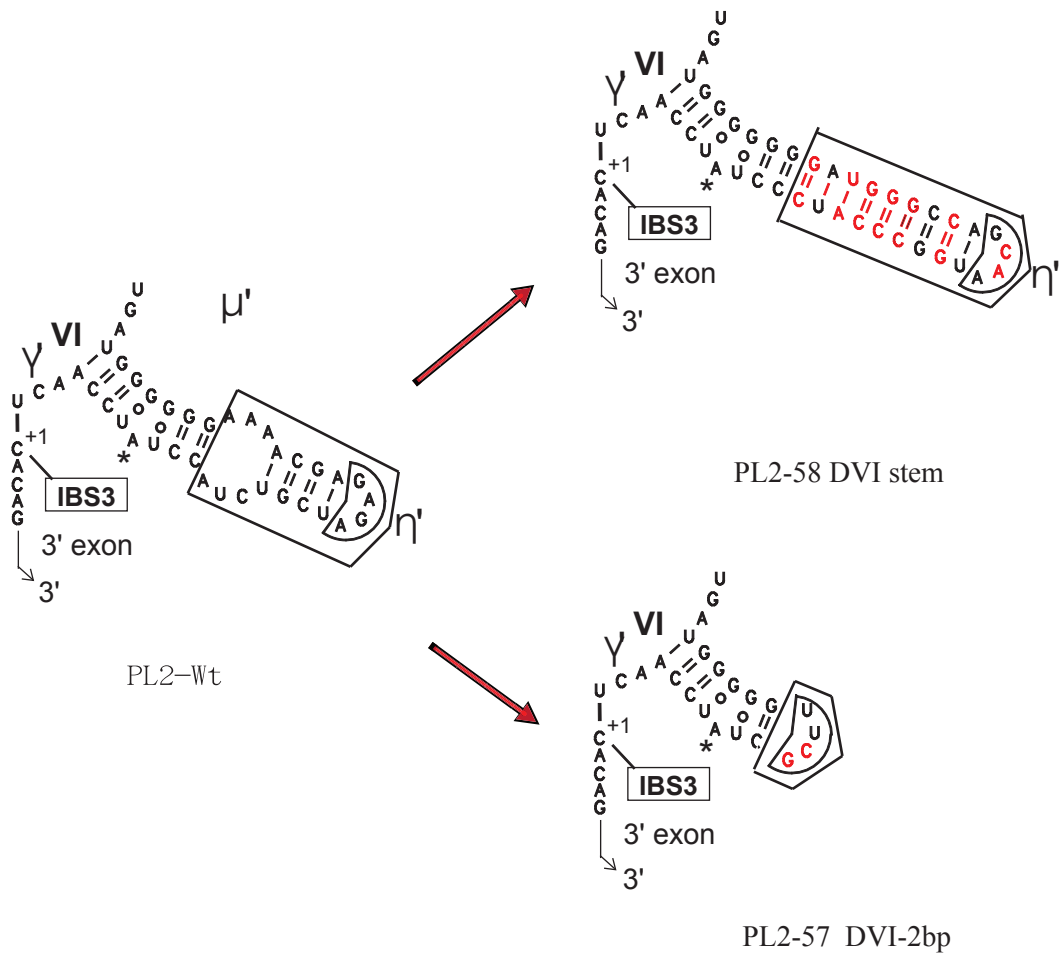
Figure 10-1. Ribozyme constructs of domain VI modification mutants. PL2-58 and PL2-57 (lab numbering) are also called DVI stem and DVI-2bp, respectively.
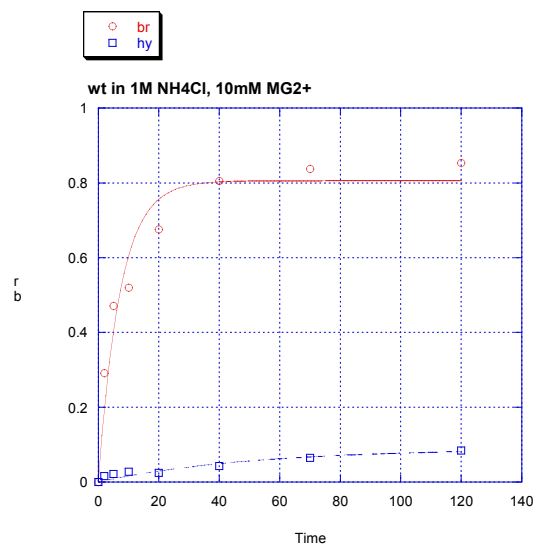
Wt
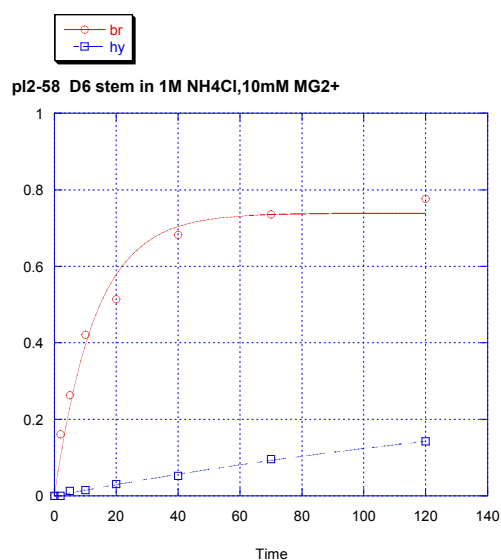
DVI stem

DVI_2bp

	K_{br} (1/min)	K_{hy} (1/min)
wt	0.139	0.0007
DVI stem	0.076	0.0015
DVI_2p	-	0.0034

Figure 10-2. Fraction reacted of branched and linear intron at 42 °C in 1 M NH₄Cl, 20 mM MgCl₂ for the PL2-wt precursor and domain VI modification mutants DVI stem and DVI-2bp mutants, respectively. The Table indicates the branching rate (k_{br}) and hydrolysis rate (k_{hy}) obtained by fitting reacted fraction data as a function of time.

*In order to compare the rate, the slope of the straight line is used to compare with initial rate of exponential curve.
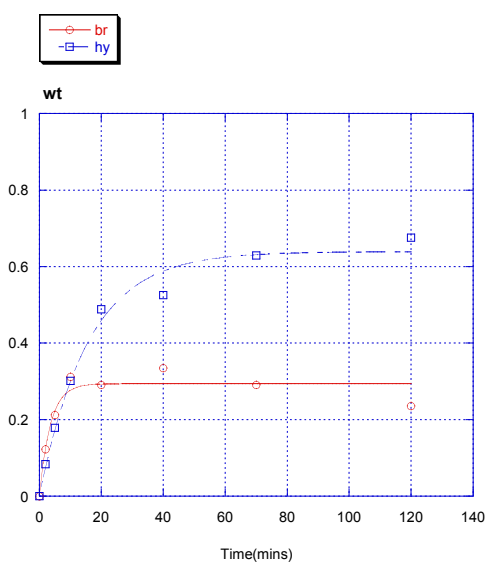
Wt
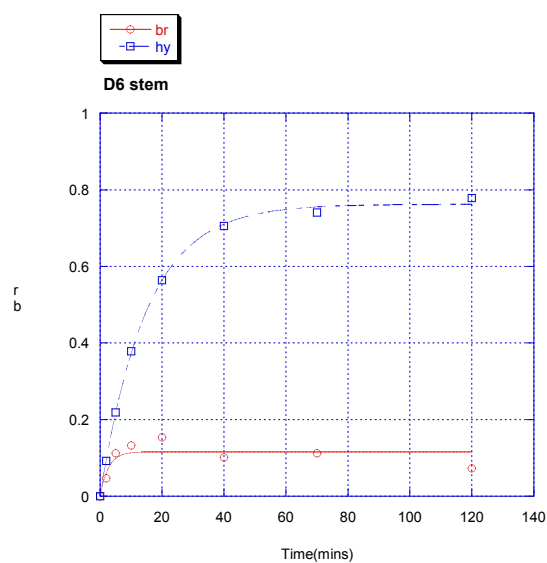
D6 stem
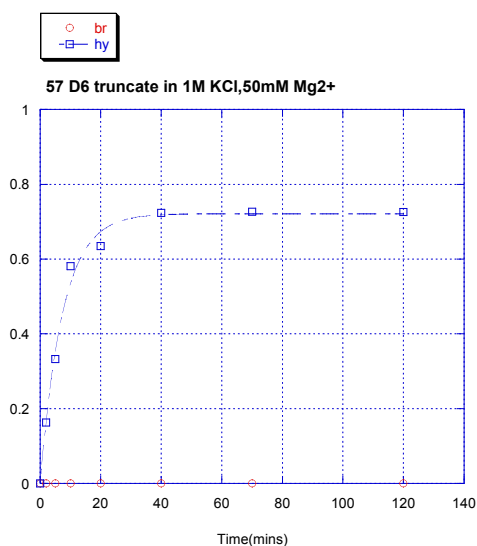
DVI-2bp

	Kbr（1/min ）	Khy（1/min ）	Kbr/Khy
wt	0.286	0.063	4.5
D6 stem	0.387	0.067	5.7
D6 truncate	0	0.1352	-

Figure 10-3. Fraction of branched and linear intron generated from PL2-wt and domain VI modification mutants DVI stem and DVI-2bp at 42 °C in 1M KCl, 50 mM MgCl₂ for. The Table indicates branching rates (k_{br}) and hydrolysis rates (k_{hy}) calculated from reacted fractions as a function of time.
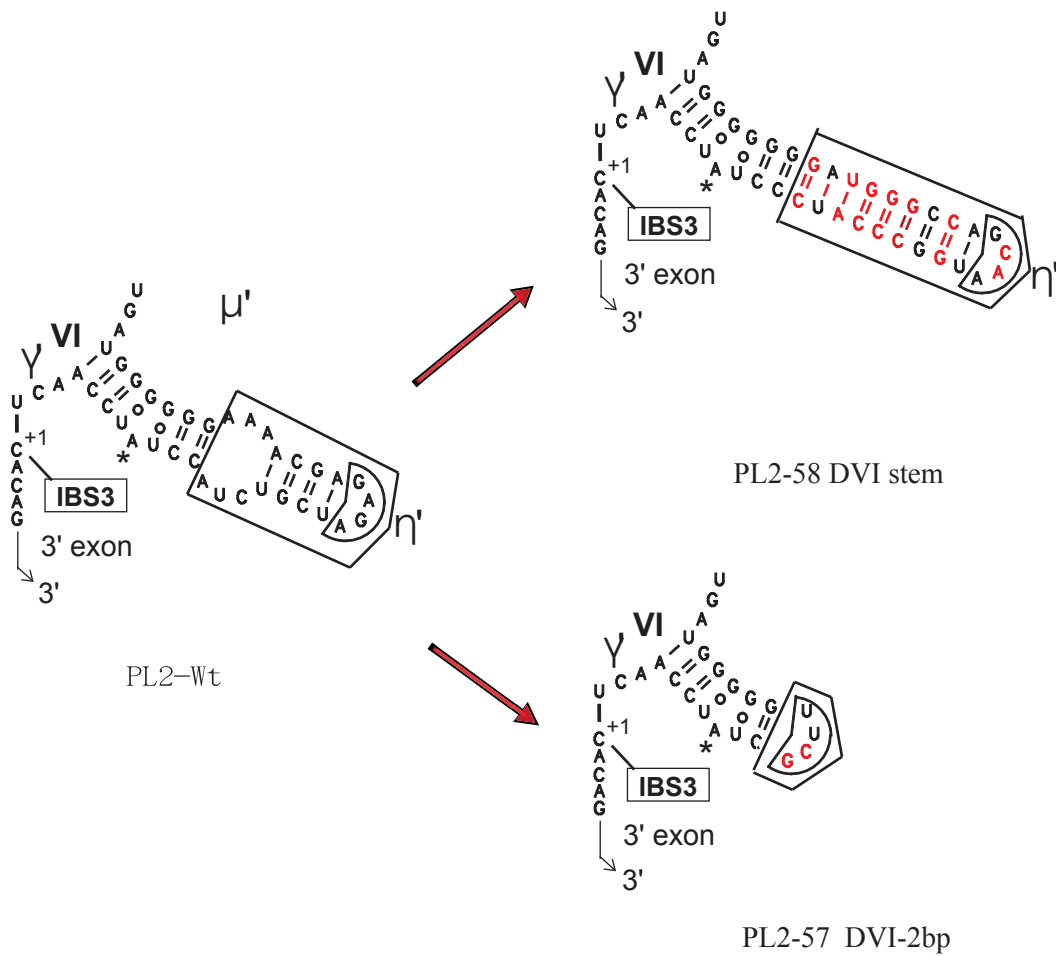
Figure 11-1. Ribozyme constructs of domain VI modification mutants. PL2-58, PL2-70 and PL2-71 (lab numbering) are also named DVI stem, DVI-4bp and DVI-7bp in our study.
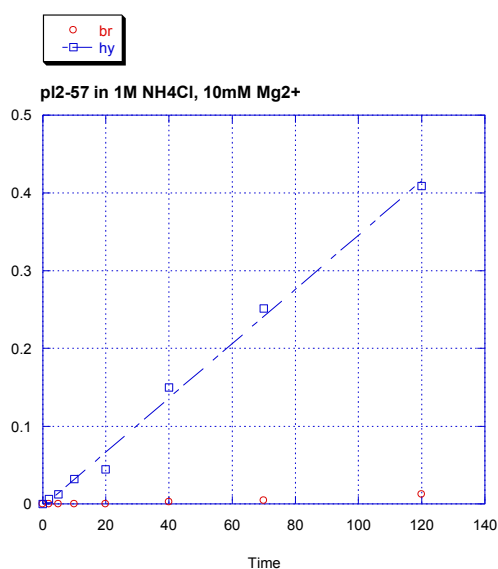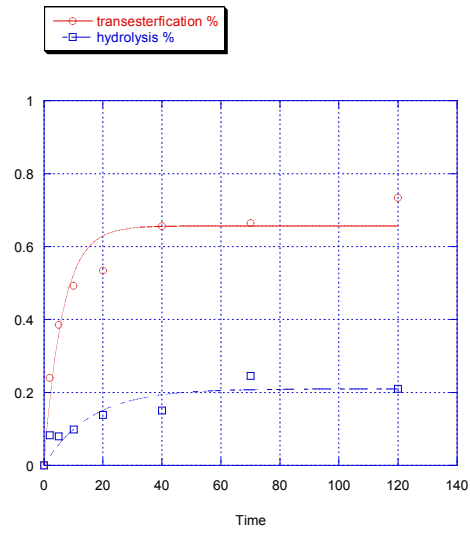
PL2-wt
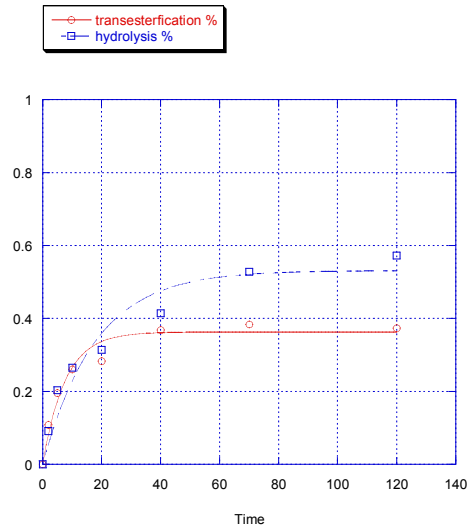
DVI-7bp
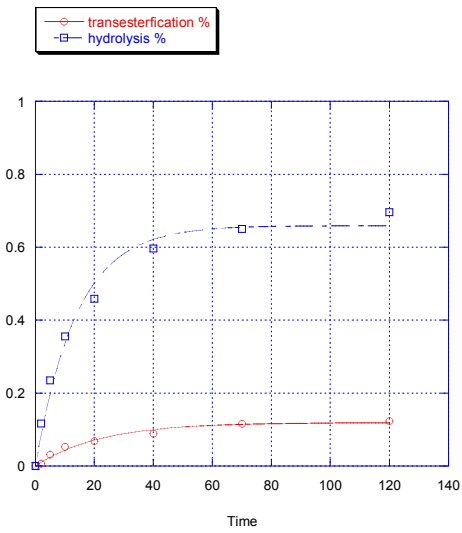
DVI-4bp

	K_{br} (1/min)	K_{hy} (1/min)	Ratio (K_{br}/K_{hy})
PL2-Wt	0.1599	0.0636	2.51
DVI-4bp	0.0045	0.0071	0.63
DVI-7bp	0.1302	0.0565	2.30

Figure 11-2. Fraction reacted of branched and linear intron from PL2-wt and domain VI modification mutants DVI stem and DVI-2bp at 42 °C in 1 M KCl, 20 mM MgCl$_2$. The Table indicates the branching rate (k_{br}) and hydrolysis rate (k_{hy}) calculated from the reacted fractions as a function of time.

PL2-Wt

PL2-4bp

PL2-7bp
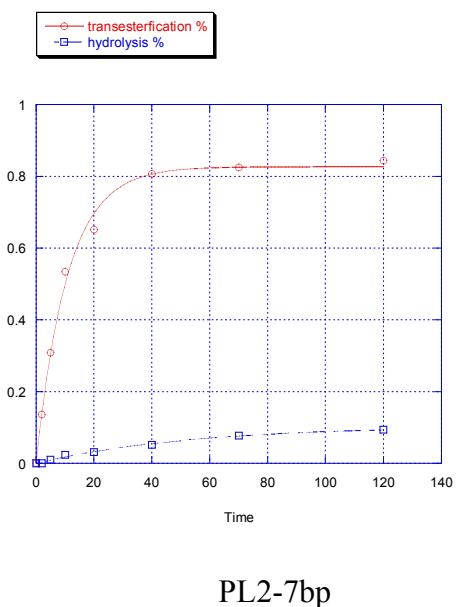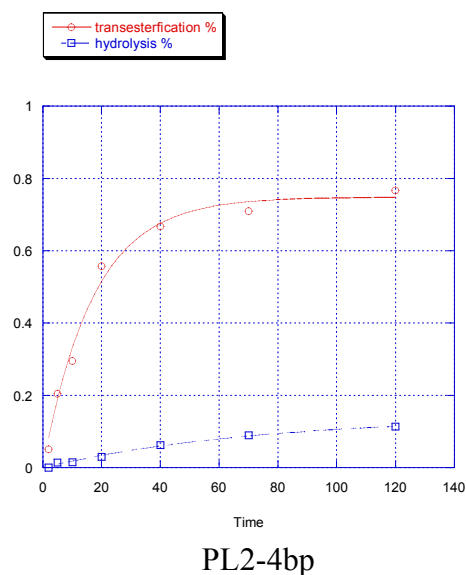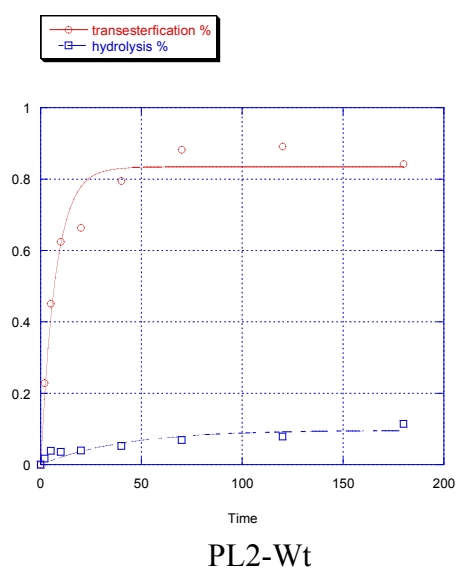
	K_{br} (1/min)	K_{hy} (1/min)	Ratio (K_{br}/K_{hy})
PL2-Wt	0.0859	0.0145	5.5
PL2-70 DVI-4bp	0.0883	0.0188	4.2
PL2-71 DVI-7bp	0.0925	0.0186	5.0

Figure 11-3. Fraction reacted of branched and linear intron from PL2-wt and domain VI modification mutants DVI stem and DVI-2bp at 42 °C in 1 M NH_4Cl, 20 mM $MgCl_2$. The Table indicates the branching rate (k_{br}) and hydrolysis rate (k_{hy}) calculated from reacted fractions as a function of time.
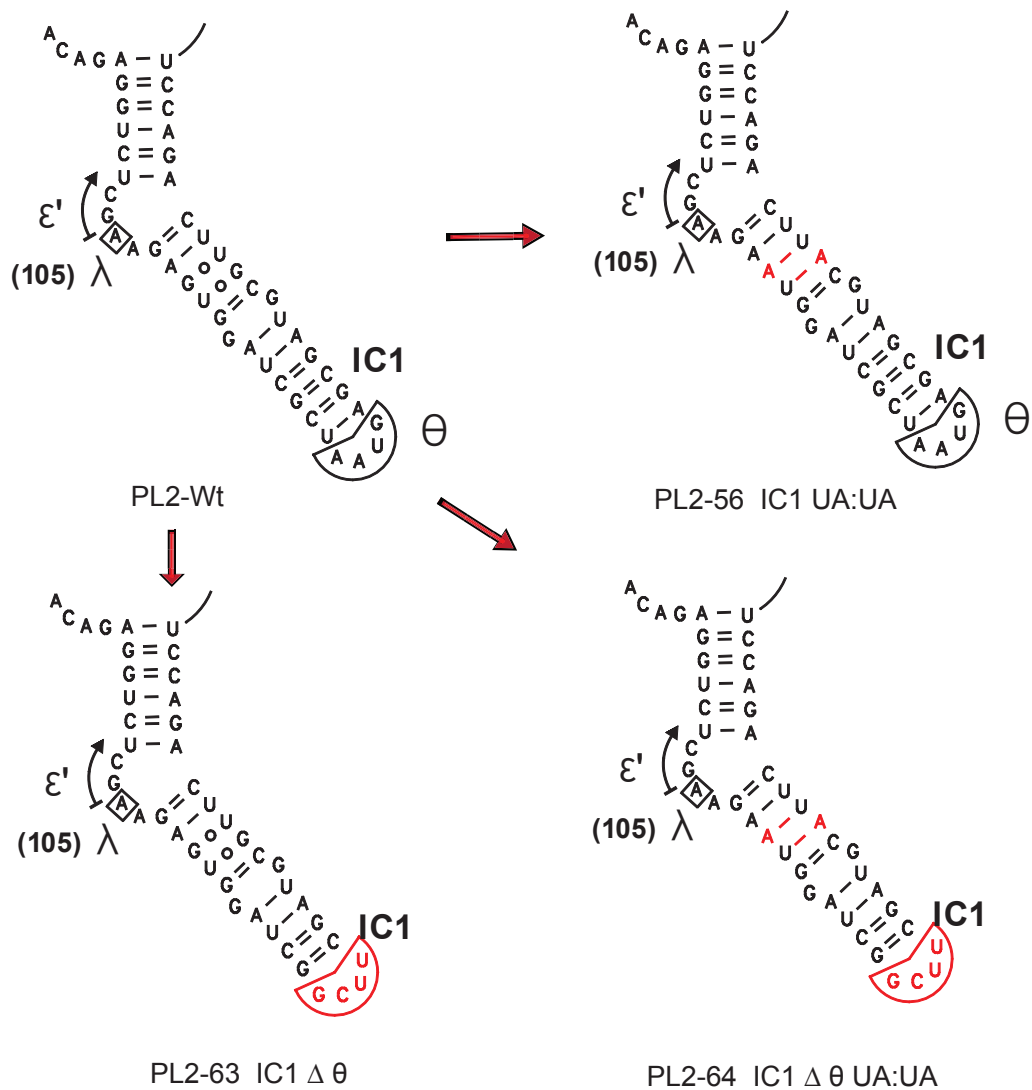
**Figure 12-1. Ribozyme constructs of domain IC1 modification mutants.** PL2-56 (IC1 UA:UA), PL2-63 (IC1 Δθ) and PL2-64 (IC1 Δθ UA:UA)
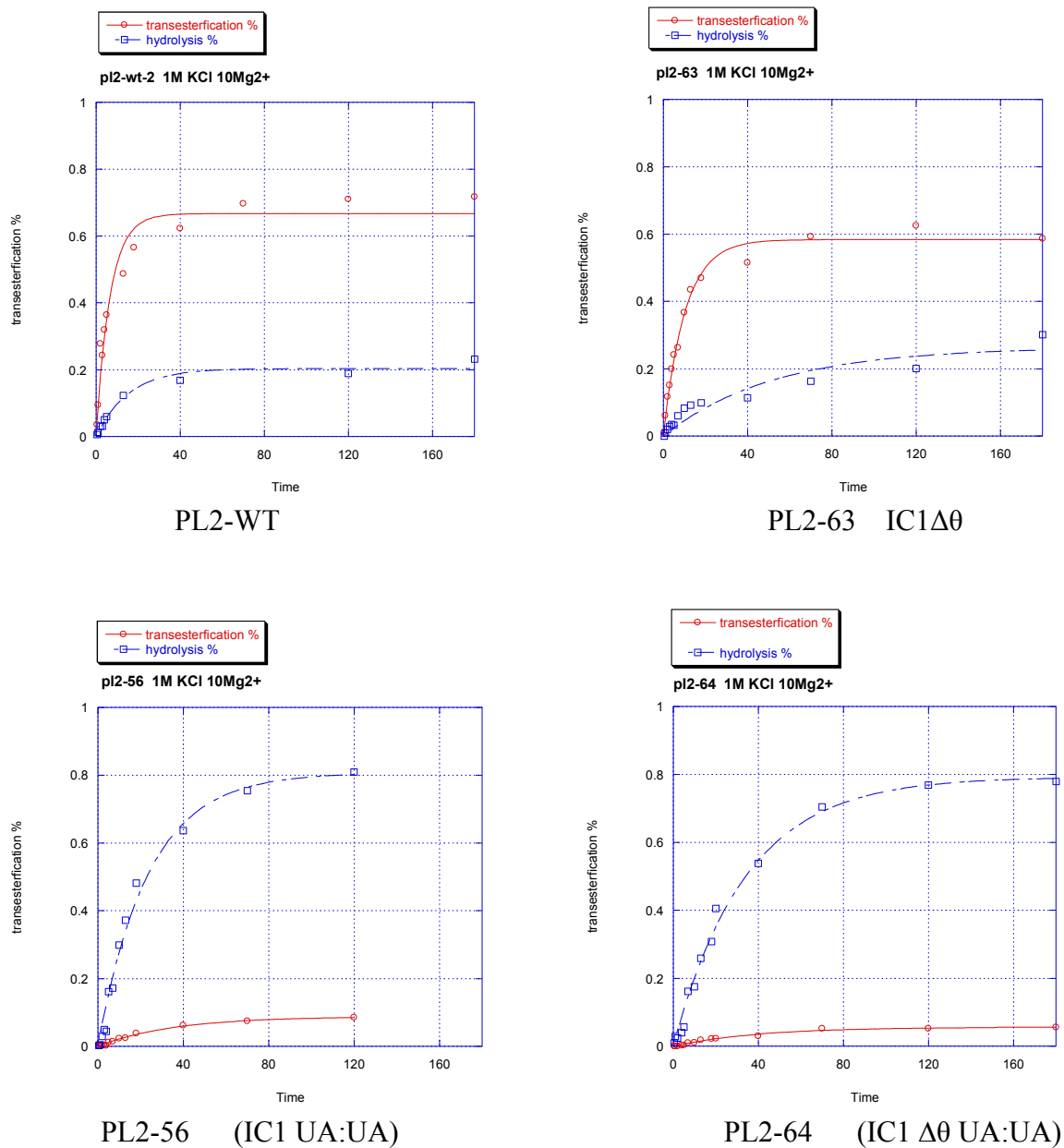
PL2-WT

PL2-63   IC1Δθ

PL2-56   (IC1 UA:UA)

PL2-64   (IC1 Δθ UA:UA)

	K_{br} (1/min)	K_{hy} (1/min)	K_{br}/K_{hy}
wt	0.149	0.064	2.3
IC1 Δθ	0.097	0.019	5.1
IC1 UA:UA	0.028	0.042	0.66
IC1 Δθ UA:UA	0.024	0.029	0.82

Figure 12-2. Fractions reacted of branched and linear intron from PL2-wt and domain IC1 modification mutants at 42 °C in 1 M KCl, 20 mM MgCl$_2$. The Table indicates the branching rate (k_{br}) and hydrolysis rate (k_{hy}) calculated from reacted fractions as a function of time.

PL2-65

PL2-66
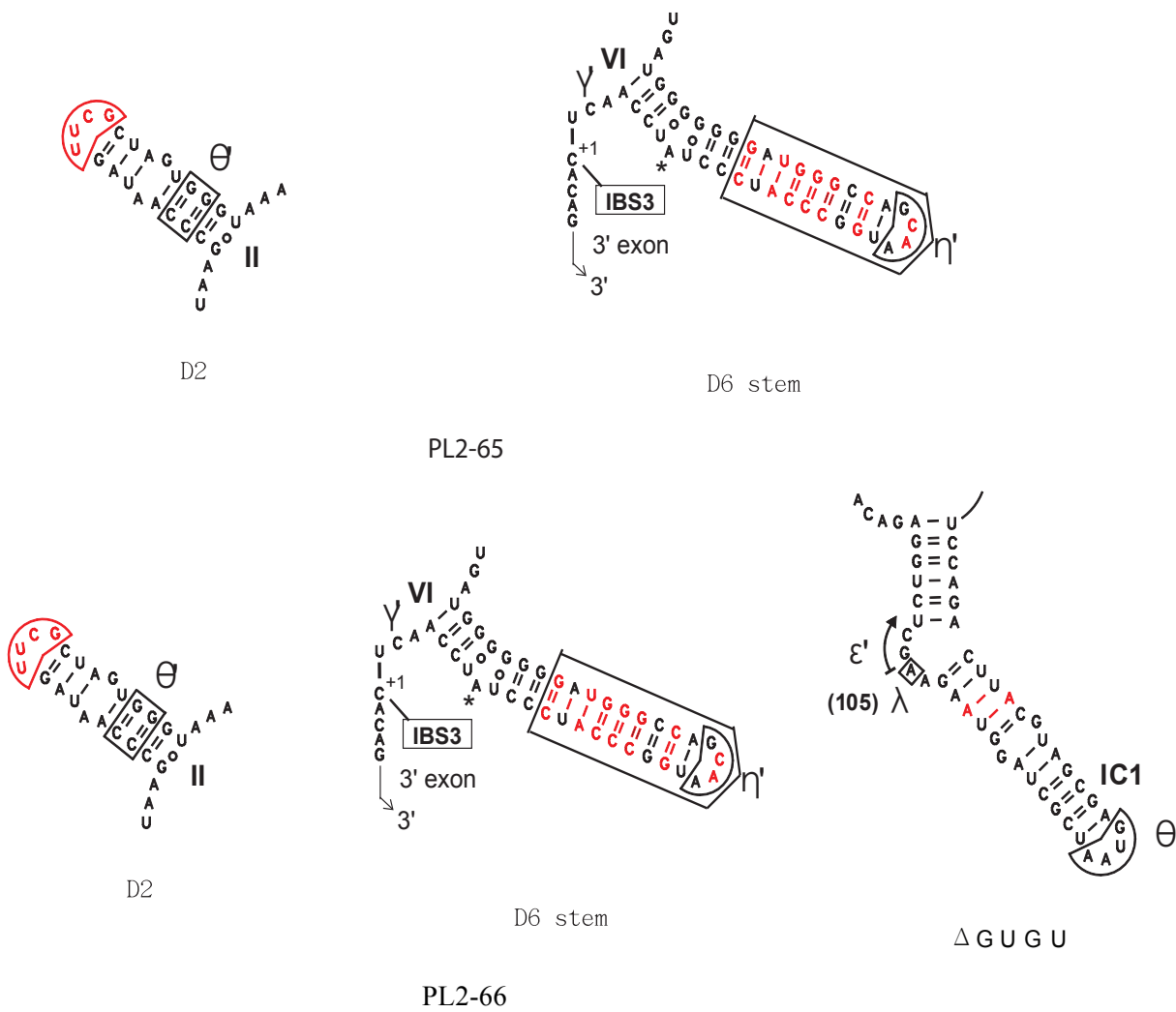
Figure 13-1. Ribozyme constructs of modification mutants used in our studies.

D2
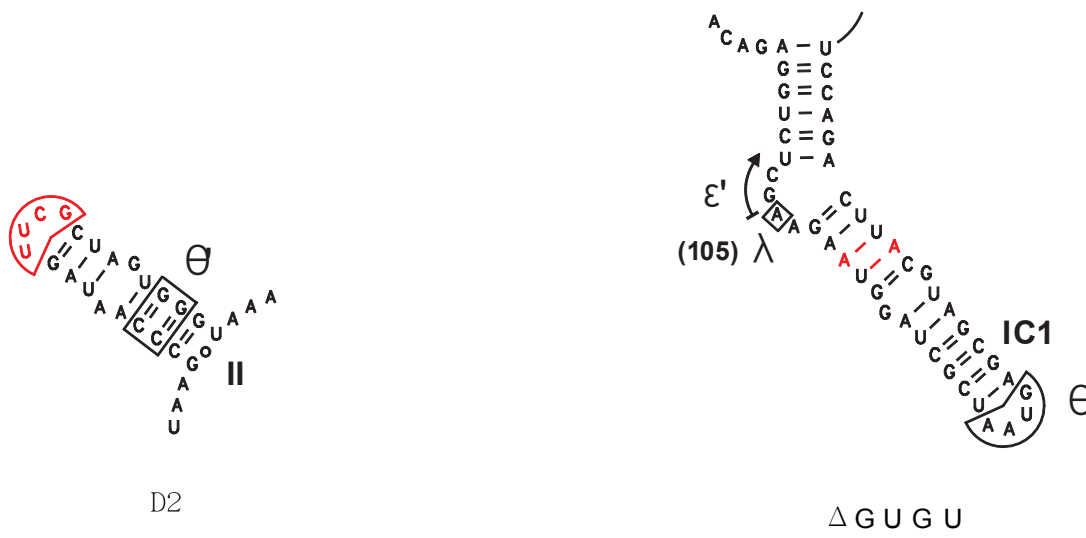
PL2-68
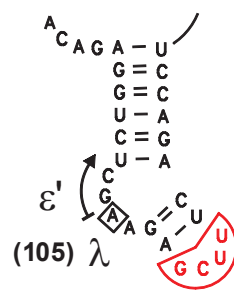
PL2−69

IC1-2bp

Figure 13-2. Ribozyme constructs of modification mutants used in our studies.
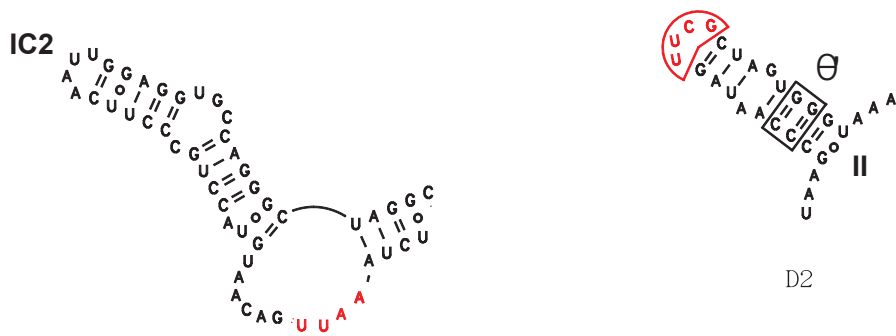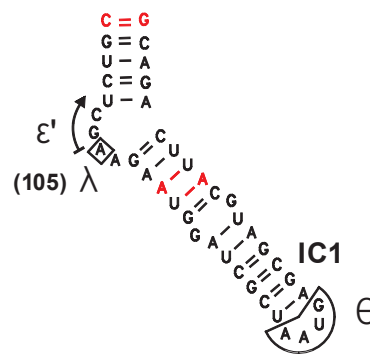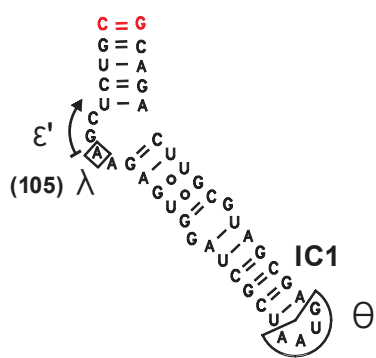
Figure 14. Ribozyme constructs used for our bimolecular setup.
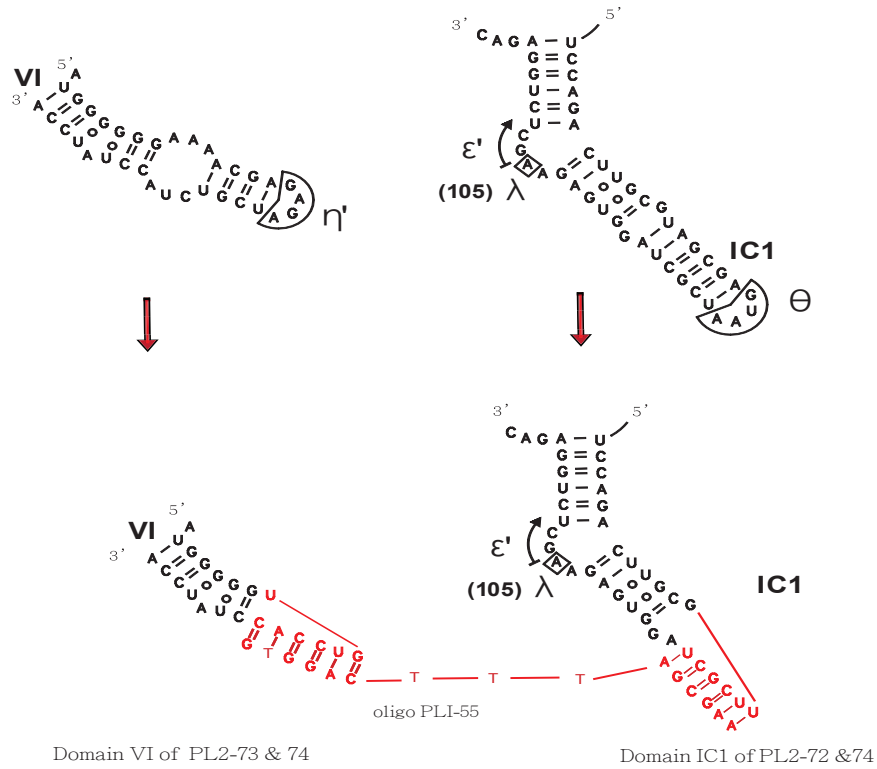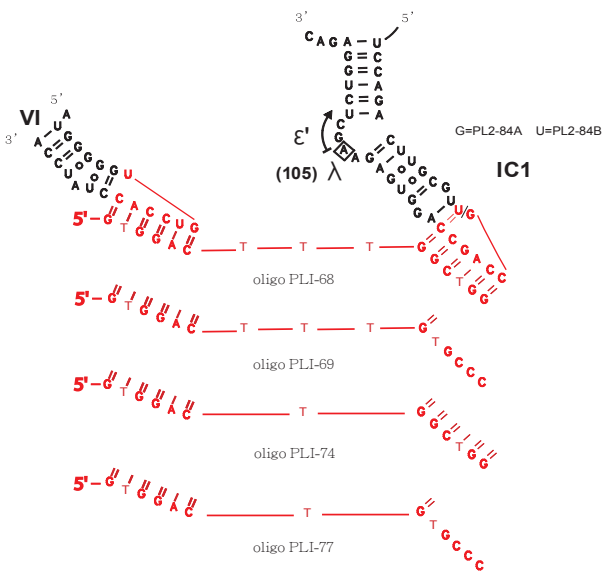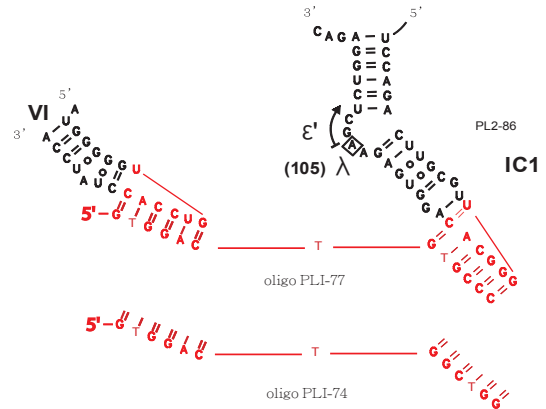
Figure 15. Principle of domain VI-IC1 oligo anchoring experiments.
A. Domain VI and IC1 sequences are modified so as to adapt to the oligonucleotide used as a linker.
B. A construct modified to allow pairing of both oligonucleotide handles provides more efficient branching of the intron. This construct was also tested with oligonucleotides with different linker lengths and a mismatched IC1 sequence. C. The domain IC1 sequence of this construct is modified to pair with the mismatched oligonucleotide used in the PL2-84B tests. In this construct, originally matched oligonucleotide PLI-74 no longer matches the sequence of domain IC1
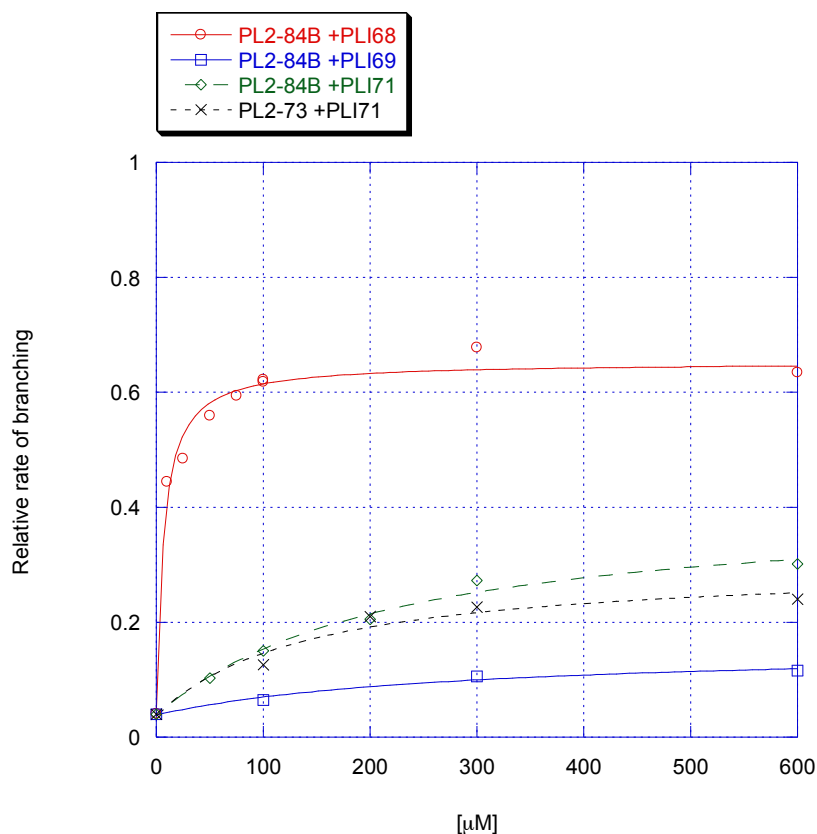
Figure 16. Relative branching fractions of Domain VI-IC1 constructs as a function of the concentration of a bridging oligonucleotide. The calculation method is described in Materials and Methods. PL2-84B show a higher relative branching rate with matched oligo PLI68 (3T) showing the highest ratio. When the oligo is replaced by mismatched oligo PLI-69, both the branching rate and fraction are found to decrease. The PL2-73 + PLI71 combination is an experimental control, in which only DVI is pairing with the oligo. PL2-73 has the same DVI sequence as PL2-84B, but a wt domain IC1. PLI-73 is a 7-mer which matches the DVI sequence of PL2-84B. Pl2-84B + PLI71 is the corresponding control.
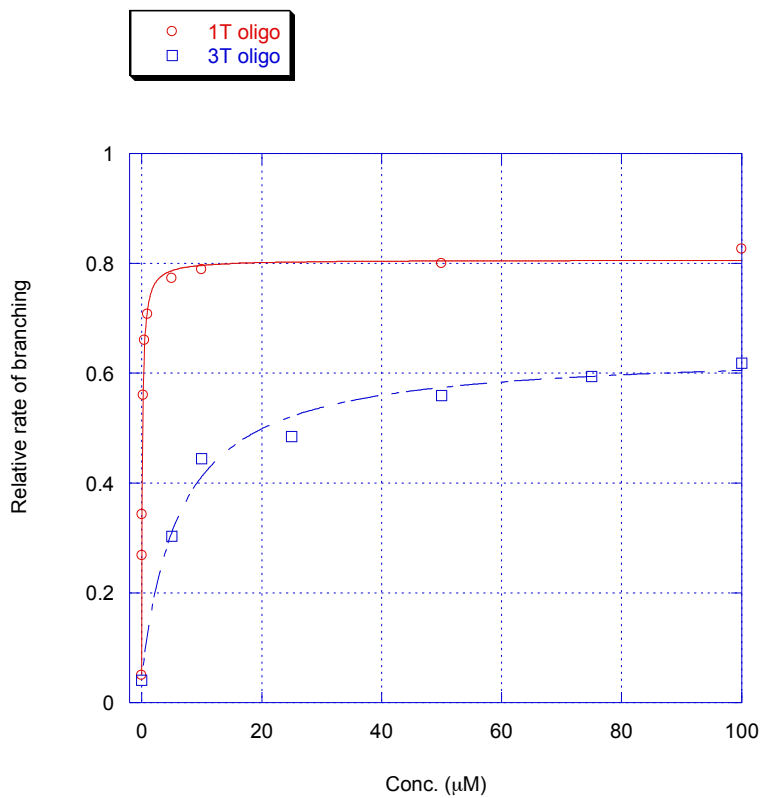
Figure 17. Relative branching fractions of Domain VI-IC1 constructs as a function of the concentration of complementary oligos. The branching reaction of a PL2-84B precursor is progressively restored with increasing concentrations of an oligonucleotide with a 1-T or a 3-T linker.
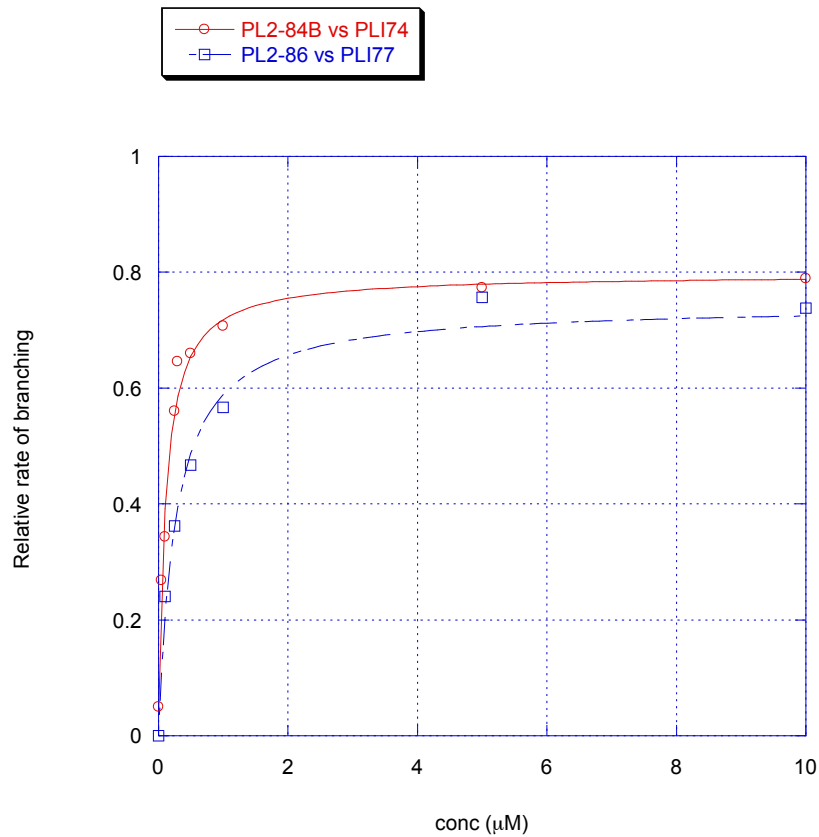
Figure 18. Relative branching fractions of Domain VI-IC1 constructs as a function of the concentration of a complementary oligo (see construct C in figure 15). PL2-84B with the IC1 sequence-mismatched oligo PLI77 failed to undergo an efficient branching reaction, but the same oligo can restore the branching reaction of PL2-86 (matched) almost as well as for the PL2-84B + PLI74 combination.
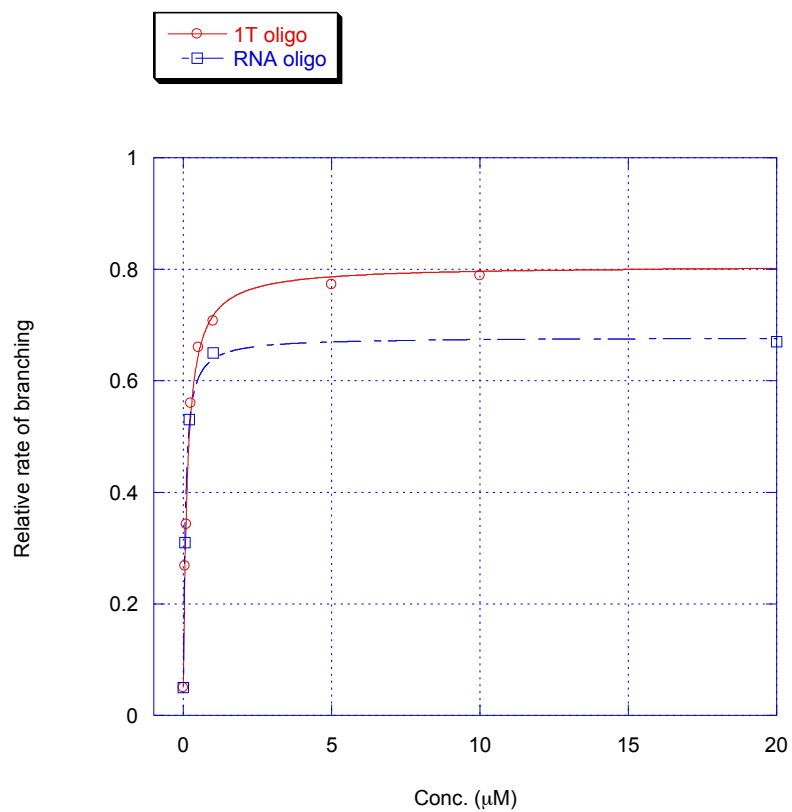
Figure 19. Relative branching fractions of Domain VI-IC1 constructs as a function of the concentration of a complementary oligonucleotide. Both a 1-T oligo (PLI-74) and an RNA oligo (PLI-79) can restore the branching reaction of PL2-84B.
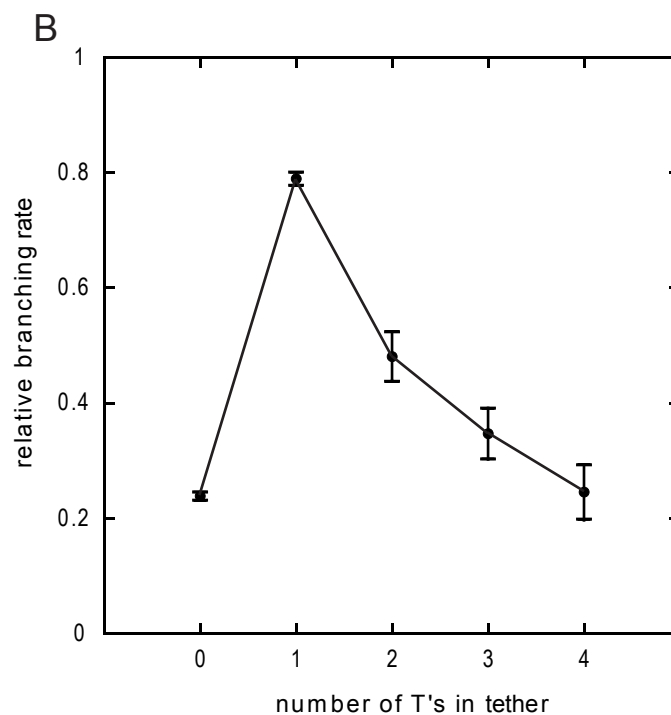
Figure 20. Relative branching rate of PL2-84B as a function of the number of linker T's in oligonucleotide 5'-GTGGAC[T]$_n$GGCTGG. The concentration of oligonucleotide was set at 5.0 μM, close to the observed K_m for the 3-T bridging oligo PLI-68.
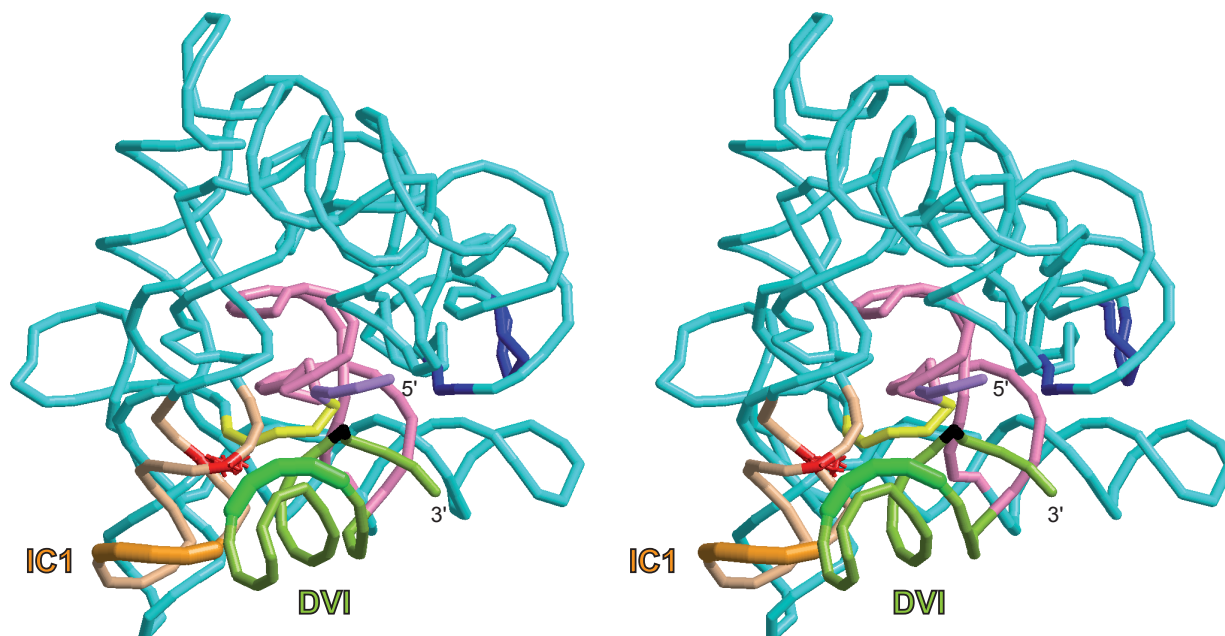
Figure 21. Proposed three-dimensional model of the interaction between ribozyme domains VI and IC1 during the branching step.

Stereo views generated from the coordinate set of Toor et al. (2010) for the *Oceanobacillus iheyensis* subgroup IIC ribozyme, except for domain VI, the 3-nt DV-DVI linker and intron residues 1-2, which were modeled de novo. Color scheme: black, branchpoint adenosine; green, domain VI; pink, domain V; violet, 5' exon; yellow, intron nt 1-5; tan, subdomain IC1; red, bp 79:100 (81:101 in the *Oceanobacillus ribozyme*); deep blue, "coordination loop".

Figure S1. A. Structure of a representative bacterial IIA1 intron, variations in IIB and IIC introns is marked in circles. The "loop" of DIV, which encodes the IEP, is depicted by dashed lines, a high-affinity binding site for the IEP. Subdomains discussed in the text are labeled. (Figure from Lambowitz and Zimmerly., 2010)

*Compared to IIA introns, major differences in other subgroups include structural features of DV (IIC introns); different $\sum''$ motifs (IIB, IIC); the number of base pairs in the κ -stem-loop (IIC); a coordination loop containing EBS3 and δ' (IIB, IIC); the absence of the DId(iii) stem-loop (IIB, IIC); the absence of a stem in the EBS2 motif (IIB, IIC); a unique $\zeta - \zeta'$ motif (IIC); and the ω−ω' interaction (IIC, some IIB).

Figre S2. Base-pairing interactions used by IIA, IIB, and IIC introns to bind the exons at the active site. EBS, exon-binding site; IBS, intron-binding site. (a) Subgroup IIA. Two terminal loops of the ribozyme secondary structure (EBS1 and EBS2) bind to the last ~12 nucleotides of the 5′ exon (IBS2 and IBS1) by canonical (Watson–Crick and wobble) base pairing. The binding site for the 3′ exon ('δ' or 'EBS3'; up to two base pairs might be involved is located immediately 5′ of EBS1. **(b)** Subgroup IIB. Binding of the 5′ exon occurs as in subgroup IIA, except that EBS2 is part of an internal, rather than terminal, loop. The first nucleotide of the 3′ exon (IBS3) is base paired to the EBS3 site. EBS3 is part of an internal loop that is tethered to the EBS1-carrying loop by the δ–δ' base pair. **(c)** Subgroup IIC. Binding occurs as in subgroup IIB, except that (with rare exceptions) IBS2 is replaced by the stem-and-loop component of a rho- independent transcription terminator.

Figure from Francois Michel et al., 2009.

161
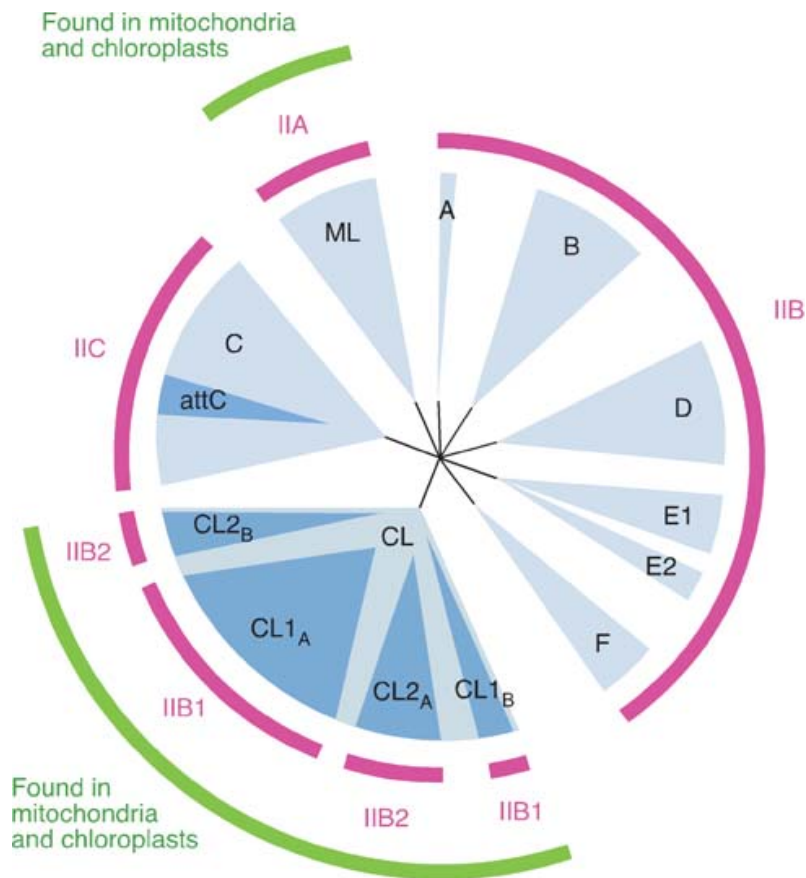
Figure S3. Group II intron lineages. The major lineages of group II intron IEPs, denoted CL (chloroplast-like), ML (mitochondrial-like), and bacterial classes A-F, are shown as blue sectors. Notable sublineages, including four subdivisions of CL and a subclass of IIC introns that inserts after attC sites, are shown as darker blue sec- tors within the major lineages. RNA structural subgroups that corre- spond to IEP lineages are shown in magenta. All group II intron lineages and RNA types are found in bacteria. Lineages and RNA types also found in organelles are delineated in green (outer circle). Note that there may be limited exceptions to the overall pattern of co-evolution within the CL group, with different sublineages possibly having exchanged IIB RNA structures. An alter- nate nomenclature for group II lineages has been proposed, which does not distinguish between IEP and ribozyme lineages or take into account exceptions to their coevolution.

Figure from A.Lambowitz & S.Zimmerly 2004.
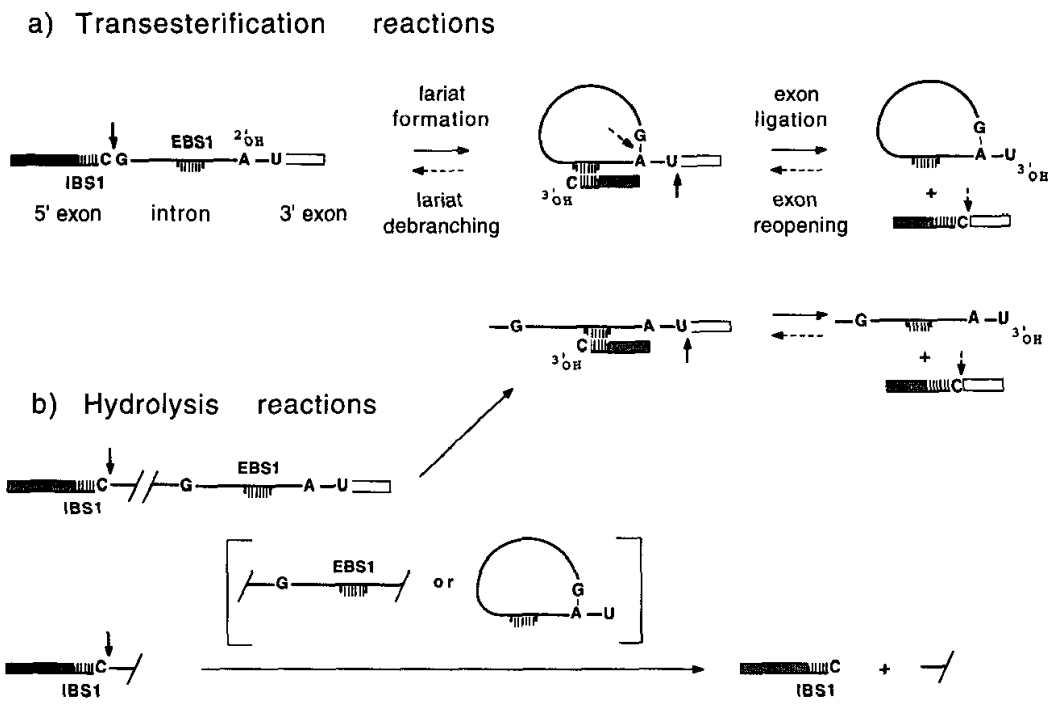
Figure S4. Two major splicing pathway of group II intron. In transesterification pathway, Group II introns splice via two sequential transesterification reactions that yield ligated exons and an excised intron lariat with a 2'-5' phosphodiester bond. In the hydrolytic pathway, water or a hydroxyl ion is used as a nucleophile in the first splicing step.

Figure is from Frangois Michel and Jean-Luc Ferat, 1995.

163
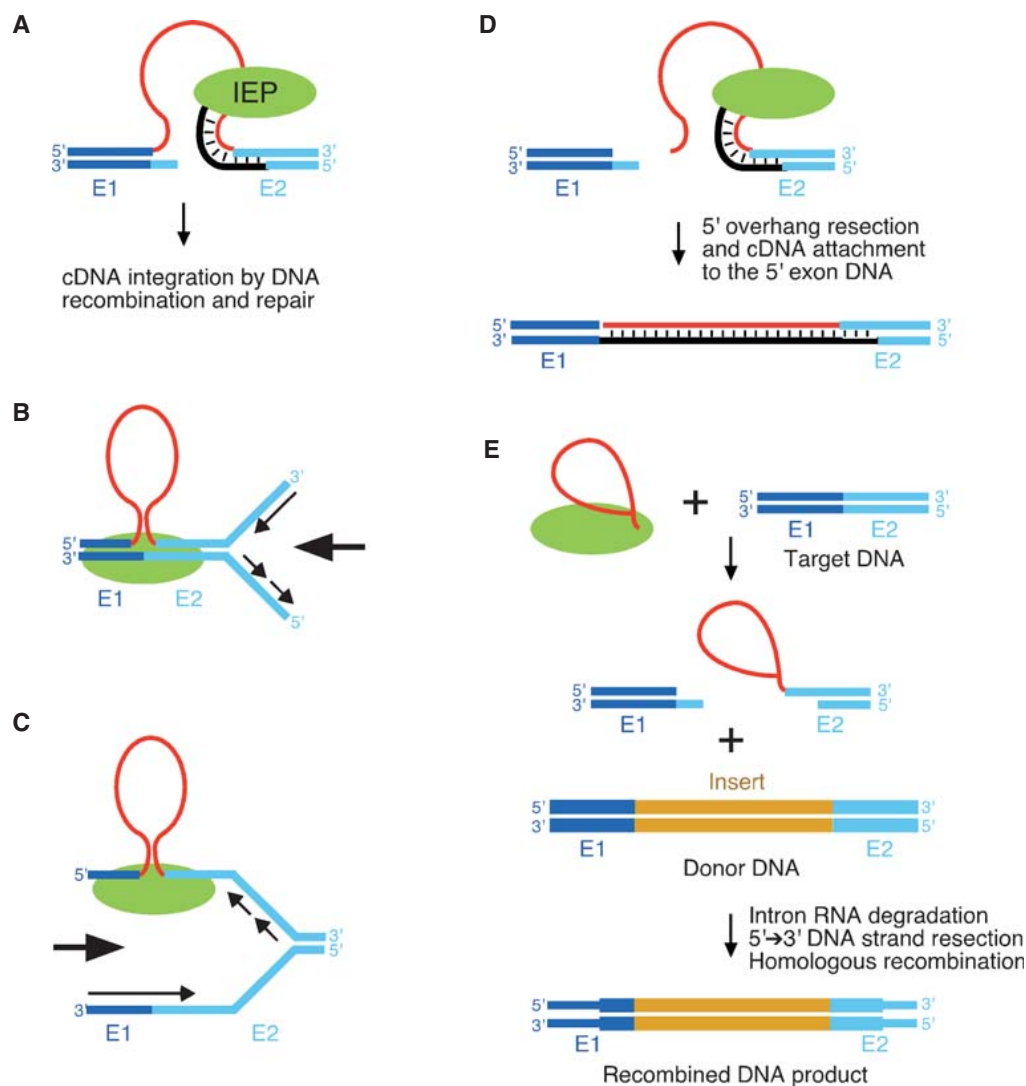
Figure S5. Group II intron mobility mechanisms. (A) Retrohoming via reverse splicing of the intron RNA into double-stranded DNA. After reverse splicing of the intron RNA into the top strand, the bottom strand is cleaved by the En domain of the IEP, and the 3′ end at the cleavage site is used as a primer for reverse transcription of the inserted intron RNA. The resulting intron cDNA is integrated by cellular DNA recombination and/or repair mechanisms. (B) Reverse splicing of the intron RNA into double-stranded DNA, with priming by the nascent leading strand of the DNA replication fork. (C) Reverse splicing of the intron RNA into single-stranded DNA, with priming by the nascent lagging strand of the DNA replication fork. (D) Retrohoming of linear intron RNA by the first step of reverse splicing, bottom-strand cleavage, reverse transcription, and attachment of the free cDNA end to the 5′ exon DNA likely by NHEJ. (E) Use of group II introns to introduce a targeted double-strand break that stimulates gene targeting by homologous recombination. Figure from Lambowitz and Zimmerly., 2010.

Table 1
The Seven Different Types of Naturally Occurring Ribozymes.

ribozyme	source	size	function	reaction products
self-cleaving RNAs				
hammerhead motif	plant viroids and satellite RNAs, salamander	≈ 40 nt	replication	5'-OH; 2', 3'-cyclic phosphate
hairpin motif	plant satellite RNAs	≈ 60 nt	replication	5'-OH; 2', 3'-cyclic phosphate
HDV	hepatitis delta virus (human)	≈ 80 nt	replication	5'-OH; 2', 3'-cyclic phosphate
VS ribozyme	Neurospora crassa mitochondria	154 nt	replication	5'-OH; 2', 3'-cyclic phosphate
RNase P RNAs	eukaryotes (nucleus, organelles), prokaryotes	140–490 nt	tRNA processing	products with 5'-phosphate and 3'OH
self-splicing RNAs				
group I introns	eukaryotes (nucleus, organelles), prokaryotes, bacteriophages	200–1500 nt	splicing	intron with 5'-guanosin and 3'-OH; 5'/3' ligated exons
group II introns	eukaryotes (organelles), prokaryotes	300–3000 nt	splicing	intron with 2'-5' lariat and 3'-OH; 5'/3' ligated exons

Table S1. List of different types of Ribozyme.
Collected from Karola Lehmann and Udo Schmidt.,2003

Article n° 1 :

Cheng-Fang Li, Maria Costa, Gurminder Bassi, Yiu-Kay Lai and
François Michel

**Recurrent insertion of 5′-terminal nucleotides and loss of the
branchpoint motif in lineages of group II introns inserted in
mitochondrial preribosomal RNAs**

RNA

A PUBLICATION OF THE RNA SOCIETY

Recurrent insertion of 5′-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs

Cheng-Fang Li, Maria Costa, Gurminder Bassi, et al.

Supplemental Material	http://rnajournal.cshlp.org/content/suppl/2011/05/05/rna.2655911.DC1.html
P<P	Published online May 25, 2011 in advance of the print journal.
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

Recurrent insertion of 5′-terminal nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in mitochondrial preribosomal RNAs

CHENG-FANG LI,[1,2] MARIA COSTA,[1] GURMINDER BASSI,[1,3] YIU-KAY LAI,[2] and FRANÇOIS MICHEL[1,4]

[1]Centre de Génétique Moléculaire du C.N.R.S., 91190 Gif-sur-Yvette, France
[2]Department of Life Science and Institute of Biotechnology, National Tsing Hua University, Hsinchu, Taiwan 30013

ABSTRACT

A survey of sequence databases revealed 10 instances of subgroup IIB1 mitochondrial ribosomal introns with 1 to 33 additional nucleotides inserted between the 5′ exon and the consensus sequence at the intron 5′ end. These 10 introns depart further from the IIB1 consensus in their predicted domain VI structure: In contrast to its basal helix and distal GNRA terminal loop, the middle part of domain VI is highly variable and lacks the bulging A that serves as the branchpoint in lariat formation. In vitro experiments using two closely related IIB1 members inserted at the same ribosomal RNA site in the basidiomycete fungi *Grifola frondosa* and *Pycnoporellus fulgens* revealed that both ribozymes are capable of efficient self-splicing. However, whereas the *Grifola* intron was excised predominantly as a lariat, the *Pycnoporellus* intron, which possesses six additional nucleotides at the 5′ end, yielded only linear products, consistent with its predicted domain VI structure. Strikingly, all of the introns with 5′ terminal insertions lack the EBS2 exon-binding site. Moreover, several of them are part of the small subset of group II introns that encode potentially functional homing endonucleases of the LAGLIDADG family rather than reverse transcriptases. Such coincidences suggest causal relationships between the shift to DNA-based mobility, the loss of one of the two ribozyme sites for binding the 5′ exon, and the exclusive use of hydrolysis to initiate splicing.

Keywords: mitochondrial group II introns; linear intron; lariat intron branchpoint; homing endonucleases

INTRODUCTION

Bacterial group II introns result from the association of a reverse transcriptase (RT) gene with a large ribozyme: the latter catalyzes the branching and ligation reactions that result in an excised intron lariat and spliced exons (for review, see Lambowitz and Zimmerly 2004; Beauregard et al. 2008). Several lineages of these widely distributed prokaryotic retrotransposons found their way into the genomes of organelles and proliferated in diverse eukaryote clades. A majority of present-day group II members from organelles subsequently lost their RT component, but a number of individual introns have retained the potential to code for a protein and still behave as mobile elements (Kennell et al. 1993; Lazowska et al. 1994).

Even in the absence of an RT gene, identifying a group II intron in sequence data remains reasonably straightforward. Of the six secondary structure domains of the ribozyme, the small domain V tends to be sufficiently conserved in terms of structure and sequence to lend itself to the design of an efficient search engine, with relatively few false negatives and positives (e.g., Griffiths-Jones et al. 2005; Lang et al. 2007). And once a candidate domain V has been spotted, it is generally feasible to use comparative sequence analysis and start building step-by-step the potential secondary structure of the rest of the ribozyme, all the way to the 3′ and 5′ splicing junctions (Michel et al. 1989). Although there exist a few noteworthy exceptions (Michel et al. 1989; Vogel and Börner 2002; Stabell et al. 2007), the 3′ terminus normally lies 2–3 nt downstream from domain VI, while the latter carries on its 3′ side a bulging adenine that serves as branchpoint for lariat formation (Fig. 1). On the other side of the intron, the last six or so nucleotides of the 5′ exon take part in a long-range, intron–exon pairing, EBS1–IBS1 (Fig. 2A), whose ultimate base pair precedes the splice site. Moreover, the first five nucleotides of the intron tend to obey a characteristic consensus sequence, GUGYG, which is conserved as such in some 85% of known group II members (to the exclusion of the "degenerate" group II introns in the chloroplast genomes of euglenoids) (see Hallick et al. 1993); the actual extent of conservation of individual nucleotides
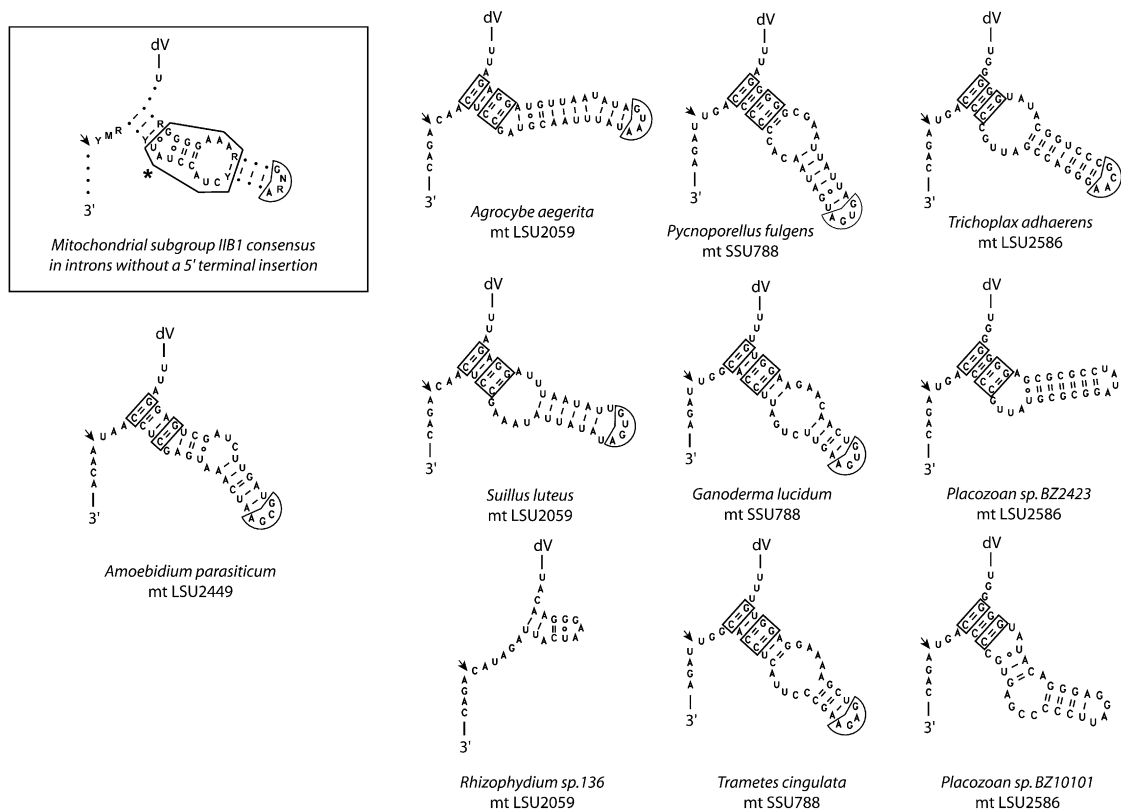
FIGURE 1. Predicted structures of domain VI in 10 introns with a 5′ terminal insertion. (dV) Domain V; (arrow) the 3′ splice site; (boxed) well-conserved G:C pairs in the basal helix of domain VI, as well as its terminal loop, when it obeys the GNRA consensus. The structures are compared with the strongly conserved consensus structure and sequence of domain VI in 34 introns devoid of additional nucleotides at the 5′ end (*inset*; the asterisk indicates the branchpoint; nucleotides and base pairs shown are at least 90% conserved; M: A or C; the highly divergent sequences of the cox1 introns of *Paracoccidioides brasiliensis* and *Candida parapsilosis* [see Table 1] were not taken into account).

varies from 92% for position three to virtually 100% for position five (Michel et al. 2009).

While taking a census of established and candidate group II introns in organelle DNA sequences, our attention was brought to a small subset of introns that appeared to diverge somewhat from these rules. In these 10 group II members, the end of the IBS1 sequence and the GUGYG consensus sequence are separated from one another by one to as many as 33 intervening nucleotides (Fig. 2B; Table 1). Moreover, at the other intron end, the potential secondary structure of domain VI lacks a bulging A at the expected location for the branchpoint (Fig. 1). These introns, which happen to belong to the same ribozyme structural subgroup (IIB1) (Michel et al. 1989) and are inserted in ribosomal RNA precursor transcripts, exhibit additional remarkable features (Table 1): They all lack the EBS2–IBS2 pairing between the ribozyme and the 5′ exon, which is potentially present in most group II introns, and several of them code for a homing endonuclease, rather than a reverse transcriptase (see also Michel and Ferat 1995; Toor and Zimmerly 2002; Monteiro-Vitorello et al. 2009).

We have cloned one of the members of this peculiar subset of intron sequences, as well as a closely related, but apparently "normal" intron inserted at the same genomic site in another host, and we now show that the self-splicing reaction of the former (but not latter) molecule is initiated by hydrolysis, resulting in excision of the intron in linear form, rather than by transesterification, which generates a lariat structure (as is normal for group II introns) (for review, see Michel and Ferat 1995). More generally, we propose that the loss of the ability to form a branched structure should be regarded as an ultimate consequence of the recently documented (Mullineux et al. 2010) evolutionary conversion of some mitochondrial group II introns into DNA transposons (the class II mobile elements of Wicker et al. [2007] that move at the DNA level, contrary to retrotransposons that change location as RNA).

RESULTS

Distribution of 5′-terminal inserts in mitochondrial subgroup IIB1 introns

Subgroup IIB1 is widespread both in bacteria and organelles and includes two members whose ribozyme is used as a model system (*Saccharomyces cerevisiae* cox1/5γ and
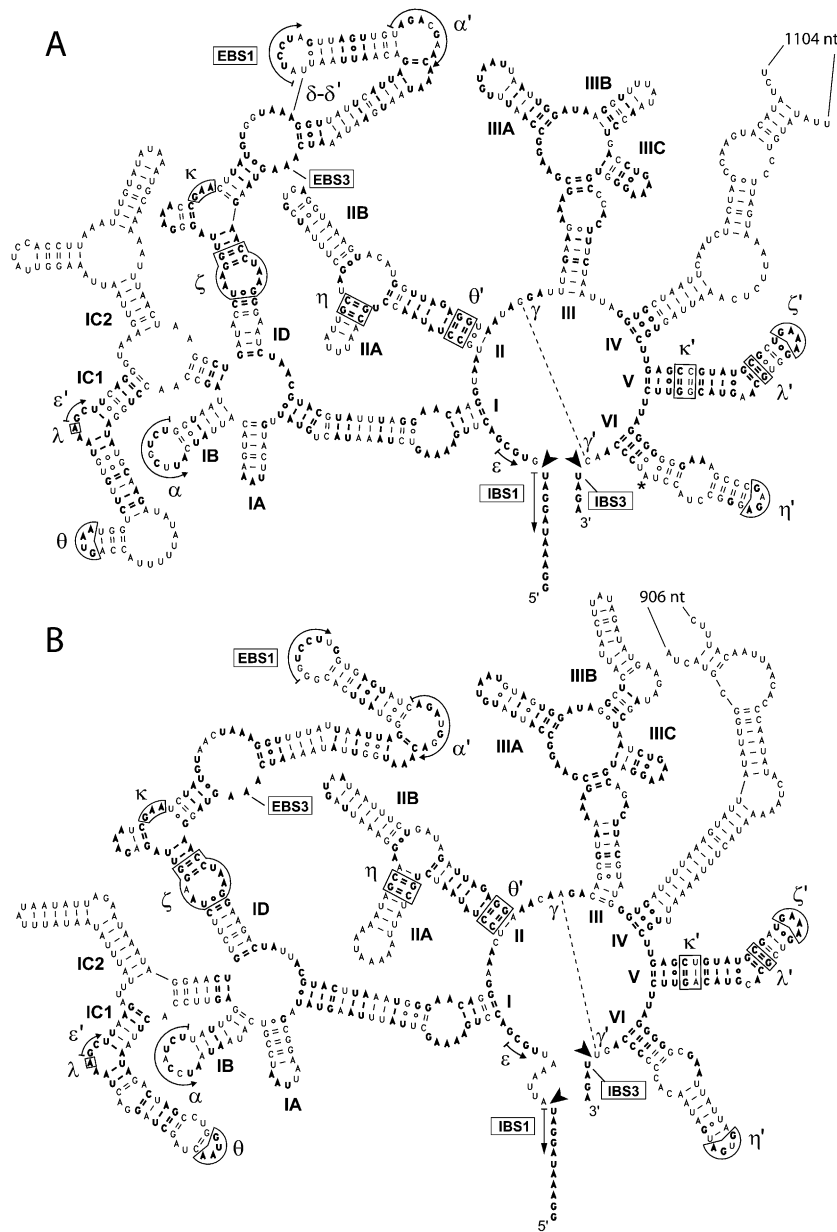
FIGURE 2. Secondary structure models of (*A*) the *G. frondosa* SSU788 intron, (*B*) the *P. fulgens* SSU788 intron. (Boldface) Nucleotides in common between the two introns. (Arrowheads) Point to splice junctions; (asterisk) points to the branchpoint of the *Grifola* intron. Labeling of secondary structure components and tertiary interactions is as in Michel et al. (2009). In domain IV, nucleotides not shown were replaced by the sequence CTCGAG in "ORF-less" constructs.

Pylaiella littoralis LSU1787, also known as Pl.LSU/2). A list of published sequences of mitochondrial subgroup IIB1 members, which contains the 10 organelle group II introns we found to possess additional nucleotides at their very 5′ extremity, is provided in Table 1 (several subgroup IIB1 members from land plants other than *Marchantia* are missing from this list; they were excluded from our analyses because of the likelihood of [partly] undocumented editing of intron nucleotides prior to splicing) (see Bonen 2008).

To ascertain that a candidate intron contained additional nucleotides at its 5′ end, the location of splice junctions was first inferred by comparison with intron-less copies of the host gene in related organisms. We then looked for entries in which the 5′-terminal consensus sequence (GUGCGAC in the case of subgroup IIB1 introns) was separated from the predicted 5′ splice site by one or more nucleotides. Finally, a complete secondary structure model was generated for each candidate ribozyme (Fig. 2B; Supplemental Data Set), and the EBS1 terminal loop in domain I was verified to base-pair with the last nucleotides of the inferred 5′ exon, rather than with the sequence preceding the 5′-terminal consensus.

An alternative interpretation to the existence of an insert at the intron 5′ end could be that the additional nucleotides are not removed during splicing. However, all 10 candidate introns happen to be in ribosomal RNA genes, and their inferred sites of insertion (SSU788, LSU2059, LSU2449, and LSU2586) (see Table 1) lie within segments of sequence that are extremely conserved and most unlikely to tolerate insertions in mature, functional molecules (the three LSU segments are either part of, or lie next to, the catalytic site for peptide synthesis, while the SSU site is part of the loop that separates the tRNA P and E sites) (Nissen et al. 2000; Schuwirth et al. 2005). As for the possibility that the sequence under scrutiny was that of a pseudogene, it can be ruled out for entries that correspond to completely sequenced genomes (the three placozoan sequences and those of *Amoebidium*, *Rhizophydium*, and *Trametes*) and have a single, intron-containing copy of the ribosomal RNA gene of interest.

The 10 introns with a 5′-terminal insert also stand out in that the sequence and predicted secondary structure of ribozyme domain VI is strikingly variable, even among closely related introns, and departs in multiple ways from the consensus domain VI structure shared by all other mitochondrial members of the IIB1 subgroup (Fig. 1). Not only is the branchpoint adenine missing at its expected location, but the well-conserved 3-bp helix and (GAA:CUA) internal loop immediately distal of it are unrecognizable. This is all the more striking since the

TABLE 1. A list of mitochondrial subgroup IIB1 introns

Organism, gene, and intron	Accession number	Intron coordinates[b]	5′ insert[c]	EBS2	ORF product[d] and location[j]
Fungi					
Paracoccidioides brasiliensis cox1/2[e]	AY955840	41,071–43,890			RT (IV)
Candida parapsilosis cox1[e]	X74411	12,690–15,605			RT (IV)
Saccharomyces cerevisiae cox1/5γ	V00694	8746–9632[*]			
Saccharomyces pastorianus cox1/5	EU852811	53,565–54,476			
S. cerevisiae cytb/1	EU004203	38,472–39,239			
Candida zemplinina LSU2584[f]	AY445918	2516–324[*]			RT (IV)
Candida ipomoeae SSU531	AY393889	176–801			
Glomus intraradices C8.3b_18 LSU1787	AM950206	1449–2257			
G. intraradices C16g1_2 LSU1787	AM950209	2541–3896			LAGLIDADG (IV)
Uncultured *Glomus* W9/1 LSU1787	FN377588	1827–3215			LAGLIDADG (IV)
Allomyces macrogynus LSU2059	U41288	2416–3192			
Rhizophydium sp. 136 LSU2059	NC_003053	3880–4564	+6 GUGCGAC	No	
Pichia angusta LSU2059	AL432964 AL434946 AL433470	625–1; 469–1; 294–879[*]		No	LAGLIDADG (IV)
Ustilago maydis SRX2 LSU2059	EU921807	3413–5372			LAGLIDADG (IV)
Agrocybe aegerita LSU2059	AF087656	9088–10,871[*]	+23 UUGCGAC	No	LAGLIDADG[k] (IV)
Suillus luteus LSU2059	L47586	2675–3341	+25 UAGCGAC	No	
Cryphonectria parasitica SSU952	AF029891	7168–9235			LAGLIDADG (III)
Leptographium truncatum 1435 SSU952	GU949593	800–2639		No	LAGLIDADG (III)
Cordyceps konnoana SSU952	AB031194	897–2724[*]			LAGLIDADG (III)
Paracoccidioides brasiliensis SSU952	AY955840	25,574–27,362			Unidentified (III)
Aleurodiscus botryosus SSU788	FR773980[g]			No	LAGLIDADG (IV)
Ceriporiopsis subvermispora SSU788	EU546103	345–907[*]		No	
Grifola frondosa SSU788	FR773978[g]			No	LAGLIDADG (IV)
Pycnoporellus fulgens SSU788	FR773979[g]		+6 UUGCGAC	No	LAGLIDADG (IV)
Ganoderma lucidum SSU788	AF214475[h]	1056–2562	+6 UUGCGAC	No	LAGLIDADG (IV)
Trametes cingulata SSU788	GU723273	39,037–40,442[*]	+6 AUGCGAC	No	LAGLIDADG[k] (IV)
Usnea antarctica SSU788	DQ990920	397–1473[*]			
C. parasitica SSU788	AF029891	2415–4596[*]		No	LAGLIDADG (IV)
Ichthyosporea					
Amoebidium parasiticum SSU788	AF538044	855–2198		No	GIY-YIG (IV)
A. parasiticum LSU2449	AF538042	5337–5909	+33 GAGCGAC	No	
Plants[a]					
Chaetosphaeridium globosum trnG(UCC)	AF494279	16,893–16,066[*]		No	
Chlorokybus atmophyticus trnS(GCU)	NC_009630	134,254–136,307			
Marchantia polymorpha trnS(GCU)	NC_001660	48,902–49,892			
M. polymorpha LSU787	NC_001660	149,541–148,824			
M. polymorpha cox3/2	NC_001660	87,980–88,911			
M. polymorpha cox2/1	M68929	81,191–82,268			
Scenedesmus obliquus LSU2455	NC_002254	28,831–29,438			
S. obliquus SSU968	NC_002254	40,780–41,627[*]			
Pedinomonas minor LSU1787	AF116775	1198–2027			
Stramenopiles					
Pylaiella littoralis LSU575	Z48620	543–2952			RT (IV)
P. littoralis LSU1787	Z48620	4052–6489			RT (IV)
P. littoralis LSU2451	Z48620[i]	7134–8304			

(*continued*)

TABLE 1. *Continued*

Organism, gene, and intron	Accession number	Intron coordinates[b]	5′ insert[c]	EBS2	ORF product[d] and location[j]
Animals					
Trichoplax adhaerens LSU2586	NC_008151	19,460–20,691[*]	+1 GUGCGAC	No	
Placozoan sp. BZ2423 LSU2586	NC_008834	16,339–17,068[*]	+1 GUGCGAC	No	
Placozoan sp. BZ10101 LSU2586	NC_008832	17,186–17,990[*]	+1 GUGCGAC	No	
Placozoan sp. BZ10101 LSU1787	NC_008832	15,476–16,215			

[a]Introns from land plants other than *Marchantia* were excluded because of the likelihood of RNA editing.
[b]Intron coordinates were inferred by comparison with closely related uninterrupted gene sequences; (asterisks) introns the coordinates of which differ from those indicated in the GenBank entry or were missing.
[c]For introns with insertions at the 5′ end, the sequence that best matches the IIB1 5′ terminal consensus is provided, together with the number of nucleotides inserted in front of it.
[d](RT) Reverse transcriptase; (LAGLIDADG and GIY . . YIG) putative homing endonucleases belonging, respectively, to these families.
[e]These introns are inserted at homologous sites in the *cox1* gene, their sequences are only distantly related to the rest of the IIB1 set, and they were used as the outgroup in Figure 5.
[f]Following Johansen and Haugen (2001), introns in ribosomal RNA genes are designated by the name of the ribosomal subunit followed by the coordinate of the nucleotide preceding the insertion site (according to the *Escherichia coli* numbering scheme).
[g]This study.
[h]The reference for this entry is Hong et al. (2002).
[i]The sequence of domain VI was from entry AB281597.
[j]Ribozyme secondary structure domain within which the ORF is located.
[k]Presence of in-frame stop codons and frameshifts.

base of the domain VI stem tends to be well-conserved—it begins with a G:C pair in nine out of 10 introns—and seven out of 10 sequences share a GNRA loop at the tip of the domain; in related introns, that loop participates in the η–η′ tertiary interaction (Chanfreau and Jacquier 1996; Costa et al. 1997a) between domains II and VI (Fig. 2).

Cloning and sequence analysis of the *Grifola frondosa* and *Pycnoporellus fulgens* SSU788 introns

Of the 10 intron sequences with 5′ terminal insertions in Table 1, that of the *P. fulgens* SSU788 intron (GenBank entry AF518690) was incomplete. We chose to clone and sequence this intron and its flanking exons, as well as two partially sequenced, insert-lacking, related SSU788 introns in the basidiomycete fungi *G. frondosa* and *Aleurodiscus botryosus* (accession numbers AF334880 and AF026646).

As shown in Figure 2, the predicted secondary structure models of the *Grifola* and *Pycnoporellus* ribozymes are very similar, and the same is true of the *Aleurodiscus* ribozyme (Supplemental Fig. S1). As expected, the identity of nucleotides at sites known to participate in intra- or inter-domain, long-range tertiary interactions (Toor et al. 2008a; Michel et al. 2009; Pyle 2010) is especially well conserved. The only exception is the δ–δ′ Watson-Crick base pair, which contributes to the stability of the EBS1–IBS1 intron–exon pairing (Costa et al. 2000): the U:A δ–δ′ pair of the *Grifola* intron is replaced by G:A in the *Pycnoporellus* molecule (the closely related *Ganoderma lucidum* and *Trametes cingulata* introns have A:A at these sites, whereas other 5′-insert-

bearing introns, and also the *A. botryosus* molecule, have diverse Watson-Crick base pairs) (data not shown). Also very well-conserved is domain III, which contributes to the efficiency of catalysis (Fedorova and Pyle 2005).

A striking feature that the two secondary structure models have in common is the lack of EBS2–IBS2, an extended canonical pairing that involves nucleotides upstream of IBS1, on the one hand, and a single-stranded loop in the distal section of subdomain ID, on the other. The EBS2–IBS2 pairing is present in a majority of group II introns, with the exception of members of subgroup IIC, whose 5′ exon displays a hairpin structure at the expected location for the IBS2 sequence (Granlund et al. 2001; Quiroga et al. 2008). What has been lost, in fact, is not only the EBS2 loop, but an entire subdomain that, in subgroup IIB, branches off the 5′ strand of the stem connecting the internal loops that contain the EBS3 and α′ nucleotides. This subdomain carries, in addition, a sequence that, in many introns, potentially participates in the β–β′ long-range interaction with subdomain IC2 (Michel et al. 1989). Interestingly (Table 1), the EBS2 loop and associated subdomain are missing from all 10 introns with 5′-terminal inserts and also all other known SSU788 introns with the exception of *Usnea antarctica* (data not shown).

Subdomains that are known (Toor et al. 2008a) or suspected (Pyle 2010) to lie at the surface of the ribozyme three-dimensional structure tend to be the most variable ones. This is especially true of domain IV, only the first three base pairs of which are conserved between the *Grifola* and *Pycnoporellus* sequences. Still, the contents of domain IV are similar in the two introns (and in the *A. botryosus*

SSU788 intron), consisting primarily of open reading frames (ORFs; 260 codons in *Grifola* and 266 in *Pycnoporellus*) that potentially encode related (38% identical at the amino acid level) members of the LAGLIDADG family of DNA double-stranded homing endonucleases (Stoddard 2005). As seen in fact in Table 1, four out of the other seven published sequences of SSU788 introns contain coding sequences for additional LAGLIDADG homing endonucleases (the gene appears defective in *T. cingulata*), while a fifth one (in *Amoebidium parasiticum*) potentially encodes a GIY-YIG protein, the second most common family of homing endonucleases in mitochondrial genomes.

Contrasting self-splicing products of the *Grifola* and *Pycnoporellus* SSU788 introns

The lack of a group II branchpoint structure in domain VI of the *Pycnoporellus* SSU788 intron suggested that splicing was initiated by hydrolysis at the 5′ splice junction, rather than by transesterification (Jarrell et al. 1988; Jacquier and Jacquesson-Breuleux 1991; Daniels et al. 1996; Podar et al. 1998; Vogel and Börner 2002). This was confirmed by incubating precursor transcripts containing the *Grifola* and *Pycnoporellus* SSU788 introns under conditions that allow in vitro self-splicing.

In vitro self-splicing of the *Grifola* SSU788 intron (Fig. 3) is reasonably efficient at 42°C in 1 M NH$_4$Cl and at a moderately high magnesium concentration (20 mM). As reported for other group II introns (Daniels et al. 1996; Costa et al. 1997a,b), reaction of precursor molecules is a kinetically complex process, converting only about half of the material to products in ∼2 min and the rest much more slowly if at all (Fig. 3B). The distribution of splicing products is also typical of most group II introns, being dominated by the lariat intron and ligated exons (Fig. 3A), the identity of which was verified by gel extraction followed by reverse transcription (Fig. 4A,C). Only small amounts of
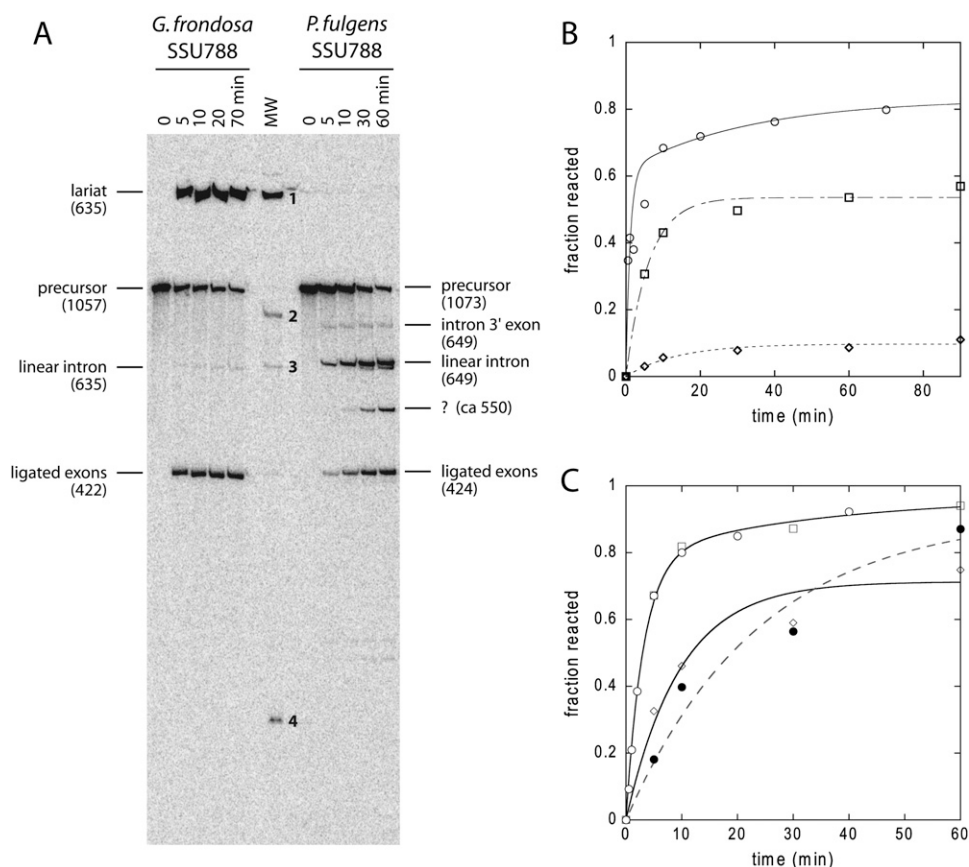
FIGURE 3. Self-splicing of the *Grifola* and *Pycnoporellus* SSU788 introns. (A) Time course of self-splicing reactions at 42°C in 1 M NH$_4$Cl, 20 mM MgCl$_2$, 40 mM Na-MES (pH 6.2). Products were identified based on (1) reverse transcription of gel-extracted molecules (see Fig. 4) and (2) their electrophoretic mobility, compared to that of known splicing products of a *P. littoralis* LSU1787 (Table 1; Costa et al. 1997b) precursor transcript (MW lane: band 1, 640 nt, lariat; band 2, 872 nt, precursor; band 3, 640 nt, linear intron; band 4, 232 nt, ligated exons). (B) Time course of self-splicing reactions of a *Grifola* SSU788 precursor RNA at 42°C in 40 mM Tris-Cl (pH 7.5 at 25°C), 20 mM MgCl$_2$, and 1 M NH$_4$Cl (circles and solid curve, generated by a biphasic exponential fit with $k_1 = 0.9 \pm 0.2$ min^{-1} and $k_2 = 0.03$ min^{-1}) (see Materials and Methods) or 1 M KCl (squares and dashed curve, lariat intron; lozenges and dotted curve, linear intron; both from single exponential fits). (C) Time course of self-splicing reactions of a *Pycnoporellus* SSU788 precursor RNA in 40 mM Tris-Cl (pH 7.5 at 25°C), 1 M NH$_4$Cl, and 10 mM MgCl$_2$ (empty squares), 20 mM MgCl$_2$ (empty circles), 50 mM MgCl$_2$ (empty lozenges), or in 40 mM Na-MES (pH 6.2) and 20 mM MgCl$_2$ (filled circles and dashed curve). Reactions at 10 and 20 mM Mg (pH 7.5) were fitted to a biphasic process ($k_1 = 0.32 \pm 0.03$ min^{-1}, $k_2 = 0.030 \pm 0.016$ min^{-1}), the other ones to simple exponentials.
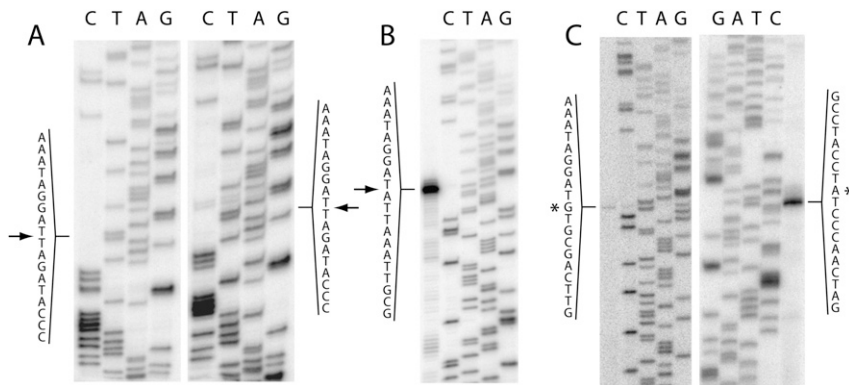
FIGURE 4. Mapping of intron–exon junctions and the branch site. Sequencing lanes are labeled by the base complementary to the dideoxynucleotide added. (*A*) Sequencing by reverse transcription of gel-extracted ligated exons; (*left* panel) *Pycnoporellus*; (*right* panel) *Grifola*. (Arrows) Splicing junctions. (*B*) Mapping of the 5′ extremity of gel-extracted linear intron molecules generated by in vitro self-splicing of a *Pycnoporellus* precursor transcript; the latter was used as a template to generate the sequencing lanes at *right* with a primer located downstream from the intron 5′ extremity. Elongation from the same primer using the excised intron molecules as template generated the strong stop in the lane at *left*; (arrow) the 5′ splice site. (*C*) Mapping of the branchpoint and 5′ extremity of gel-extracted lariat intron molecules generated by in vitro self-splicing of a *Grifola* precursor transcript. (*Left* panel) Elongation from a primer located downstream from the intron 5′ extremity; the stop (marked by an asterisk) corresponds to the first intron nucleotide; sequencing lanes (at *right*) were generated by the same primer on a precursor RNA template. (*Right* panel) elongation from a primer located in the 3′ exon (intron-3′exon branched molecules were used as template); (asterisk) the branch site (elongation stops on the nucleotide immediately 3′ of the branch site); sequencing lanes were generated by the same primer on a precursor RNA template.

a linear intron form were observed, unless ammonium ions were replaced by potassium ions (Jarrell et al. 1988). Even then, the final molar fraction of linear intron molecules—presumably generated by hydrolysis at the 5′ splice site—did not exceed 15% of intron-containing products (Fig. 3B).

Under the conditions used for the *Grifola* intron, self-splicing of the *Pycnoporellus* SSU788 intron is also a rather rapid process (Fig. 3C, solid curve), with ∼80% of precursor molecules converted to products in 10 min. However, the lariat intron is absent from reaction products, which consist primarily of the ligated exons and a linear intron form. The 5′ extremity of the latter was verified by reverse transcription (Fig. 4B) to coincide with the 5′ splice site, as determined by alignment with uninterrupted versions of the host gene, on the one hand, and sequencing of the ligated exons (Fig. 4A), on the other. Additional products include small amounts of molecules with the expected electrophoretic mobility of the linear intron–3′exon splicing intermediate and a molecule of ∼550 nt, which could have been generated by ribozyme-catalyzed, hydrolytic cleavage of the linear intron at position 110, 3′ of the sequence AGGAC. The latter offers a better match to EBS1 (GUCCU) than the IBS1 sequence (AGGAU) at the 3′ end of the 5′ exon (see Fig. 2).

Varying the concentration of magnesium (Fig. 3C) did not make it possible to observe lariat molecules among self-splicing products of the *Pycnoporellus* intron but confirmed that the optimal magnesium concentration in terms of

reaction rate and final extent of reaction is ∼10–20 mM (Fig. 3C). On the other hand, lowering the pH from 7.0 to 6.2 did lead to a substantial reduction of the reaction rate (Fig. 3C), suggesting that catalysis is at least partly rate-limiting when splicing is initiated by hydrolysis.

Phylogenetic relationships of mitochondrial subgroup IIB1 introns with and without a 5′ terminal insert

To generate a phylogenetic tree of mitochondrial subgroup IIB1 introns, their ribozyme sequences were aligned over shared components of the subgroup IIB1 secondary structure (see Materials and Methods; Supplemental Data Set). The number of sites that can be unambiguously aligned (526) is too small to resolve the complete phylogenetic relationships of all members of this subgroup (Fig. 5). Nevertheless, bootstrap analysis indicates that introns inserted at the same ribosomal site tend to form well-supported clades, consistent with a common origin. The only exception comes from introns inserted at position 1787 of the large ribosomal RNA: neither the four available sequences, from *P. littoralis*, *Pedinomonas minor*, *Glomus intraradices*, and *Placozoan* sp. BZ10101 (Fig. 5; Table 1), nor the corresponding secondary structure models (data not shown) reveal any particularly close similarity.

Provided it is assumed that the insertion of nucleotides at the 5′ intron extremity and the accompanying loss of the branchpoint structure are irreversible events, the minimal number of occurrences that gave birth to lineages of introns with 5′-terminal inserts may be estimated from the phylogeny proposed in Figure 5. The most parsimonious interpretation of the data implies at least four founding insertion events, and a fifth one would become necessary should the hypothetical relationship of the *A. parasiticum* LSU2449 intron with the *Placozoan* LSU2586 introns prove nonsignificant (in that case, two events would need to be postulated at LSU2059 and one at each of the other three ribosomal RNA sites occupied by introns with 5′-terminal inserts).

Interestingly, the presence in mitochondrial members of subgroup IIB1 of intron-contained homing endonuclease ORFs results as well from multiple, independent acquisition events (Supplemental Fig. S2). As already pointed out by Monteiro-Vitorello et al. (2009), the proteins potentially encoded by the SSU788 and SSU952 introns of *Cryphonectria parasitica* are not closely related. More generally, whereas
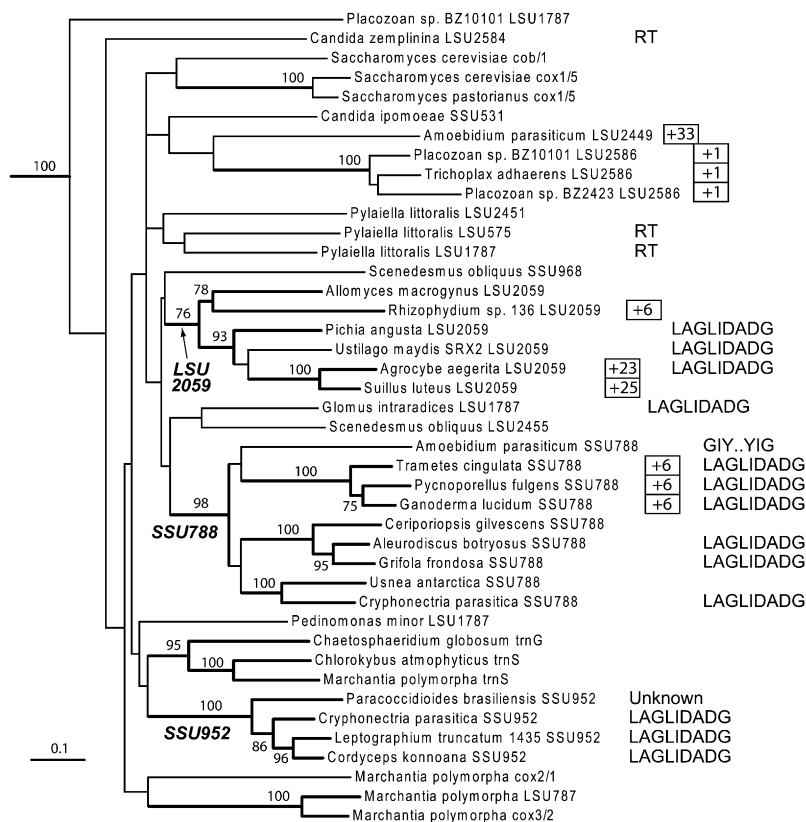
FIGURE 5. Phylogenetic relationships of mitochondrial subgroup IIB1 introns based on aligned ribozyme sequences (see Materials and Methods and Supplemental Data Set). Introns are designated as in Table 1; the cox1 introns from *P. brasiliensis* and *C. parapsilosis*, the sequences of which are markedly divergent from the rest, were used as outgroups. Numbers next to nodes are bootstrap proportions (200 replicates) ≥75% (corresponding branches are thickened). The roots of well-supported, major clades of ribosomal introns are indicated. The length of the 5′ terminal insertion, when present, is provided to the *right* of an intron name (boxed numbers preceded by + sign). RT, LAGLIDADG, GIY . . YIG, and ''Unknown'' designate proteins potentially encoded by the introns (see Table 1; note that only some versions of the *G. intraradices* LSU1787 intron include an ORF).

LAGLIDADG proteins encoded by introns inserted at the same ribosomal site tend to be rather similar—they form monophyletic groups—and may have coevolved with their intron host, introns located at different sites encode proteins that belong to separate lineages within the LAGLIDADG phylogenetic tree (Supplemental Fig. S2). Note also that in contrast to the ORFs located in introns inserted at positions LSU2059, SSU788, and SSU952, which contain two LAGLIDADG motifs, the much shorter *Glomus* LSU1787 intron ORFs (200 and 208 codons; accession numbers AM950209 and FN377588, respectively) contain a single LAGLIDADG element, so that the corresponding homing endonuclease must be a homodimer, rather than a monomer (see Stoddard 2005).

The following facts provide further evidence that mitochondrial subgroup IIB1 introns acquired ORFs for proteins other than reverse transcriptases through independent insertion events: (1) in introns inserted at the SSU952 site, the ORF is inserted in ribozyme domain III, rather than in

domain IV (Table 1; Mullineux et al. 2010); (2) the ORF in the *A. parasiticum* SSU788 intron encodes a protein belonging to the GIY-YIG family of homing endonucleases, rather than a member of the LAGLIDADG family; (3) the proteins possibly encoded by the LSU2059 introns are closely related (Supplemental Fig. S2) to the protein specified by a group I intron inserted at position 2066 of the mitochondrial LSU gene of *Tuber melanosporum*, just 7 nt 3′ of position 2059.

The latter observation is obviously in keeping with a model (Bonocora and Shub 2009) in which mobilization of an intron by a homing endonuclease precedes the acquisition of the endonuclease gene by the intron. Should the proteins encoded by the LSU2059 and LSU2066 introns eventually be found to share the same site of cleavage, as appears likely, this would constitute an additional instance (after those reported by Zeng et al. 2009 and Bonocora and Shub 2009) of an endonuclease-coding sequence being translocated without the endonuclease changing its cleavage specificity. Note that even though available phylogenetic data provide no indication as to whether translocation was from the group I to the group II intron subclade or vice versa in this particular case, the much greater abundance of endonuclease-encoding group I introns in fungal mitochondrial genomes makes it far more likely that they act as donors. The occasional transposition of an entire group I intron into domain IV (or the periphery of domain III), followed by the rapid degeneration of the group I ribozyme sections, constitutes an obvious way for an endonuclease-coding gene to invade a group II intron: The resulting genomic arrangement should be readily selected whenever the recipient intron already happens to lie within the recognition sequence of the endonuclease.

DISCUSSION

Additional nucleotides at the intron 5′ end and the inability to initiate splicing by transesterification

We have shown that under in vitro self-splicing conditions, the SSU788 intron of *P. fulgens* generates only linear intron forms, in contrast to its close relative in *G. frondosa*, the excision of which yields the expected lariat (branched)

intron. Either the absence of a bulging A at the expected location for the branchpoint or the presence of an insert at the intron 5′ end could be invoked to account for the inability of the *Pycnoporellus* intron to perform the branching reaction. Deletion or base-pairing of the branchpoint adenosine has long been known to inhibit branching of the *S. cerevisiae* cox1/5γ intron (van der Veen et al. 1987; Chu et al. 1998), although splicing remains possible via hydrolysis at the 5′ splice site. Similarly, the insertion of additional nucleotides at the intron 5′ end was reported by Jacquier and Jacquesson-Breuleux (1991) to result in the loss of the cox1/5γ branching reaction in vitro; splicing could be initiated only by hydrolysis, and they showed the 5′ splice site to coincide with the 3′ end of the IBS1 sequence, rather than with the 5′ end of the GUGCG intron consensus sequence, just as we now report for the *Pycnoporellus* intron.

The existence of natural group II introns that lack a bulging A on the 3′ side of domain VI was noted long ago (Michel et al. 1989), and one of these introns, in the gene encoding the tRNA^{Val} (UAC) of plant chloroplasts, was later shown to be excised without forming lariats (Vogel and Börner 2002). On the other hand, this is the first time that the existence of group II introns with additional nucleotides at the 5′ end is explicitly reported (the presence of a 5′ terminal insert in the *Agrocybe aegerita* LSU2059 intron was apparent in the secondary structure model in Figure 3 of Gonzalez et al. [1999] but was not discussed in the text). 5′-Terminal inserts can be surprisingly long: In *Paxillus atrotomentosus* isolate TDB-782 (Bruns et al. 1998), an intron closely related to the LSU2059 intron of *Suillus luteus* has no fewer than 48 additional nucleotides inserted between the presumed 5′ splice site and the UAGCGAC sequence motif that these two introns share on the 5′ side of domain I (this sequence [accession number AD001614] was not listed in Table 1 because it stops 73 nt within the intron). At the other end of the length spectrum, only 1 nt separates the inferred 5′ splice site from the canonical GUGCG sequence motif in the three placozoan LSU2586 sequences. Whether a single-nucleotide insert is sufficient to abolish branching is questionable: Insertion of just 1 nt at the 5′ end of *S. cerevisiae* intron cytb/1 does not prevent branching, even though it results in a marked shift toward initiation of splicing by hydrolysis (Wallasch et al. 1991). Still, the noncanonical secondary structure of domain VI in the placozoan LSU2586 introns (Fig. 1) makes it unlikely that these ribozymes would succeed in catalyzing branch formation.

As inferred from experiments in which phosphodiester bonds were replaced by phosphorothioates (Steitz and Steitz 1993; also, for review, see Michel and Ferat 1995; Jacquier 1996), the geometry of the reactive bond in the branching step must differ from the one that prevails during reversal of exon ligation, and also in 5′ hydrolysis. Introns in which the end of the IBS1 sequence is not directly connected to the GUGCG consensus sequence are unable to catalyze branch-

ing, probably because interactions between the ribozyme and nucleotides bordering the 5′ splice site on both its 5′ and 3′ sides are necessary to drive the phosphodiester bond between the intron and 5′ exon into the appropriate, presumably highly constrained conformation required for first-step transesterification. In contrast, the two exons are believed to be maintained in helical continuity by the EBS1–IBS1 and EBS3–IBS3 interactions in the ligation step (Costa et al. 2000; Toor et al. 2008b). Now, since 5′ hydrolysis has the same phosphorothioate requirements as the reversal of exon ligation, one would expect EBS3 to base-pair not only with the first nucleotide of the 3′ exon, but also with the first nucleotide of a 5′-terminal insert, when present. Jacquier and Jacquesson-Breuleux (1991) did observe that for *S. cerevisiae* cox1/5γ constructs with a 5′ insert, hydrolysis was facilitated when the nucleotide following the 5′ splice site was an A (which could base-pair with the U at the site that would come to be known as EBS3). In nature, however, while the EBS3–IBS3 interaction is maintained in all 10 introns with 5′-terminal inserts, the first intron nucleotide forms U:U mismatches with EBS3 in the placozoan introns and an A:A mismatch in the *Pycnoporellus* intron. Thus, the identity of the nucleotide following (or to be linked to) the IBS1 sequence may be less important for 5′ splice site hydrolysis (see also Su et al. 2001) than it is for exon ligation (Costa et al. 2000) or retrotransposition (Jimenez-Zurdo et al. 2003).

Loss of the ability to initiate splicing by branching entails only limited degeneration of ribozyme domain VI

The diversity of domain VI structures in introns with a 5′-terminal insert (Fig. 1), which stands in striking contrast to the well-conserved structure and sequence of this domain in the rest of the mitochondrial IIB1 subset, is strongly suggestive of rapid, unconstrained divergent evolution. Still, in all but the *Rhizophydium* intron, apparent degeneration is limited to sections in the middle part of domain VI that have been shown to matter to the efficiency and specificity of the branching reaction. Specifically affected are (1) the branchpoint bulging A, of which the deletion or base-pairing inhibit branching (Schmelzer and Muller 1987; van der Veen et al. 1987; Chu et al. 1998); (2) the two G:U pairs flanking the branchpoint, whose replacement by G:C pairs specifically decreases the rate of branching compared to hydrolysis (Chu et al. 1998); (3) the AAA:CUA internal loop (and its closing base pairs). Replacement of this loop by base pairs has moderate, yet significant, effects on the efficiency of branching relative to hydrolysis under stringent conditions (Chu et al. 1998). Moreover, atomic group substitutions on the 5′ side of the loop were found to interfere with branching (Chanfreau and Jacquier 1994; Boudvillain and Pyle 1998), while its deletion was reported to have a marked effect on the accessibility to the solvent of

the branchpoint nucleotide in a magnesium-bound domain VI construct (Schlatterer and Greenbaum 2008).

In contrast, both the base and tip of domain VI remain highly constrained in introns with a 5′ terminal insertion. All but the *Rhizophydium* intron retain a 3–4-bp helix at the base of domain VI, which is connected by 3-nt joining segments to domain V on one side and the 3′ splice site on the other (Fig. 1). Complete deletion of domain VI has long been known to interfere with the choice of the proper 3′ splice site (Jacquier and Jacquesson-Breuleux 1991). Moreover, shortening and, to some extent, lengthening of the segment connecting domains V and VI in *S. cerevisiae* intron cox1/5γ (Boulanger et al. 1996) not only interfere with branching, but can lead to mis-splicing, even when the reaction is initiated by hydrolysis: In deletion mutants, only a fraction of molecules used the correct 3′ splice site, despite both the 5′ and 3′ flanking nucleotides of the latter being involved in tertiary interactions (γ–γ′ and EBS3–IBS3) (Fig. 2). Interestingly, the data in Figure 1 suggest that the identity of base pairs in the basal helix of domain VI is important as well for efficient and faithful exon ligation: in introns with a 5′-terminal insert, G:C (not C:G) base pairs predominate at positions 1, 3, and 4 of the basal helix, being present in nine, eight, and seven sequences, respectively.

At the other, distal end of domain VI, seven of the 10 intron sequences with additional nucleotides at the 5′ extremity have retained a 4-nt terminal loop of the GNRA family, like nearly all mitochondrial and bacterial members of subgroup IIB1 (Fig. 1). The GUAA loop (η) that caps domain VI of the *S. cerevisiae* cox1/5γ intron was shown by Chanfreau and Jacquier (1996) to interact with a specific receptor (η′) in ribozyme subdomain IIA: Such a receptor potentially exists in all the intron sequences in Figure 1 that share a GNRA loop at the tip of domain VI (Fig. 2; data not shown). Binding of domain VI to domain II after branch formation was proposed to drag the first-step product—i.e., the 2′–5′ bonded A-G dinucleotide—out of the catalytic site, so as to make way for the 3′ splice site (Chanfreau and Jacquier 1996). However, persistence of the η–η′ interaction in introns that have lost the branchpoint structure and, presumably, the ability to carry out the branching reaction implies that formation of this interaction does not merely sequester domain VI (see Pyle 2010), but contributes also to the specific positioning of the 3′ splice site for exon ligation. In fact, disruption of η–η′ impairs specifically the second step of splicing in vitro (Chanfreau and Jacquier 1996). Even though the strikingly diverse structures of the middle part of domain VI in introns with 5′-terminal inserts (Fig. 1) may not all be capable of positioning precisely the proximal and distal ends of domain VI with respect to one another, formation of η–η′ may favor correct exon ligation simply by reducing the complexity of the conformational space to be explored to bring the 3′ splice site into the catalytic center of the ribozyme.

Endonuclease-mediated homing and the loss of the lariat structure

Compared to the thousands of group II introns that have been sequenced from hundreds of organisms (Rfam database) (Griffiths-Jones et al. 2005), the number of group II introns that possess 5′-terminal inserts is quite small. Moreover, these introns have a limited distribution, being confined to ribosomal RNA precursor molecules transcribed from mitochondrial genomes, and they belong to a single subgroup of ribozyme structures (IIB1). This makes it all the more striking that the 10 known instances of 5′-terminal inserts should result from no fewer than four to five independent insertion events (Fig. 5).

Just as remarkable, insertion of additional nucleotides at an intron 5′ end is not the only process that has been at play specifically in mitochondrial ribosomal genes and that recurrently led to the creation of novel lineages of unusual group II introns. All known examples of group II introns encoding proteins completely unrelated to reverse transcriptases also come from mitochondrial genes encoding ribosomal RNA precursor transcripts (Toor and Zimerly 2002; Monteiro-Vitorello et al. 2009; Mullineux et al. 2010); moreover, these introns belong again to subgroup IIB1, and multiple events of the insertion of an ORF (at least six of them) (Fig. 5; Supplemental Fig. S2) need as well be postulated to account for the phylogenetic distribution of ORF–ribozyme pairs. There exist, in fact, introns—the SSU788 introns of *Pycnoporellus*, *Trametes*, and *Ganoderma* and the LSU2059 intron of *Agrocybe*—that possess a 5′-terminal insert and encode a protein of the LAGLIDADG family at the same time (Fig. 5) (the ORF of the *Agrocybe* intron is defective, but a closely related, apparently intact, ORF exists in the *Ustilago maydis* SRX2 LSU2059 intron, which belongs to the same ribozyme lineage). Such a coincidence inevitably raises suspicion that some causal relationship may exist between the acquisition of a 5′-terminal insert and that of a non-RT ORF, encoding a protein with proven (in the case of *Leptographium truncatum*) (Mullineux et al. 2010) or putative endonuclease activity.

Admittedly, six out of 10 introns with 5′ inserts lack any significant protein-coding potential, while a majority of the introns that contain non-RT ORFs are devoid of 5′ inserts and have a normal domain VI, which was shown to support efficient branching in the case of the *G. frondosa* SSU788 (this study) and *L. truncatum* SSU952 (Mullineux et al. 2010) introns. However, whereas degeneration of the middle part of domain VI, which closely precedes or follows the insertion of nucleotides at the 5′ splice site, must be irreversible, acquisition of the coding sequence of a homing endonuclease is likely temporary. The reason is that, in a panmictic host population, the selective advantage provided by homing decreases rapidly as previously empty insertion sites become filled by a copy of the intron (Goddard and Burt 1999), so that the coding sequence of the endonuclease

should soon begin to accumulate deleterious mutations and degenerate beyond recognition. That is, unless the protein has become essential to its host by acquiring "maturase" activity. Maturase function, by which the intron-encoded protein participates in the splicing process, typically by helping the ribozyme to fold into an active structure, is commonplace in LAGLIDADG proteins encoded by group I introns (Ho et al. 1997; Bassi et al. 2002), but has not been detected so far for their counterparts in group II introns (Mullineux et al. 2010; G Bassi, unpubl., experiments with the *Grifola* SSU788 intron). To summarize, it is not unreasonable to hypothesize that not only the intron clades at the SSU788 and LSU2059 sites, but also those at LSU2586 and LSU2449, experienced invasion by the coding sequences of homing endonucleases and that sequencing of other group II introns inserted at these sites will eventually reveal their presence in some organisms.

Assuming then that all group II introns with 5′-terminal inserts had ancestors that encoded LAGLIDADG or other DNA endonucleases, why should relying on these proteins for homing eventually lead to the loss of branching? On the one hand, the lariat structure appears essential for retrotransposition by inverse splicing; the linear intron molecules that result from hydrolysis at the 5′ splice site are unable to perform the second transesterification reaction (reverse of branch formation) and to complete their integration into a DNA target by themselves. On the other hand, the intron RNA, whether branched or linear, does not play any part in the homing process mediated by DNA endonucleases of the LAGLIDADG and GIY-YIG families, which rests on resealing of a double-strand break by general, homologous recombination, using the intact, intron-carrying copy as template. Thus, once a group II retrotransposon has been converted into a DNA transposon (class II mobile element) (Wicker et al. 2007) by the loss of its reverse transcriptase and the acquisition of the coding sequence of a homing endonuclease, a 2′–5′ phosphodiester bond should no longer be required for mobility: The ability to generate this bond could become lost through mutations at, or next to, the branchpoint or else, the insertion of nucleotides at the 5′ splice site.

Why a branched intron structure in the absence of retrotransposition?

While the branching reaction would no longer appear necessary in introns that have lost retrotransposition, the data in Table 1 point to a much more complex reality. The mere fact that a majority of mitochondrial subgroup IIB1 ribozymes have retained a canonical branchpoint means that a branched structure remains somehow important for introns that do not encode an RT gene. Furthermore, since the branchpoint and, presumably, branching have survived the acquisition of a homing endonuclease gene in more than half of the introns expected to propagate (or to have propagated) as DNA, initiation of splicing by transesterification may remain advantageous even in this subset.

At the same time, one might question the need for a branched structure even in retrotransposition. In fact, correct integration of a linear intron that has undergone partial reverse splicing followed by reverse transcription should still be possible, by recombination with the intron-carrying DNA copy: This is how the unidirectional conversion of upstream exon sequences that accompanies insertion of *S. cerevisiae* intron cox1/2 into its intron-less target has been accounted for (Lazowska et al. 1994; Eskes et al. 2000). However, the 5′ exons of the intron-carrying donor and recipient molecules must be homologous, as is the case, indeed, when, but only when, homing—as opposed to ectopic transposition—is involved. Such situations in which an intron is transmitted partly by retrotransposition and partly by homologous recombination may actually reflect transition from one mode of propagation to the other.

Admittedly, retrotransposition even of exon-less, linear intron molecules was recently reported in heterologous systems (see Zhuang et al. 2009). However, that process, which involves nonhomologous end-joining at the 5′ intron extremity, is orders of magnitude less efficient than lariat retrohoming. In fact, imprecise recombination at an intron 5′ end can generate 5′-terminal inserts, whose presence, and the resulting loss of branching, would trigger rapid degeneration of the branchpoint structure. Alternatively, since even linear intron molecules may retain the ability to attack suitable targets with their 3′ extremity and generate partially reverse spliced molecules, it may be argued that the loss of branching should be followed by insertion events at the 5′ splice site. Whatever the actual mechanism, degeneration of the branchpoint structure and the acquisition of a 5′-terminal insert must be closely coupled in subgroup IIB1, for evolutionary intermediates have not been found so far.

Coming back to the possible significance of branching for introns devoid of an RT gene, the overall coevolution of the ribozyme and protein components of group II introns (Toor et al. 2001) makes it unlikely that the intimate molecular interactions at its root could form back once they have been lost. Still, RT-less, lariat-forming introns may manage to transpose by diverting, whether on an occasional or more lasting basis, a group II-encoded reverse transcriptase that happens to be synthesized in the cellular compartment in which they reside. (Mitochondrial members of subgroup IIA, another subclass of group II introns that is widely distributed in organelles [Michel et al. 1989; Toor et al. 2001] generally encode reverse transcriptases and at least some of them are indeed mobile [Lazowska et al. 1994].)

A more subtle justification for retaining the ability to form lariats takes its roots in experimental evidence pointing to (some levels of) indiscriminate reverse transcription by group II–encoded reverse transcriptases. In yeast mitochondria, the presence of RT-encoding group II introns has been shown to promote genomic deletion of both group II

and group I introns (Gargouri et al. 1983), presumably via cDNA synthesis from spliced mRNAs. Such occasional reverse transcription could lead as well to the genomic gain of an intron that had happened to reverse-splice into an ectopic RNA site.

The branching reaction may also remain advantageous because it is liable to be more efficient than hydrolysis. This is certainly the case in vitro (Jacquier and Jacquesson-Breuleux 1991; see also Fig. 3), at physiological pH values, and could also be true in vivo, unless folding of precursor molecules were to remain rate-limiting even when compared to hydrolysis. Yet another potential advantage of making lariats is that it provides resistance to digestion by exonucleases. Stabilization of the intron would, in turn, stabilize the mRNA for the intron-encoded protein, as was argued, for instance, to account for the production of mini-, 5′-terminal lariats by a group I–derived ribozyme (Nielsen et al. 2005). However, some alternative mechanisms must exist to allow efficient production of homing endonucleases from linear intron molecules: The extensive, long-range RNA–RNA pairings that flank the ORFs of the *Pycnoporellus* molecule and other introns with 5′-terminal inserts (Fig. 2B; data not shown) may substitute for the 2′–5′ phosphodiester bond of the lariat and slow down the progression of exonucleases.

Possible implication of the EBS2–IBS2 pairing in branch formation

Another feature that shows correlation with the loss of branching is the absence of EBS2 (Table 1). As already emphasized (see Results), all 10 introns with a 5′ insert actually lack the entire subdomain that the EBS2 segment is normally part of (Michel et al. 1989; Dai et al. 2003). The EBS2–IBS2 pairing is known to be important for insertion of group II introns by reverse-splicing into double-stranded DNA, presumably because it helps stabilize interactions between the intron and its target relative to DNA:DNA base-pairing, but it does not appear to be required for transposition into single-stranded nucleic acids (Coros et al. 2005). Still, that this interaction should persist not only in introns that have lost the coding sequence for a reverse transcriptase, but in several of those that encode a LAGLIDADG homing endonuclease (Table 1), implies that the EBS2–IBS2 pairing has some significant function in splicing as well. Partial disruption of that interaction in *S. cerevisiae* intron cox1/5γ decreases the stability of the complex between the 5′ exon and intron, resulting in accumulation of the intron–3′exon reaction intermediate in vitro (Jacquier and Michel 1987). Somewhat more unexpectedly, it also appears to affect the chemical step of the reaction by which oligonucleotides that mimic the intron target site are cleaved (Xiang et al. 1998).

Close examination of the data in Table 1 and, in particular, the absence of EBS2 from all but one of the SSU788 introns, suggests that loss of the EBS2–IBS2 pairing precedes, and might even be a necessary step for, the loss of branching. One possibility is that the deletion of EBS2 somehow facilitates hydrolysis at the 5′ splice site (although for the *Grifola* ribozyme, hydrolysis-initiated self-splicing was found to constitute but a minor reaction pathway even in the presence of potassium) (Fig. 3B). This would not be without precedent, for the only intron subclass—subgroup IIC—in which these components are systematically missing is noteworthy for (most of) its members initiating self-splicing in vitro by hydrolysis (Granlund et al. 2001; Toor et al. 2006).

Unfortunately, possible ways in which EBS2 and the structures that surround it might affect the balance between transesterification and hydrolysis remain difficult to think of at present. The only currently available group II crystal structure (Toor et al. 2008a,b) happens to be that of a subgroup IIC intron, and it lacks not only domain VI, but the EBS2–IBS2 interaction and a number of additional RNA subdomains and devices that a majority of other lineages of group II ribozymes have opted to conserve (for review, see Pyle 2010). Additional group II structures, in which domain VI and the branchpoint can be visualized in interaction with the rest of the ribozyme, are a prerequisite if we are eventually to reach a complete understanding of why the branching reaction has been so stubbornly, although not universally, retained during the diversifying evolution of group II introns.

MATERIALS AND METHODS

Sequence analyses of mitochondrial subgroup IIB1 ribozymes

Published sequences of mitochondrial introns that possessed characteristic sequence and secondary structure features of subgroup IIB1 (Michel et al. 1989) were collected (Table 1), and their ribozyme sections were manually aligned (Supplemental Data Set) based on conservation of both sequence and potential secondary structure (the distal sections of stems IC2, ID2, IIA, IIIB, and IV [see Fig. 2; Michel et al. 2009] could not be reliably aligned and were discarded). Starting from this alignment, a phylogenetic tree was generated by PAUP* 4.0b10 (Swofford 2002) using the Neighbor-Joining algorithm and a matrix that had been obtained by using the LogDet measure of distance, which is insensitive to differences in base composition (Lockhart et al. 1994). Note that (1) to avoid biasing the tree-building procedure in favor of subsets constituted by introns that share homologous insertion sites, the EBS1, EBS2, and EBS3 sites were removed, leaving 526 sites in the final alignment; (2) of the three closely related *G. intraradices* LSU1787 intron sequences in Table 1, only the first one, which does not include an ORF, was retained for the tree-building process.

Sequence analyses of LAGLIDADG proteins

To investigate the phylogenetic relationships of LAGLIDADG proteins potentially encoded by subgroup IIB1 introns, apparently

intact or defective sequences generated from the intron nucleotide sequences (Table 1) were compared to the NCBI nonredundant protein data set, and for each comparison, the 10 target sequences with the highest BLASTP scores were retained. The resulting sequence data set was then aligned together with the 91 sequences in the Pfam LAGLIDADG 1 (PF00961) "seed" set. After manual refinement, the final alignment (available from the authors) consisted of 146 sites and 174 sequences, 26 of which were of presumably dimeric proteins (Stoddard 2005), with a single LAGLIDADG motif, while the rest corresponded to monomeric proteins (in which case, each of the two sections following a LAGLIDADG motif was aligned separately).

Because of the rather large number of sequences in this data set, we resorted to the efficient Neighbor-Joining algorithm, using distances generated by the program PROTDIST (Felsenstein 2004), to generate a phylogenetic tree and calculate bootstrap percentages (Supplemental Fig. S2). Even though the resulting phylogeny is far from being completely resolved, our main conclusions regarding phylogenetic relationships of the proteins encoded by group II introns (see above) are supported by relatively high bootstrap percentages and/or the fact that groupings were found to be the same whether the first or second pseudo-repeat of monomeric proteins was used for comparisons (Supplemental Fig. S2).

Amplification, sequencing, and cloning of fungal introns

DNA extracts from *A. botryosus* CBS195.91, *G. frondosa* CBS 480.63, and *P. fulgens* T-325 were obtained from David Hibbett (Clark Fungal Database at Clark University). PCR amplifications of the SSU788 intron and surrounding exons were performed in 50 μL with 1 μM primers BMS65MOD and BMS103E (Supplemental Table S1) using 1 unit of high-fidelity Phusion polymerase in HF buffer (Finnzymes) and 33 cycles (10 sec at 98°C, 45 sec at 60°C, 90 sec at 72°C). Sequencing of amplification products was carried out on both strands by GATC Biotech using the same primers as well as species-specific primers listed in Supplemental Table S1. Accession numbers for the assembled sequences are FR773978, FR773979, and FR773980.

For cloning into *Escherichia coli*, amplification products were reamplified with primers BMS65MODT7 and BMS103EZ, digested with BamHI and XmaI, and ligated into the pUC19 vector plasmid. For deletion of ORF sequences from ribozyme domain IV of the *G. frondosa* and *P. fulgens* introns, primers GRXHOREV (or PYXHOREV) and GRXHOFWD (or PYXHOFWD) (see Supplemental Table S1) were used in combination with vector-specific primers ANT7 and 24mer, respectively, to generate PCR products. These products were digested with XhoI and either BamHI or XmaI, and cloned back into pUC19. The resulting constructs, pUC19-GR1ΔORF and pUC19-PY1ΔORF, in which most of domain IV has been replaced by an XhoI site (Fig. 2, legend), were verified by sequencing.

In vitro transcription and purification of precursor RNA

Templates for synthesis of the *Grifola* and *Pycnoporellus* precursor RNAs were obtained by digestion of plasmids pUC19-GR1ΔORF and pUC19-PY1ΔORF with SmaI. RNA synthesis and purification were carried out as described in Costa et al. (1997b), except that the transcription mixture contained 10% DMSO so as to avoid premature transcription stops and a 1.55 molar concentration ratio of magnesium over nucleotides was used to prevent premature intron splicing.

Self-splicing reactions

Precursor transcripts internally labeled with ^{32}P-UTP were denatured in water at 90°C prior to cooling to reaction temperature. Reactions were started by addition of an equal volume of 2×-concentrated splicing buffer. The final concentration of precursor molecules was 20 nM. Reactions were stopped by addition of an equal volume of a solution of formamide containing EDTA at a concentration appropriate to complex all of the magnesium, and products were separated on a denaturing polyacrylamide gel (50% [w/v] urea, 4% total acrylamide, 0.2% bis-acrylamide). Radioactivity was quantitated on fixed, dried gels using a PhosphorImager (MolecularDynamics), and the molar fraction of each product was calculated. Reaction time courses were fitted to either single $\{m_1[1 - \exp(-kt)]\}$ or double $\{m_1[1 - \exp(-k_1t)] + m_2[1 - \exp(-k_2t)]\}$ exponentials (m_1 or $m_1 + m_2$ are final fractions of reacted product).

Reverse transcription of splicing products

Preparative self-splicing reactions were carried out in 40 mM Tris-Cl (pH 7.5), 20 mM MgCl$_2$, and 1 M NH$_4$Cl at 42°C (the *G. frondosa* intron–3′exon lariat molecule was isolated from a splicing reaction that included 20 mM CaCl$_2$). Purification of splicing products from preparative denaturing polyacrylamide gels and their reverse transcription with ^{32}P-labeled, gel-purified oligonucleotides were performed essentially as described by Costa et al. (1997b). The following oligonucleotides (see Supplemental Table S1) were used for reverse transcription: BMS103B, to sequence ligated exons and determine the branchpoint of the *G. frondosa* intron–3′exon lariat; Gr-R2, to determine the 5′ splice of the *G. frondosa* intron lariat; Py-R2, to determine the 5′ extremity of *P. fulgens* linear intron molecules.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

REFERENCES

Bassi GS, de Oliveira DM, White MF, Weeks KM. 2002. Recruitment of intron-encoded and co-opted proteins in splicing of the bI3 group I intron RNA. *Proc Natl Acad Sci* **99:** 128–133.

Beauregard A, Curcio MJ, Belfort M. 2008. The take and give between retrotransposable elements and their hosts. *Annu Rev Genet* **42:** 587–617.

Bonen L. 2008. *Cis-* and *trans-*splicing of group II introns in plant mitochondria. *Mitochondrion* **8:** 26–34.

Bonocora RP, Shub DA. 2009. A likely pathway for formation of mobile group I introns. *Curr Biol* **19:** 223–228.

Boudvillain M, Pyle AM. 1998. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J* **17:** 7091–7104.

Boulanger SC, Faix PH, Yang H, Zhuo J, Franzen JS, Perlman PS. 1996. Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3′ splice site selection. *Mol Cell Biol* **16:** 5896–5904.

Bruns TD, Szaro TM, Gardes M, Cullings KW, Pan JJ, Taylor DL, Horton TR, Kretzer A, Garbelotto M, Li Y. 1998. A sequence database for the identification of ectomycorrhizal basidiomycetes by phylogenetic analysis. *Mol Ecol* **7:** 257–272.

Chanfreau G, Jacquier A. 1994. Catalytic site components common to both splicing steps of a group II intron. *Science* **266:** 1383–1387.

Chanfreau G, Jacquier A. 1996. An RNA conformational change between the two chemical steps of group II self-splicing. *EMBO J* **15:** 3466–3476.

Chu VT, Liu Q, Podar M, Perlman PS, Pyle AM. 1998. More than one way to splice an RNA: Branching without a bulge and splicing without branching in group II introns. *RNA* **4:** 1186–1202.

Coros CJ, Landthaler M, Piazza CL, Beauregard A, Esposito D, Perutka J, Lambowitz AM, Belfort M. 2005. Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol* **56:** 509–524.

Costa M, Deme E, Jacquier A, Michel F. 1997a. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol* **267:** 520–536.

Costa M, Fontaine JM, Loiseaux-de Goër S, Michel F. 1997b. A group II self-splicing intron from the brown alga *Pylaiella littoralis* is active at unusually low magnesium concentrations and forms populations of molecules with a uniform conformation. *J Mol Biol* **274:** 353–364.

Costa M, Michel F, Westhof E. 2000. A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* **19:** 5007–5018.

Dai L, Toor N, Olson R, Keeping A, Zimmerly S. 2003. Database for mobile group II introns. *Nucleic Acids Res* **31:** 424–426.

Daniels DL, Michels WJ Jr, Pyle AM. 1996. Two competing pathways for self-splicing by group II introns: A quantitative analysis of in vitro reaction rates and products. *J Mol Biol* **256:** 31–49.

Eskes R, Liu L, Ma H, Chao MY, Dickson L, Lambowitz AM, Perlman PS. 2000. Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol* **20:** 8432–8446.

Fedorova O, Pyle AM. 2005. Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. *EMBO J* **24:** 3906–3916.

Felsenstein J. 2004. *PHYLIP (Phylogeny Inference Package) version 3.6.* Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Gargouri A, Lazowska J, Slonimski PP. 1983. DNA-splicing of introns in the gene: A general way of reverting intron mutations. In *Mitochondria 1983: Nucleo–mitochondrial interactions* (ed. RJ Schweyen et al.), pp. 259–268. de Gruyter, Berlin.

Goddard MR, Burt A. 1999. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci* **96:** 13880–13885.

Gonzalez P, Barroso G, Labarère J. 1999. Molecular gene organisation and secondary structure of the mitochondrial large subunit ribosomal RNA from the cultivated Basidiomycota *Agrocybe aegerita*: a 13 kb gene possessing six unusual nucleotide extensions and eight introns. *Nucleic Acids Res* **27:** 1754–1761.

Granlund M, Michel F, Norgren M. 2001. Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J Bacteriol* **183:** 2560–2569.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res* **33:** D121–D124.

Hallick RB, Hong L, Drager RG, Favreau MR, Monfort A, Orsat B, Spielmann A, Stutz E. 1993. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res* **21:** 3537–3544.

Ho Y, Kim SJ, Waring RB. 1997. A protein encoded by a group I intron in *Aspergillus nidulans* directly assists RNA splicing and is a DNA endonuclease. *Proc Natl Acad Sci* **94:** 8994–8999.

Hong SG, Jeong W, Jung HS. 2002. Amplification of mitochondrial small subunit ribosomal DNA of polypores and its potential for phylogenetic analysis. *Mycologia* **94:** 823–833.

Jacquier A. 1996. Group II introns: Elaborate ribozymes. *Biochimie* **78:** 474–487.

Jacquier A, Jacquesson-Breuleux N. 1991. Splice site selection and role of the lariat in a group II intron. *J Mol Biol* **219:** 415–428.

Jacquier A, Michel F. 1987. Multiple exon-binding sites in class II self-splicing introns. *Cell* **50:** 17–29.

Jarrell KA, Peebles CL, Dietrich RC, Romiti SL, Perlman PS. 1988. Group II intron self-splicing. Alternative reaction conditions yield novel products. *J Biol Chem* **263:** 3432–3439.

Jimenez-Zurdo JI, Garcia-Rodriguez FM, Barrientos-Duran A, Toro N. 2003. DNA target site requirements for homing in vivo of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. *J Mol Biol* **326:** 413–423.

Johansen S, Haugen P. 2001. A new nomenclature of group I introns in ribosomal DNA. *RNA* **7:** 935–936.

Kennell JC, Moran JV, Perlman PS, Butow RA, Lambowitz AM. 1993. Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell* **73:** 133–146.

Lambowitz AM, Zimmerly S. 2004. Mobile Group II introns. *Annu Rev Genet* **38:** 1–35.

Lang BF, Laforest MJ, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet* **23:** 119–125.

Lazowska J, Meunier B, Macadre C. 1994. Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstream-located markers. *EMBO J* **13:** 4963–4972.

Lockhart PJ, Steel MA, Hendy MD, Penny D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol Biol Evol* **11:** 605–612.

Michel F, Ferat JL. 1995. Structure and activities of group II introns. *Annu Rev Biochem* **64:** 435–461.

Michel F, Umesono K, Ozeki H. 1989. Comparative and functional anatomy of group II catalytic introns—a review. *Gene* **82:** 5–30.

Michel F, Costa M, Westhof E. 2009. The ribozyme core of group II introns: a structure in want of partners. *Trends Biochem Sci* **34:** 189–199.

Monteiro-Vitorello CB, Hausner G, Searles DB, Gibb EA, Fulbright DW, Bertrand H. 2009. The *Cryphonectria parasitica* mitochondrial *rns* gene: Plasmid-like elements, introns and homing endonucleases. *Fungal Genet Biol* **46:** 837–848.

Mullineux ST, Costa M, Bassi GS, Michel F, Hausner G. 2010. A group II intron encodes a functional LAGLIDADG homing endonuclease and self-splices under moderate temperature and ionic conditions. *RNA* **16:** 1818–1831.

Nielsen H, Westhof E, Johansen S. 2005. An mRNA is capped by a 2′, 5′ lariat catalyzed by a group I-like ribozyme. *Science* **309:** 1584–1587.

Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. 2000. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289:** 920–930.

Podar M, Chu VT, Pyle AM, Perlman PS. 1998. Group II intron splicing in vivo by first-step hydrolysis. *Nature* **391:** 915–918.

Pyle AM. 2010. The tertiary structure of group II introns: Implications for biological function and evolution. *Crit Rev Biochem Mol Biol* **45:** 215–232.

Quiroga C, Roy PH, Centron D. 2008. The S.ma.I2 class C group II intron inserts at integron attC sites. *Microbiology* **154:** 1341–1353.

Schlatterer JC, Greenbaum NL. 2008. Specificity of Mg^{2+} binding at the Group II intron branch site. *Biophys Chem* **136:** 96–100.

Schmelzer C, Muller MW. 1987. Self-splicing of group II introns in vitro: Lariat formation and 3′ splice site selection in mutant RNAs. *Cell* **51:** 753–762.

Schuwirth BS, Borovinskaya MA, Hau CW, Zhang W, Vila-Sanjurjo A, Holton JM, Cate JH. 2005. Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310:** 827–834.

Stabell FB, Tourasse NJ, Ravnum S, Kolsto AB. 2007. Group II intron in *Bacillus cereus* has an unusual 3′ extension and splices 56 nucleotides downstream of the predicted site. *Nucleic Acids Res* **35:** 1612–1623.

Steitz TA, Steitz JA. 1993. A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci* **90:** 6498–6502.

Stoddard BL. 2005. Homing endonuclease structure and function. *Q Rev Biophys* **38:** 49–95.

Su LJ, Qin PZ, Michels WJ, Pyle AM. 2001. Guiding ribozyme cleavage through motif recognition: The mechanism of cleavage site selection by a group II intron ribozyme. *J Mol Biol* **306:** 655–668.

Swofford DL. 2002. *PAUP*: Phylogenetic analysis using parsimony* [*and other methods*]. *4.0 Beta*. Sinauer Associates, Sunderland, MA.

Toor N, Zimmerly S. 2002. Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA* **8:** 1373–1377.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7:** 1142–1152.

Toor N, Robart AR, Christianson J, Zimmerly S. 2006. Self-splicing of a group IIC intron: 5′ exon recognition and alternative 5′ splicing

events implicate the stem loop motif of a transcriptional terminator. *Nucleic Acids Res* **34:** 6461–6471.

Toor N, Keating KS, Taylor SD, Pyle AM. 2008a. Crystal structure of a self-spliced group II intron. *Science* **320:** 77–82.

Toor N, Rajashankar K, Keating KS, Pyle AM. 2008b. Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol* **15:** 1221–1222.

van der Veen R, Kwakman JH, Grivell LA. 1987. Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J* **6:** 3827–3831.

Vogel J, Börner T. 2002. Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J* **21:** 3794–3803.

Wallasch C, Morl M, Niemer I, Schmelzer C. 1991. Structural requirements for selection of 5′- and 3′ splice sites of group II introns. *Nucleic Acids Res* **19:** 3307–3314.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8:** 973–982.

Xiang Q, Qin PZ, Michels WJ, Freeland K, Pyle AM. 1998. Sequence specificity of a group II intron ribozyme: Multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37:** 3839–3849.

Zeng Q, Bonocora RP, Shub DA. 2009. A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. *Curr Biol* **19:** 218–222.

Zhuang F, Mastroianni M, White TB, Lambowitz AM. 2009. Linear group II intron RNAs can retrohome in eukaryotes and may use nonhomologous end-joining for cDNA ligation. *Proc Natl Acad Sci* **106:** 18189–18194.

SUPPLEMENTARY MATERIAL

TABLE S1. List of oligonucleotides used (sequences 5' to 3')

```
BMS65MOD       GGTGCCAGAAGACTCGGTAAGA
BMS65MODT7     TGCTGGATCCTTTAATACGACTCACTATAGGGTGCCAGAAGACTCGGTAAGA
BMS103E        CACTCCGTTTGCTTCGAGACCGAC
BMS103EZ       ATGCGAGCTCGGTACCCGGGCACTCCGTTTGCTTCGAGACCGAC
BMS103B        GTACTCACAAGGCGGAATGG
Al-S1          TTAATTAGACAAGTAATTGTCCT
Al-R1          CACTTGTAATATCTATACGATAAG
Al-R2          TATATTCTAGTATAGGTTTATCC
Gr-S1          GGATGTGCGACTTGAAAAG
Gr-S2          CTGAAAGGGTGCCCACTTT
Gr-S3          TAAAAGACTTAAATAGTATTATTC
Gr-R1          TCTAGAATGAACTTTTGGATTT
Gr-R2          AAGCTTAACACAAGACCATTACTGG
Py-S1          ACAGGGAAACTCCTATAATC
Py-S2          TTTTATTAATGGGGAAGGTT
Py-S3          GTCGCACGTACAGTTCTTAG
Py-R1          AAACATTTGGAGTTAAATCTAATC
Py-R2          TAAATCTTAAGCTTTAGTCCTAG
GRXHOFWD       AGCTACGTCACTCGAGTTATATGTCCTTAGTAAATTCTCAAA
GRXHOREV       GTATCGTCTTCTCGAGAGATATGTACTTGGCTAGTAGATTGA
PYXHOFWD       AGCTACGTCACTCGAGCTTTACAATTAACACCCAATATACTA
PYXHOREV       TATCGTCTTCCTCGAGTAGTACGCCAATATAATACTT
ANT7           AACAGCTATGACCATGATTACG
24mer          CGCCAGGGTTTTCCCAGTCACGAC
```

Supplementary Figure Legends

FIGURE S1. Secondary structure model of the *Aleurodiscus botryosus* SSU788 intron. Arrowheads point to splice junctions and an asterisk, to the branchpoint. Labeling of secondary structure components and tertiary interactions as in Michel et al. (2009).

FIGURE S2, parts A, B and C. Phylogenetic relationships of LAGLIDADG proteins potentially encoded by subgroup IIB1 introns. This phylogenetic tree was built by the Neighbor-Joining algorithm from a distance matrix generated by program PROTDIST (Felsenstein, 2004) using a set of 174 sequences aligned over 146 sites (see Materials and Methods; bootstrap percentages, when at least equal to 70, are indicated next to nodes). The tree was rooted by choosing proteins encoded by the LSU2593 introns as outgroup. For monomeric proteins, I- and II- refer to sections following the first and second LAGLIDADG motif, respectively (thickened branches indicate subclades in common between sections of the tree generated from the first and second LAGLIDADG pseudo-repeat). For organelle introns, mt stands for mitochondrial and cp for chloroplastic. Introns are named according to their location in host genes, see Table 1. Accession numbers for protein sequences (exceptionally, for nucleotide sequences, when a protein accession number is lacking) are indicated between square brackets. Names of proteins encoded by group II introns are in red, the corresponding clades are named according to the intron location in ribosomal RNA genes; except for three genes designated 'ORF', all of the other coding sequences are located within group I introns.

Aleurodiscus botryosus
mit SSU788

Li *et al.*   Fig. S2 part A

Li *et al.*   Fig. S2 part C
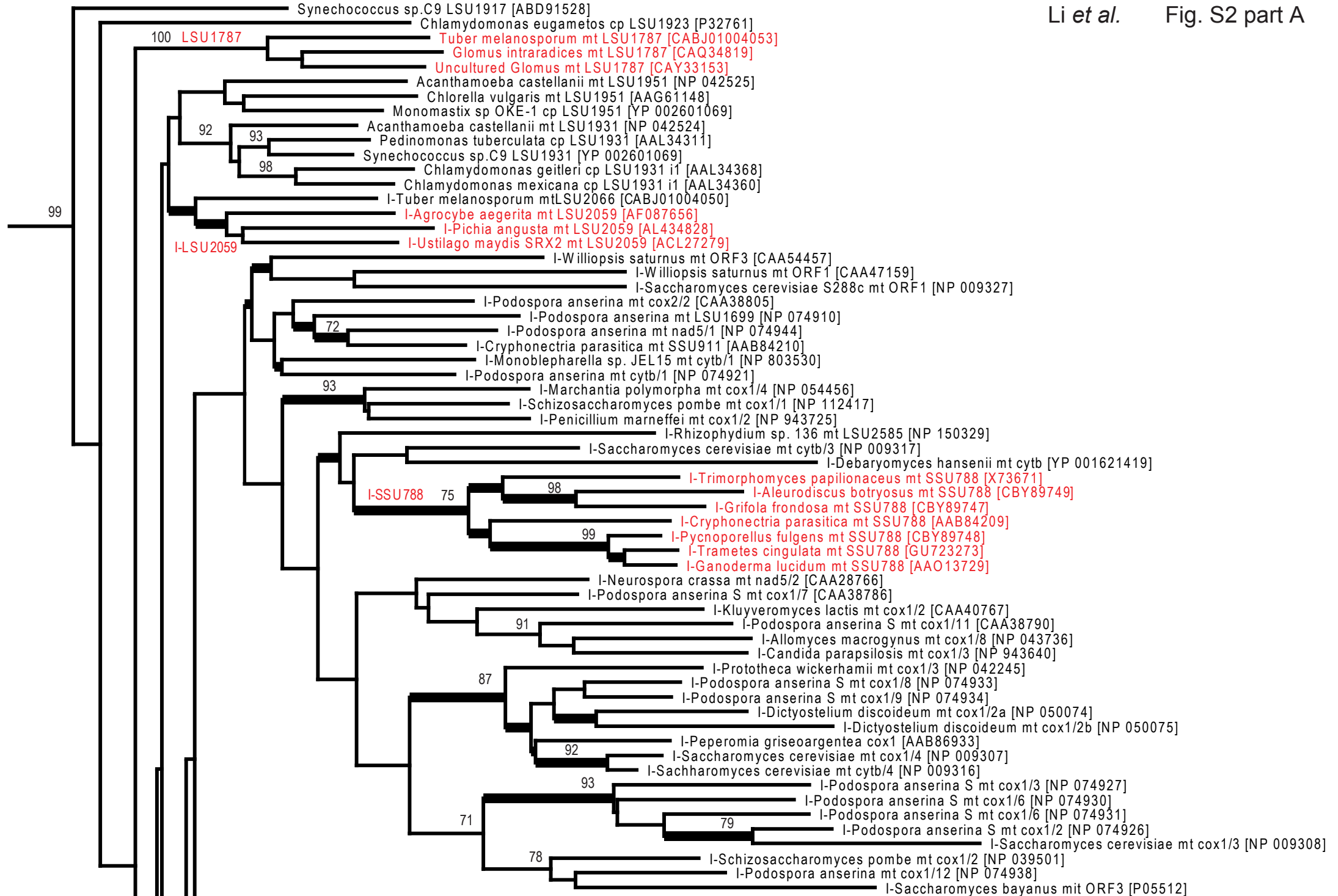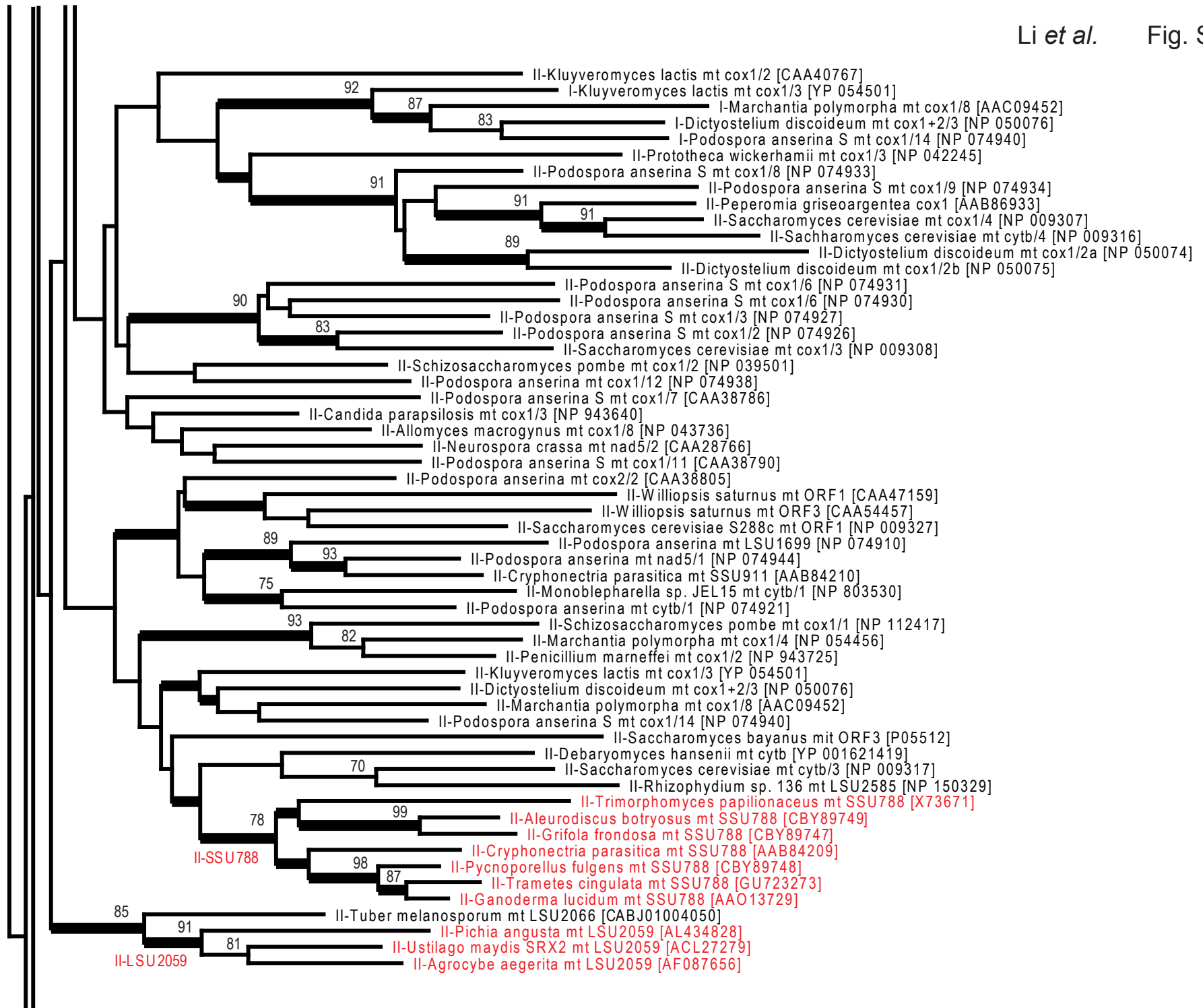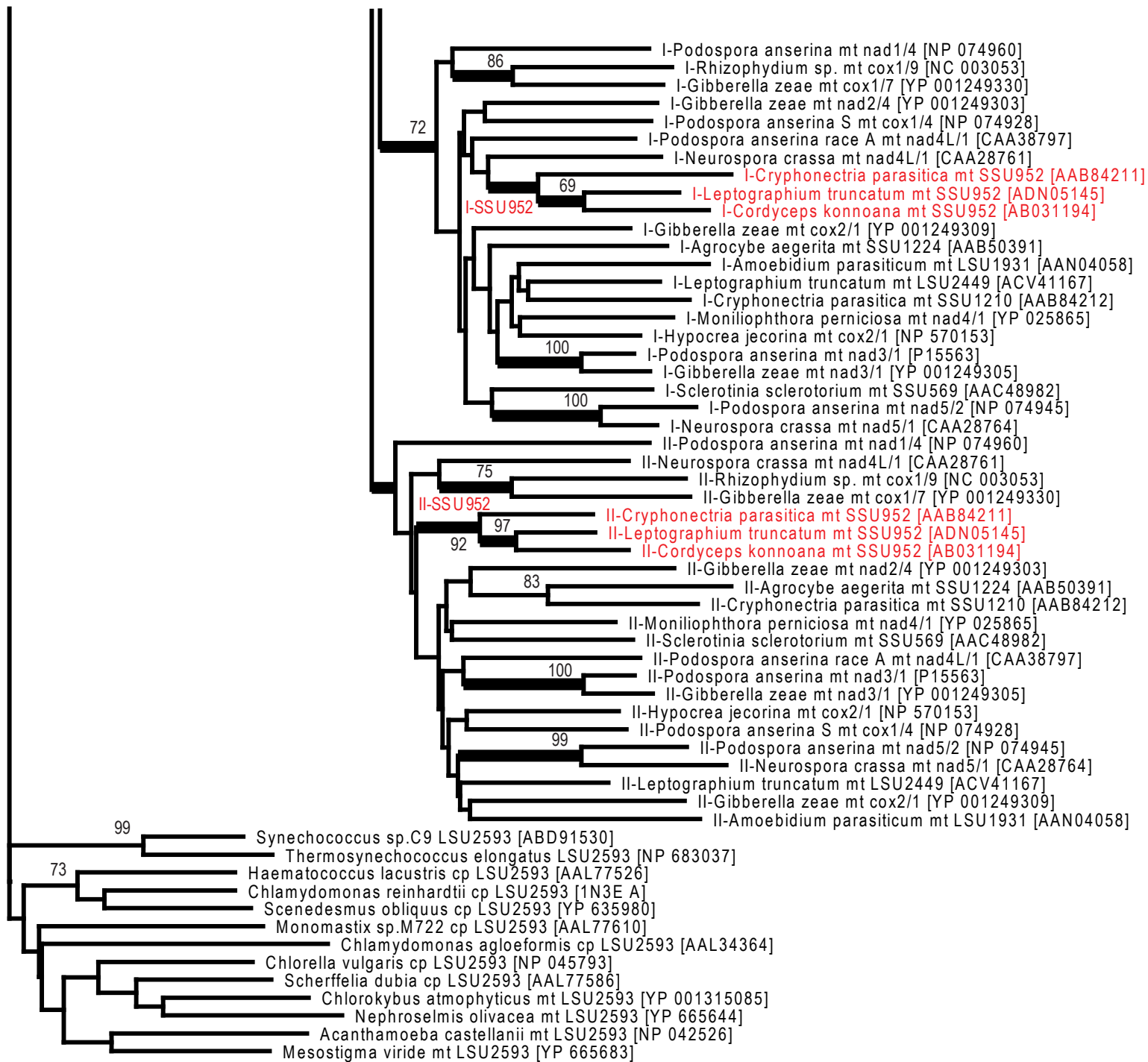
Article n° 2 :

Cheng-Fang Li, Maria Costa and François Michel

Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron

Linking the branchpoint helix to a newly found receptor allows lariat formation by a group II intron

Cheng-Fang Li[1,2], Maria Costa[1] and François Michel[1,*]

[1]Centre de Génétique Moléculaire du C.N.R.S., 1 Avenue de la Terrasse, 91190 Gif-sur-Yvette, France,

[2]Department of Life Science and Institute of Biotechnology, National Tsing Hua University, Hsinchu, Taiwan

*Corresponding author

Tel. : +33 1 69 82 31 88; Fax: +33 1 69 82 43 86: E-mail: michel@cgm.cnrs-gif.fr

Running Title: RNA receptor for group II intron branchpoint helix

Total character count, including spaces: 51649

Abstract

Like spliceosomal introns, the ribozyme-containing group II introns are excised as branched, lariat structures: a 2'-5' bond is created between the first nucleotide of the intron and an adenosine in domain VI, a component which is missing from available crystal structures of the ribozyme. Comparative sequence analysis, modeling and nucleotide substitutions point to the existence, and probable location, of a specific RNA receptor for the section of domain VI that lies just distal to the branchpoint adenosine. By designing oligonucleotides that tether domain VI to this novel binding site, we have been able to specifically activate lariat formation in an engineered, defective group II ribozyme. The location of the newly identified receptor implies that prior to exon ligation, the distal part of domain VI undergoes a major translocation, which can now be brought under control by the system of anchoring oligonucleotides we have developed. Interestingly, these oligonucleotides, which link the branchpoint helix and the binding site for intron nucleotides 3-4, may be viewed as counterparts of U2-U6 helix III in the spliceosome.

2

Introduction

Group II introns, when fully functional, are retrotransposons composed of a large ribozyme and the coding sequence of a reverse transcriptase. The ribozyme catalyzes splicing of the intron-containing precursor transcript and reverse splicing of the excised intron into DNA targets, while the intron-encoded protein is essential to copy the inserted intron RNA into DNA (Lambowitz and Zimmerly, 2004). Both the ribozyme components of group II introns and the eukaryotic spliceosome excise introns as branched, lariat structures. Lariats result from a 2'-5' phosphodiester bond being formed between an adenosine internal to the intron and the first intron nucleotide. In group II introns, the adenosine whose 2'OH group will attack the 5' splice site during the first step of splicing bulges out of ribozyme domain VI, on its 3' side (Figure 1A). After the branching reaction, the newly formed 2'-5' dinucleotide is removed from the (apparently) single ribozyme catalytic center and replaced by the 3' splice site in order for exon ligation to take place (Chanfreau and Jacquier, 1994; 1996).

Except for its branchpoint adenosine, the rather small domain VI is poorly conserved between subgroups of group II ribozymes (e.g. Michel *et al*, 2009) and its sequence and secondary structure may vary even within sets of closely related introns. Nevertheless, an RNA tertiary contact involving domain VI and domain II (η-η' in Figure 1A), which had been identified by Chanfreau and Jacquier (1996) in a screen for interactions specific to the exon ligation step, was subsequently shown to be present in both major subdivisions, IIA and IIB, of the group II intron family. In addition to dramatically reducing the rate of exon ligation, disruption of η-η' promotes branching: it increases the rate of first step transesterifications (branching and its reverse reaction, debranching) and, in a subgroup IIA intron (Costa *et al*, 1997a), it was shown to favor branching over hydrolysis at the 5' splice site. The latter is usually a minor reaction which only prevails when the branchsite is missing

or mutated (Van der Veen *et al*, 1987), when the 5' splice site is separated from the rest of the intron (Jacquier and Jacquesson-Breuleux, 1991) or else, in the presence of potassium ions (Jarrell *et al*, 1988). These data were rationalized by postulating that group II ribozymes exist in two conformations, one in which η-η' contributes to the specific positioning of the 3' splice site for exon ligation and another one in which domain VI and the branch site are somehow poised for branching.

By contrast to the identification of η-η', the search for interactions that, by being specific to the branching step, could contribute to our understanding of the mechanism by which formation of the lariat bond is activated, proved particularly frustrating. Only in 2006 was a candidate receptor for the domain VI branchpoint finally proposed by Hamill and Pyle, based on crosslinking experiments. This receptor consists of a subdomain ID internal loop which had previously been shown to contain the binding site for the 3' exon of subgroup IIB introns and to be indirectly involved as well in the binding of the 5' exon (Costa *et al*, 2000); it was accordingly dubbed the 'coordination loop' by Hamill and Pyle (Figure 1A). However, no counterpart for the subgroup IIB coordination loop can be discerned in secondary struture models of subgroup IIA ribozymes (see Michel *et al*, 2009), which is surprising, given the nearly universal conservation of the branchpoint adenosine and bulge. Also, some nucleotide substitutions in the coordination loop do reduce dramatically the rate at which precursor molecules react (Hamill and Pyle, 2006), but they have not been shown to affect branching specifically (that is, with respect to hydrolysis).

The first atomic-resolution structure of a group II ribozyme, by Toor *et al* (2008a), lacked both the coordination loop and domain VI. Subsequent refinements of this structure have made it possible to visualize the coordination loop and its predicted interactions with both the 3' exon and the EBS1 loop which binds the 5' exon (Toor *et al*, 2010; Wang, 2010),

but domain VI remains invisible, possibly because its flexibility leads to its degradation (see discussion in Pyle, 2010).

This situation, and our recent finding that the ability to initiate splicing by branching was recurrently lost during the evolution of a subclass of natural group II introns (Li *et al*, 2011) prompted us to reexamine the sequences of group II ribozymes in search for a potential receptor site that would bind the nucleotides that surround the branchpoint, in the middle part of domain VI. We now show that there exists such a candidate site, located in subdomain IC1 (Figure 1), at which nucleotide substitutions specifically affect branching, rather than hydrolysis. In a second stage, by taking advantage of the currently available group II ribozyme structure, we were able to model the possible interaction of domain VI with this receptor and from there, to create an allosteric ribozyme (Tang and Breaker, 1997), whose ability to form the lariat bond depends on oligonucleotides that anchor domain VI to its binding site.

Results

Comparison of introns with and without branchsites points to a potential first-step receptor for domain VI

It has long been known that some rare group II introns in organelles lack a bulging A on the 3' side of domain VI (Michel *et al*, 1989; Li-Pook-Than and Bonen, 2006) and at least one of these introns, in the tRNA^{Val} (UAC) gene of plant chloroplasts, is excised indeed as a linear molecule, rather than a lariat (Vogel and Börner, 2002). Such cases used to be regarded as oddities but recently, an evolutionary process that recurrently created intron lineages with additional nucleotides at the intron 5' extremity, and no apparent branchpoint, was shown to be at play in mitochondria (Li *et al*, 2011; one member of this subset was confirmed to be

unable to generate other than linear excised intron molecules in vitro). In these lineages, not only is the branchpoint adenosine missing, but the middle part of domain VI next to it, which normally consists, in the IIB1 intron subclass from which these lineages originated, of a 3-bp helix and a well-conserved 6-nt internal loop (Figure 1A), is highly variable, in contrast to the basal and distal sections of the same domain VI (Li *et al*, 2011). This suggests that not merely the branchpoint and its two flanking G:U base pairs (Chu *et al*, 1998; Figure 1), but the entire middle part of domain VI could be involved in branching, presumably by binding to one or several specific receptor sites. We sought to identify candidate sites for such receptors by taking advantage of the fact that their sequences and structures may no longer be constrained in molecules that have lost the ability to carry out branching.

Only 10 sequences of introns with a 5' terminal insert are currently known, but these sequences belong to four to five independent lineages (Li *et al*, 2011, and Figure 1C), which should ensure some measure of statistical significance in comparisons. In fact, when those ten sequences are aligned with 32 sequences of mitochondrial introns that belong to the same intron subclass, but lack a 5' terminal insert (and possess a potential branchpoint), and the sequence entropy in each subset is systematically compared site by site (Figure 1B and Materials and Methods), a small number of intron positions at which the difference in sequence entropy (ΔE) lies well beyond the main distribution stand out from the rest. In simple terms, these sites are very well conserved as long as the branchpoint is present, but very poorly so otherwise.

Among the 20 sites with the highest ΔE scores, two were discarded because their nucleotide composition was too variable (entropy above 0.3) in the no-5'-insert subset. Out of the remaining 18 sites (Figure 1), 12 are concentrated in the middle part of domain VI, which, as already emphasized, is quite variable in the 5'-insert subset; one corresponds to the first intron nucleotide, that no longer forms a 2'-5' bond in that subset; another one is at position

2389 (generally an A), which, in the crystal-derived atomic-resolution model of the

Oceanobacillus ribozyme (Toor *et al*, 2008a), lies next to the 5' splice site and binds two

metal ions that have been proposed to be critical for catalysis; and yet another one, at position

104, is also known to be part of the catalytic core. Remarkably, however, the remaining three

sites (positions 78, 79, 100) belong to two consecutive G:U pairs in the IC1 distal helix, a

component whose terminal loop (θ) is known to play an important structural role by

contacting domain II, but which lies rather far away from the reaction center and had not yet

been proposed to be implicated in catalysis.

The distribution of bases at positions 79 and 100 is especially striking. These

nucleotides form a G:U pair in all but one of the 32 intron sequences with a recognizable

branchpoint, whereas nine out of the 10 sequences with a 5'-terminal insert have a Watson-

Crick pair instead and one has an A:A mismatch (Figure 1C). Such a nearly perfect

correlation suggests that presence of a G:U pair at positions 79:100 is particularly important

for the initiation of splicing by branching, whereas in the absence of a functional branchpoint,

the type of base pairing at that site affects only the overall stability and precise geometry of

the IC1 stem.

Nucleotide substitutions in domain VI and its IC1 candidate receptor site

In constructs that lack domain VI or have an altered branchpoint, hydrolysis at the 5' splice

site substitutes for branching and the intron is excised in linear, rather than lariat form. By

contrast, introns with fully functional, well-folded ribozymes are expected to initiate splicing

almost exclusively by transesterification. As seen in Table I, that is the case for the *Pylaiella*

L1787 intron (Pl.LSU/2; Costa *et al*, 1997b), which we have been using as a model subgroup

IIB1 molecule: about 90 percent of excised intron products are lariats when the in vitro self-

splicing reaction takes place in the presence of ammonium and magnesium counterions.

Assuming, as suggested by sequence analyses, that both the middle section of domain VI and the 79:100 and 78:101 G:U pairs are specifically involved in the branching reaction, nucleotide substitutions at these sites should shift the balance between transesterification and hydrolysis towards the latter process. However, replacement of the internal loop of domain VI by canonical base pairs and trimming of the resulting, extended helix down to four base pairs (Figure 2) have barely detectable effects on the fraction of products branched when reactions are carried out in the presence of 1M ammonium chloride (Table I). The observed rate constant for branching ($k_{branching}$) does decrease (by less than 3-fold), but so does that for hydrolysis, so that their ratio is barely affected. Only by bringing the length of the helix distal to the branchpoint down to two base pairs (mutant dVI-2bp) do consequences suddenly become dramatic, with splicing proceeding almost exclusively by hydrolysis (Table I, line 4).

One possibility was that under optimal in vitro self-splicing conditions, processes other than the positioning of domain VI are rate-limiting for transesterification at the 5' splice site of precursor molecules. Among monovalent counterions, potassium has long been known to favor hydrolysis relative to branch formation (Jarrell *et al*, 1988). Compared indeed to the situation in ammonium, the observed rate constant for hydrolysis is increased by almost 3-fold for the wild-type Pl.LSU/2 ribozyme, and there is also a significant decrease in the fraction of lariats among intron excision products (Table I). Interestingly, all of the mutant domain VI (dVI) constructs in Figure 2 are further affected in their ability to react when assayed in the presence of potassium. Removal of the dVI internal loop significantly decreases the fraction of molecules that initiate splicing by branching compared to the wild-type, while truncation of the resulting helix to four base pairs not only reduces this fraction further, but specifically affects the observed rate constant for branching, by about 4-fold relative to the wild-type.

Since any nucleotide substitution in the IC1 stem might affect the relative position in three-dimensional space of the ε' and θ loops, we deemed it preferable to try and delete the latter component before assessing our IC1 constructs in potassium. Fortunately, removal of θ turned out to be without severe consequences on the ability of precursor molecules to react under the experimental conditions we had chosen; in fact, transesterification is even less affected than hydrolysis, resulting in an elevated $k_{branching}/k_{hydrolysis}$ ratio (Table I). By contrast, when the two G:U pairs at positions 79:100 and 78:101, which 16 out of 32 mitochondrial IIB1 introns with a recognizable branchpoint share, are simultaneously substituted by A:U pairs, the observed rate constant of branching and the fraction of intron molecules excised in lariat form are both markedly affected and this, whether in a wild-type or Δθ context (Table I; in ammonium, $k_{branching}$ is specifically affected as well, but the fraction of molecules that react by branching is left unchanged; see also Figures S1 and S2). Interestingly also, trimming of the IC1 helix down to only two base pairs is without further effects on kinetic parameters. Thus, these experiments are consistent with the conclusions of comparative sequence analyses, which pointed to the tandem G:U pairs in IC1 as major potential contributors to the ability to perform branching.

Modeling of the interaction between domain VI and its proposed IC1 receptor

The G79:U100 pair is highly conserved in a majority of group II intron subclasses (Dai *et al*, 2003), including the somewhat divergent subgroup IIC, to which the *Oceanobacillus* intron belongs. We have explored the possibility that this pair constitutes part of the first-step receptor site for domain VI by attempting to model the missing domain VI (Figure 3A) into the latest atomic-resolution models (Toor *et al*, 2010; Wang, 2010) of the *Oceanobacillus* group II ribozyme.

Currently available structures of the *Oceanobacillus* ribozyme reflect the final stage of splicing, after exon ligation. They lack domain VI and the last three intron residues as well as the first intron nucleotide (G1). The latter must move away from the catalytic center after the first step of splicing in order to make way for the 3' splice site and the segment at the intron 5' extremity that gets relocated may include also U2 (although not G3, for the ε-ε' interaction – Jacquier and Michel, 1990 – is believed to persist throughout splicing). As first pointed out by Steitz and Steitz (1993; see also Michel and Ferat, 1995; Jacquier, 1996), the best way to reconcile data on the inhibition of individual splicing steps by phosphorothioate stereoisomers of the reactive phosphate group with the generally accepted existence of a single catalytic site is to postulate that the O3'-P-O5' dihedral angle at the 5' splice site undergoes a 120° rotation away from the helical geometry that prevails at the splice junction of the intron-bound ligated exons (Toor *et al*, 2008b; Costa *et al*, 2000). In the predicted structure of the *Oceanobacillus* precursor RNA, such a sharp bend is required anyway in order to ensure connectivity within the segment that extends between the last nucleotide of the 5' exon, which is expected to remain bound to EBS1 throughout the splicing process, and G3 (see Wang, 2010). Modeling of the phosphodiester bond at the 5' splice site then makes it possible to position precisely the attacking 2'OH group of the branchpoint adenosine, which sets in turn the stage for placing the basal and distal helices of domain VI.

We found that in order for the 5' strand of the basal dVI helix to bridge the distance between the branchpoint and domain V, the first two base pairings at the base of the latter in Figure 1 of Toor *et al* (2008a) need to be disrupted: these pairings, the existence of which is not supported by comparative sequence analysis (note their absence in Figure 1A), may owe their presence in the *Oceanobacillus* ribozyme structure to the absence of domain VI. As for the section of domain VI that lies distal to the branchpoint, we chose to model it as a continuous helix despite the presence of a very well conserved internal loop (Figure 1A and

Li *et al*, 2011) in mitochondrial subgroup IIB1 introns. The reasons for this are (i) most bacterial members of this subclass lack an internal loop in their distal dVI stem, even though they share tandem IC1 G:U pairs with their mitochondrial counterparts; (ii) substitution of canonical base pairing for the internal loops of introns Sc.cox1/5γ (Chu *et al*, 1998) and Pl.LSU/2 (Figure 2 and Table I) has limited effects on their ability to carry out branching.

As shown in Figure 3A, it is possible indeed to position a continuous dVI distal helix in such a way that its base is connected to, and stacked on, the proximal section of the domain (consisting of the basal dVI helix and branchpoint adenosine), while its 5' backbone fits neatly into the shallow ('minor') groove of the IC1 stem. This model is consistent with our comparative sequence analysis and nucleotide substitution experiments, since the section of IC1 that is specifically contacted by domain VI encompasses the G79:U100 base pair (G81:U101 in the *Oceanobacillus* intron). For the sake of consistency with η-η', we propose to name ι-ι' (iota-iota') this novel interaction between the IC1 shallow groove at, and immediately distal to, positions 79 and 100 (ι) and the middle part of the dVI distal stem (ι').

Activation of lariat formation by oligonucleotides that anchor domain VI to its binding site

As apparent from Figure 3A, optimal positioning of the dVI distal helix into the shallow groove of helix IC1 results in placing IC1 nucleotides A83 to A87 (Pl.LSU/2 numbering) in near continuity of A2413 in the 5' strand of domain VI. This peculiar arrangement suggested to us that it might be possible to replace part of the 5' strands of the dVI and IC1 helices by an oligonucleotide that would at the same time restore the dVI helical structure and anchor it to its proposed receptor. The complete setup, consisting of such an 'anchoring' DNA oligonucleotide with segments ('handles') that are complementary to the terminal loops of the

truncated dVI and IC1 stems and are connected with one another by a tether made out of deoxythymidines, is shown in Figure 3B.

As expected from the data in Table I, the construct in Figure 3B, in which the dVI distal helix has been truncated down to two base pairs, with a 7-nt terminal loop, has only residual branching activity (Figure 4A, intercept with the y axis). However, the same precursor transcript, when incubated in the presence of increasing concentrations of an oligonucleotide capable of restoring base pairing in both the dVI and IC1 stems (Figure 3B), gradually recovered the ability to initiate splicing by transesterification, with up to *ca* 58 percent of reaction products consisting of the lariat intron at 200 μM oligonucleotide (not shown). A plot of the fractional rate of branching (observed rate of branching relative to total rate of conversion of precursor into products) as a function of the concentration of oligonucleotide can be fitted indeed to a saturation curve (see Materials and Methods) with an estimated K_m equal to 58 ± 20 μM (Figure 4A).

Subsequent experiments showed that this K_m could be decreased by playing with both the geometry of the IC1 terminal loop and its sequence. Among the combinations we tried, the one shown in Figure 3C turned out to be optimal, with a K_m of 5.4 ± 1.0 μM (Figure 4A; a G which had been introduced at position 82 so as to leave unspecified the junction between the IC1 and anchoring helices proved suboptimal). As a control, reactions in the presence of increasing concentrations of a 7-mer, no-anchor oligonucleotide that merely restored the dVI helix resulted in only minimal recovery of branching activity (Figure 4A). Additional controls (Table II) performed in the presence of oligonucleotide concentrations (100 μM) well above the observed K_m for the combination in Figure 3C demonstrate that: (i) whether the structure of IC1 is wild-type (setup 1), truncated (setup 3) or (presumably) restored by a complementary 7–mer oligonucleotide (setup 5), only residual branching activity is observed as long as the terminal loop of the truncated dVI stem is left unpaired; (ii) restoration of the

dVI stem by a complementary 7-mer, whether in a wild-type (setup 2) or IC1 mutant context (setup 4; Figure 4A) only slightly improves branching activity; (iii) simultaneous restoration of base pairing in both the dVI and IC1 stems by two 7-mers (setup 6) is not sufficient: branching activity remains very modest unless anchoring is achieved by creating a covalent link between these oligonucleotides (setup 7).

The next step in optimizing this system consisted in keeping the sequence of the anchor in Figure 3C constant and varying the length of the tether from zero to four T's (Figure 4B) at an oligonucleotide concentration (5 µM) about equal to the K_m determined for a 3-T tether (Figure 4A, full curve). A sharp optimum was observed for a tether consisting of just one T, with a relative rate of branching equal to 0.790 ± 0.011. The latter value should be close to saturation, as was verified indeed by determining the corresponding K_m (0.073 ± 0.009 µM; Figure 4C).

Final proof that complementarity between the IC1 terminal loop and an anchoring oligonucleotide is both necessary and sufficient to activate branching was obtained by nucleotide substitutions (Figures 3D and 4C): whereas mismatched combinations devoid of potential for base pairing exhibit no detectable branching activity, restoration of complementarity by substitution of both the oligonucleotide anchor and the IC1 terminal loop was found to result in almost complete recovery of the ability to initiate splicing by branching (relative rate of branching at saturation, 0.744 ± 0.022; K_m equal to 0.270 ± 0.047 µM). Finally, it should be noted that for the setup of Figure 3C, we verified the oligonucleotide-induced branching reaction to be an authentic one, in the sense that the same branchpoint is used as in a wild-type molecule and the resulting ligated exons have the same sequence (see Materials and Methods and also the analytical gel in Figure S3).

Discussion

A first-step-specific receptor for the branchpoint-carrying domain VI

We have shown that by using oligonucleotides that bring together domain VI and what we propose to be a first-step RNA receptor for this domain, in subdomain IC1 (Figure 5), it is possible to specifically activate the branching reaction in a defective precursor molecule that is otherwise essentially unable to initiate self-splicing, except by 5' splice site hydrolysis. The dVI and IC1 helices must truly come in contact in the active first-step complex, for we found the optimal connecting segment between the dVI and IC1 handles of the anchoring oligonucleotide to consist of just one thymidine residue (Figure 4B). The use of longer tethers leads to a gradual decrease in the efficiency of branching, as would be predicted by a random-coil model (Jacobson and Stockmayer, 1950), whereas, conversely, when the single connecting nucleotide is removed, restoration of branching is much less efficient, presumably because the anchoring oligonucleotide and its targets must give up one or several base pairs in order to release the resulting strain.

While compatible with all available data, our modeling of the interaction between domain VI and the IC1 distal helix was dictated by our identification of the G79:U100 base pair as a likely receptor for domain VI. Current ignorance of the exact configuration of the branchpoint adenosine, which has alternatively been proposed to be extrahelical (Schlatterer *et al*, 2006), to be stacked between two base pairs (Erat *et al*, 2007) or to be part of a two-nucleotide bulge (Zhang and Doudna, 2002), is such that in fitting the middle part of domain VI optimally into the shallow groove of IC1, we opted to care primarily about the need to retain connectivity to the dVI proximal helix: the two dVI helices are actually stacked on top of one another in Figure 3 and in connecting the branchpoint ribose to its immediate neighbors, we chose to bulge it out from the helical stem, without taking stands on its exact geometry.

In this context, our finding that the optimal dVI-IC1 tether consists of only one

nucleotide is important and clearly pleads in favor of our own working model of the ribozyme

first-step configuration (Figure 3A), when compared with another recently proposed

arrangement of domain VI (Wang, 2010), which attempted to meet previous claims that the

coordination loop serves as receptor for the branchpoint (Hamill and Pyle, 2006). In the latter

model (Figure 10 of Wang, 2010), which includes a hypothetical 'mispair' between the

universal branchpoint adenine and A393 (*Oceanobacillus* ribozyme numbering), a nucleotide

that is poorly conserved by evolution, domain VI is oriented right towards the coordination

loop, away from IC1. In yet another recently published sketch of a possible first-step

conformation (Figure 13 of Pyle, 2010), the location of domain VI, which is represented only

as a cylinder, is somewhat intermediate between ours and Wang's since it is placed in

between IC1 and the coordination loop, though in a position that would still not allow it to

contact our proposed IC1 receptor. It is also important to note that even though they clearly

differ, Wang's, Pyle's and our own modeling of the ribozyme first-step conformation all

imply a major rotation of domain VI after the branching step in order for its tip to dock into its

domain II, second-step receptor (inasmuch as the position of the latter can be modeled

precisely, see Figure 3A and its legend).

When interpreting crosslinks between the dVI branchsite and the coordination loop as

evidence that the latter constitutes the binding site for the former, Hamill and Pyle (2006)

implicitly assumed that domain VI should be stably docked in its first-step receptor prior to

the branching reaction. However, it seems more likely that domain VI keeps toggling between

different states, as initially proposed by Costa et al. (1997a), based on kinetic analysis of

mutant ribozymes, and, more recently, by Toor et al. (2010) to account for the absence of that

domain in the crystal structure of the *Oceanobacillus* intron. In fact, when Hamill and Pyle's

sites of crosslinking are mapped on the atomic resolution model of ribozyme domains I to V

(Toor et al., 2008), it becomes apparent that essentially every nucleotide that would have been accessible to the branch site and its two flanking nucleotides in a dVI molecule that could freely rotate around the dV-dVI junction did give rise to a crosslink (only residues in the proximal helix of domain V are missing, since crosslinks at those sites are internal to the D56 piece and, therefore, could not be recovered in the experiment). When domain VI is bound to its IC1 receptor, photoactivable bases at the branchpoint and its two neighbors are predicted from our model to crosslink neither to the coordination loop, nor to stem IC1, but to the first two nucleotides of the intron and the last nucleotide of the 5' exon: these three positions were indeed among those recovered by Hamill and Pyle. Moreover, among the latter crosslinks, those to the G1 nucleotide (and perhaps also to the second residue of the intron) are liable to be compatible with splicing, which provides a ready explanation for the reactivity of part of the XL1 material of Hamill and Pyle (2006).

Interestingly, some published pieces of data in the literature already hinted at the possible involvement of the IC1 distal helix in the branching process. Stabell *et al* (2009) noted that in a paraphyletic subset of group II introns that share additional secondary structures 3' of domain VI, the section of the IC1 stem that lies immediately distal to the ε' loop is unexpectedly conserved. Several nucleotide substitutions were introduced, among which was the replacement of the (counterpart of the) 79:100 G:U pair by A:U. That mutation was found to markedly decrease the rate of reaction of precursor molecules, but in the absence of 5' splice site hydrolysis, branching could not be shown to be specifically affected.

Much earlier, Boudvillain and Pyle (1998) had published a map of domains I to III of the subgroup IIB1 Sc.a5γ ribozyme (a close relative of Pl.LSU/2) that showed, based on NAIM (Nucleotide Analog Interference Mapping; see Strobel, 1999), which nucleotides were important for a branching reaction with domains V and VI (unfortunately, the authors' setup did not make it possible to discriminate between nucleotides required specifically for

branching and those involved in catalysis in general or in binding of domain V by domains I-III). Removal of the NH_2 at position 2 of G79 (Pl.LSU/2 numbering) and also of the 2'OH groups of U78 and U100 was reported to interfere with activity, thus pointing to the importance of the shallow groove in this section of the IC1 distal helix; remarkably, these three residues are none other than the ones that generate a statistical signal when molecules with and without a recognizable branchpoint are compared (Figure 1; it is also worth noting that no hit was found in the coordination loop proper, whether by NAIM or our comparative sequence analyses, despite its claimed function as a receptor for domain VI – Hamill and Pyle, 2006). In fact, our phylogenetic approach may rightly be regarded as related to NAIS (Nucleotide Analog Interference Suppression, also called 'chemogenetics'; Strobel, 1999), a method in which nucleotide interference maps (rather than sequence conservation maps) are compared for the wild-type and a molecule that includes a specific defect.

Towards atomic resolution

It is now generally agreed that group II ribozymes exist in at least two major states (Figure 5), one in which domain VI is prepositioned for the branching reaction and another one in which it interacts with domain II (whether the latter interaction helps positioning the 3' splice site for exon ligation is still a matter of debate – see Pyle, 2010 – despite the fact that disruption of η-η' was found to impair specifically the second step of splicing – Chanfreau and Jacquier, 1996). The identification of a second-step-specific receptor for domain VI (Chanfreau and Jacquier (1996) was a breakthrough, if only since it made it possible, by playing with the strength of the interaction between diverse loops of the GNRA family and their RNA receptors (Costa and Michel, 1997), to place introns into a well-defined configuration that could be probed by biochemical and biophysical methods. Our use of anchoring oligonucleotides that force domain VI and its IC1 first-step receptor to interact should

similarly open the way to trapping the ribozyme into its branching-ready configuration, something which could presumably be achieved by replacing our current DNA 'handles' by higher-affinity, RNA or perhaps LNA (Locked Nucleic Acid; Petersen *et al*, 2002) counterparts (the affinity of even our best anchoring oligonucleotides for their targets – see Legend to Figure 4 and Materials and Methods for estimated K_d values – is still too low to prevent 'breathing' of helices, which also explains why we did not observe accumulation of the lariat-3'exon reaction intermediate – not shown). This approach might even make it possible to obtain crystals and visualize at last the ribozyme branchpoint and its molecular context at atomic resolution.

One possible objection to the use of anchoring oligonucleotides for biochemical and biophysical probing is that despite the fact that the authentic branchpoint is being used (Materials and Methods) the resulting arrangement in space of domain VI and subdomain IC1 might be an unnatural one. However, because the segment of IC1 that was engineered to interact with the oligonucleotide anchor is located distal to the section that we believe to constitute the natural receptor for domain VI (Figures 3 and 5), that receptor is likely to remain structurally intact in the complex (our initial choice of a 3-nucleotide tether reflected our concern that shorter connecting segments might distort proximally located contacts). It may prove possible also to reconstruct an authentic middle dVI section by replacing our current DNA handle by an RNA counterpart with the appropriate sequence to generate the characteristic internal loop of mitochondrial subgroup IIB1 introns (Figure 1A). This would open the way to the substitution of individual chemical groups in the 5' strand of that loop, which we propose to be the site of contact with the IC1 receptor (in this respect, it is interesting to note that besides the branchpoint adenosine, the only other sites in domain VI to give rise to interference signals in the NAIM experiments of Boudvillain and Pyle (1998) were positions 2411-2413 (Pl.LSU/2 numbering), which are precisely the ones that should

contact the IC1 shallow groove according to the model in Figure 3A). Up to now, the introduction of atomic substitutions and, therefore, the use of NAIS to explore interactions in this section of the ribozyme was made difficult (though not impossible) by the fact that domain VI cannot be supplied alone in a two-piece intron system, but needs to be covalently connected to domain V in order to be bound by the rest of the ribozyme (Jarrell *et al*, 1988).

Conclusion

Now that a tertiary contact between the branchpoint-carrying component of group II introns and the rest of the group II ribozyme has been found and shown to be essential for the efficiency of lariat formation, the stage is set at last to explore the atomic surroundings of the branchpoint itself. In the meantime, pending a high-resolution structure of an entire intron, our newly acquired ability to control at will the conformation of the ribozyme through the use of oligonucleotides should prove particularly useful for detailed mechanistic investigations of individual steps in the splicing and transposition processes carried out by the sophisticated molecular machinery that we call a group II intron. Finally, it did not escape our notice that in tinkering with the architecture of the group II ribozyme, we may have been preceded by nature: U2-U6 helix III (Sun and Manley, 1995) which, in the spliceosome, links together the branchpoint helix and the segment of U6 that, like ε', binds the first intron nucleotides, may be regarded as a counterpart of our dVI-anchoring oligonucleotides.

Materials and Methods

Sequence analyses

The set of 42 subgroup IIB1 mitochondrial intron sequences collected and aligned by Li *et al* (2011; the aligned set is accessible at http://rnajournal.cshlp.org/content/suppl/2011/05/05/rna.2655911.DC1.html) was divided into a subset of 32 intron sequences in which the 5' splice site is followed by the GUGCG consensus at the intron 5' end and a subset of 10 intron sequences with a 5' terminal insert. Entropy (as defined in BioEdit – Hall, 1999: $H(l) = -\Sigma f(b,l)ln(f(b,l))$, where $f(b,l)$ is the frequency of base b at position l) was calculated for each subset at each of the 577 positions of the alignment and values for the no-insert subset were subtracted from those for the insert-carrying subset in order to generate a 'Δ Entropy' measure, the distribution of which is plotted in Figure 1C. In the phylogenetic tree of Fig. 1A, host genes were abbreviated as follows: L and S designate the large and small subunit rRNA genes, respectively, and the following number corresponds to the site of insertion, according to *E. coli* numbering – see Johansen and Haugen, 2001; cob: cytochrome b; cox1, 2, 3: subunits 1, 2, 3 of cytochrome c oxidase.

Modeling

Modelling and refinement were carried out with Rastop 2.2 and the Assemble 1.0 software (Jossinet *et al*, 2010).

DNA constructs and precursor transcripts

Wild-type precursor transcripts were generated from plasmid pPl.LSU2 (Costa *et al*, 1997b), a pBluescript II KS (-) (Stratagene) derivative. All mutant constructs in Figures 2 and 4 were

verified by sequencing the entire length of the insert. Transcription and RNA purification were carried out as in Costa *et al* (1997b).

Kinetic analyses

Monomolecular reactions of the wild-type and mutant constructs listed in Table I were initiated by addition of 2X-concentrated splicing buffer (final concentrations: 40 mM Tris-HCl pH 7.5 at 25°C, 1M NH_4Cl or KCl, 10 mM $MgCl_2$, 0.02% sodium dodecyl sulfate) to an equal volume of a water solution of ^{32}P-labelled precursor RNA molecules (final molar concentration 20 to 40 nM) which was preequilibrated at the reaction temperature (45°C) after having been denatured for 2 min at 90°C. Reactions were stopped by addition of an equal volume of formamide loading buffer containing Na_2EDTA (final concentration 20 mM; each time point – from 0.5 to 180 min – was generated from a separate initial mix). Samples were run on denaturing polyacrylamide gels (50% urea w:v, 4% total acrylamide, with 1:20 bis-acrylamide), and bands associated with the precursor and reaction products were quantitated with a PhosphorImager (Molecular Dynamics).

Accumulation of branched and linear intron products was fitted (with Kaleidagraph 3.6) to simple exponentials,

$[Lar] = [Lar]_\infty (1 - exp(-k_{br}.t))$ and

$[Lin] = [Lin]_\infty (1 - exp(-k_{hy}.t))$

where [Lar] and [Lin] are the molar fractions of branched and linear molecules at time t, $[Lar]_\infty$ and $[Lin]_\infty$, the corresponding, estimated final values, and k_{br} and k_{hy}, the observed rate constants for branching and hydrolysis. As already noted by others (e.g. Chu *et al*, 1988), values obtained for k_{br} and k_{hy} typically differ (Table I), which means that refolded precursor molecules do not form a single population, but rather exist in multiple conformations that do not readily interconvert during the time course of experiments. In ammonium buffer, about 90

per cent of molecules remain committed to forming lariats, even in mutants with a 10-fold reduced rate constant for branching (the only exception is the dVI-2bp mutant). In potassium buffer, however, changes in rate constants for branching and hydrolysis tend to be reflected in correspondingly altered proportions of branched and linear molecules among reaction products (bottom part of Table I). Importantly, measurements were found to be highly reproducible, whether for the wild-type (Table I) or mutant constructs.

For reactions in the presence of an oligonucleotide (Sigma-Aldrich), the latter was added to concentrated splicing buffer (final concentrations: 40 mM Tris-HCl pH 7.5 at 25°C, 1M NH_4Cl, 10 mM $MgCl_2$, 0.02% sodium dodecyl sulfate) prior to mixing with the solution of purified precursor molecules (final molar concentration 20 nM) at reaction temperature (37°C). Reaction time courses were modeled according to the following scheme, in which pre:oligo is the unreacted complex between a precursor and an oligonucleotide molecule (whereas hydrolysis at the 5' splice site is irreversible, transesterification is expected to be reversible; however, the intron-3'exon lariat intermediate was either absent or barely detectable, even at short reaction times, for all construct and oligonucleotide combinations we tested, so that in this experimental system, branching may be regarded as irreversible for all practical purposes).

$$
\begin{array}{ccc}
& \xrightarrow{k_{hy,B}} & \text{linear intron} \\
\text{pre:oligo} & & \\
& \xrightarrow{k_{br,B}} & \text{lariat intron} \\
k_{on} \uparrow \downarrow k_{off} & & \\
& \xrightarrow{k_{hy,U}} & \text{linear intron} \\
\text{precursor} & & \\
& \xrightarrow{k_{br,U}} & \text{lariat intron}
\end{array}
$$

Provided k_{off} and k_{on} are much larger than the rate constants for reactions, the rates of formation of lariat and linear intron products become:

$$d[Lar]/dt = [Pre] (k_{br,U} + k_{br,B}.[OLI]/K_d) \quad (1)$$

$$d[Lin]/dt = [Pre] (k_{hy,U} + k_{hy,B}.[Oli]/K_d) \quad (2)$$

where [Pre] is the molar fraction of unbound precursor molecules at time t; $k_{br,U}$, $k_{hy,U}$, $k_{br,B}$ and $k_{hy,B}$ are rate constants for branching (br) and hydrolysis (hy) in the absence (Unbound) and presence of a bound (B) oligonucleotide, respectively; $K_d = k_{off}/k_{on}$; and [OLI] is the molar concentration of oligonucleotide. Let f be the fractional (relative) rate of formation of lariat intron (f_0 and f_{max} are initial and final values of f):

$$f = (d[Lar]/dt)/(d[Lar]/dt + d[Lin]/dt)$$

$$= f_0 + (f_{max} - f_0) / (1 + K_m/[OLI]) \quad (3) \quad \text{with}$$

$$f_0 = k_{br,U}/(k_{br,U} + k_{hy,U}) \quad (4)$$

$$f_{max} = k_{br,B}/(k_{br,B} + k_{hy,B}) \quad (5)$$

$$K_m = K_d (k_{hy,U}/ k_{br,B}) (f_{max}/(1 - f_0)) \quad (6)$$

In practice, (i) the accumulation of lariat and linear intron forms for a given oligonucleotide concentration was fitted to a simple exponential or, exceptionally, when reaction was both slow and limited, to a linear function; (ii) initial rates at $t = 0$ and their standard errors were obtained from these fits, f was calculated and plotted as a function of oligonucleotide concentration (the relative error of f was estimated by adding the relative errors of branching and total reaction rates, which were calculated from standard errors associated with initial rates); (iii) the resulting plot was fitted with equation (3) to determine f_0, f_{max} and K_m; (iv) K_d

was extracted from equation (6) after $k_{hy,U}$ and $k_{br,B}$ had been obtained from initial reaction rates in the absence and at saturating concentrations of the oligonucleotide, respectively.

Verification of splice junctions and the branchpoint

The identity of splice junctions and the branchpoint were verified for the construct-oligonucleotide combination shown in Figure 3C by purifying the ligated exons and intron lariat from a denaturing polyacrylamide gel prior to reverse transcription, as described in Costa *et al* (1997b; an analytical version of that gel is shown in Figure S3). After reverse transcription of the ligated exons with primer 5'-GAGGTCGACGGTATCGATAA (which matches positions 70-89 of the 3' exon), PCR amplification was carried out with the same primer and 5'-AGCTTTTATCTTTGACACAAAATCGGGGGTG (positions -19 to – 49 of the 5' exon) and products cloned with the pGEM-T vector system (Promega): all clones examined had the expected sequence for the ligated exons. After reverse transcription with primer 5'-GCAGGTACATTGTCTCCAGA (complementary to intron positions 58-77) and PCR amplification with the same primer and 5'-GAAAGGCTGCAGACTTATTA (corresponding to part of ribozyme domain III), five clones were sequenced and found to contain the intron sequence preceding the branchpoint followed by the beginning of the intron, as expected. However, in three clones, an A rather than a T had been incorporated by the reverse transcriptase at the position facing the adenine of the branchpoint, one clone lacked both the branchpoint A and the preceding T and the fifth one lacked that T: these are typical of the errors made by the Superscript II reverse transcriptase when trying to bypass a 2'-5' branched structure (Vogel and Börner, 2002).

Supplementary information is available at The EMBO Journal Online

Acknowledgements

Author Contribution

C.-F.Li and M.C. designed, carried out and analyzed experiments. F.M. carried out comparative sequence analysis and structural modeling, conceived some experiments and contributed to their interpretation. All authors contributed to drafting the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Boudvillain M, Pyle AM (1998) Defining functional groups, core structural features and inter-
domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis.
EMBO J **17:** 7091-7104

Chanfreau G, Jacquier A (1994) Catalytic site components common to both splicing steps of a
group II intron. *Science* **266:** 1383-1387

Chanfreau G, Jacquier A (1996) An RNA conformational change between the two chemical
steps of group II self-splicing. *EMBO J* **15:** 3466-3476

Chu VT, Liu Q, Podar M, Perlman PS, Pyle AM (1998) More than one way to splice an RNA:
branching without a bulge and splicing without branching in group II introns. *RNA* **4:**
1186-1202

Costa M, Deme E, Jacquier A, Michel F (1997a) Multiple tertiary interactions involving
domain II of group II self-splicing introns. *J Mol Biol* **267:** 520-536

Costa M, Fontaine JM, Loiseaux-de Goër S, Michel F (1997b) A group II self-splicing intron
from the brown alga Pylaiella littoralis is active at unusually low magnesium
concentrations and forms populations of molecules with a uniform conformation. *J Mol
Biol* **274:** 353-364

Costa M, Michel F (1997) Rules for RNA recognition of GNRA tetraloops deduced by in
vitro selection**:** comparison with in vivo evolution. *EMBO J* **16:** 3289-3302

Costa M, Michel F, Westhof E (2000) A three-dimensional perspective on exon binding by a
group II self-splicing intron. *EMBO J* **19:** 5007-5018

Dai L, Toor N, Olson R, Keeping A, Zimmerly S (2003) Database for mobile group II introns.
Nucleic Acids Res **31:** 424-426

Erat MC, Zerbe O, Fox T, Sigel RK (2007) Solution Structure of Domain 6 from a Self-
Splicing Group II Intron Ribozyme: A Mg(2+) Binding Site is Located Close to the
Stacked Branch Adenosine. *Chembiochem* **8:** 306-314

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
program for Windows 95/98/NT. *Nucl Acids Symp Ser* **41:**95-98

Hamill S, Pyle AM (2006) The Receptor for Branch-Site Docking within a Group II Intron
Active Site. *Mol Cell* **23:** 831-840

Jacobson H, Stockmayer WH (1950) Intramolecular reaction in polycondensations. i. the
theory of linear systems. *J Chem Phys* **18:** 1600-1606

Jacquier A (1996) Group II introns: elaborate ribozymes. *Biochimie* **78:** 474-487

Jacquier A, Jacquesson-Breuleux N (1991) Splice site selection and role of the lariat in a
group II intron. *J Mol Biol* **219:** 415-428

Jacquier A, Michel F (1990) Base-pairing interactions involving the 5' and 3'-terminal
nucleotides of group II self-splicing introns. *J Mol Biol* **213:** 437-447

Jarrell KA, Peebles CL, Dietrich RC, Romiti SL, Perlman PS (1988) Group II intron self-
splicing. Alternative reaction conditions yield novel products. *J Biol Chem* **263:** 3432-
3439

Johansen S, Haugen P (2001) A new nomenclature of group I introns in ribosomal DNA. *RNA*
7: 935-936

Jossinet F, Ludwig TE, Westhof E (2010) Assemble: an interactive graphical tool to analyze
and build RNA architectures at the 2D and 3D levels. *Bioinformatics* **26:** 2057-2059

Lambowitz AM, Zimmerly S (2004) Mobile group II introns. *Annu Rev Genet* **38:** 1-35

Li C-F, Costa M, Bassi G, Lai Y-K, Michel F (2011) Recurrent insertion of 5'-terminal
nucleotides and loss of the branchpoint motif in lineages of group II introns inserted in
mitochondrial preribosomal RNAs. *RNA*, in press.

Li-Pook-Than J, Bonen L (2006) Multiple physical forms of excised group II intron RNAs in wheat mitochondria. *Nucleic Acids Res* **34:** 2782-2790

Michel F, Costa M, Westhof E (2009) The ribozyme core of group II introns: a structure in want of partners. *Trends Biochem Sci* **34:** 189-199

Michel F, Ferat JL (1995) Structure and activities of group II introns. *Annu Rev Biochem* **64:** 435-461

Michel F, Umesono K, Ozeki H (1989) Comparative and functional anatomy of group II catalytic introns – a review. *Gene* **82:** 5-30

Petersen M, Bondensgaard K, , Jacobsen JP (2002) Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA**:**RNA hybrids. *J Am Chem Soc* **124:** 5974-5982

Pyle AM (2010) The tertiary structure of group II introns: implications for biological function and evolution. *Crit Rev Biochem Mol Biol* **45:** 215-232

Schlatterer JC, Crayton SH, Greenbaum NL (2006) Conformation of the Group II intron branch site in solution. *J Am Chem Soc* **128:** 3866-3867

Stabell FB, Tourasse NJ, Kolstø AB (2009) A conserved 3' extension in unusual group II introns is important for efficient second-step splicing. *Nucleic Acids Res* **37:** 3202-3214

Steitz TA, Steitz JA (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* **90:** 6498-6502

Strobel SA (1999) A chemogenetic approach to RNA function/structure analysis. *Curr Opin Struct Biol* **9:** 346-352

Sun JS, Manley JL (1995) A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev* **9:** 843-854.

Tang J, Breaker RR (1997) Rational design of allosteric ribozymes. *Chem Biol* **4:** 453-459

Toor N, Keating KS, Fedorova O, Rajashankar K, Wang J, Pyle AM (2010) Tertiary architecture of the *Oceanobacillus iheyensis* group II intron. *RNA* **16:** 57-69

Toor N, Keating KS, Taylor SD, Pyle AM (2008a) Crystal structure of a self-spliced group II intron. *Science* **320:** 77-82

Toor N, Rajashankar K, Keating KS, Pyle AM (2008b) Structural basis for exon recognition by a group II intron. *Nat Struct Mol Biol* **15:** 1221-1222

van der Veen R, Kwakman JH, Grivell LA (1987) Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J* **6:** 3827-3831

Vogel J, Börner T (2002) Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *EMBO J* **21:** 3794-3803

Wang J (2010) Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr D Biol Crystallogr* **66:** 988-1000

Zhang L, Doudna JA (2002) Structural Insights into Group II Intron Catalysis and Branch-Site Selection. *Science* **295:** 2084-2088

Table I Kinetic parameters of dVI and IC1 mutants.

Construct	Fraction of products branched	$k_{branching}$ (min^{-1})	$k_{hydrolysis}$ (min^{-1})	k_{br}/k_{hy}
		ammonium		
wt[1]	0.90 ± 0.07	0.136 ± 0.019	0.024 ± 0.010	5.5
	0.88 ± 0.11	0.166 ± 0.032	0.023 ± 0.008	7.2
dVI -7 bp	0.89 ± 0.04	0.092 ± 0.006	0.019 ± 0.002	5.0
dVI -4 bp	0.84 ± 0.06	0.058 ± 0.006	0.014 ± 0.002	4.2
dVI -2 bp	0.02[2]	<0.008 ± 0.002[3]	0.013 ± 0.002	<0.62
IC1 Δθ	n.d.	n.d.	n.d.	n.d.
IC1 UA:UA	0.89 ± 0.09	0.028 ± 0.003	0.024 ± 0.004	1.3
IC1 Δθ / UA:UA	n.d.	n.d.	n.d.	n.d.
IC1-2bp	0.90 ± 0.11	0.016 ± 0.004	0.024 ± 0.004	0.69
		potassium		
wt[1]	0.76 ± 0.08	0.160 ± 0.030	0.064 ± 0.023	2.5
	0.77 ± 0.06	0.149 ± 0.020	0.065 ± 0.009	2.3
dVI -7 bp	0.41 ± 0.04	0.132 ± 0.021	0.057 ± 0.012	2.3
dVI -4 bp	0.15 ± 0.01	0.045 ± 0.006	0.072 ± 0.008	0.63
dVI -2 bp	0	0	[0.135 ± 0.011][4]	0
IC1 Δθ	0.69 ± 0.05	0.097 ± 0.006	0.019 ± 0.004	5.1
IC1 UA:UA	0.10 ± 0.007	0.028 ± 0.002	0.042 ± 0.005	0.67
IC1 Δθ / UA:UA	0.067 ± 0.005	0.025 ± 0.003	0.029 ± 0.002	0.84
IC1-2bp	0.063 ± 0.025	0.026 ± 0.013	0.031 ± 0.011	0.85

n.d. : not determined

[1] determinations from different RNA preparations

[2] observed value at 180 min

[3] estimated from the fraction branched at 180 min

[4] determined at 50 mM Mg

Table II Rate of branching relative to total reaction rate in the presence of a 15-mer

anchoring oligonucleotide and 7-mer controls

Setup	IC1[1]	Oligonucleotide(s) (100 µM)	anti-dVI handle	anti-IC1 handle	relative rate of branching
1	wt	no			0.040 ± 0.009
2	wt	7-mer	**GTGGAC**T		0.126 ± 0.012
3	Fig.3C	no			0.040 ± 0.009
4	Fig.3C	7-mer	**GTGGAC**T		0.145 ± 0.021
5	Fig.3C	7-mer		T**GGCTGG**	0.068 ± 0.017
6	Fig.3C	7-mer + 7-mer	**GTGGAC**T	T**GGCTGG**	0.150 ± 0.037
7	Fig.3C	15-mer	**GTGGAC**T-T-T**GGCTGG**		0.530 ± 0.045

[1] Domain VI of all constructs was truncated as in Figure 3B.

Figure Legends

Figure 1 Identification of a candidate site for binding the branchpoint-carrying domain of a group II intron. (**A**) Schematic secondary structure of the Pl.L1787 (Pl.LSU/2) ribozyme, a representative mitochondrial member of subgroup IIB1. Only the sequences of domains V and VI and the distal part of subdomain IC1 are shown, the asterisk next to domain VI indicates the branchpoint. Greek letters and arrows correspond to prominent tertiary interactions, which are generally conserved in group II introns (see Michel *et al*, 2009). Sites in red and orange are those at which the difference in sequence entropy between the set of introns with and without a 5' terminal insert exceeds 1.0 or is included in the 0.70-1.0 range, respectively (see panel B). (**B**) Statistical distribution over aligned ribozyme sites of the difference in sequence entropy between sets of introns with and without a 5' terminal insert. Ordinates: number of sites; abscissa: difference in sequence entropy at homologous sites between the two intron sets, calculated as in Materials and Methods (numbers are positives when site entropy is larger for the set of introns with a 5' insert). The arrow points to the 0.70 differential entropy threshold (for sites highlighted in panel A; red and blue rectangles correspond to sites in domains VI and IC1, respectively). (**C**) Phylogenetic relationships of mitochondrial subgroup IIB1 introns based on an alignment of their ribozyme sequences (the tree is redrawn from Li *et al*, 2011). Introns and intron clades are designated by their host gene (Li *et al*, 2011). Thick red lines correspond to lineages of introns that possess a 5' terminal insert, the length of which is indicated at right (boxed numbers). When not G and U, the nucleotides at positions 79 and 100 (of the Pl.L1787 ribozyme) are indicated at the far right.

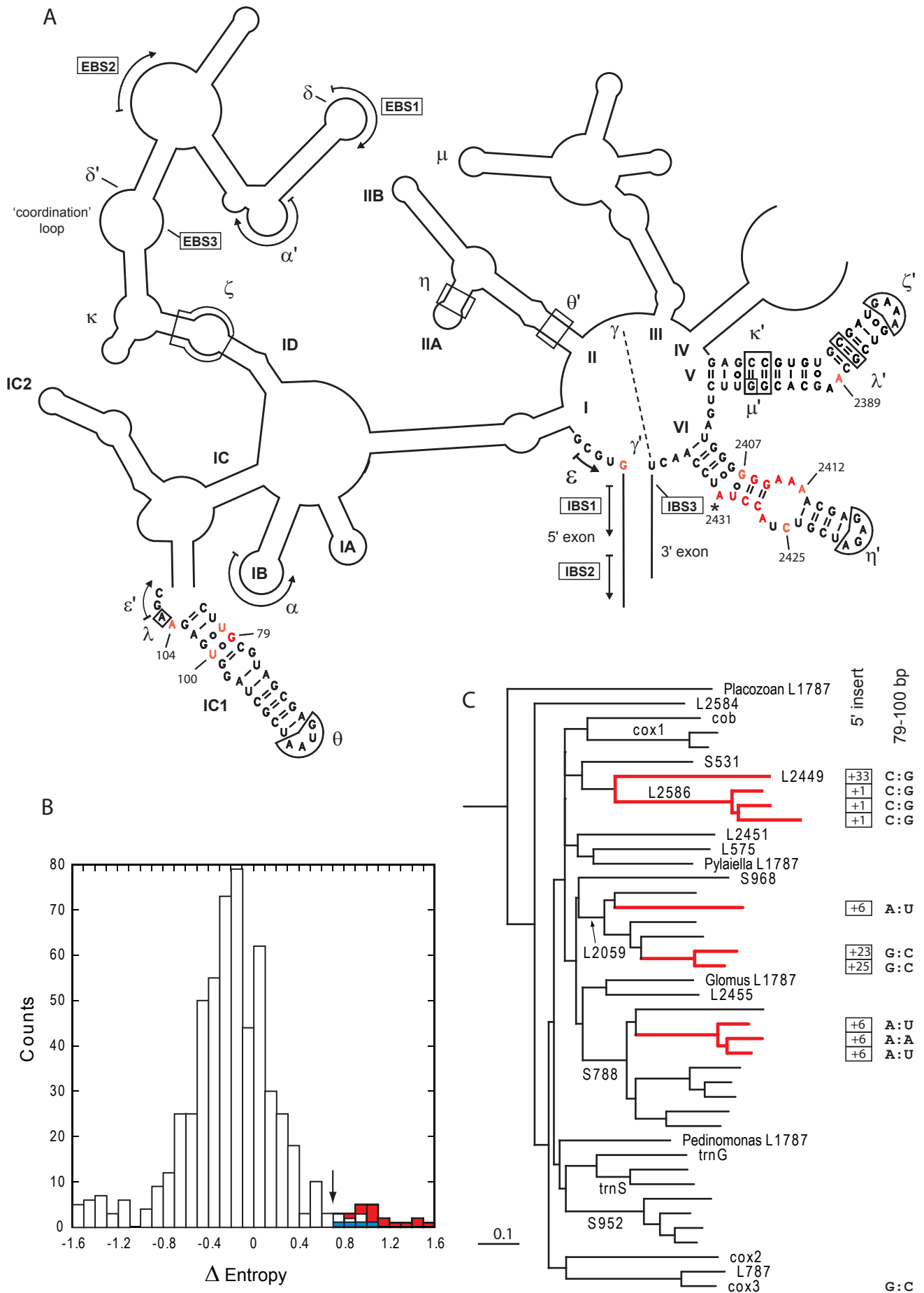**Figure 2** Ribozyme constructs with altered dVI and IC1 structures.

Figure 3 Three-dimensional model of the interaction between ribozyme domains VI and IC1 and optimization of oligonucleotides anchoring dVI to IC1. (**A**) Stereo views were generated from the coordinate set of Toor *et al* (2010) for the *Oceanobacillus iheyensis* subgroup IIC ribozyme (PDB accesion number 3IGI). Only domain VI, the 3-nt dV-dVI linker and intron residues 1-2 were modeled de novo; the last three nucleotides of the intron and the 3' exon are missing (see Results and Materials and Methods). Color scheme: black, branchpoint adenosine; green, domain VI; pink, domain V; violet, 5' exon; yellow, intron nt 1-5; tan, subdomain IC1; red, bp 79:100 (81:101 in the *Oceanobacillus ribozyme*); deep blue, 'coordination loop'. Thickened sections of dVI (G2409 to A2413) and IC1 (A83 to A87) correspond to the base-paired segments ('handles') of our anchoring oligonucleotides (panels B-D). The arrow points to the location in the Pl.LSU/2 ribozyme of the η receptor (see Figure 1A); assuming stems II and IIA are stacked, the latter should be situated about one helical turn beyond the tip of what was left of domain II in the molecule crystallized by Toor et al. (2008). (**B**) Scheme for anchoring dVI to IC1, showing IC1 anchor 1 with a 3-T tether. (**C**) anchor 2, with a 3-T tether; at position 82, G was introduced before switching back to U. (**D**) anchor 3, with a 1-T tether.

Figure 4 Branching by dVI-IC1 constructs as a function of the concentration of anchoring oligonucleotides, their tether length and their complementarity to the terminal loop of truncated IC1 stems**.** See Materials and Methods for calculation of relative branching rates and standard errors. (**A**) Optimization of IC1 anchors. Relative branching rate as a function of oligonucleotide concentration for individual construct-oligonucleotide combinations (Figure 3B-C) was fitted to equation (3) of Materials and Methods. Empty squares and dashed curve, construct in Figure 3B with matched oligonucleotide 5'-GTGGAC-TTT-AGCGAA, $K_m = 58 \pm 20$ µM, Pearson's R=0.9969; empty circles and solid curve, construct in Figure 3C with

33

matched oligonucleotide 5'-GTGGAC-TTT-GGCTGG, $K_m = 5.4 \pm 1.0 \ \mu M$ ($K_d = 7.5 \ \mu M$),

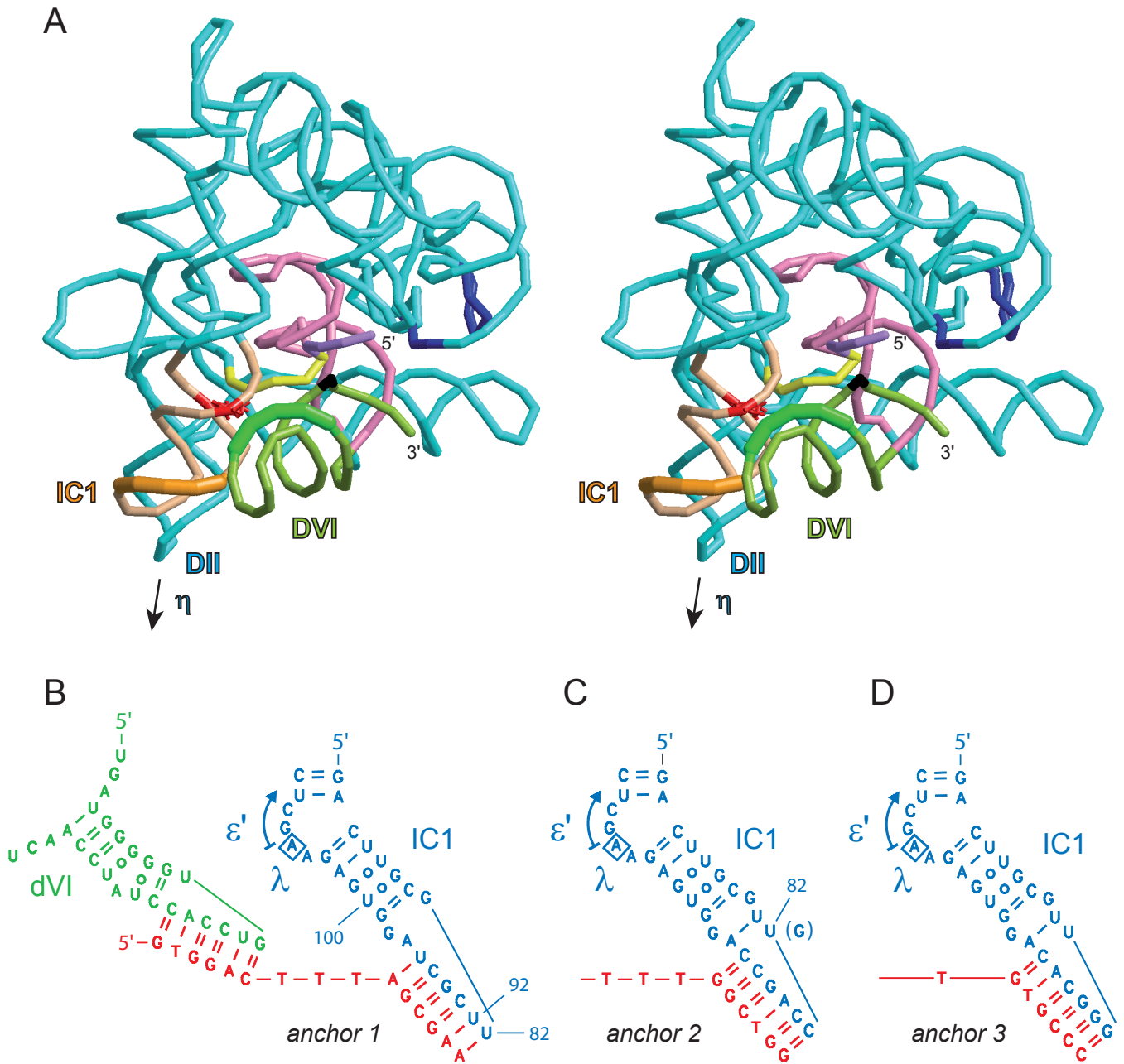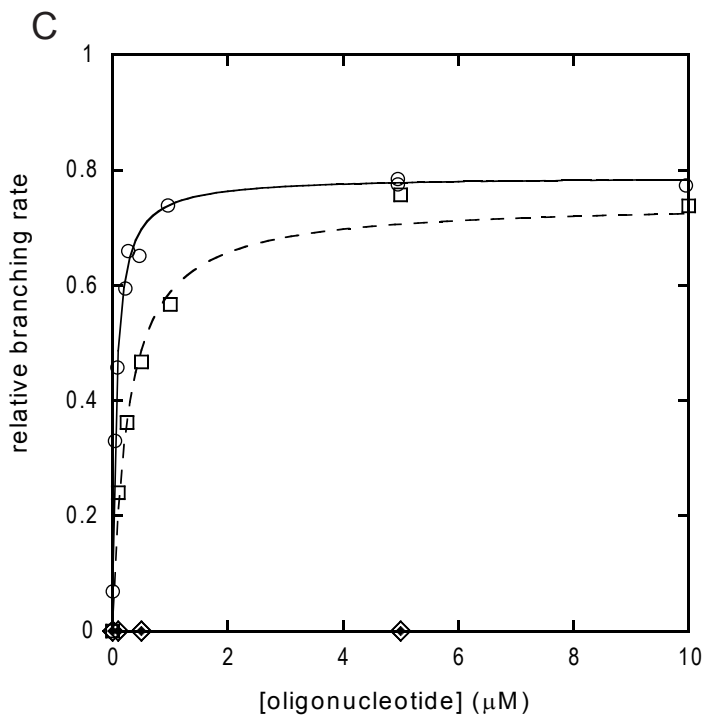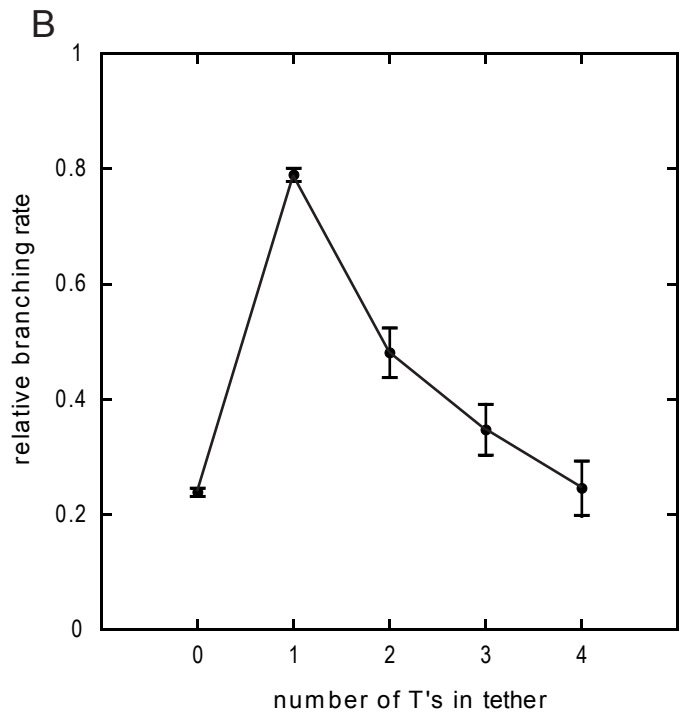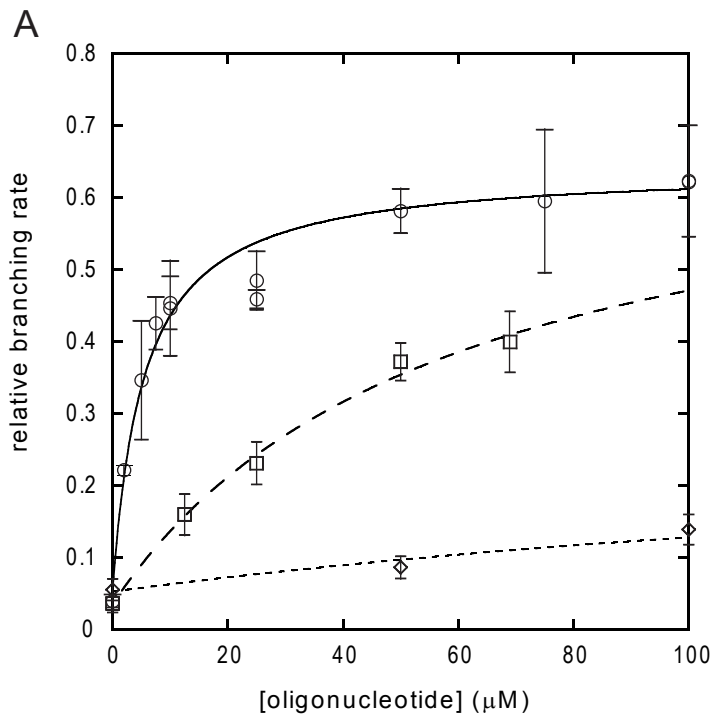R=0.9825; lozenges and dotted curve, construct in Figure 3C with 5'-GTGGACT (no

anchor). (**B**) Relative branching rate of construct in Figure 3C as a function of the number of

T's in the tether of oligonucleotide 5'-GTGGAC[T]$_n$GGCTGG. The concentration of

oligonucleotide was set at 5.0 μM, close to the observed K_m for a 3-T tether (see panel A). (**C**)

Abscissa and ordinates as in panel A. Empty circles and solid curve, construct in Figure 3C

with matched oligonucleotide 5'-GTGGAC-T-GGCTGG, $K_m = 0.073 \pm 0.009 \ \mu M$ ($K_d = 0.17$

μM), R=0.9946; empty squares and dashed curve, construct in Figure 3D with matched

oligonucleotide 5'-GTGGAC-T-GTGCCC, $K_m = 0.27 \pm 0.05 \ \mu M$ ($K_d = 0.55 \ \mu M$), R=0.9938;

filled lozenges, construct in Figure 3C with mismatched oligonucleotide 5'-GTGGAC-T-

GTGCCC; empty lozenges, construct in Figure 3D with mismatched oligonucleotide 5'-

GTGGAC-T-GGCTGG.

Figure 5 Conformational rearrangements and tertiary interactions involving domain VI.

Tentative delimitation of the ι and ι' motifs is based on our modeling of the interaction in

Figure 3A. During the splicing process, domain VI is successively bound by ribozyme

subdomain IC1 (ι-ι' interaction – this work – which positions domain VI for the branching

step) and subdomain IIA (η-η' interaction – Chanfreau and Jacquier, 1996 – which positions

domain VI for exon ligation; a 90 degree rotation was chosen for convenience of drawing, the

actual value must be less, see Figure 3A). In reverse splicing into a DNA or (possibly) RNA

target, formation of ι-ι' should follow that of η-η' (dashed arrow). Bases shown are

consensus ones for mitochondrial subgroup IIB1 introns (Li *et al*, 2011). Curved arrows

symbolize reactions.

Figure 1   Li et al. (2011)

Figure 2   Li et al. (2011)

Figure 3   Li et al. (2011)

Figure 4   Li et al. (2011)

Figure 5   Li et al. (2011)

branching   *conformational rearrangement*   *exon ligation*
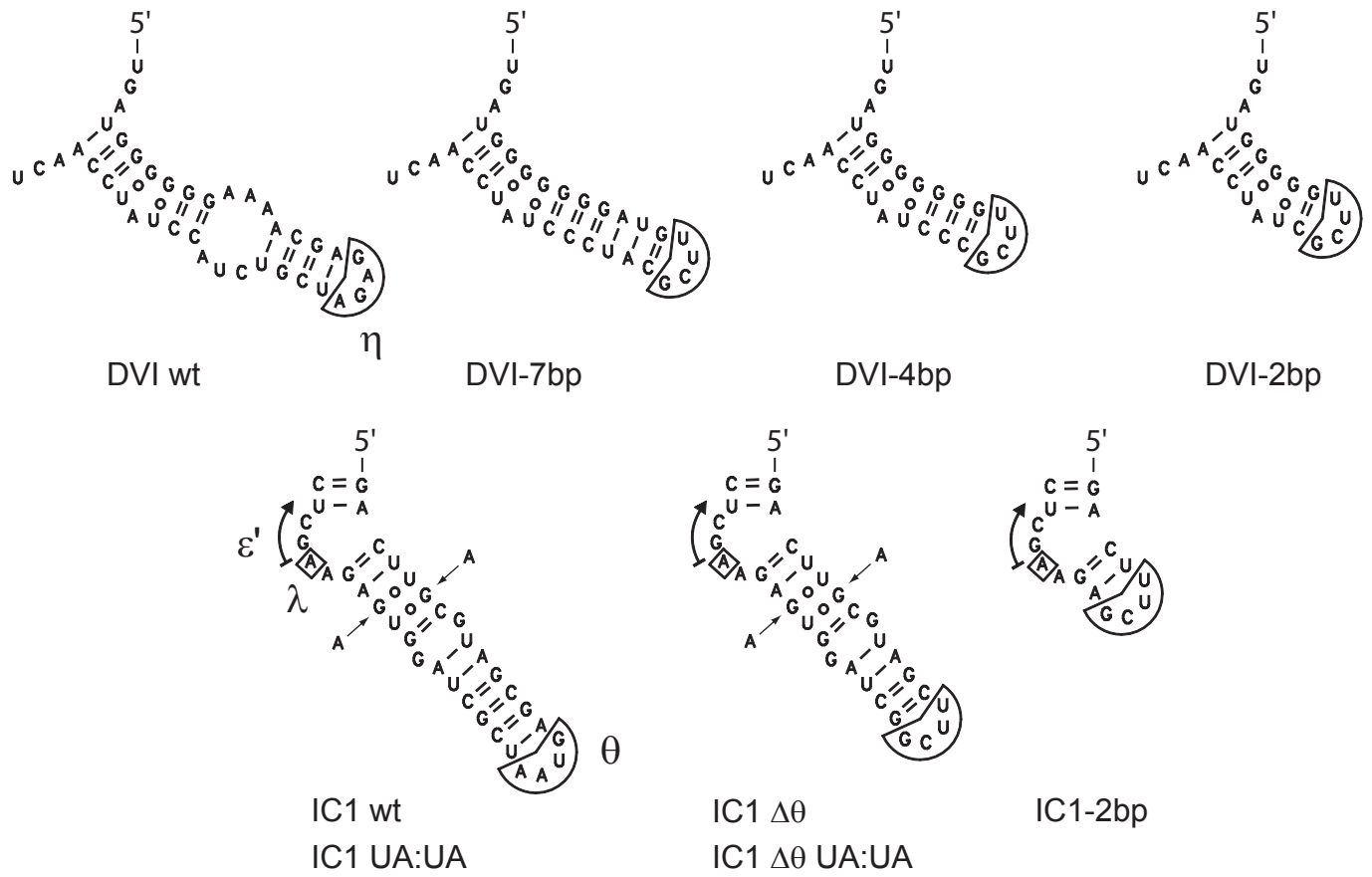
Supplementary Figure Legends

Figure S1 Splicing reactions of internally labeled precursor transcripts with a wild-type or IC1 UA:UA mutant sequence. Products were separated on a denaturing 4% polyacrylamide gel which was fixed and dried prior to exposure and quantitation with a PhosphorImager (Molecular Dynamics). For expected lengths and identification of splicing products, see Costa et al. (1997b).

Figure S2 Time courses of splicing reactions of wild-type and IC1 UA:UA mutant transcripts. **(A) (B)** ammonium-containing buffer; **(C) (D)** potassium-containing buffer. For kinetic parameters and their determination, see Table I and Materials and Methods.

Figure S3 Splicing reactions of construct 3C in the presence of 1 M of an oligonucleotide with a 1T tether and either a matched or mismatched anti-IC1 anchor, compared with a wt splicing reaction (see legend to Fig. S1 for methods). Expected lengths for construct 3C: precursor, 850 nt; intron-3'exon, 724 nt; lariat and linear intron, 618 nt; ligated exons, 232 nt. For the wild-type, all intron-containing forms are 22 nt longer.

Supplementary Figure S1 (Li et al.)
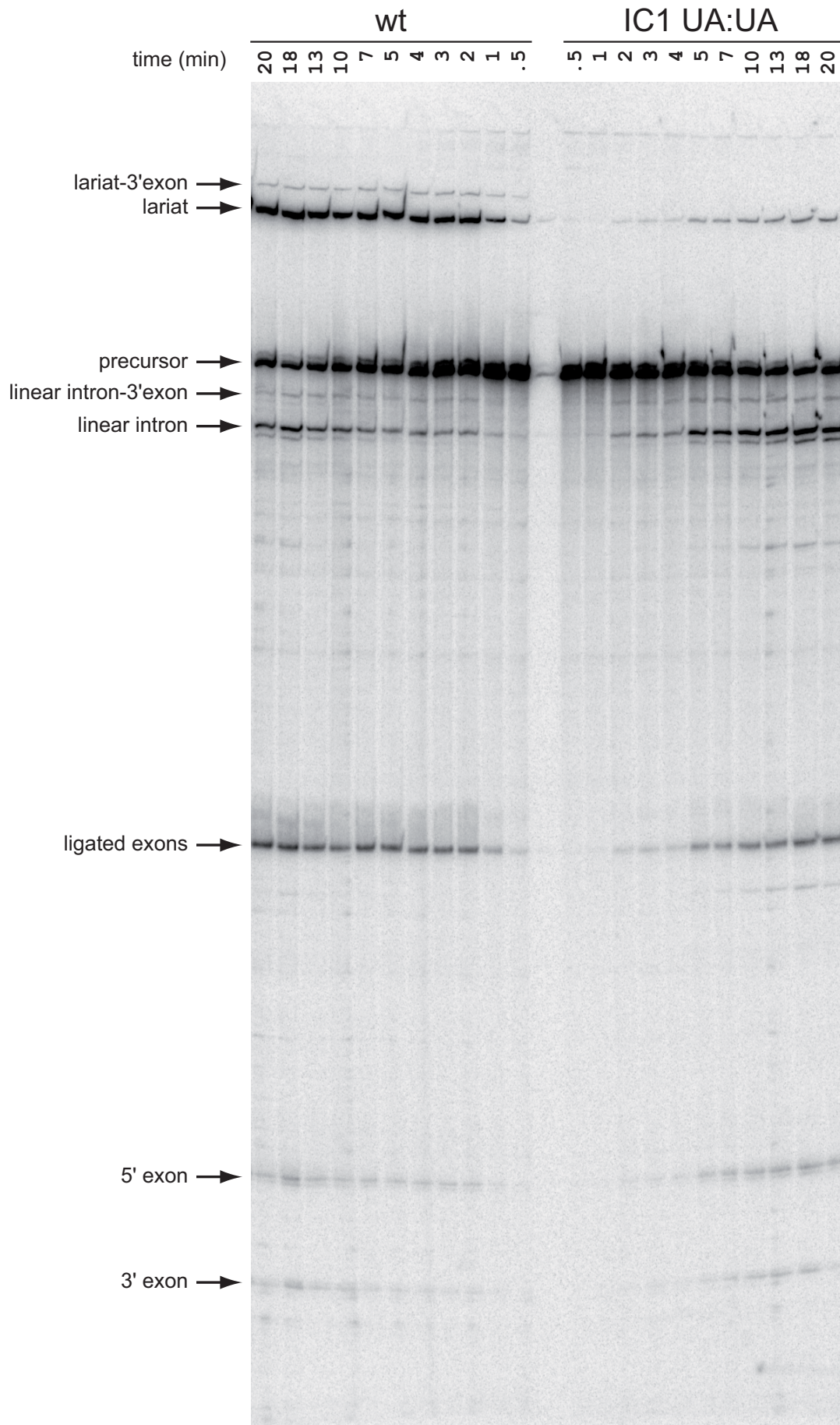
Supplementary Figure S2 (Li et al.)
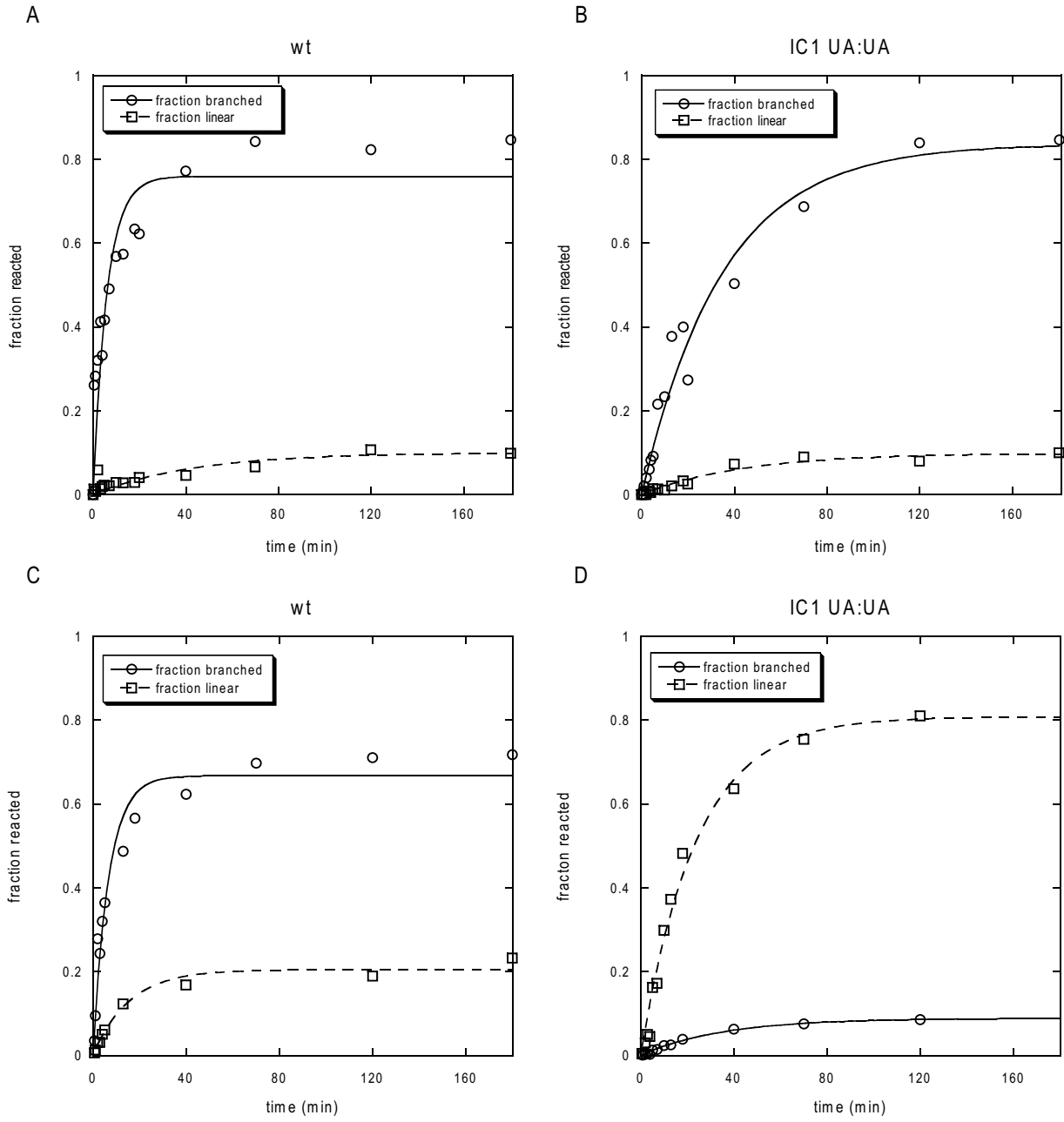
Supplementary Figure S3 (Li et al.)
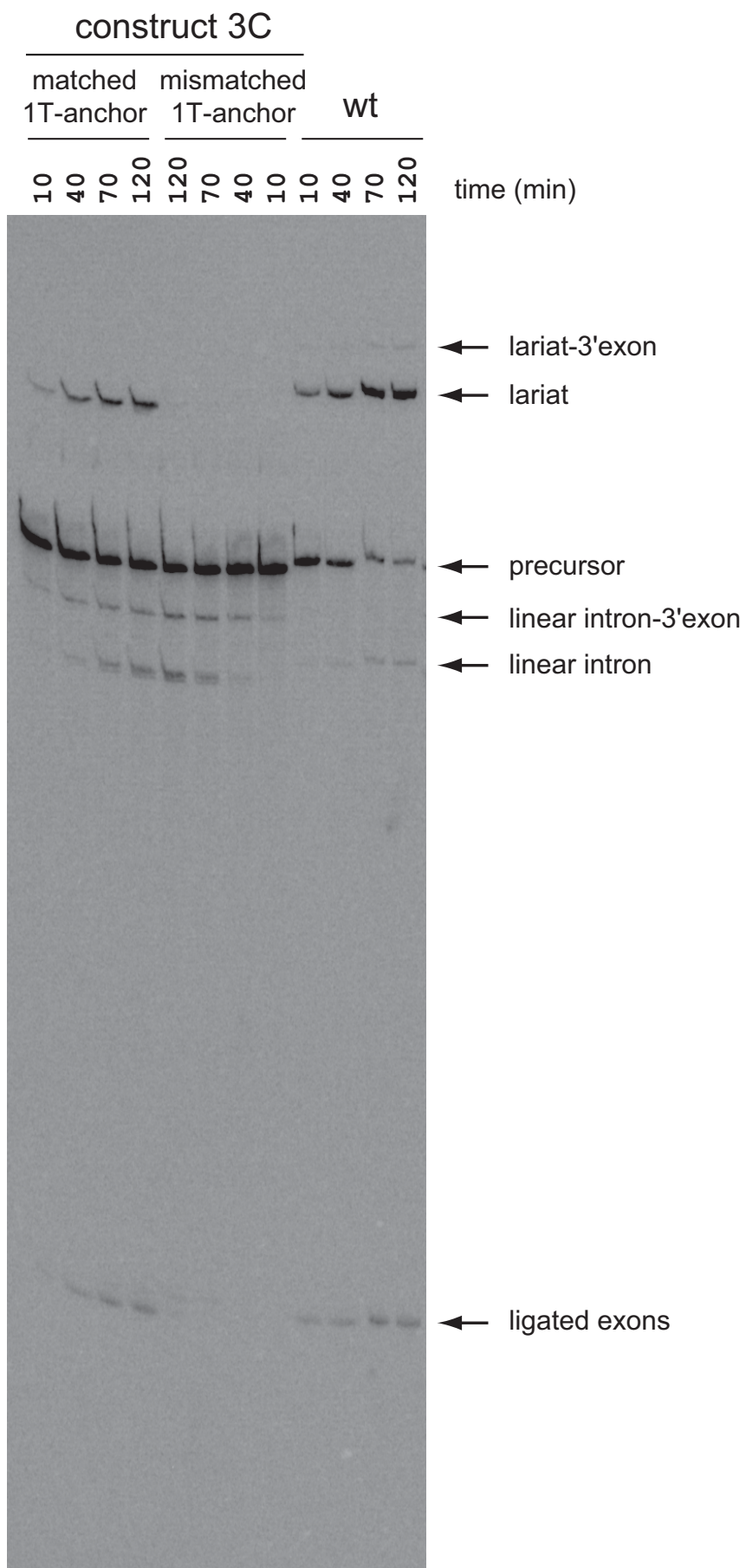
Referee comments and authors' answers for Li, Costa and Michel (EMBO J.)

Fisrt round:

Referee #1

This is an important study because it raises questions about a generally accepted structural feature of group II introns, the coordination loop. The coordination loop was proposed by the Pyle lab to be the receptor for the branch point A in the first step of splicing. However, the coordination loop does not have a clear analog among different intron types, as one would expect for such a conserved element. The present manuscript proposes a different position for domain VI during the branching reaction that is incompatible with the purported role of the coordination loop. This study will probably not convince everyone immediately, but it sets forth a plausible alternative that is supported by at least as much data as the coordination loop.

The study begins with the examination of a subset of mitochondrial introns that have lost the branch site motif and an adjacent bulge motif. Conservation profiles suggested that when the branch site is lost, the residues in domain IC1 are free to mutate. This led to the hypothesis that the bulge motif in domain VI forms an interaction with two base pairs of domain IC1, which is named iota-iota'. Domain VI and IC1 were mutated to demonstrate that the predicted motifs have a role in the branching reaction. The interaction was modeled, which led to the successful testing of the juxtaposed arrangement of domains VI and IC1 using an oligonucleotide splint assay.

Comment: *Francois Michel has an impressive record in identifying interactions using such methods, and this alone makes the study and conclusions compelling. On the other hand, the conclusions would be strengthened by additional experimental evidence, because there are weaknesses in some of the lines of evidence. For example, the strength of the statistical arguments is unclear because an alignment is not provided.*

Answer: As stated in 'Materials in Methods', the alignment that we used for the statistical analyses in Figure 1 is available in Li et al. (RNA Journal, 2011, in press); the final version of that paper contains a 'Supplementary Dataset', which consists of our entire alignment of mitochondrial subgroup IIB1 ribozyme sequences in FASTA format. Please note that 577 sites were actually aligned and analyzed; the value wrongly quoted in the original version of this manuscript – 526 – was the number of sites used to build the phylogenetic tree in Fig.1C, after removal of the IBS and EBS nucleotides and a few highly variable segments (see Li et al., 2011).

Comment: *Mutation of the iota-iota' motifs do not have the consequences one would expect: mutation of the loop motif of domain VI had no discernible effect, and progressive deletions of domain VI showed specific disruptions for branching only for the most extreme deletions where splicing is nearly abolished, raising the possibility that other factors may influence the ratios.*

Answer: In our potassium-containing buffer, replacement of the internal loop motif by canonical base pairs or truncation of the distal section of domain VI down to four base pairs (which removes part – but only part – of what we propose to be the iota' motif) does have significant effects: the fraction of branched products is reduced from at least 70 percent in the wild-type down to 41 and 15 percent, respectively (bottom part of Table I). Moreover, the latter change affects the rate constant for branching both significantly and specifically (hydrolysis is not affected). In order to explain why relatively mild structural changes had no discernible effect in the ammonium-containing buffer, we proposed that in that buffer, docking of domain VI into its IC1 receptor may not be rate-limiting in wild-type precursor molecules.

Comment: *The strongest evidence is definitely the oligonucleotide splint assay. It validates the logic that led to the hypothesis for the interaction. It is difficult to rationalize how the oligonucleotide data could be produced unless domain VI has the proposed position during the branching reaction.*

Other:
For the domain IC1 UA:UA and IC1-2bp mutants, the branching rate decreases dramatically while hydrolysis does not, yet the ratio of lariat:linear products remains nearly the same as wild type. Why?

Answer: That is true only in ammonium buffer, in which about 90 per cent of molecules remain committed to forming lariats, and unable to hydrolyze the 5' splice site, even in mutants with a 10-fold reduced rate constant for branching (the only exception is the dVI-2bp mutant). In potassium buffer, however, changes in rate constants for branching and hydrolysis tend to be reflected in correspondingly altered proportions of branched and linear molecules among reaction products (see bottom part of Table I): what we believe to be the ability to dock domain VI into its receptor affects both observed rates of reaction and the partition of refolded precursor molecules into branching-compatible and hydrolysis-compatible conformations. As we pointed out in the 'Kinetic analyses' section of Materials and Methods, it already was noted by others that refolded precursor molecules do not form a single population with respect to the ability to initiate splicing by branching versus hydrolysis.

Comment: *In the entropy analysis, why were the bulged A in domain V and A104 identified? Is it reasonable that they may contact the iota components in domain VI?*

Answer: The ten intron sequences that lack a discernible branchpoint also have additional nucleotides at their 5' extremity; none of these sequences begins with G and in only five of them does the GUGCG group II conserved motif actually start with G – see Supplementary Dataset). The loss of both G1 and the ability to form lariats may have relieved in turn constraints on the base of A2389 (A376 in the Oceanobacillus ribozyme), whose location in Toor et al.'s crystal structure is compatible with a contact with G1 during the branching reaction. On the other hand, we have no plausible explanation for the statistical signal at A104, unless that nucleotide were to be oriented quite differently in the branching-ready configuration of the ribozyme.

Comment: *What exactly are the nucleotides involved in the iota-iota' interaction? In Figure 5, the interaction is shown as four base pairs in domain IC1 and four nucleotides in domain VI. This is based presumably on modeling rather than experimental evidence. The entropy analysis suggested that a larger portion of domain VI interacts with two base pairs of domain IC1. Please specify what the proposed interaction consists of, and the reasons.*

Answer: Yes, our tentative delimitation of the iota-iota' interaction in Figure 5 is based primarily on modeling, as we now make clear in the legend to that figure: the nucleotides we included are the ones that may directly contact one another according to our provisional model. In domain VI, sequence conservation is expected to extend beyond these nucleotides: most mutations within and immediately next to the internal loop are likely to have been counterselected because they would affect the conformation of that loop and, indirectly, the optimal positioning of nucleotides involved in the iota-iota' interaction (inversely, involvement of the IC1 helical backbone distal to the two G:U pairs is expected to remain undetected by our type of statistical analysis as long as only non-canonical base pairs need to be avoided).

Comment: *In Figure 5, a 90 degree rotation of domain VI is depicted between the steps of splicing.*

Answer: Figure 5 is intended as a mere sketch; as now explicitly stated, a 90 degree rotation was chosen for convenience of drawing and in order to convey the impression that this is a major translocation (although not quite as extensive as in J. Wang's model or A.M. Pyle's latest sketch, as indicated in our Discussion); see also below.

Comment: In the model, however, the positions of the eta-eta' components in DII and DVI seem to be close enough that both eta-eta' and iota-iota' might form simultaneously even without DVI movement.

Answer: Figure 3A is to be viewed in three dimensions and it should be noted that in the a5gamma and Pylaiella LSU2 ribozymes, the eta receptor is located far beyond the tip of what was left of domain II – hardly more than a stump – in the molecule that was cristallized by Toor et al. and whose atomic coordinates we used to build our model. We now point to the location of eta in Figure 3A and mention, in the legend to that Figure, the need to extend mentally the helix.

Comment: Do the authors propose a twist of the DVI helix rather than a rotation? Please explain in more detail what conformational change is considered likely to occur for domain VI.

Answer: Precise modeling of domain VI in its second-step conformation is beyond the scope of this work, if only since it involves some speculations. Nevertheless, let us assume that stems II and IIA in Figure 1A are coaxially stacked, domain VI (assuming it is rigid) would then need to rotate by about 50 degrees and undergo a 150-160 degree twist after the branching reaction in order to dock into its eta receptor; as a consequence, its tip should move by some 55 angstroms, which is a major translocation indeed (note that this value should not depend too much on the exact angle between helices II and IIA, since the latter is far shorter than the former).

Minor suggestions:
Comment: p. 2 line 8 and throughout manuscript. Change "distal of" to "distal to"

Answer: OK

Comment: p. 2 line 13 Change "Noteworthingly" to "It is noteworthy that".

Answer: We changed it to 'Interestingly' in order to remain at the 175 word limit

Comment: p. 3 line 7 Change "ribozyme component" to "ribozyme components"

Answer: OK

Comment: p. 3 line 13 What is meant by "seemingly unique"? Unique among ribozyme active sites? A single active site within the group II ribozyme?

Answer: We have replaced 'unique' by 'single'

Comment: p. 4 line 14 Change "dubbed coordination loop" to "dubbed the coordination loop"

Answer: OK

Comment: p. 6 line 8 Change "sequence and structure" to "sequences and structures"

Answer: OK

Comment: p. 7 seven lines from bottom Change "with an altered" to "have an altered"

Answer: OK

Comment: p. 8 line 4 Change "Still, replacement" to "Unexpectedly" or "Contrary to this hypothesis"

Answer: We replaced 'Still' by 'However'

Comment: p. 10 line 3 Change "3' splice" to "3' splice site"

Answer: Yes

Comment: p. 10 line 4 Change "though" to "although"

Answer: OK

Comment: p. 11 line 11 and throughout manuscript. Change "shallow groove" to "minor groove" .

Answer: In an A-type helix, the counterpart of the minor groove of the DNA B-type helix is actually somewhat wider than the so-called 'major' groove. We, and a number of our colleagues – just search for 'RNA helix shallow groove' on the Web – rather believe that the two grooves of an RNA double helix should be designated by the terms 'shallow' and 'deep'. We have added 'minor' between brackets after our first mention of 'shallow' [groove].

Comment: p. 14 line 5 Change "must come truly" to "must truly come"

Answer: OK

Comment: p. 14 *line 13 What does "prior modeling" refer to? If it is published it should be cited. If it is not published it should be "data not shown." If it is the modeling in this manuscript, then omit "prior."*

Answer: Yes, we were referring to the modeling in this manuscript, 'prior' was inappropriate and we removed it.

Comment: p. 15 line 17 Change "supernumary" to something else, perhaps "supernumerary". "Appended" would be clearer.

Answer: We changed this word to 'additional'

Comment: p. 16 lines 17-20. The sentence is contradictory because it says there is both general agreement and debate.

Answer: Yes, there is agreement on the fact that domain VI interacts at some stage with domain II, but debate on whether this interaction contributes to positioning the 3' splice site for exon ligation. We have rewritten that sentence accordingly.

Comment: p. 18 line 5 What is meant by "costly"? Expensive? Experimentally difficult?

Answer: We replaced 'costly' by 'difficult'

Comment: p. 32 Fig. 4 legend. Is the error measurement the standard deviation or standard error of the mean?

Answer: All errors are standard errors of parameters, which were estimated by fitting

experimental data (see Figure S2) to equations, as explained in Materials and Methods (we have added 'and their standard errors' after 'initial rates at t=0').

Comment: *p. 33 Fig. 4 legend Change "full curve" to "solid curve"*

Answer: Yes

Referee #2

In this work Li et al., aimed at identifying a receptor site for the branch-point adenosine in group IIB1 introns. By combining phylogenetic and mutational analysis with a rational design of a "molecular tether", the authors were able to identify such a receptor site for D6. Strikingly, this site is located close to the ε' and λ sites within stem c1 in domain 1. Importantly, this receptor site is specific for the first transesterification step of splicing. The proposed novel interaction was termed ι-ι' and it has been suggested that this interaction is disrupted after branching and D6 has to undergo a significant rearrangement for the η-η' contact to occur, which is known to be essential for performing the second step of splicing. Impressively, the authors went on and demonstrated that D6 can be modeled into the crystal structure the group IIC intron from Oceanobacillus iheyensis to allow the ι-ι' contact. In light of previous findings, this manuscript presents a very interesting, elegant, but also in part controversial study on an aspect of group II intron splicing that finds its parallels in spliceosomal intron splicing.

Major comments

Comment: *1. As mentioned by the authors, Pyle and coworkers demonstrated a few years ago that the branch-point adenosine is coordinated to the asymmetric internal loop composed of Jd''/d''' and Jd'''/d'' in the ai5γ group IIB1 intron (Hamill, 2006). The Pyle lab had applied cross-linking to identify residues in spatial proximity to the branch-point. These residues included G1 and C-1 together with two nucleotides in J2/3 (G588 and U590; the counterpart of former and of A589 are part of the triple helix in the active site of the Oi. Intron (Toor et al., 2008).) All other cross-linked residues were located in the coordination loop, which harbors EBS 3 as well. Since Pyle and coworkers used a trans-branching system, the obtained data should have been specific for the branching pathway of splicing. While cross-links to G1, C-1, G588 and U590 can be readily explained by the fact that they are active-site constituents, the phylogenetic data in the current study do not seem to support the coordination loop as docking site for the branch-point. How can the cross-links from A880 (ai5γ branch-point) to the coordination loop be reconciled? Also, looking at Fig. 3A the branch-point A in D6 is not even remotely close to the coordination loop in the model. On the other hand, Hamill and Pyle did not observe any cross-links to stem c1 in D1. Is there any explanation for this apparent discrepancy between the two studies?*

Answer: What Hamill and Pyle actually observed is (1) when D56 molecules with a photoactivable group at either the branch point or one of the two flanking nucleotides are incubated with the rest of the ribozyme (exD123) under conditions compatible with a folded structure, they crosslink to a diversity of sites, including the coordination loop; (2) when the unreacted, 'XL1' mixture of crosslinks is reincubated under conditions conducive to splicing, some of it reacts, and yields the free 5' exon, as well as another product, which might be a branched molecule, although that was not established. Importantly, crosslinks were mapped before, not after reincubation, so that it is not known which of the crosslinks in the mixture were compatible with activity (the same is true of the fraction of molecules that were able to carry out both steps of splicing – Fig. 5B).
When attempting to interpret these data, the authors implicitly assumed that domain VI

should be stably docked into its first-step binding site in unreacted molecules. However, in Costa et al. (1997a) we had previously shown that yeast intron Sc.cox1/1 precursor molecules exist in (at least) two distinct, about equally abundant conformations, one which leads to branching and the other one to 5' splice site hydrolysis; these conformations must differ by the location of domain VI, since disrupting the eta-eta' interaction between domains II and VI suppressed hydrolysis, whereas molecules in which the same interaction was reinforced reacted exclusively by that mechanism. In Fig. 10 of that paper, domain VI of unreacted molecules was accordingly depicted as toggling between a DII-bound state and another one in which it was poised for the branching reaction (that is, bound to its then hypothetical, first-step receptor; obviously, domain VI should also remain unbound for some length of time in between two docking events). Our working model, which remains compatible with all data we know of (and in particular, the absence of domain VI in Toor et al.'s structure; see discussions in Pyle, 2010) was that as a general feature of group II self-splicing, domain VI always keeps toggling between its first- and second-step binding sites, even though the exact equilibrium between the two conformations must depend on each particular intron and intron form. These ideas may now be put to test by mapping Hamill and Pyle's sites of crosslinking on Toor et al.'s atomic resolution model of ribozyme domains I to V. When that is done, it becomes apparent that essentially every nucleotide that would have been accessible to the branch site and its two flanking nucleotide in a dVI molecule that could freely rotate around the dV-dVI junction did give rise to a crosslink. Only residues in the proximal helix of domain V are missing in the list: since crosslinks at these sites would be internal to the D56 piece, they were not recovered in Hamill and Pyle's experiments.

It is also apparent from three-dimensional modeling that the IC1 helix is essentially out of reach of the branchpoint and its immediate neighbors (this can be checked with the help of the stereo drawings in Fig. 3A of this manuscript, even though the angle of view is not ideal). When domain VI is bound to its IC1 receptor, photoactivable bases at the branchpoint and its two neighbors are predicted to crosslink instead to the first two nucleotides of the intron and the last nucleotide of the 5' exon: these three positions were indeed among those recovered by Hamill and Pyle. Moreover, among the latter crosslinks, those to the G1 nucleotide (and perhaps also to the second residue of the intron) are liable to be compatible with splicing, which provides a ready explanation for the (limited) reactivity of the XL1 material.

To summarize, as long as they are not overinterpreted, Hamill and Pyle's data do not contradict in any way our own findings and conclusions.

Comment: *Since the ai5γ intron has often been referred to as "weirdo" among group II introns, do the authors consider it a possibility that the coordination loop functions as receptor for the branch-point in the ai5γ intron only? Or, is more likely that cross-linking possibly produced in part erroneous data (as it had happened before).*
Since there is this controversy, this reviewer is of the opinion that it would be an critical control experiment to mutate the corresponding tandem GU wobble pairs in stem c1 of the ai5γ intron and test for its ability to perform branching (despite data on the Bc. intron from Stabell et al., 2009).

Answer: As just stated, any possible controversy should not be about data, but their interpretation. In spite of its high A:U content and elevated magnesium requirements for in vitro activity, which do not make it such a good model system, the ai5γ intron looks fairly typical of organelle members of subgroup IIB1, and we see no reason why mutation of the tandem G:U pairs in the IC1 stem of that intron would not affect the ability of the ribozyme to carry out branching, as reported in this work for the Pl.LSU/2 ribozyme and also by Stabell et al., using a molecule from a different structural subgroup (the experimental setup must of course ensure that branching is rate-limiting).

Comment: *2. Hydroxyl radical footprinting has been done on the Pylaiella intron (Costa et al., 2000). Have the authors also mapped D6 and stem c1? At least from the ai5γ footprinting data it appears that both GU wobble pair are internalized (except G87;*

Swisher et al., 2001). In other words, how does the model in Fig. 3A correlate with such footprinting data?

Answer: Our data for the IC1 stem were published in Fig. 7 of that paper. In the lariat intron, there is a small zone of partial protection from hydroxyl radicals centered on G79, whereas nucleotides around position 100 are moderately accessible. However, these data should be regarded as irrelevant to our model as long as there is no evidence that domain VI is stably docked into its first-step receptor. As we explained on p.17 l.4-9, it is our hope that our system of anchoring nucleotides will make it possible to lock the ribozyme into its first-step conformation, which would in turn make 'footprinting' pertinent.

Comment: *3. Assuming that ι-ι' takes place in the ai5γ intron, the available NAIM data on this yeast mitochondrial intron support that the minor groove of stem c1 is involved in the ι-ι' interaction: the exocyclic amine of G87 and the 2'hydroxyl groups of U86 and U110 were described to be important for branching (Boudvillain et al., 1998) - as stated by the authors. In the same paper a 2-aminopurine and 7-deaza effect were observed for A861 and A863, suggesting an involvement of the major groove (N6, N7). Can one infer any interaction from the available data and your model?*

Answer: The information provided by NAIM experiments may reasonably be interpreted in terms of specific, direct atomic contacts only as long as one is dealing with components the structure of which is known (or believed to be so), as is the case for the IC1 distal helix. We believe that the structure, either in isolation or in interaction, of the AAA:CUA, DVI internal loop of mitochondrial subgroup IIB1 introns can not be predicted from currently available data and accordingly, we do not wish to take stands on what it might be. As already stated in our Text, we provisionally modeled the distal part of domain VI as a continuous helix because (i) that is by far the most commonly encountered situation in intron subgroups that have a conserved G:U pair in IC1 at positions equivalent to Pl.LSU/2 79 and 100; (ii) replacement of that loop by canonical base pairs is compatible with branching (even though it is not optimal; this work and Chu et al., 1998).

Comment: *The presented data unambiguously demonstrate that D6 branch site and stem c1 are spatially very close, however, in order to definitively state that these structural elements are in physical contact (i.e. a novel interaction), it is preferable to have an idea about potential H-bonds in the ι-ι' pairing (in addition to the phylogenetic evidence).*

Answer: As can be checked by modeling, the fact that a single-nucleotide tether is optimal (and the lack of it is tolerated) is hardly compatible with anything but a direct contact.

Comment: *As stated* (**Authors:** observed, in fact) *by the authors, the D6 internal loop can be replaced with canonical base pairs without abolishing branching. What is the advantage of maintaining an internal loop throughout evolution?*

Answer: Admittedly, the internal loop of mitochondrial members of subgroup IIB1 is replaced by a continous helix in the vast majority of bacterial members of this subgroup. However, close examination of secondary structure models reveals that the location of the predicted eta receptor may not be exactly the same in the two subsets and in fact, the total length of the distal DVI stem differs (by two base pairs): the need to simultaneously ensure efficient docking into the IC1 receptor and maintain the geometry between the base and tip of domain VI appropriate for the eta-eta' interaction may be the key to structural conservatism in the middle part of domain VI.

Comment: *4. The oligonucleotide tether is a very elegant way to further support the*

spatial proximity of D6 bulge and the tandem GU pairs in stem c1. The different variants were compared for their relative branching rates. However, it would be helpful to enlist the absolute kobs values together with the wt activity from Table I.

Answer: At a 5 μM concentration of anchoring oligonucleotide, k_{obs} values (calculated from the fraction of unreacted precursor molecules) ranged from 0.0042 ± 0.0002 min^{-1} (for a 4-nucleotide tether) to 0.0093 ± 0.0005 min^{-1} (for a 1–nucleotide tether).

Comment: *Along the same line, I urge the authors to show a representative gel for splicing of the wt and at least the IC1 UA::UA mutant and the anchor 2 with 1 T only.*

Answer: Yes, we now have three supplementary figures: Figure S1 shows a representative gel autoradiograph of wt and IC1 UA:UA splicing reactions in the presence of potassium; in Figure S2, reaction time courses in ammonium- and potassium-containing buffers are compared; and Figure S3 shows a gel with splicing reactions of the construct in Figure 3C in the presence of oligonucleotides with a 1T-tether and either a matched or mismatched handle for binding IC1; next to these lanes a wt splicing reaction was run on the same gel for comparison purposes.

Comment: *5. It is my understanding that the coordination loop is poorly conserved among group II introns (Michel et al., TIBS 2009), but what about the tandem GU pairs in stem c1? How well are these and in turn the ι-ι' contact conserved among group II introns (possessing a branch-point) of different phylogenetic families.*

Answer: The (counterpart of the) G79:U100 IC1 base pair is generally conserved in ribozyme structural subgroups IIB1, IIB3, IIB4 and IIC (see Toro, 2003, Environ. Microbiol. 5, 143-151, for nomenclature). Most importantly, in subgroup IIA, lengthening of the epsilon' loop from 4 to 11 nucleotides generates a ready candidate for a domain VI receptor, whereas the section of the ribozyme that would be expected to host a counterpart to the coordination loop (rather, to what would be left of that loop after migration of EBS3 to the δ position, next to EBS1) is not conserved, whether in terms of sequence or structure.

Comment: *6. Another tertiary contact has been proposed by Pyle and coworkers a few years ago (Fedorova et al., 2005): μ-μ' in ai5γ. As this contact has not been included into the schematic drawing of Fig. 1A, I am wondering whether there is a specific reason for it. Is such a contact not supported by phylogeny in the Pylaiella intron?*

Answer: Loop IIIA, with a GUAAU consensus sequence (the two adenines were proposed to constitute the μ site), has a scattered distribution in subgroups IIB and IIC – there is no evidence of its, or a counterpart of it, being present in the Oceanobacillus ribozyme, or in most of the many subgroup IIC members. In fact, μ-μ' is often omitted from secondary structure models (for instance from Fig. 3 of Hamill and Pyle, 2006). However, we agree that that is no good reason to overlook this proposed contact and since it is potentially present in the Pl.LSU/2 ribozyme, we have now included it in Figure 1A.

Minor comments:

Comment: *I suggest highlighting Domain 2 in Fig. 3A to be able to imagine the conformational switch of D6 shown in Fig. 5.*

Answer: Yes, we now point to the location of the eta receptor in Fig. 3A, and also stress in the legend to that figure that it is necessary to extend mentally the DII helix by about one helical turn when trying to imagine the position of the tip of domain VI when it is bound by the eta-eta' interaction (see also our answer to a similar comment by referee #1).

Comment: *In the methods section, please name and cite the program used for the alignment.*

Answer: The usefulness of alignments destined to be exploited in comparative sequence analyses depends for a large part on human expertise and accordingly, they should best be generated manually (see discussions in Michel and Costa, 'Inferring RNA structure by phylogenetic and genetic analyses', in 'RNA Structure and Function', Cold Spring Harbor Laboratory Press, R. Simons and M. Grunberg-Manago eds., pp. 175-202, 1998). We have made our alignment of mitochondrial subgroup IIB1 introns available as a Supplementary Dataset in Li et al. (RNA Journal, 2011).

Comment: *In Table II setup 2, please explain the 3-fold enhancement compared to setup 1.*

Answer: Yes, the anti-DVI 7-mer included in setup 2 (and also in setups 4 and 6) provides limited compensation, and only so at very high oligonucleotide concentrations; in fact, the relative rate of branching remains too small to allow a Km to be estimated (see Figure 4A, lozenges and dotted curve). We have added a sentence to make it explicit that that is true not merely in a mutated IC1 context, but also in the presence of the wild-type IC1 sequence.

Comment: *In the last section of the Discussion the authors mention "costly". I suggest removing or rephrasing this sentence. There is no need to explain why the authors have not performed NAIS (yet), as this would go beyond the scope of this manuscript.*

Answer: We replaced 'costly' by 'difficult'.

Second round:

Referee #1 (Remarks to the Author):

Comment: *The manuscript is improved from the previous version, although some issues were not fully addressed.*
-The alignment upon which the comparative data are based is stated to be available as supplementary data for a manuscript in press; however, the reviewers have not had access to it.

Answer: As now indicated in Materials and Methods, in the 'Sequence analyses' subsection, the alignment of intron sequences that we used for our comparative analyses can now be downloaded at
http://rnajournal.cshlp.org/content/suppl/2011/05/05/rna.2655911.DC1.html

Comment: *I still think the domain VI mutagenesis experiments provide weak evidence. If the domain VI motif is so noticeably conserved over evolution, then one would expect a significant effect when it is mutated. Instead the significant effects seem to be for mutations in domain IC. It would be preferable to provide an additional type of experimental evidence for the contact between DVI and DIC. Cross-linking experiments would be an obvious choice, or protection experiments, which the Michel lab has used quite successfully in the past. The oligonucleotide experiment provides the only direct evidence for a contact. An additional source of data might cement the existence of the iota interaction, rather than making a strong case.*

Answer: With regard to mutational effects, it is essential to distinguish between ammonium and potassium ions, as was done in Table I. Even in subdomain IC1, point mutations have only limited effects as long as ammonium-containing solutions are used for self-splicing tests: substitution of the two consecutive G:U pairs by U:A pairs does not

change detectably the fraction of molecules that will react by branching (Table 1) and our interpretation (p.8 l.12-14) was that docking of domain VI into its proposed IC1 receptor is not limiting for the ability to carry out branching under these conditions. In contrast, when branching is challenged by the presence of potassium ions, the fraction of molecules that react by branching becomes quite sensitive to structural alterations in both the iota and iota' motifs: precise removal of the DVI internal loop (dVI-4b) brings that fraction down to 0.15, a value that is close indeed to that observed (0.10) for the IC1 UA:UA mutant (Table I; the kbr/khy ratio is also the same for the two mutants).

Cross-linking or protection experiments on the wild-type intron are unlikely to detect the contact between DVI and IC1, since the molecule is predominantly in a second-step conformation, as revealed by the fact that disruption of eta-eta' greatly increases either the rate of branching (Chanfreau and Jacquier, 1996) or the fraction of molecules that react by branching (Costa et al., 1997). Performing similar experiments on the complex between one of our mutated ribozymes and an oligonucleotide that restores branching through complementarity to both DVI and IC1 will merely confirm the existence of this complex (provided its stability is sufficient to make it the new ground state of the system). For future experiments to be truly informative, they should be able to probe the active state of the branching complex, and as we already suggested (now on p.19 l.1-5), we believe that the best possible approach at present is NAIS (Strobel, 1999). However, as previously pointed out by referee #2, 'this would go beyond the scope of this manuscript'.

Comment: *p. 8 "Only by bringing the length of the helix distal to the branchpoint down to two base pairs do consequences suddenly become dramatic, with splicing proceeding almost exclusively by hydrolysis." A dramatic difference is not seen for the ammonium conditions, but for the potassium conditions, which have not yet been explained in the text.*

Answer: Our sentence referred to mutant dVI-2bp (Table I, line 4), which dramatically differs from the wild-type indeed, since splicing is seen to proceed almost exclusively by hydrolysis, even in ammonium conditions. In order to remove any possible source of confusion, we have now added an explicit reference to the mutant and corresponding line in Table I within the sentence in question.

Comment: *The explanation in Table 1 for the discrepancy between the fraction of product branched and the kbr/khy ratio is reasonable, but it would be helpful to include this explanation in the manuscript, perhaps as a footnote to the table.*

Answer: Yes, we have now inserted our explanation for this observed discrepancy at what we believe to be the most appropriate place – in the 'Kinetic analyses' subsection of Materials and Methods.

Comment: *My point about the position of the eta interaction was not that it should be modeled, but that it is nearby. Given the uncertainties in modeling, it seems possible that one could model the eta and iota interactions simultaneously. This is unlikely to occur of course, but it suggests that the conformational change for domain VI may be subtle rather than dramatic.*

Answer: As we previously explained, reasonable assumptions about the position of the eta receptor imply a large translocation. Nevertheless, we have now added bried notes of caution in the Discussion and legend to Fig. 3A so as to warn the reader that the exact location of eta is a bit uncertain.

Comment: *The detailed explanation accounting for data in the Hamill et al paper is informative, and in my opinion should be included in the Discussion. The present study directly contradicts the conclusions of the Hamill paper, and for the sake of clarity in the*

literature, it would be helpful for this to be acknowledged more directly with a rational explanation provided, since the Hamill data are not disbelieved.

Answer: We have inserted in the Discussion an additional paragraph in which we explain that we believe Hamill and Pyle's data are compatible with our work indeed, as long as it is not assumed that domain VI is stably docked into its receptor site prior to branching.

Comment: *The added three supplementary figures are all improvements. I suggest including Supplementary Figure 1 in the main text.*

Answer: We are grateful to referee #2 for pointing out the necessity to give interested readers access to representative examples of our raw data and for helping us to select the gels to be provided as Supplementary Materials. On the other hand, we do not think it necessary to include alongside our main text gels and graphics that were not conceived for illustrative purposes, but for quantitation, and the main information content of which we believe to be appropriately summarized in the Tables and Figure 4.

Comment: *The phrase "fully consistent" is used in several places to describe how the experimental data relates to the comparative data. This seems an overstatement, because it implies that the theoretical inferences were entirely correct. There is noise in the comparative data though. "Consistent" would be better.*

Answer: Ok, we removed 'fully' every time it appeared in front of 'consistent'.

Comment: *p. 5 six lines from bottom "chloroplasts" not "chloroplats"*

Answer: Thank you.