CHARACTERIZATION OF THE SPLICING REACTION OF THE Sinorhizobium meliloti RmInt1 GROUP II INTRON

ISABEL CHILLÓN GÁZQUEZ
TESIS DOCTORAL
2011
CHARACTERIZATION OF THE SPLICING REACTION OF THE *Sinorhizobium meliloti* RmInt1 GROUP II INTRON

Memoria que presenta la licenciada en Biología
Dña. Isabel Chillón Gázquez
como aspirante al grado de Doctor

Fdo: *Isabel Chillón Gázquez*

VºBº de los Directores de la Tesis Doctoral

Fdo: Dr. Nicolás Toro García
Doctor en Ciencias Biológicas
Profesor de Investigación del CSIC

Fdo: Dr. Francisco Martínez-Abarca Pastor
Doctor en Ciencias Biológicas
Investigador Científico del CSIC

Universidad de Granada
2011
Este trabajo de Tesis Doctoral ha sido realizado en el Departamento de Microbiología y Sistemas Simbióticos (Grupo de Ecología Genética de la Rizosfera) de la Estación Experimental del Zaidín (CSIC-Granada).

Para la realización del siguiente trabajo, la Lda. Isabel Chillón Gázquez fue financiada mediante las siguientes fuentes:

- Beca predoctoral I3P-CSIC-2006, de enero de 2007 a diciembre de 2010.
- Beca de movilidad para estancias breves del CSIC, disfrutada en el Departamento de Biofísica y Bioquímica Molecular de la Universidad de Yale (USA), bajo la dirección de la Profesora Anna Marie Pyle. Periodo: Abril-Junio 2010.

Parte de los resultados presentados en esta Tesis Doctoral han sido publicados en revistas internacionales o están en preparación:


**Chillón I, Fedorova O, Martínez-Abarca F, Pyle AM and Toro N** (2011) In vivo and in vitro characterization of the splicing of RmInt1 group II intron. Implications of tertiary interactions in the ribozyme for the maintenance of the catalytic properties (In preparation).
Abstract

Group II introns are both auto-catalytic RNAs and retrotransposable elements. They were initially found in organelar genomes, and later in bacterial and archaeal genomes. Several data suggest that they are the ancestors of eukaryotic spliceosomal introns. They consist of a catalytically active intron RNA (ribozyme) and an optional intron-encoded protein (IEP), which assists the intron in the splicing and mobility processes. The secondary structure of the ribozyme consists of six helical domains radiating from a central wheel, which in turn interact to each other creating a complex tertiary networking. Group II introns are spliced via a lariat o linear intermediate, and can retrohome in cognate intronless target sites they recognized through base pairing.

RmInt1 is an efficient mobile intron found within the ISRm2011-2 insertion sequence in the symbiotic bacterium Sinorhizobium meliloti. This group II intron is excised, in vivo and in vitro, as both intron lariats and putative intron circles. However, until date, the complete splicing reaction in vivo remains to be elucidated. A lacZ reporter gene system, northern blotting and real-time reverse transcription were carried out to investigate RmInt1 splicing activity (Chapter 1 and 2). Splicing efficiency of 0.07 ± 0.02 % was recorded. These findings suggest that bacterial group II introns function more like retroelements than spliceosomal introns.

Previous studies on the in vitro self-splicing of RmInt1 intron showed an inefficient reaction characterized by the formation of unconventional side products. By testing different RmInt1-transcript context and reaction conditions, we obtained a RmInt1-construction by eliminating an IBS1-like sequence present in the first 10 or 11 nucleotides of the 3’ exon from the natural target IS2011-2 insertion sequence (Chapter 4). The new construct prevented the formation of unusual truncated side products previously observed, producing an increment of the amplitude of the reaction.

Finally, we analyzed the effect of point mutations in conserved positions of the ribozyme of RmInt1 on the efficiency of first and second splicing steps in vivo and in vitro (Chapter 3 and 4). We identified the occurrence of tertiary interactions common to other group IIB introns, such as the γ-γ’, EBS3-IBS3 and the non Watson-Crick interaction between first and penultimate nucleotides of the intron. In addition, we proved the existence of structural motifs conserved in group II introns as the triplex between domain
V and the nucleotides within the linker region J2/3. Mutations disrupting interaction between components of this motif resulted in the alteration of both steps of splicing \textit{in vivo} and \textit{in vitro}, which supports the occurrence of a unique active site catalyzing both steps of the splicing reaction.
Resumen

Los intrones del grupo II son ARNs autocatalíticos y elementos retrotransponibles. Fueron descubiertos inicialmente en los genomas de orgánulos, y más tarde en genomas bacterianos y de arqueas. Según diversos datos, se ha sugerido que podrían ser los antecesores de los intrones espliceósómicos eucarióticos. Constán de un componente de ARN que es activo catalíticamente (ribozima) y de una proteína codificada por el intrón (del inglés Intron-Encoded Protein, IEP) que es opcional y asiste al intrón en los procesos de splicing y movilidad. La estructura secundaria de la ribozima consiste en seis domíniuos helicoidales que radian de una estructura central, y que contactan entre sí dando lugar a una compleja red de interacciones terciarias. Los intrones del grupo II se escinden mediante un intermediario en forma de lazo (lariát) o lineal, y pueden retrotransponerse (retroblomning o homing) en dianas apropiadas libres de intrón que son reconocidas mediante apareamiento de bases.

RmInt1 es un intrón móvil eficiente encontrado en el interior de la secuencia de inserción ISRm2011-2 de la bacteria simbiótica Sinorhizobium meliloti. Este intrón del grupo II se escinde, in vivo e in vitro, en forma de lazos y de círculos. Sin embargo, hasta la fecha, no se ha dilucidado el proceso completo de splicing in vivo. Con el fin de determinar la eficiencia del splicing de RmInt1, se elaboró un sistema informador basado en el gen lacZ y se realizaron ensayos de Northern Blotting y PCR cuantitativa (Capítulos 1 y 2). La eficiencia de splicing fue calculada en 0.07 ± 0.02 %. Estos resultados sugieren que los intrones bacterianos del grupo II funcionan de un modo más parecido a los retroelementos que a los intrones espliceósómicos.

Estudios previos sobre la reacción de self-splicing de RmInt1 mostraron que esta reacción, caracterizada por la aparición de co-productos no convencionales, es ineficiente. Mediante el análisis de diferentes transcriptos de RmInt1 en diversos contextos y distintas condiciones de reacción, obtuvimos una construcción en la que se eliminó una secuencia parecida a IBS1 presente en los primeros 10 u 11 nucleótidos del exón 3’ de la diana natural localizada en la secuencia de inserción IS2011-2 (Capítulo 4). Esta nueva construcción soslayó la formación de productos truncados e inusuales observados previamente, lo que se tradujo en un incremento de la amplitud de la reacción.

Finalmente, analizamos los efectos de diversas mutaciones puntuales en posiciones conservadas de la ribozima de RmInt1 sobre la eficiencia del primer y segundo paso de la
reacción de splicing in vivo e in vitro (Capítulos 3 y 4). Así, identificamos la existencia de interacciones terciarias comunes a otros intrones del grupo II, tales como γ-γ', EBS3-IBS3 y la interacción no Watson-Crick entre el primer y el penúltimo nucleótidos del intrón. Además, probamos la existencia de motivos estructurales conservados en los intrones de grupo II como la triple hélice entre el dominio V y los nucleótidos de la región interdominios J2/3. Las mutaciones que interrumpen la interacción entre componentes de este motivo resultaron en la alteración de ambos pasos de la reacción de splicing in vivo e in vitro, lo que apoya la existencia de un único sitio activo que cataliza ambos pasos de la reacción de splicing.
Table of contents

Abstract vii
Resumen ix
Table of contents xi
List of figures xv
List of tables xvii

Introduction 1

I.1 General Features 1
I.2 Structural Elements of Group II intron RNA 2
I.3 Intron-Encoded Proteins (IEPs) and Intron-Recruited Maturases 7
I.4 Group II Intron Lineages and Evolution 11
I.5 The Mobility of Group II introns 12
   I.5.1 The homing (or retrohoming) process 12
   I.5.2 The retrotransposition process 14
   I.5.3 Host-encoded proteins for mobility 15
I.6 The Splicing of Group II introns 16
   I.6.1 The chemical mechanism of group II intron catalysis 16
   I.6.2 A single-active site for group II intron catalysis 17
   I.6.3 Definition of targets for activity 18
   I.6.4 Reactions catalyzed by group II introns RNA 20
I.7 Group II Introns as Biotechnological Tools 23
I.8 The Sinorhizobium meliloti RmInt1 Group II Intron 24
   I.8.1 Intron-Encoded Protein (IEP) promoted mobility of RmInt1 intron 25
   I.8.2 Splicing of the RmInt1 intron 27
   I.5.3 Dispersion and evolution of the RmInt1 intro 28

Objectives 33

Materials and Methods 35
   M.1 Bacterial Strains 37
   M.2 Bacterial Cultures 37
M.2.1 Culture media
M.2.2 Antibiotics
M.2.3 Growth conditions and storage
M.3 Plasmid and Cloning Vectors
  M.3.1 Commercial plasmids and vectors
  M.3.2 pICG plasmids
  M.3.3 Other described pICG and pICGTer plasmids
  M.3.4 pLM plasmid series
M.4 Nucleic Acids Extraction
  M.4.1 Plasmid DNA extraction by magnesium salts
  M.4.2 Plasmid DNA extraction by alkaline lysis
  M.4.3 Total RNA extraction
M.5 Cloning and Enzymatic Manipulation of DNA
  M.5.1 DNA digestion with restriction endonucleases
  M.5.2 Conversion of protruding to blunt-ended restriction fragments
  M.5.3 Dephosphorilation of DNA fragments
  M.5.4 Nucleic acids purification
  M.5.5 Ligation of DNA fragments
M.6 Bacterial Transformation and Conjugation
  M.6.1 Competent E.coli cells and transformation
  M.6.2 Plasmid conjugation between E.coli and S.meliloti strains
M.7 DNA Amplification
  M.7.1 The Polymerase Chain Reaction (PCR)
  M.7.2 Quantitative real-time RT-PCR
M.8 Electrophoresis
  M.8.1 Non-denaturing agarose gel electrophoresis
  M.8.2 Denaturing agarose gel electrophoresis
  M.8.3 Denaturing polyacrylamide gel electrophoresis
M.9 Nucleic Acids Hybridization
  M.9.1 DNA-DNA hybridization (Southern Blotting)
  M.9.2 RNA-RNA hybridization (Northern Blotting)
M.10 Mobility Assay (Homing) Of RmInt1 Intron
M.11 β-Galactosidase Assay
M.12 Primer Extension Assay
M.13 In Vitro Splicing Assays
  M.13.1 RNA synthesis and purification
  M.13.2 Self-splicing reaction
Results

Chapter 1: Design of a reporter system for the quantification of exon ligation during the in vivo splicing of RmInt1 intron

R.1.1 Background 69
R.1.2 Design of a genetic assay bases on a luxZ gene reporter system for the evaluation of the RmInt1 group II intron splicing efficiency 69
R.1.3 Introduction of the RmInt1 intron in the reporter system of pICG plasmids 75
R.1.4 Introduction of the Lactococcus lactis II.ltrB intron in the reporter system of pICG plasmids 78
R.1.5 Potential of the pICG reporter system in the study of the biology of S.meliloti 79

Chapter 2: In vivo quantification of the S.meliloti RmInt1 group II intron splicing. Comparison with the L.lactis II.ltrB group IIA intron expressed in E.coli

R.2.1 Background 83
R.2.2 The in vivo excision of RmInt1 84
R.2.3 RmInt1 splicing in S.meliloti 85
R.2.4 Quantitative analysis of RmInt1 expression and splicing 88
R.2.5 Analysis of the L.lactis II.ltrB intron splicing in E.coli and S.meliloti 92
R.2.6 Discussion 94

Chapter 3: Effect of point mutations in conserved positions of the RmInt1 group II intron on the splicing reaction in vivo

R.3.1 Background 101
R.3.2 Map of point mutations introduced in the ribozyme of RmInt1 intron 102
R.3.3 Quantitative analysis of splicing efficiency in wild-type and mutant RmInt1 intron 104
R.3.4 Quantification of the first step of the in vivo splicing in wild-type and mutant RmInt1 intron derivatives 106
R.3.5 Overview of the complete splicing reaction of mutant-derivatives and wild-type RmInt1 ΔORF intron in *S. meliloti* 107
R.3.6 Discussion 110

Chapter 4: Self-splicing of RmInt1 group II intron: a qualitative, quantitative and mutational analysis

R.4.1 Background 117
R.4.2 Constructions for self-splicing assays 120
R.4.3 Optimization of reaction conditions 121
R.4.4 Qualitative analysis of branching and hydrolytic self-splicing pathways of RmInt1 intron 123
R.4.5 Characterization of self-splicing products by DNAzyme cleavage 127
  R.4.5.1 Design of DNAzymes for splicing product analysis 127
  R.4.5.2 Self-splicing product characterization 130
R.4.6 Kinetic analysis of branching and hydrolytic self-splicing pathways with high concentrations of monovalent ion 133
  R.4.6.1 Kinetics of self-splicing in the LT1 transcript 134
  R.4.6.2 Kinetics of self-splicing in the LT2 transcript 138
  R.4.6.3 Kinetics of self-splicing in the ST2 transcript 139
R.4.7 Analysis of point mutants for the study of self-splicing mechanism in the RmInt1 intron 141
R.4.8 Discussion 149

Bibliography 165
List of figures

Introduction

Figure I1: Group II intron RNA structure 11
Figure I2: Folding pathways of ribozymes derived from the ai5γ group II intron under high-salt (top) and near-physiological (bottom) conditions 14
Figure I3: Three-dimensional model of L.1trB-ΔORF Lariat RNA and Ligated exons 15
Figure I4: IEPs, host-encoded proteins and endonucleases 16
Figure I5: Group II intron lineages 20
Figure I6: Models of RT-mediated group II intron mobility 22
Figure I7: The chemical mechanism of catalysis by group II introns 25
Figure I8: Three pathways for group II intron splicing and intron excision 27
Figure I9: Three-dimensional organization of the catalytic core in group IIB introns 30
Figure I10: A model for RmInt1 mobility 34
Figure I11: DNA target site recognition by RmInt1 group II intron 35

Materials and Methods

Figure M1: Construction of broad host range expression vector pICG-E1E2 40
Figure M2: Construction of pICG plasmids from pICG-E1E2 42
Figure M3: Construction of pICG-Ter series plasmids from pICG-E1E2 44
Figure M4: Generation of a standard curve to assess reaction optimization 51
Figure M5: Annealing temperature optimization 52

Results

Chapter 1: Design of a reporter system for the quantification of exon ligation during the *in vivo* splicing of RmInt1 intron

Figure R1.1: β-Galactosidase activity of laiZ gene in pICG constructs carrying the pSyn promoter 70
Figure R1.2: Strategy used in the construction of a genetic assay for the splicing of RmInt1 linked to the expression of β-galactosidase protein 71
Figure R1.3: Constructions used in the *in vivo* splicing analysis of *S.meliloti* RmInt1 and *L.lactis* L.1trB group II introns 73
Figure R1.4: Homing assay of wild-type RmInt1 in the pICG reporter system 75
Chapter 2: In vivo quantification of the S. meliloti RmInt1 group II intron splicing. Comparison with the L. lactis L1.ltrB group IIA intron expressed in E. coli

Figure R2.1: In vivo excision of the RmInt1 intron in S. meliloti and E. coli 85
Figure R2.2: Attempt of detection of ligated exons in the RmInt1 splicing 86
Figure R2.3: RmInt1 splicing products in S. meliloti 87
Figure R2.4: Quantitative analysis of RmInt1 in S. meliloti 90
Figure R2.5: In vivo excision of L1.ltrB intron in E. coli 92
Figure R2.6: Quantitative analysis of L1.ltrB intron in E. coli and S. meliloti hosts 94

Chapter 3: Effect of point mutations in conserved positions of the RmInt1 group II intron on the splicing reaction in vivo

Figure R3.1: Diagram of point mutations and deletions of the ΔORF RmInt1 group II intron flanked by short exons (-20/+5) 103
Figure R3.2: In vivo excision of the wild-type and mutant ΔORF intron in S. meliloti 106
Figure R3.3: Wild-type and mutant RmInt1 ΔORF intron splicing products in S. meliloti 108

Chapter 4: Self-splicing of RmInt1 group II intron: a qualitative, quantitative and mutational analysis

Figure R4.1: Schematic representation of the observed reactions during in vitro self-splicing of group II introns 117
Figure R4.2: Self-splicing of pLM1 (ΔORF RmInt1 construct) at 45°C 119
Figure R4.3: Constructs for RmInt1 self-splicing analysis 121
Figure R4.4: Optimization of RmInt1 self-splicing reaction conditions 122
Figure R4.5: Self-splicing of ΔORF RmInt1 from pLM constructs 124
Figure R4.6: DNAzymes for the analysis of self-splicing products 128
Figure R4.7: Identification of RmInt1 self-splicing products by DNAzyme treatment 131
Figure R4.8: Kinetic analysis of RmInt1 self-splicing under high concentrations of monovalent ions 135
Figure R4.9: Semi-log plots of reaction in LT1, LT2 and ST2 transcripts 138
Figure R4.10: Timecourses of RmInt1 splicing mutants under high salt conditions 142
List of tables

Materials and Methods

Table M1: Bacterial strains used in this work 37
Table M2: Plasmids and vectors used in this work 38
Table M3: Primers used for the construction of mutants in the ribozyme of RmInt1 intron 45
Table M4: Described pICG and pICGTer plasmids used in this work 45
Table M5: Primers used in real-time qRT-PCR 51
Table M6: Real-time PCR assay characteristics 52
Table M7: Probes used in Northern Blotting assay 55
Table M8: Primers used to amplify probe oligonucleotides 56
Table M9: Sequences of DNAzymes used in this work 60

Results

Chapter 2: In vivo quantification of the \textit{S.meliolii} RmInt1 group II intron splicing. Comparison with the \textit{L.lactis} L.ltrB group IIA intron expressed in \textit{E.coli}

Table R2.1: RmInt1 intron splicing measured as β-galactosidase activity 89
Table R2.2: RmInt1 and L.ltrB splicing activity in \textit{S.meliolii} and \textit{E.coli} measured by relative quantification in qRT-PCR 91

Chapter 3: Effect of point mutations in conserved positions of the RmInt1 group II intron on the splicing reaction \textit{in vivo}

Table R3.1: RmInt1 splicing activity in ΔORF and ΔORF mutant constructs expressed in \textit{S.meliolii} 105

Chapter 4: Self-splicing of RmInt1 group II intron: a qualitative, quantitative and mutational analysis
Table R4.1: Rate constants for self-splicing 137
Table R4.2: Rate constants for self-splicing in wild-type and mutant transcripts 146
Table R4.3: Relative self-splicing efficiencies of mutants in the ribozyme of RmInt1intron 148
INTRODUCTION
I.1 GENERAL FEATURES

Group II introns are one of the seven classes of natural ribozymes described to date (RNA molecules with catalytic activity), some of them showing capacity for self-splicing in the form of lariat RNA molecules (Peebles et al., 1986) [for a review, see (Lehmann and Schmidt, 2003)]. They are also retroelements that encode reverse transcriptases (RTs) and can insert themselves into new localizations (Robart and Zimmerly, 2005). For these reasons, group II introns are thought to be the ancestors of nuclear spliceosomal introns and non-Long Terminal Repeat (LTR) retrotransposons (Sharp, 1985; Cech, 1986; Sharp, 1991; Cavalier-Smith, 2009; Koonin, 2009).

The concept of group II intron was created in 1982, when the sequences of several fungal mitochondrial introns were arranged into two unrelated families (Michel et al., 1982). Until 1993, group II introns had been found only in precursor transcripts of essential organelar intron, encoding proteins, transfer RNAs or ribosomal RNAs (Michel et al., 2009). Nevertheless, group II introns were then found in Cyanobacteria and Proteobacteria, two groups containing members considered to be the ancestors of chloroplasts and mitochondria, respectively (Ferat and Michel, 1993) and, subsequently, they were also described in Archeobacteria (Dai and Zimmerly, 2003).

Typically, group II introns consist of a conserved RNA structure and a RT ORF (optional). The RNA structure is organized in six domains, of which domain V is considered an essential part of the catalytic core (Pyle, 2010). The protein is commonly encoded in domain IV and forms RNP (ribonucleoprotein) particles with the RNA, serving as a splicing and mobility factor for the invasion of new DNA target sequences (Lambowitz and Zimmerly, 2010). These properties made group II introns suitable as biotechnological tools in which a group II intron is retargeted to insert into desired DNA by modifying the base pairing sequences in the intron RNA (Guo et al., 2000; Zhuang et al., 2009a). Finally, the RmInt1 intron, which deserves a special attention for being the subject of this work, is a IIB3 member of group II introns, found in the genome of the bacterium Sinorhizobium meliloti (Martínez-Abarca et al., 1998), characterized by a low efficiency splicing but however a high mobility frequency.
1.2 STRUCTURAL ELEMENTS OF GROUP II INTRON RNA

Group II introns are characterized by a conserved secondary structure which varies in size from 100 nt up to about 3000 nt (Lehmann and Schmidt, 2003). The first model was established on the basis of a phylogenetic data comparison, looking for potential base pairings that had been preserved by evolution in spite of divergence in primary sequence (Michel et al., 1982; Michel et al., 1989). In fact, the only strongly conserved sequences are intron boundaries (GUAG in the 5’ exon junction and AG in the 3’ counterpart), resembling those of spliceosomal introns (GU...AG) and few nucleotides dispersed in the rest of the structure (Michel and Ferat, 1995).

It is organized in six domains, DI-DVI, radiating from a central core. The structure is defined as a set of double helices resulted from Watson-Crick and Crick Wobble base pairs (Dunny and McKay, 1999; Lambowitz and Zimmerly, 2004) (Figure II). According to differences in the RNA structure, three subclasses of group II introns: IIA, IIB and IIC, and further subdivisions: A1, A2, B1, B2 and B3 have been defined (See below) (Michel et al., 1989; Toor et al., 2001). The six domains fold into a catalytically active tertiary structure aided by a series of conserved motifs involved in long-range tertiary interactions. Some interactions involve Watson-Crick base pairs (α-α’, β-β’, γ-γ’, δ-δ’, ε-ε’, IBS1-EBS1, IBS2-

![Figure II](image.jpg)

**Figure II.** Group II intron RNA structure. The schematic structure of a IIA1 group II intron, with differences from IIB and IIC subgroups shown in circles. Tertiary interactions (Greek letters, EBSs and IBSs) are indicated in boxes. Adapted from Lambowitz and Zimmerly (2010).
EBS2 and IBS3-EBS3), whereas other are tetraloop-receptor interactions of known geometries ($\zeta$, $\zeta'$, $\eta$, $\eta'$ and $\theta, \theta'$), or other types of less defined non-Watson-Crick interactions ($\lambda, \lambda'$, $\kappa, \kappa'$ and $\mu, \mu'$) (Dai et al., 2008; Pyle, 2010).

Domain I (DI) is the largest domain and it is essential for catalysis, along with domain V (DV) (Fedorova and Zingler, 2007). Four conserved interactions linking DI with catalytically important sequences have been described (Figure 11): $\zeta, \zeta'$ (Costa and Michel, 1995) and $\kappa, \kappa'$ (Boudvillain and Pyle, 1998), which anchors DV binding face into the DI scaffold, so that orienting the latter properly in the active site and; $\varepsilon, \varepsilon'$ (Jacquier and Michel, 1987) and $\lambda, \lambda'$ (Boudvillain et al., 2000), which form together a composite interaction involving DI, the 5'-splice site and the chemical face of DV, close to the catalytic triad, a critical trinucleotide for catalysis in group II introns. DI also establishes contacts with DII and DIII by means of $\theta, \theta'$ (Costa et al., 1997). Domain I is also indispensable for exonic substrate recognition. In classes IIA and IIB of group II introns, the tertiary interaction aimed to join the secondary structure of the intron with the 5'-splice junction is mediated by the base-pairing between the two exon-binding sites 1 and 2 (EBS1 and EBS2), located in two distinct regions in the DI, and the two corresponding intron-binding sites 1 and 2 (IBS1-IBS2), located in the extremity of the 5'-exon, adjacent to each other (Figure 11) (Qin and Pyle, 1998). However, in class IIC introns, usually there is no EBS2/IBS2 interaction, instead recognizing an upstream transcriptional terminator motif important for the 5'-splice site definition (Fedorova and Zingler, 2007). The recognition of the 3'-splice site is mediated by an EBS3/IBS3 interaction in IIB and IIC subclasses, whereas a $\delta, \delta'$ interaction is the counterpart for IIA introns (Costa et al., 2000). Other internal tertiary interactions localized in DI, which are essential for the correct folding and stability of the intron are the $\alpha, \alpha'$ interaction (Michel et al., 1989), the less important $\beta, \beta'$ interaction (Michel and Ferat, 1995) and the $\delta, \delta'$ interaction for members of the class IIIB introns, which facilitates the anchor of DI to the 5'-splice site and to the branching A sited in DVI, through the “coordination loop” (Figure 11) (Hamill and Pyle, 2006).

Domain II (DII) forms two essential long-range tertiary contacts with DI ($\theta, \theta'$) (Costa et al., 1997) and with DV ($\eta, \eta'$) (Chanfreau and Jacquier, 1996) (Figure 11). The $\eta, \eta'$ interaction has been proposed to mediate a rearrangement of DVI between the first and the second splicing step aimed to approximate both ends of the intron and so intervening in the second splicing step efficiency (Lehmann and Schmidt, 2003). To the contrary, the $\theta, \theta'$ interaction seems to be important for an efficient folding of the group II intron ribozyme molecule prior to the initiation of the self-splicing reaction (Costa et al., 1997).
Introducción

The linker region between DII and DIII, also known as J2/3, contains important positions for group II intron catalysis, especially the G588 and A589 (in the mitochondrial α5γ intron of yeasts) proposed to be important in both the first and the second step of splicing (Chanfreau and Jacquier, 1996; Mikheeva et al., 2000) and; the G587, which interacts with the last nucleotide of the intron by means the γ-γ′ (Michel et al., 1989). These positions have been also shown to be close to the active site during catalysis (de Lencastre et al., 2005; de Lencastre and Pyle, 2008).

Domain III (DIII) is considered to function as a catalytic effector, since it enhances the chemical rate of group II-derived ribozyme constructs (Qin and Pyle, 1998; Fedorova and Zingler, 2007). DIII establishes a tertiary interaction with DV, the μ-μ′ (Fedorova and Pyle, 2005) whereas the A-rich bulge in DIII also interacts, and potentially stabilizes the ε-ε′ motif, bringing it closer to DV (Fedorova and Pyle, 2008).

Domain IV (DIV) is the region, in some group II introns, where it is located an open reading frame (ORF). Proteins codified by ORFs, also known as intron-encoded proteins (IEP), assist the intron during the folding and splicing in vivo due to its maturase activity and permit the invasion of DNA by its retrotransposase (RT) activity (Fedorova and Zingler, 2007). DIV also contains the primary binding site for these proteins (Wank et al., 1999) and for other host-encoded splicing factor factors (Ostertreter et al., 2005; Fedorova and Zingler, 2007).

Domain V (DV) is, together with DI, absolutely required for catalysis in group II introns. It is divided in a “binding site”, serving as a docking for other components of the intron interacting with DV and; a “chemical side”, in which it is located the catalytic functionalities (Fedorova and Zingler, 2007). Interactions involving the binding site are the above described ζ-ζ′ and κ-κ′ whereas interactions involving the chemical face are the also mentioned λ-λ′ and μ-μ′. It contains the catalytic triad, AGC, in the base of DV, in which G is an invariant residue critical for both in vivo and in vitro splicing (Koch et al., 1992; Boulanger et al., 1995; Peebles et al., 1995). Another important catalytic motif is the bulge AC of DV. It has shown that the bulge AC establishes tertiary contacts with a conserve G (G588 in α5γ intron) in the linker region J2/3 and with the fifth nucleotide of the intron (l position), bring them together during catalysis (de Lencastre et al., 2005; de Lencastre and Pyle, 2008; Michel et al., 2009). It is also known that the AC bulge binds two metal ions within the core, which includes one with the AGC triad and another with the G288 (in the α5γ intron) of J2/3, forming a metal ion binding platform that participates directly in chemical catalysis (Pyle, 2002; Toor et al., 2009). The metal binding site is supported by a triple helical structure that results from binding of the conserved J2/3 region to an
Figure 12. Folding pathways of ribozymes derived from the a5γ group II intron under high-salt (top) and near-physiological (bottom) conditions. Adapted from Fedorova and Zingler (2007).

invariant region within the major groove of the DV lower stem (Keating et al., 2010; Pyle, 2010).

Domain VI (DVI) contains a highly conserved bulged adenosine, which constitutes the branch point for lariat formation during splicing (Qin and Pyle, 1998). Other structural constraints determining the bulging point are the G-U wobble pairs flanking the bulged A, the length of the linker region between DV and DVI and the 4 bp basal stem of DVI (Qin and Pyle, 1998; Chu et al., 2001; Fedorova and Zingler, 2007). DVI also takes part of the long-range interaction η-η′, mentioned above, between the terminal loop of DVI and an internal helix of DII, reported to be important for transesterification at the 3′-splice site (Lehmann and Schmidt, 2003).

The six domains of group II introns need to fold into a correct native conformation in order to carry out the catalytic activity (Fedorova and Zingler, 2007). It has shown that this process occur in vitro through an apparent two-step mechanism, in which the folding of D1 is a rate-limiting step and it constitutes an on-pathway intermediate (Swisher et al., 2002; Su et al., 2005; Waldsich and Pyle, 2008)(Figure 12) Once D1 has folded, catalytic Domains 3, 5 and 6 rapidly dock into respective receptor sites within the D1 scaffold, thereby completing the assembly of the intron (Pyle et al., 2007).
Under near-physiological conditions (30 ºC and low Mg$^{2+}$ concentration) a slow compact on-pathway folding intermediate (the near-native state) is formed (Fedorova et al., 2007). Thus, the change to native state and the stabilization of the structure is achieved by the increment of Mg$^{2+}$ concentration or the addition of a protein co-factor both in vitro (Dai et al., 2008; Fedorova et al., 2010) and in vivo (Liebeg et al., 2010) (Figure 12).

Once the intron has folded up, the DV is found between the two halves of DI, joining them together by α-α’ and β-β’ tertiary interactions on one side, and with the DV, through the χ-χ’ and ζ-ζ’, on the other side (Dai et al., 2008) (Figure 13). DIII is internalized, with contacts to DV and both halves of DI. The catalytic face of DV is oriented inward, where it interacts with the catalytically important elements J2/3, γ-γ’, ε-ε’ and the branch A. While the exon 5’ binds EBS1 and EBS2 in the surface of the ribozyme, the exon junction penetrates internally into the active site (Dai et al., 2008).

Group II introns show several types of structural variations. In organellar introns, it is frequently found degenerate introns, which show mispairs, insertions and deletions in DV and DVI or the absence of the bulged A (Lambowitz and Zimmerly, 2004). It is thought that the splicing of these introns require trans-acting RNAs or adjuvant proteins supplying the absence of RNA structures (Lambowitz and Zimmerly, 2004). Twintrons are

![Figure 13](image)

**Figure 13.** Three-dimensional model of L1litrB-ΔORF Lariat RNA and Ligated exons. **A** Schematic of the secondary structure of the L1litrB intron highlighting important tertiary interactions for the compaction of the intron structure. **B** Three-dimensional model of intron domains after folding of L1litrB intron. Domains are colored as in A. Adapted from Dai et al (2008).
structures in which one intron has inserted into another; thus, if the outer intron interrupts essential structures of the inner intron, the former must be spliced first (Lambowitz and Zimmerly, 2010). *Trans*-splicing group II introns occurs when an intron has been split into two or more separate segments which must reassociate via tertiary interactions between domains in order to splice (Lambowitz and Zimmerly, 2010).

1.3 INTRON-ENCODED PROTEINS (IEPs) AND INTRON-RECRUITED MATURASES

Intron-encoded proteins (IEPs) are those proteins that have been acquired by introns, assisting them during its *in vivo* activity. They are about 2 kb in size and are, generally, encoded within intron DIV, outside the catalytic core (Lehmann and Schmidt, 2003). Functions associated with IEPs include RNA folding, acting as maturases that help the intron fold into a catalytically active structure; *in vivo* splicing, by stabilizing the active
Figure 14. IEPs, host-encoded proteins and retrotransposons. A LtrA protein encoded by the *Lactobacillus lactis* intron. B IEP encoded by the *Sinorhizobium meliloti* RmlInt1 intron, which lacks the En domain. The “class D motif” is a conserved sequence that is required for splicing and mobility functions in lineage D IEPs. C Matk protein encoded by the *Arabidopsis thaliana* trnKII intron. This protein only retains block RT5-RT7 of the RT domain and lack the En domain. D nMat-1a protein encoded by a nuclear gene in *Arabidopsis thaliana*. nMat-1 protein contains complete RT and X domains, but have mutations expected to inhibit RT activity. E LAGLIDAGL protein encoded by *Cryptococcus parapsitica* rnr inton, which belongs to a familia of endonucleases typical of group I introns. F HIV-1 RT. Insertions between RT sequence block are denoted 2a, 3a, 4a and 7a. The locations of the predicted α-helices of characteristic thumb domains are shown above domain X. Adapted from Lambowitz and Zimmerly (2010).

structure and; retrohoming and retrotransposition through target-primed reverse transcription to insert introns into new DNA target genes (Fedorova and Zingler, 2007). All characterized IEPs are intron-specific factors, although closely related IEP’s may have some cross-reactivity (Lambowitz and Zimmerly, 2004).

Group II intron IEPs consist of several domains: RT, reverse transcriptase; X, maturase; D, DNA binding and; En, endonuclease (Figure 14). The RT domain is comprised of seven conserved blocks common to all retroelements, which correspond to the palm and finger regions of the HIV-RT (Figure 14, F) (Lehmann and Schmidt, 2003). The RT5 contains the highly conserved sequence YADD, which is part of the RT active site (Lambowitz and Zimmerly, 2010). Moreover, group II introns and non-LTR-retrotransposons also harbor an amino-terminal extension (RT0) and insertions (Figure 14 A, light red blocks) between the RT sequence subdomains (Lambowitz and Zimmerly, 2010). The domain X is also known as “maturase” due to its essential role in the folding and splicing of the intron ribozyme. It is analogous to the thumb motif of HIV-RT, comprised of conserve sequences and three α-helices (Lambowitz and Zimmerly, 2010). Many IEPs, especially those of organellar intron, have lost some RT blocks required for retrotransposasase activity, as that of *Arabidopsis thaliana* trnKII intron (Figure 14 C) or have accumulate mutations that also impair the RT activity, as the nuclear-encoded nMat proteins (Figure 14 D), both continuing to function in RNA splicing (Bonen and Vogel, 2001). The C-terminal region of IEPs is comprised of domains D and En. D is a DNA binding domain, characterized by the presence of a pair of cysteine residues which are part of a major class of zinc fingers and have been also proposed to play a critical role in maintaining the structure of the En domain, while a short region upstream would be responsible for DNA-binding (San Filippo and Lambowitz, 2002). The endonuclease domain (En) cleaves a target DNA strand to generate the primer for reverse transcription
(Lambowitz et al., 2005; Marcaida et al., 2010). It usually belongs to the H-N-H family, which contains a tight bound Mg$^{2+}$ ion (San Filippo and Lambowitz, 2002). However, some fungal mitochondrial group II introns encode IEPs belonging to a family of DNA endonucleases with the conserved motif LAGLIDADG, typical of group I introns (Figure 14 E) (Toor and Zimmerly, 2002; Mullineux et al., 2010). Many group II introns, particularly in bacteria, encode proteins lacking the En domain, and at least some are mobile, using alternate mechanisms for priming reverse transcription, as the RmInt1 intron of Sinorhizobium meliloti (Figure 14, B) (see above).

The IEP interacts with the intron RNA, producing ribonucleoprotein (RNP) particles. This interaction begins when the IEP binds unspliced precursor RNA, promoting its folding and splicing by stabilizing the catalytically active RNA structure (Lambowitz and Zimmerly, 2004). It is known, from studies on the bacterial LtrA protein coded in the L.ltrB intron form Lactococcus lactis that the IEP binds the intron in the form of a dimer, similar to HIV-1 and other RTs (Saldanha et al., 1999; Rambo and Doudna, 2004). The IEP has a high-affinity binding site in the intron subdomain DVIA, a small stem-loop structure at the beginning of DIV that contains the Shine–Dalgarno (SD) and initiation codon at the beginning of the LtrA ORF (Figure 11), and makes additional contacts with conserved catalytic core regions, including parts of DI, DII, and DV1, which stabilize the active RNA structure (Lambowitz et al., 2005). In the IEP, two regions have been identified to be required for the binding to the intron RNA: one corresponding to the N-terminal extension up to RT-0, which specifically binds the DIVa region and, a second, extending from RT-0 to RT-2 (Gu et al., 2010). Another region, which extends to RT-3 to RT-4, seems to stabilize the protein or assist it during its folding, also contributing to splicing by binding other regions of the intron RNA in a DIVa-dependent manner (Gu et al., 2010). Other regions, such as that extending from RT-2 to RT-3 and the X domain are presumably required for LtrA binding to core regions of the intron RNA (Gu et al., 2010).

Some group II introns, especially those from mitochondria and chloroplast, lack ORFs and rely on host-encoded proteins to promote their splicing and homing processes and on chaperones to resolve stable inactive or intermediate structures that limit the rate of RNA folding (Lambowitz and Zimmerly, 2010). In general, group II intron splicing factor are proteins with additional cellular functions which can support splicing directly, interacting with specific regions of the RNA or indirectly, if they are involved in a generalized splicing mechanism (Solem et al., 2009).
Host-encoded proteins known to assist chloroplast group II introns directly belong to a few families of proteins based on RNA-binding domains they contain: CRM (chloroplast RNA splicing and ribosome maturation), PPR (pentatricopeptide repeat), PORR (plant RNA recognition), OPR, ψ-uridine synthase, PTH (peptidyl-rRNA hydrolases), ribonuclease III or NAD binding (Lambowitz and Zimmerly, 2010). In these cases splicing factors play accessory roles without participating in catalysis directly. They may either stabilize the structure of the active intron ribozyme or assist in its proper folding to form the catalytically structure and avoid the formation of competing inactive folds (Stern et al., 2010).

In mitochondrial group II introns, several host-encoded factors have been shown to participate in the splicing process: the MSS proteins (mitochondrial splicing system), specially Ms51, which is required for the splicing of the yeast a15γ intron; MRF proteins (mitochondrial peptide chain release factor), which also participate in the termination of translation of mitochondrial proteins; NAM proteins (nuclear accommodation of mitochondria), some members of which have been shown to function in the maintenance of the stability of the intron-containing transcripts and; MRS (mitochondrial RNA splicing), which plays an essential role in mitochondrial Mg²⁺ homeostasis (Lehmann and Schmidt, 2003).

In addition to splicing factors, some host-encoded chaperones have been showed to be crucial for the correct RNA folding and splicing. Thus, the DEAD-box protein Mss116 from Saccharomyces cerevisiae is essential for the in vivo splicing of the group II intron a15γ and sufficient alone to stimulate its self-splicing in vitro under near-physiological conditions (Fedorova and Zingler, 2007). Mss116 can be replaced in vivo and in vitro by other DEAD-box proteins that function in mitochondrial group I intron splicing, such as CYT-19 from Neurospora crassa, or by other DEAD-box proteins which do not usually function in splicing, such as Ded1p from S.cerevisiae or SmB from E.coli (Lambowitz and Zimmerly, 2010). DEAD-box proteins can assist RNA folding through three major ways in an ATP-dependent manner: facilitating the collapse of the RNA, preventing or resolving misfolded structures or by stabilizing correctly folded structures (Solem et al., 2009). Finally, it has been proposed that DEAD-box proteins can promote splicing in several ways: from the resolution of misfolded structures through its helicase activity to the stabilization of a correctly folded structure, both of them dependent on the ATP hydrolysis (Solem et al., 2009).
I.4 GROUP II INTRON LINEAGES AND EVOLUTION

IEPs bind specifically to the intron RNA that encodes them. For that reason it is though that ribozymes and IEPs have coevolved, forming phylogenetic lineages (Toor et al., 2001; Lambowitz and Zimmerly, 2010).

Phylogenetic analyses have identified eight lineages of group II introns IEPs: bacterial classes A-F, ML (mitochondrial-like) and CL (chloroplast-like) (Figure 15). Classes B, ML and CL all contain an En domain in their IEP, whereas classes C, D, E and F do not (Simon et al., 2009). Each IEP lineage is associated with a RNA subgroup: ML with IIA, bacterial class C with IIC, CL with IIB1 and IIB2, bacterial class D with IIB3, bacterial class E with IIB5 and the remainder (bacterial A, B and F) with less typical IIB RNAs, (Toro et al., 2007; Lambowitz and Zimmerly, 2010).

According to the RNA phylogenetic studies, IIA (mitochondrial-like) and IIC (class C) intron RNA structures are monophyletic groups, while IIB introns do not (chloroplast-like and bacterial classes) (Simon et al., 2009). Moreover, according to the coevolution of the RNA and ORF, the IIA ribozyme structure has a relative recent origin compared with IIB, whereas the IIC RNAs would be the oldest (Toro et al., 2007; Simon et al., 2009).

The general correlation observed support the retroelement ancestor hypothesis, which predicts that the major RNA structural forms of group II introns developed through coevolution with the intron-encoded protein rather than as independent catalytic

Figure 15. Group II intron lineages. The major lineages of group II intron IEPs are shown as blue sector. Sublineages are shown as darker blue sector within the major lineages. RNA structural subgroups that correspond to IEP lineages are shown in magenta. All group II intron IEP lineages and RNA types are found in bacteria. IEP lineages and RNA types also found in organelles are delineated in green (outer circle). Adapted from Lambowitz and Zimmerly (2010).
RNAs, and that most ORF-less introns are derivatives of ORF-containing introns (Toor et al., 2001). However, CL subclasses of ORFs (CL1a, CL1B, CL2A and CL2B), which do not strictly correspond to the two RNA structural subtypes IIB1 and IIB2, are exception to this rule (Simon et al., 2009).

I.5 THE MOBILITY OF GROUP II INTRONS

Group II introns are mobile elements that can spread within a genome or invade new genomes (Bonen and Vogel, 2001). The general mechanism consists on a process by which the intron RNA in the RNP particles uses its ribozyme activity to reverse splice into a DNA target site and is then reverse transcribed by the IEP (Lambowitz et al., 2005). Two strategies are exploited for the movement, varying if the movement occurs into homologous sites (homing or retrohoming) or into ectopic sites (retrotransposition).

I.5.1 The homing (or retrohoming) process

Homing is initiated by the recognition of the DNA target site by the RNP. Target sites required for efficient homing were found to comprise 31 bp from positions -21 to +10 for yeast intron a12, 35 bp from pos. -26 to +9 for the L.lactis L.ltrB intron and 25 bp from pos. -20 to +5 for the S.meliloti RmInt1 (Jiménez-Zurdo et al., 2003; Lehmann and Schmidt, 2003). In the well-studied L.ltrB intron, the target sequence is mainly recognized by base-pairing with the RNA through the EBS2-IBS2 (-12 to -8 positions), EBS1-IBS1 (-6 to -1 positions) and δ-δ' (+1 to +3 positions) interactions (Toro et al., 2007). Most of the interactions with the LtrA protein are in the distal 5'-exon region of the DNA target site (positions -23 to -14), with includes the critical bases T-23, G-21 and a-20.

Figure I6. Models of RT-mediated group II intron mobility. Schematic showing RNP composed of excised intron lariat RNA (orange line) and intron-encoded protein (grey oval), along with several possible mobility pathways. Exon sequences are indicated by dark green lines for donor gene, light green lines for homologous recipient gene and blue lines for ectopic site. After transcription of the donor gene and RNP assembly, movement can be into a homologous site by retrohoming (A, B, C) or into an ectopic site by retrotransposition (D). Retrohoming follow different pathways for mitochondrial (A), bacterial with En domain (B) or bacterial without En domain (C) group II intron. Adapted from Bonen and Vogel (2001).
The model for *L. LtrB* retrohoming, through target DNA-primed reverse transcription (TPRT) is shown in the Figure 16 B. The process initiates when the lariat intron reverse-splices in the sense strand of the target site complete reverse splicing (CRS) (Cousineau et al., 1998; Edgell et al., 2000). The antisense strand is then recognized by the LtrA protein, although bases flanking the T+5 appear to make smaller contributions (Lambowitz et al., 2005). The antisense strand is then cleaved by the endonuclease activity of the LtrA between positions +9 and +10, and the 3’-end of the cleaved antisense strand is then used as a primer for reverse transcription of the reverse-spliced intron RNA.
(Cousineau et al., 1998). After digestion of the RNA template, second strand synthesis could be primed by the 3'-end of the 5'-exon. Ligation of both strands completes the retrohoming event, which is RecA independent and occurs without co-conversion of the donor exons (Cousineau et al., 1998). By contrast, retrohoming in yeasts is accompanied of co-conversion of upstream exon sequences, even the retrohoming initiates via complete or partial reverse splicing (PRS) (Figure 6 A) (Eickbush, 1999).

In group II introns lacking the endonuclease domain in the IEP, as the bacterial RmInt1 intron of Sinorhizobium meliloti, an En-independent retrohoming mechanism is used (Martínez-Abarca et al., 2000) (see Section IX, below). In LLItrB, this mechanism is also found in conditions in which second-strand cleavage is blocked by mutations in the endonuclease domain or DNA target site (Lambowitz and Zimmerly, 2004). Such mutations decreased the mobility frequency, which showed a replication dependence and strand bias, suggesting that a nascent leading strand at a DNA replication fork was used to primer reverse transcription (Lambowitz and Zimmerly, 2004). The preference for leading strands could reflect that, after reverse-splicing into double-stranded DNA, the intron RNP is positioned to use a leading strand primer directly (Lambowitz and Zimmerly, 2004). In LLItrB intron retrohoming, some mobility also occurred in the opposite strand, but a lower frequency, whereas natural occurring group II intron that lack the En domain are preferentially found in the lagging strand (Lambowitz and Zimmerly, 2004). The latter bias could reflect that En- introns mainly reverse splice into single-stranded DNA after the passage of the replication fork, using then the lagging strand primers (Lambowitz and Zimmerly, 2004).

**1.5.2 The retrotransposition process**

Group II introns can also invade ectopic sites in their natural hosts at a low frequency (typically 10^4 to 10^5), by a retrotransposition mechanism (Lambowitz and Zimmerly, 2004). The retrotransposition pathway studied in the bacterial LLItrB intron occurs preferentially via DNA target site in the absence of homologous DNA recombination (Ichyanagi et al., 2002). Moreover, this process does not produce co-conversion of the 5'-exon, as observed in retrohoming (Figure 6). The IBS1 and 8' (for IIA introns)/IBS3 (for IIB introns) sequences in the target DNA match well the intron, whereas IBS2 shows only weak conservation (Cousineau et al., 2000). Moreover, the positions -23 to -13 and +5 to +16, which are recognized by the LtrA moiety of the RNP particle, are only weakly conserved in the retrotransposition sites, which is consistent with a majority retrotransposition process independent of endonuclease activity (Cousineau et
In this process, the intron reverse-splices into transiently single-stranded DNA at a replication fork, followed by priming from a nascent lagging DNA strand. However, in *E. coli*, insertion occurs preferentially by an En-dependent manner, indicating that retrotransposition may be largely influenced by the host (Coros et al., 2005).

### 1.5.3 Host-encoded proteins for mobility

Mobility of group II introns is influenced by cellular interactions, host factors and stress responses. In the initial steps of retrohoming, the group II intron recruits multiple host functions and enters the cellular replication and repair pathways (Smith et al., 2005; Lambowitz and Zimmerly, 2010). Thus, after reverse transcription, the intron RNA must be removed for the synthesis of the second DNA strand using the cDNA as template, a process in which the *E. coli* RNAse H1 (*rmbA*) and the 5’-3’ endonuclease activity of Pol I (*polA*) are involved (Smith et al., 2005; Beauregard et al., 2008). During the synthesis of the second DNA strand, it has been shown that the α-subunit of the polymerase Pol III and 3’-5’ exonuclease ε proof-reading subunit of Pol III (*dnaQ*) are required in *E. coli*, helping the lower processive intron-encoded RT (Smith et al., 2005; Beauregard et al., 2008). In addition, the repair polymerases, Pol II (*polB*), Pol IV (*dinB*), and Pol V (*umuDC*) have a stimulatory effect on retrohoming in *E. coli*, participating in the repair of nicks on the DNA or in the DNA-RNA junction (Smith et al., 2005; Beauregard et al., 2008). 5’-3’ ssDNA exonuclease RecJ is thought to participate resection of the 5’ exon DNA, to allow the nascent cDNA to pair with the top strand, while the *E. coli* DNA ligase is required to seal the nicks produced after retrohoming (Beauregard et al., 2008). Among factor inhibiting the retrohoming, both RNases (RNase E, RNase I and Exo III) and DNsases are included, which may represent a cellular response against the group II intron mobility.

The retrohoming is also influenced by global regulators and environmental factor. Two factors described, guanosine tetraphosphate (ppGpp) and cAMP, are of special interest (Beauregard et al., 2008; Coros et al., 2009). Polyphosphates accumulate when cells undergo amino acid starvation, while cAMP is activated under carbon source starvation, suggesting that these kinds of nutritional stresses can activate intron mobility (Coros et al., 2009). In both cases, retrotransposition to ectopic sites was favored over retrohoming.
I.6 THE SPLICING OF GROUP II INTRONS

The splicing of group II introns was first described in 1980 in the mitochondria of yeast (Arnberg et al., 1980; Halbreich et al., 1980), when it was realized that group II introns accumulated as circular RNA species. Later, a self-splicing reaction was developed for the a15γ and b11 introns from the mitochondrial genome of the yeast *Saccharomyces cerevisiae*, showing that group II introns were able to splice autocatalytically in the form of RNA lariats, resembling those of spliceosomal introns (Peebles et al., 1986; Schmelzer and Schweyen, 1986; van der Veen et al., 1986). These latter experiments, which required unphysiologically high salt concentrations and temperatures, provided highlights on group II intron catalytic activity (Fedorova and Zingler, 2007).

I.6.1 The chemical mechanism of group II intron catalysis

The basic chemical reaction catalyzed by group II introns is the in-line, nucleophilic attack of an alcohol moiety or a water molecule on an activated phosphodiester linkage, resulting in release of a 3'-hydroxyl leaving group and a 5'-phosphate or new phosphodiester linkage (Figure 17) (Pyle, 2010).

![Figure 17](image-url) The chemical mechanism of catalysis by group II introns. Nuc can represent the 2'-OH group of the branch-site (for the first splicing step), the 3'-OH group of the 5'-exon (for the second splicing step) or the terminus of the intron (during circle formation), and it can also represent water (for the hydrolytic pathway). Note that the 2'-group on the departing sugar moiety can be a 2'-hydroxyl group (for RNA) or a 2'-hydrogen (for DNA), since both can be substrates of the intron RNA. Metal ions are indicated with red balls. Adapted from Pyle (2010).
Metal divalent ions are essential for the catalytic mechanism of group II introns. As protein phosphoester transferase enzymes, group II intron contain two metal ions within their active sites, helping to position the reactants, activate the nucleophile and stabilize the charge on the leaving group and phosphoryl group in the transition state (Gordon et al., 2007). In particular, it has been seen that one Mg$^{2+}$ ion interacts with the 3’ oxyanion leaving group during the first and second splicing steps, while a second one interacts with the nucleophile during the second exon-ligation step (Figure 17) (Gordon et al., 2000; 2007). This is consistent with the group II intron crystal structure, which reveals that the catalytic triad and the bulge of DV, composing the active site, are proximal and coordinate two divalent metal ions to the scissile phosphate and the exon junction (Toor et al., 2009). Additionally, it has been observed other metal ions within the core, including one that binds the G288 (of the ai5γ intron) of the J2/3 region and the bulge in DV, helping to link the active sites of the catalytic core (Toor et al., 2009).

### I.6.2 A single active-site for group II intron catalysis

The active-site of group II introns are composed of several elements, in addition to the well-known DV, which play important roles in catalysis and are positioned appropriately. These are the bulge and AGC triad regions of DV, the J2/3 linker region, the ε-ε’ nucleotides and the coordination loop in DI (de Lencastre et al., 2005; de Lencastre and Pyle, 2008). All of them are interconnected through long-range interactions, creating a network of tertiary contacts which compose the core of the catalytic properties (Figure I8).

It has been a major interest in elucidating the presence of either a unique or two active-sites catalyzing the first and second steps of the splicing reaction, which has been related with the precise identification of elements composing the catalytic core of group II introns.

It has been shown that certain functional groups at conserved positions of DV, such as the triad AGC in the base of the domain, are required for catalysis of both splicing steps, indicating the presence of a single active-site (Chanfreau and Jacquier, 1994). If this is true, then the active-site should rearrange its components to accommodate the successive substrates. In favor of the latter statement, it was shown the formation of the γ-γ’ tertiary contact between DII and DVI was involved in a conformational change occurring before the second splicing step aimed to the displacement of the first-step product out of the active site (Chanfreau and Jacquier, 1996). More recently, it has been demonstrated that all reactants important for both steps of splicing are proximal before
the first step (de Lencastre et al., 2005). Thus, the branch site, both exons, the catalytically essential regions of DV and J2/3, and ε-ε’ are in close proximity before the first step of splicing occurs, which indicates the existence of a single active-site (de Lencastre et al., 2005).

1.6.3 Definition of targets for activity

There are three targets or “reactants” that must be brought together before the initiation of the splicing reaction: the 5’ and 3’-splice site (the substrates) and the branch-site (the nucleophile in the first step) (Pyle, 2008).

The 5’ exon and the adjacent 5’ splice site are mainly determined by the EBS-IBS interactions (EBS1-IBS1 and EBS2-IBS2), in which the sequences only need to co-vary and maintain the base-pairing, no existing sequence constraints. Then, the intron core specifically targets the phosphodiester junction between single and double stranded RNA (Pyle, 2008). In addition, the ε-ε’ interaction, which occurs downstream of the 5’ splice
junction, acts together with the EBS-IBS pairings to position correctly exon 1 and activate the appropriate phosphodiester bond (Jacquier and Jacquesson-Breuleux, 1991). These requirements are also true for the reverse splicing and the rest of reactions catalyzed by group II introns, in which the EBS1-IBS1 interaction has a major role for the definition of the 5’-splice site (Jacquier and Jacquesson-Breuleux, 1991; Michel and Ferat, 1995). By contrast, in group IIC introns, the 5’-splice site definition occurs through the EBS1-IBS1 interaction along with the recognition of transcription terminator stem-loop structures after which these group II introns insert (Toor et al., 2006).

The 3’-splice site is determined by the contribution of several structural components. The γ-γ’ interaction is important for the efficiency and precision of 3’-cleavage, particularly when some flexibility is introduced by the deletion of DVI (Jacquier and Jacquesson-Breuleux, 1991). It has also been observed that cryptic splice sites generated downstream in the 3’ exon were at positions that allow the restoration of the stable wild-type pairing (Schmidt et al., 1993). Moreover, in the group II intron B.a12 from Bacillus anthracis, an alternative splicing site is defined as a result of the presence of naturally occurring atypical structural motifs, in which the 3’ splice site can be shifted as far away as four nucleotides from the wild-type site, thus restoring missing γ-γ’ and EBS3-IBS3 pairings (Robart et al., 2004). Another factor involved in the selection of the 3’-splice site is the length in the joining segment between DV and DVI [J(56)3], in which lengthening of the J(56)3 segment results in a reduced rate but with a correct branching, while shortening of the same produces an inefficient and inaccurate second splicing step (Boulanger et al., 1996). In addition, the replacement of the first nucleotide of the 3’ exon can result in a loss of fidelity and a downstream 3’-splice site selection (Su et al., 2001), whereas mismatches incorporated at the EBS1/IBS1 terminus lead to a truncated interaction and a cleavage site occurring upstream the regular 3’ splice site (Su et al., 2001).

The two exon ends are brought together via three tertiary interactions: δ-δ’, EBS1-IBS1 and EBS3-IBS3. While the δ’ position is close to the EBS1 sequence in a loop of the D1 stem, the δ position is located in the coordination loop (see below) along with the EBS3 sequence. This tripartite interaction allows the approximation of D1 to each of the intron flanking exons through the EBS-IBS interactions whereas the δ-δ’ join both regions, and so both exon ends (Costa et al., 2000).

The branch site, acting as the nucleophile in the splicing reaction, has been showed to be composed by the bulged adenosine along with the noncanonical pairs that flank the nucleophilic A and the polypurine sequence that anneals on the opposite strand (Chu et al., 2001). The branch-site receptor for group IIB introns is found in a stem-loop structure
within intron DI, also containing nucleotides that participate in the δ-δ’ and EBS3-IBS3 pairings that contribute to specification of both splice sites (Hamill and Pyle, 2006). The branch-site receptor, also known as coordination loop, positionates the nucleophilic adenosine to react with the 5’-splice site. Moreover, ζζ’ and κ-κ’ interactions, which are separated from the coordination loop by a short helical stem of 6 base pairs, positionate the branch-site receptor close to major active-site components located in DV. This indicates that DV and branch-site receptor function together to place the nucleophile in the correct orientation for promoting catalysis (Hamill and Pyle, 2006). Finally, since branch-site receptor sequences are located next to EBS3-IBS3 and δ-δ’ interactions, the branch site is brought into close proximity with both of the phosphodiester linkages that are cleaved during self-splicing (Hamill and Pyle, 2006).

I.6.4. Reactions catalyzed by group II intron RNAs

By employing the chemical reaction described above, group II intron can catalyze different reactions following diverse pathways (Figure I9):

a) Branching pathway

Group II intron splicing proceeds through two sequential transesterification reactions. In the first step, the 2’ OH of the bulged adenosine in domain VI act as the nucleophile to attack the 5’-splice site, producing an intron lariat/3’-exon intermediate. In the second step, the 3’ OH of the cleaved 5’ exon, which remains tightly bound to the intron via IBS1 and IBS2 base-pairing interactions, is the nucleophile to attack the 3’-splice site, resulting in exon ligation and excision of an intron lariat RNA (Figure I9 A) (Lambowitz and Zimmerly, 2010). In group II intron self-splicing, the first step is typically rate-limiting, and the lariat/3’-exon intermediate is only detected under conditions inhibiting the second step (Pyle and Lambowitz, 2006).

Elimination of the bulging A leads to a predominance of linear intron forms (see below), although the global splicing efficiency is not affected (Chu et al., 1998). Despite its importance, it seems that not only the bulging A is behind the branching process efficiency but lariat formation is in the base of several constraints, promoted by different substructures in RNA and DNA (Chu et al., 1998). Thus, G-U base-pairings surrounding the bulging A also contribute to the splicing efficiency (Chu et al., 1998), while the length of the segment joining DV to DVI has been reported to affect both the splicing and the accuracy of the 3’ splice site (Boulanger et al., 1996).
**Figure 19.** Three pathways for group II intron splicing and intron excision. **A** The branching pathway. **B** The hydrolytic pathway. **C** A possible circle formation pathway. 5'-exons are shown in blue and 3'-exons are shown in yellow. The intron is indicated by a thin black line. Adapted from Pyle (2010).

**b) Hydrolytic splicing**

This pathway involves the attack of a molecule of water as the nucleophile during the first step of splicing, while the second step proceeds as in the branching pathway, being the products of this reaction ligated exons and a linear intron (Figure 19 B) (Fedorova and Zingler, 2007; Pyle, 2010).

Although the branching pathway is more efficient in most in vitro studies, the hydrolytic route becomes dominant in the presence of KCl buffer (Jarrell et al., 1988; Daniels et al., 1996) and when the branch point is inactivated or deleted (Michel and Ferat, 1995; Pyle and Lambowitz, 2006). The hydrolytic pathway has been also observed in vivo for mutants blocking the branching pathway (Podar et al., 1998) and for some chloroplast group II introns lacking the naturally occurring bulging A in domain VI (Vogel and Borner, 2002).

**c) Excision as circles**

Group II introns can also be excised as a true circle in vitro (Murray et al., 2001; Molina-Sánchez et al., 2006). The circle is created by the formation of a 2'-5'
phosphodiester bond that links the first residue of the intron to the last residue (Murray et al., 2001). The 2'-5' linkage is generated when the 2' hydroxyl of the terminal nucleotide of the intron attacks the phosphodiester bond located at the 5' splice site (Murray et al., 2001).

It has been suggested that circle formation is initiated by the attack of free 5'- exon molecules, which are the product of a spliced-exon reopening (SER) (Jarrell et al., 1988). SER is a side-reaction catalyzed by linear and lariat group II introns in which the ligated exons are hydrolytically cleaved into free 5’ and 3’ exons (Figure I9 C)(Pyle, 2010).

d) **Trans-splicing**

Some organelle group II introns from chloroplasts and mitochondria of lower eukaryotes and plants are discontinuous, consisting of two or more segments encoded in different parts of the genome. Transcripts of these segments reassociate in vivo via tertiary interactions to form a functional intron, resulting in “trans-splicing” of the associated exons (Lambowitz and Zimmerly, 2004). Most fragmented group II introns are interrupted within DIV, with the IEP located either upstream or downstream the fragmentation site and some examples of fragmentation within DIII and tripartite group II introns have been also found (Lambowitz and Zimmerly, 2004).

Despite no discontinuous group II introns have been found in Prokaryotes, the Ll.LtrB intron from *L.lactis* shows the in vivo ability to splice in trans in a process dependent of the intron-encoded LtrA protein when fragmented at locations of various naturally occurring fragmentation sites (Belhocine et al., 2007) or other locations not found in the nature (Belhocine et al., 2008).

e) **Reverse splicing**

The splicing reaction can occur in the forward and reverse directions (Augustin et al., 1990; Morl and Schmelzer, 1990). The reverse reaction is initiated by binding the free intron lariat via the EBS elements (EBS1, EBS2 and EBS3) located in DI with their counterparts IBSs (IBS1, IBS2 and IBS3) in the 5’ exon region next to the exon ligation site (Lehmann and Schmidt, 2003). The 3’-OH of the intron attacks the exon ligation junction and the intron is integrated via two consecutive transesterifications, resulting in the original pre-RNA (Lehmann and Schmidt, 2003).

Excised linear introns are also able to carry out the reversal of the second splicing step, leading to the formation of intron/3'-exon intermediates. For the reversion of the first step, the energy of the 2'-5' phosphodiester bond not present in the linear molecule is
essential (Lehmann and Schmidt, 2003). Although this reaction was considered to be inefficient, it has been recently shown that linear introns can catalyze splicing efficiently and with high precision (Roitzsch and Pyle, 2009). Moreover, the reverse splicing occurs with similar rates for linear and lariat intron, but the majority of lariat intron molecules locks in a conformation deficient for reverse splicing, leading to the a prevalence of integrated linear intron intermediates (Roitzsch and Pyle, 2009).

1.7 GROUP II INTRONS AS BIOTECHNOLOGICAL TOOLS

Group II introns have a set of characteristics that make them suitable for their use as biotechnological tools (Toro et al., 2007):

- They are mobile elements with the ability of integration into DNA target with high insertion frequency and specificity, by a homologous recombination-independent process;
- The recognition of DNA target is mediated by base-pairing of EBS-IBS sequences, which can be modified to change the target specificity;
- They can mobilize foreign genetic information inserted within the intron;
- Accessory functionalities are provided by the intron-encoded protein (IEP) or by conserved housekeeping genes, which allow its use in diverse background hosts.

In the basis on these properties, the Ll.ltrB group II intron from L. lactis was reprogrammed to insert into new target sequences, leading to a biotechnological tools known as “targetron” (Guo et al., 2000). This tool has been used for conditional targeted gene disruption, in which the expression of the LtrA protein under an inducible promoter could lead to the expression of the interrupted gene (Guo et al., 2000). Targetrons have also been developed for other bacterial group II introns, such as EcI5 of E. coli (Zhuang et al., 2009a) and RmInt1 from S. meliloti (see below) (Garcia-Rodriguez et al., 2011) and in eukaryotes (Mastroianni et al., 2008; Zhuang et al., 2009b). Introns integrated into the target DNA can be selected using a selectable marker, which can be coded in DIV after removal of the IEP sequence (Toro et al., 2007). A development in this strategy is the retrotransposition-activated selectable marker (RAM), which includes a tAl group I intron interrupting the selectable marker, so this can be only expressed if retrohoming occurs (Zhong et al., 2003; Toro et al., 2007).

Moreover, targetrons can be used to generate a targeted double-strand DNA break (resulted from partial reverse splicing and second-strand cleavage by the endonuclease
activity of the intron-encoded protein), that stimulates targeted DNA integration by homologous recombination, a favored mode of gene targeting in higher organisms (Karberg et al., 2001; Lambowitz and Zinnerly, 2010).

1.8 THE Sinorhizobium meliloti RmInt1 GROUP II INTRON

RmInt1 was the first group II intron described in the family Rhizobiaceae. RmInt1 is a group IIB3 intron originally found by sequence analysis in the Sinorhizobium meliloti strain GR4, inserted within the insertion sequence ISRm2011-2 of the IS630-Tc1/IS3 retroposon superfamily (Martínez-Abarca et al., 1998). Later, it was shown that RmInt1 can be also associated to IS elements closely related to ISRm2011-2 and to the ectopic site of the axi1 gene with a 20-fold lower efficiency than into the homing site (Martínez-Abarca and Toro, 2000).

1.8.1 Intron-encoded protein (IEP) promoted mobility of RmInt1 intron

RmInt1 intron was found to possess an intron-encoded protein (IEP) within its DIV RNA structure, with homology to reverse transcriptases (Martínez-Abarca et al., 1998). The RmInt1 IEP has a complete RT domain followed by a putative RNA-binding domain (domain X) associated with RNA splicing or maturase activity and an atypical C-terminal region which does not harbor the Zn motif characteristic of the endonuclease domain (En). However, RmInt1 has shown to be highly mobile in vivo, with a mechanism independent of endonuclease cleavage and homologous recombination (Martínez-Abarca et al., 2000; Martínez-Abarca and Toro, 2000).

The conserved reverse transcriptase (RT) of the RmInt1 encoded-protein has been shown to be active in vivo and in vitro (Muñoz-Adelantado et al., 2003). Like other group II intron IEPs, its function is completely abolished when replacing the conserved D230 and D231 residues with H230 and H231 of the active site in the RT domain, whereas binding to RNA and self-splicing promoting functions remain unaltered (Muñoz-Adelantado et al., 2003).

The C-terminal region of the RmInt1 intron-encoded protein (IEP) has been found to be related to bacterial ORF class D (Molina-Sanchez et al., 2010). Thus, the C-tail of the RmInt1 IEP, located in a predicted α-helical region, contributes to the maturase
function of the reverse transcriptase domain of the protein, whereas a region located downstream of the former and also taking part of the α-helix, contains conserved residues specifically required for the insertion of intron RNA into DNA targets in the leading strand of the replication fork (Molina-Sanchez et al., 2010). These novel characteristics may have evolved to facilitate the mobility of ORF-less introns by using either leading or lagging-oriented DNA targets (Molina-Sanchez et al., 2010).

Like other group II intron, RmIntr1 is able to reverse-spliced into single-stranded and double-stranded DNA target sites (Muñoz-Adelantado et al., 2003), using two distinct mechanisms depending on the orientation of the target site relative to the direction of DNA replication (Martínez-Abarca et al., 2004). A major pathway implies the insertion of intron RNA into single-stranded DNA once the replication fork has passed the DNA target and before use of the nascent lagging DNA strand as a primer (Martínez-Abarca et al., 2004), similarly as ectopic transposition of group II introns (Figure I10 A). The minor retrohoming pathway involves reverse splicing into either double-stranded DNA or
transiently single-stranded DNA target sites. Mechanisms for priming include random non-specific opposite strand nicks, a nascent leading strand or de novo initiation priming (Figure I10 B) (Martínez-Abarca et al., 2004).

Retrohoming in RmInt1 is very efficient, with all recipient cells undergoing homing event (100 % homing frequency) and up to 50 % of the recipient targets are invaded (Martínez-Abarca et al., 2004). The sequence requirements for efficient RmInt1 homing in vivo have been deciphered, showing that a DNA fragment extending from position -20 to +5, relative to the intron insertion site, supports the homing efficiency (Figure I11) (Jiménez-Zurdo et al., 2003). This segment includes the IBS1 and IBS1 sequences, extending from positions -13 to -9 and -7 to -1, respectively, in the 5’ exon, with a spacing of one nucleotide that it is not essential. In the 3’ exon, the IBS3, corresponding to the first nucleotide, is also required. 2nt 5’ upstream the IBS2 are sufficient to support homing, although wild-type levels also require the segment extending from -16 to -20 positions, being the -15 nucleotide highly conserved (Jiménez-Zurdo et al., 2003). This constraint seems to rely on a novel pairing generated by shifting the IBS2 sequence from -15 to -18, which has been observed to be conserved in all published sequences (Costa et al., 2006b). The 3’ distal exon extends 4 nt downstream from IBS3 and only the G+4 is compulsory required (Jiménez-Zurdo et al., 2003).

![Figure I11](image)

**Figure I11.** DNA target site recognition by RmInt1 group II intron. The most critical positions conserved in 5’ and 3’ distal exons are indicated by green and yellow shadowing, respectively. The IBS regions recognized by base-pairing of the intron RNA’s EBS sequences are shaded in pink. Adapted from Jiménez-Zurdo et al (2003) and Lambowitz and Zimmerly (2004).

Likewise other group II introns, such as L.ltrB from _L.lactis_ and ai5γ from _S.terrevisiae_, RmInt1 can be retargeted at high frequency in genes present in plasmids or as single copies in the chromosome of _E.coli_ (García-Rodríguez et al., 2011). Potential targets were selected to retain only critical nucleotide residues T-15, C+1 and G+4, whereas the
intron was engineered in order to EBS/IBS base pairings occur. Targeting frequencies of
the retargeted introns varied from 0.02 % for the gene *ompT2* to 30 % for *trxB1* (Garcia-
Rodriguez *et al.*, 2011). In addition, the retrohoming efficiency in *E. coli* is lower than in its
natural host, and in the former, RmInt1 mobility is higher at 28 °C than at 37 °C,
suggesting that host factors may play an important role in the process (Garcia-Rodriguez *et
al.*, 2011).

**I.8.2 Splicing of the RmInt1 intron**

The *in vivo* splicing of RmInt1 was firstly assessed by RT-PCR in *E. coli* and
*Sinorhizobium meliloti*, showing that the intron is able to splice from the transposase gene of the IS2011-
2 where it is inserted, and that this process it is dependent of the maturase protein

The wild-type RmInt1 intron efficiently excises from its precursor ARN in a lariat
form (Muñoz-Adelantado *et al.*, 2003). Mutations in the conserved YADD motif to
YAH in the RT domain just produce a slight reduction, showing that the RT activity is
not required for splicing. However, the truncation of the RT domain or the disruption in
the catalytically critical conserved pairing AGC-GUU of the DV completely abolish the
excision capacity (Muñoz-Adelantado *et al.*, 2003). In addition, RmInt1 is able to excise as
circles *in vivo*, even in not mobile mutant intron-derivatives, indicating that intron dimers
arranged head-to-tail are not generated (Molina-Sánchez *et al.*, 2006). It has been also
shown that the maturase domain of the IEP is required for the formation of both RNA
circle and intron lariat *in vivo* and that it controls in some way the balance of intron
excision as lariat or intron circles (Molina-Sánchez *et al.*, 2006).

The excision of the RmInt1, through the branching pathway or in the form of
circles, has been shown to be regulated by the branch-point and the coordination loop
structures (Molina Sanchez *et al.*, 2011). Mutations in the branch-point prevent branching,
producing an enrichment on the circular form, whereas mutation in the coordination loop
do not abolish the formation of lariat intron, unlike other group II introns, which show a
similar phenotype when both structures are altered (Molina Sanchez *et al.*, 2011). Moreover,
the mutation in the EBS3 position (G to C) results in a great accumulation of
circular and head-to-tail intron molecules, resulted from circularization of the intron and
the formation of intron dimers, respectively (Molina Sanchez *et al.*, 2011).
The exon sequence requirements for RmInt1 excision in vitro have been elucidated, showing that the EBS1-IBS1 interaction is required and sufficient for RmInt1 excision and that, in contrast to the critical positions of the EBS-IBS interactions in the DNA target, neither the distal 5’-exon region, the IBS2-IBS2, the putative IBS2*-IBS2, the IBS3-IBS3 pairings nor the distal 3’-exon region are required for intron excision (Barrientos-Durán et al., 2011).

RmInt1 intron self-splices in vitro in the absence of the intron-encoded protein when incubated under permissive conditions (Costa et al., 2006a). However, the reaction is very inefficient, which show low reactivity and reaction rates. In addition, unconventional products truncated in the 3’- exon accumulate along with the expected ones, a fact resulting from the presence of a IBS1-like sequence in the 3’-exon better matching the authentic IBS1 sequence in the 5’-exon (Costa et al., 2006a).

I.8.3 Dispersion and evolution of the RmInt1 intron

The ectopic retrotransposition to novel (ectopic) sites is considered a key mechanism by which group II introns have dispersed in the nature. The RmInt1 intron has been shown to invade the awt1 gene independently of recombinase in its natural bacterial population (Martínez-Abarca and Toro, 2000; Muñoz-Adelantado et al., 2001). However, the RmInt1 intron propagates into the S. meliloti genome primary by retrohoming upon acquisition by conjugative transfer, invading unoccupied RmInt1 homing sites in the genome of some S. meliloti strains with high efficiency (Nisa-Martínez et al., 2007). According with previous results (Martínez-Abarca et al., 2004), the homing sites located on the template for the lagging strand synthesis are the first to be invaded by the intron, although the available homing sites on the leading strand template are eventually occupied (Nisa-Martínez et al., 2007). Once the homing sites are saturated, retrotransposition could contribute to further intron dispersion (Nisa-Martínez et al., 2007).

The RmInt1 intron is also present in other Sinorhizobium and rhizobia species (Fernandez-Lopez et al., 2005). The intron-homing sites in these species are also IS elements of the IS2011-2 group, as in S. meliloti, although ectopic sites of insertion have been observed as well (Fernandez-Lopez et al., 2005). RmInt1 is present predominantly in the genus Sinorhizobium, where vertical inheritance seems to occur from a common ancestor. In other rhizobia species, the intron only appears as fragmented copies, which show the existence of a horizontal transfer among closely related strains and species with a possible lost in some of them (Fernandez-Lopez et al., 2005). Then, the RmInt1 intron is
able to propagate by vertical and lateral transfer, retrohoming or ectopic transposition, and show that they tend to evolve into fragmented introns, mechanisms underlying the origin of eukaryotic spliceosomal introns (Fernandez-Lopez et al., 2005).
OBJECTIVES
On the basis of the previous knowledge of the RmInt1 group II intron biology, our goal is to deepen in the study of the molecular processes involved in the \textit{in vivo} and \textit{in vitro} splicing of the \textit{Sinorhizobium meliloti} RmInt1 intron.

This general objective will be developed through the following specific aims:

1. Creation of a reporter system for the quantitative assessment of the \textit{in vivo} splicing of the RmInt1 intron.

2. Quantitative determination of the complete splicing reaction (intron excision and exon ligation) efficiency \textit{in vivo} of the RmInt1 intron in its natural host \textit{Sinorhizobium meliloti} and in the heterologous host \textit{Escherichia coli}.

3. Improvement on the efficiency of the \textit{in vitro} self-splicing reaction of RmInt1. Quantitative analysis of the reaction rates, products and pathways in different RmInt1-transcript contexts.

Study of the effects on the first and second splicing step efficiency of point mutations in the ribozyme of RmInt1 intron previously reported to be part of conserved tertiary interactions involved in the maintenance of the RNA intron structure and function.
MATERIALS AND METHODS
M.1 BACTERIAL STRAINS

Bacterial strains used in this work are described in Table M1.

TABLE M1. Bacterial strains used in this work.

<table>
<thead>
<tr>
<th>Bacterial Strain</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinorhizobium meliloti</td>
<td>RmInt1 intronless strain. Nod(^+), Fix(^+).</td>
<td>Villadas et al, 1995</td>
</tr>
<tr>
<td>RMO17</td>
<td>F(^-), o80ΔlacZΔM15, (Δ(lacZYA-argF))U169, deoR, recA41, endA1, bsdR17(rK(^-), mK(^+)), phoA, supE44, λ(^-), thi-1, gyrA96, relA</td>
<td>Bethesda Research Lab</td>
</tr>
<tr>
<td>Escherichia coli DH5α</td>
<td>supE44, (Δ(mcrC-mrr)), recA413, ara-14, proA42, lacY1, galK2, rpsL20, xyl-5, mtl-1, leuB6, thi-1</td>
<td>Promega Corporation</td>
</tr>
<tr>
<td>E.coli HB101</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

M.2 BACTERIAL CULTURES

M.2.1 Culture media

E.coli strain was grown in Luria-Bertani (LB) broth (Sambrook et al., 1989): NaCl 5 g/l, tryptone 10 g/l and yeast extract 5 g/l, made with desionized water (MQ water) and adjusted to pH 7. Solid medium was supplemented with 1.6% w/v agar. It was sterilized by autoclaving at 120 °C for 20 minutes.

The selective solid culture medium of coliform bacteria Endo Agar (Difco) was used for the detection of contaminant E.coli cells in S.meliloti trans-conjugants. It was made with 51 g/l of powder in MQ water, heating to dissolve and sterilizing by autoclaving at 120 °C for 20 minutes.

S.meliloti strain was grown in TY medium (Beringer, 1974): CaCl\(_2\)·2H\(_2\)O 0.9 g/l, tryptone 5 g/l and yeast extract 3 g/l, prepared with MQ water and adjusted to pH 7. Solid medium was supplemented with 1.6% w/v agar. It was sterilized by autoclaving at 120 °C for 20 minutes.

S.meliloti trans-conjugants were grown in minimal medium (Robertsen et al. 1981: K\(_2\)HPO\(_4\) 0.3 g/l, KH\(_2\)PO\(_4\) 0.3 g/l, CaCl\(_2\) 0.05 g/l, MgSO\(_4\)·7H\(_2\)O 0.15 g/l, FeCl\(_3\) 0.006 g/l,
NaCl 0.05 g/l, sodium glutamate 1.1 g/l, mannitol 10g/l, biotin 0.0002 g/l and calcium pantothenate 0.0001 g/l. Solid medium was supplemented with 1.6% w/v agar. It was sterilized by autoclaving at 120 °C for 20 minutes.

M.2.2 Antibiotics

Antibiotics were made as 100X stock solutions in MQ water (Km, Ap) and filter sterilized through 0.2 μm membrane Minisart® NML (Sartorius). Tetracycline and Ampicillin were used at 10 g/l and 200 g/l final concentration, respectively, for both *E.coli* and *S.melioti*. Kanamycin was used at a final concentration of 180 g/l for *S.melioti* and of 50 g/l for *E.coli*.

M.2.3 Growth conditions and storage

*S. melioli* was routinely grown at 28°C in shaker at 190 rpm, having a generation time of 2.5 h, approximately. *E.coli* was grown at 37°C in the same conditions as above, with a generation time of 30 min.

Bacterial cultures grown until late exponential growth phase were deposited in cryotubes with added 20% (v/v) final concentration glycerol and preserved at -80 °C for long-term storage.

M.3 PLASMIDS AND CLONING VECTORS

M.3.1 Commercial plasmids and vectors

**TABLE M2.** Plasmids and vectors used in this work

<table>
<thead>
<tr>
<th>Name</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pGEM-T easy</td>
<td>Vector for the cloning of PCR products. High copy number. 3015 bp. Ampicillin resistant. Recombinant cloned identified by color screening.</td>
<td>Promega</td>
</tr>
</tbody>
</table>

38
### Materials and Methods

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pBB0.6</strong></td>
<td>Receptor plasmid in homing assays. pBBR1MS-2 derivative (Kovach et al., 1995). 5820 bp. Kanamycin resistant. Comprised of the -175/+466 of the RmInt1 target site of ISRm2011-2. (Martinez-Abarca et al., 2000)</td>
</tr>
<tr>
<td><strong>pBBΔ129</strong></td>
<td>Derivative of pBB0.6, where a region of 129 bp (-125/+4) in the ISRm2011-2 has been deleted. (Martinez-Abarca et al., 2000)</td>
</tr>
<tr>
<td><strong>pKG0IEP</strong></td>
<td>Plasmid producing constitutively the intron-encoded protein (IEP) for trans-complementation in splicing. 6533 bp. Kanamycin resistant. (Muñoz-Adelantado et al., 2003)</td>
</tr>
<tr>
<td><strong>pMP220</strong></td>
<td>Broad-host range transcriptional fusion vector. Precursor of pICG plasmids. It contain a multicloning site sequence, a Shine-Dalgarno sequence from the <em>E. coli</em> CAT gene and the β-galactosidase <em>laiz</em> gene of <em>E. coli</em>. 10.5 Kb. Tetracycline resistant. (Spaing et al., 1987)</td>
</tr>
<tr>
<td><strong>pKG2.5</strong></td>
<td>Intron construct used as a donor in homing assays. Flanking exons of -174/+466 bp. 7859 bp. Kanamycin resistant. (Martinez-Abarca et al., 2000)</td>
</tr>
<tr>
<td><strong>pKG2.5DV-CGA</strong></td>
<td>Splicing defective intron construct. Derivative of pKG2.5. 7859 bp. Kanamycin resistant. (Muñoz-Adelantado et al., 2003)</td>
</tr>
<tr>
<td><strong>pKGEMA4</strong></td>
<td>IEP-ΔORF intron construct, from which most nucleotides of the ORF has been deleted. Flanking exons of -20/+5 bp. 7318 bp. Kanamycin resistant. (Nisa-Martinez et al., 2007)</td>
</tr>
<tr>
<td><strong>pKGEMA4DV</strong></td>
<td>Splicing defective intron construct. Derivative of pKGEMA4. 7859 bp. Kanamycin resistant. (Nisa-Martinez et al., 2007)</td>
</tr>
<tr>
<td><strong>pLM1</strong></td>
<td>ΔORF intron construct under T7 promoter. Template for in vitro transcription. pUC derivative. Flanking exons of -15/+146 bp. Ampicillin resistant. (Costa et al., 2006)</td>
</tr>
<tr>
<td><strong>pLI1</strong></td>
<td>Contains the L.ltrB intron and flanking exons (-58/+38) cloned behind the phage T7 promoter in the expression vector pET-11a. 8477 bp. Ampicillin resistant. (Matsuura et al., 1997)</td>
</tr>
</tbody>
</table>

#### M.3.2 pICG plasmid series

pICG plasmids are derivative of pMP220 (Spaing et al., 1987), to which the pSyn promoter (Giacomini et al., 1994) has been cloned upstream from the *laiz* gene and the ribosome binding site, to ensure strong constitutive expression. In addition, the *Bln*I and polylinker *Bam*HI sites have been removed (Figure M1).
**Figure M1.** Construction of broad host range expression vector pICG-E1E2. A The pSyn promoter (Giacomini *et al.*, 1994) from pGEMTSSynX was cloned in the *EcoRI* cloning site in the polylinker of pMP220. *BamHI*, *PstI* and *SphI* were first removed of the polylinker by digestion with *KpnI* and *SphI*, treated with T4 DNA-polymerase (Roche) and auto-ligated with T4 DNA ligase (Roche). After removal of *BlnI* by digestion and filling in with Klenow fragment of DNA
Materials and Methods

polymerase I, a fragment containing the minimal RmInt1 target site (20 bp of the 5’ exon (E1) and 5 bp of the 3’ exon (E2)) (Nisa-Martínez et al., 2007) from the ISRm2011.2 insertion sequence and flanked by BamHI restriction sequences was inserted into the BamHI site of the lacZ gene to create pICG-E1E2. Te8 and Ap8 indicate resistance markers. Plasmids and polylinkers have not been drawn to scale. H, HindIII; Bg, BglII; E, EcoRI; K, KpnI; B, BamHI; P, PstI; S, SphI; Bl, Blal; Pl, PlaI. B Schematic diagram of the nucleotide sequence of the lacZ mRNA in pICG-E1E2 and the ligated exons resulting from RmInt1 splicing, presented with the corresponding amino-acid sequence. This region consists of 42 nucleotides, including 25 nucleotides corresponding to the ISRm2011.2 flanking exons and 17 corresponding to the additional PmaI, BhiI and BamHI cloning sites linked in-frame to the lacZ gene. Upper case letters indicate the lacZ sequence, whereas lower case letter indicate the exon flanking sequences. Additional amino acids due to the addition of nucleotides to the residual lacZ remaining after intron excision are identified by asterisks. A vertical arrow indicates the site of intron insertion (wt or □ ORF) in the intron-containing constructs.

pICG-E1E2 contains 25 nucleotides corresponding to the intron flanking exons (20 bp of the 5’ exon (E1) and 5 bp of the 3’ exon (E2)) (Nisa-Martínez et al., 2007) from the ISRm2011.2 insertion sequence inserted into the BamHI site of the lacZ gene (Figure M1 B). pICG-WT contains the wild-type RmInt1 intron, generated in vivo by the invasion of pICG-E1E2 by the RmInt1 intron from pKG2.5 (Martínez-Abarca et al., 2000). pICG-DV contains a variant of the intron with a mutation in catalytic domain V, obtained from pKG2.5D5-CGA (Muñoz-Adelantado et al., 2003) (Figure M2 A).

The construct used to produce the IEP protein in trans was pKGIEP (Muñoz-Adelantado et al., 2003) (Figure M2 B). The series of constructs in trans generated in this study are ΔORF constructs (Nisa-Martínez et al., 2007), from which most of the intron ORF has been deleted. pICG-ΔORF was constructed by ligating the 0.8 kb PmaI/BhiI fragment of pKGMA4 between the corresponding sites of pICG-E1E2. pICGΔORF DV contains the ΔORF fragment with a mutation in the catalytic domain, as in pKG2.5D5-CGA (Muñoz-Adelantado et al., 2003), inserted between the PmaI and BhiI restriction sites (Figure M2 B).

The plasmids in the cis series are also based on ΔORF constructs, with the sequence encoding the IEP protein located upstream from the 5’ exon on the same plasmid (Figure M2 C). This IEP sequence was obtained by amplification from pKGIEP, with the primers 5’-GGGGAGTACTGGAACAGGATGACTTCGGA and 5’-GGGGGCATCTCAAGTAAACGTGTTCGTTCC, and was inserted between the ScaI and SphI sites added to the polylinker of pICG-E1E2, pICG-ΔORF and pICG-ΔORF DV, to give pICG-IEP-E1E2 (without intron), pICG-IEP-ΔORF (intron ΔORF) and pICG-IEP-ΔORF DV (CGA-mutated intron), respectively (Figure M2 C).
**Figure M2.** Construction of pICG plasmids from pICG-E1E2. A The intron donor plasmid pKG2.5 (Martínez-Abarca *et al.*, 2000) was mobilized to a *S. meliloti* RMO17 strain carrying the intronless construct pICG-E1E2, invading the later and creating pICG-WT. The fragment ΔORF DV from the splicing-defective pKG2.5 DV plasmid (Muñoz-Adelantado *et al.*, 2003) was amplified and cloned in the *BlnI* and *PmlI* sites of pICG-E1E2. B pKMA4 and pKEGA4DV plasmids were used to construct pICG-ΔORF and pICG-ΔORF DV plasmids, respectively, in which the fragment containing the ΔORF or the ΔORF DV sequences was cloned into the *BlnI* and *PmlI* sites of pICG-E1E2. C The transient plasmid pICG-SeaSph was created by introducing a mutated KpnI-KpnI oligonucleotide into the KpnI site of the pICG-ΔORF plasmid, thus inactivating this site and adding new SeaI, KpnI and SphI sites to the poly linker. The pICG-SeaSph plasmid was subsequently used to create the pICG-IEP-ΔORF plasmid in which the IEP fragment from pKG1E2P was amplified and cloned in the *SeaI*-SphI sites. The ΔORF DV fragment from pKMA4DV and the flanking-exon oligonucleotide were exchanged with the ΔORF fragment to create pICG-IEP-ΔORF DV and pICG-IEP-E1E2, respectively. Plasmids and poly linkers have not been drawn to scale. H, *Hind*III; Bg, *BglII*; E, *EcoRI*; K, *KpnI*; B, *BamHI*; P, *PstI*; S, *SphI*; Bl, *BlnI*; Pl, *PmlI*, Sc, *SeaI*.

The plasmids of the ter series contain a synthetic Rho-independent transcription terminator (Kingsford *et al.*, 2007) (Figure M3). In these constructs, oligonucleotides 5’-GATCCCGTACGCAAGTAGATCCGTTTT and 5’-GATCCCGTACGCAAGTAGATCCGTTTTGTCGTTTCGACTAGGG were annealed and ligated downstream from the intron-flanking exons, into the *BlnI/BamHI* sites of pICG-E1E2, pICG-ΔORF and pICG-ΔORF DV, generating pICGTer-E1E2 (without intron), pICGTer-ΔORF (intron ΔORF) and pICGTer-ΔORF DV (CGA mutated intron), respectively.

pICG plasmids expressing the *Lactococcus lactis* Ll.ltrB intron derived from pICG12 and pL1 (Matsuda *et al.*, 1997), pICG-Ll.ltrB (Ll.ltrB intron in sense orientation) and pICG-Ll.ltrBA (in antisense orientation) were constructed by using primers 11P_ltrBE1 (5’-GGGGGATCAGGTGGCAGAATA) and 12PR_ltrBE2 (5’-GGGGGATCAGGTGGCAGAATA) to amplify a 2.6 kb region of pL1, which was then cloned in the *BamHI* site of pICG12. The resulting constructions were confirmed by sequencing and both orientation of the insert were selected by restriction analysis. pICG-ltrB contained a 96-nucleotide sequence of ligated ltrB exons 1 and 2, extending from 58 nucleotides upstream to 38 nucleotides downstream of the intron insertion site, cloned in the *BamHI* site of pICG-E1E2. The insert was generated from cDNA of reverse transcribed pL1 RNA by PCR with the primers 11P_ltrBE1 and 12PR_ltrBE2.

pICG plasmid series containing mutations in the ribozyme of RnInt1 intron derived from pICG-ΔORF plasmid via site-directed mutagenesis. Mutations located in the termini of the intron sequence was incorporated into one of the primer used in the PCR.
reaction (Scharf et al., 1986). Mismatched and overlapping primers were used when mutations were located anywhere else along the intron length (Higuchi et al., 1988). In that case, two primary PCR reactions produced two overlapping DNA fragments, which were subsequently mixed, serving as templates for the secondary reamplification using only the outermost two of the four primers used in the primary PCR reaction. The recombined PCR product was then cloned in PmlI-BlnI sites of pICG-E1E2 or pICGTer -E1E2 plasmids. Recombinant plasmids were sequenced in order to verify constructions. A summary of mutant constructions and the primers used are listed in Table M3.

Figure M3. Construction of pICG-Ter series plasmids from pICG-E1E2. The synthetic Rho-independent transcription terminator (Kingsford et al., 2007) was introduced in the BamHI sites of pICG-E1E2, replacing the 25nt intron flanking exon fragment and creating the pICGTer plasmid. The SpeI-BlnI fragment of the later was exchanged with pICG-E1E2, pICG-ΔORF and pICG-ΔORFDV, resulting in pICGTer-E1E2, pICGTer-ΔORF and pICGTer-ΔORFDV, respectively. Plasmids and polylinkers have not been drawn to scale. B, BamHI; Bl, BlnI; PmlI, Sp, SpeI.
### Materials and Methods

**TABLE M3.** Primers used for the construction of mutants in the ribozyme of RmInt1 intron

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Template</th>
<th>PCR Primer(^1)</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1883A</td>
<td>pICG-ΔORF</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>22PR_C1883A</td>
<td>CCCCTAGGCGAGGGTGGTTGAGTAGGG</td>
</tr>
<tr>
<td>C1883G</td>
<td>pICG-ΔORF</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>23PR_C1883G</td>
<td>CCCCTAGGCGAGGGTGGTTGAGTAGGG</td>
</tr>
<tr>
<td>C1883T</td>
<td>pICG-ΔORF</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24PR_C1883T</td>
<td>CCCCTAGGCGAGGGTGGTTGAGTAGGG</td>
</tr>
<tr>
<td>ΔGA</td>
<td>pICG-ΔORF</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4OR_ΔGA</td>
<td>TACGGGCGGAGCCACGTCTCCGTGATCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4O_ΔGA</td>
<td>GACTGGGGCGTGCCCCCGTATGAGCGCGTCAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6QR_LACZ</td>
<td>GATGTGCTGCAAGGCGATT</td>
</tr>
<tr>
<td>Δγ</td>
<td>pICG-ΔORF</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5OR_Agamma</td>
<td>TACGGGCGGATCGCCCCAGTCTCTCCGTGATCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5O_Agamma</td>
<td>GACTGGGGCGTGCCCCCGTATGAGCGCGTCAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6QR_LACZ</td>
<td>GATGTGCTGCAAGGCGATT</td>
</tr>
<tr>
<td>Δγ-GEBS3C</td>
<td>pICG-GEBS3C</td>
<td>5Q_CAT</td>
<td>TGTACCTATAACCAGACCGTTCCAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5OR_Agamma</td>
<td>TACGGGCGGATCGCCCCAGTCTCTCCGTGATCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5O_Agamma</td>
<td>GACTGGGGCGTGCCCCCGTATGAGCGCGTCAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6QR_LACZ</td>
<td>GATGTGCTGCAAGGCGATT</td>
</tr>
</tbody>
</table>

\(^1\) Bold letters indicate outer primers used in secondary PCR reaction.

#### M.3.3 Other described pICG and pICGTer plasmids

**TABLE M4.** Described pICG and pICGTer plasmids used in this work

<table>
<thead>
<tr>
<th>Name</th>
<th>Characteristics</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GγC</td>
<td>Position G452 mutated to C. Disrupts γ-γ’ interaction.</td>
<td></td>
</tr>
<tr>
<td>Cγ’G</td>
<td>Position C1884 mutated to G. Disrupts γ-γ’ interaction.</td>
<td>(Molina Sanchez et al., 2011)</td>
</tr>
<tr>
<td>GγCγ’G</td>
<td>Double mutation G452 to C and C1884 to C. Restoration of γ-γ’ interaction.</td>
<td></td>
</tr>
<tr>
<td>GEBS3C</td>
<td>Position G329 mutated to C. Disrupts EBS3-IBS3 interaction.</td>
<td></td>
</tr>
</tbody>
</table>
M.3.4 pLM plasmid series

pLM plasmids were derivative of pLM1 (Costa et al., 2006). To create pLM-WT, primers 5Q_CAT and 27PR_lacZtranscript (5’-GGAGATCTCATATGACTGTGGGAAAGGCGA) were used to amplify a 1009 bp fragment of pICG-ΔORF, which was then cloned in Clal-BamHI sites of pLM1.

Mutant pLM-derivative plasmids were generated by cutting and cloning a Clal-BamHI fragment from mutated pICG plasmids in pLM-WT.

M.4 NUCLEIC ACIDS EXTRACTION

M.4.1 Plasmid DNA extraction by magnesium salts

This procedure, based on that of Studier (Studier, 1991), was used for rapid plasmid extraction from E.coli cultures. An overnight-grown culture of E.coli was collected in 1.5 ml microcentrifuge tubes and centrifuged at 13000 rpm for 2 min. The supernatant was removed by aspiration and the pellet suspended in 100 μl of MQ water. To this suspension it was added NaOH 0.1 M, EDTA 10 mM and 2% SDS and mixed immediately by vortexing. Tubes were heated for 2 min in a boiling water-bath. 50 μl of 1 M MgCl₂ was mixed in by vortexing and the tube was placed on ice for 2 min. The precipitate was pelleted by centrifuging for 1 min. 50 μl of 5 M potassium acetate was mixed into the supernatant in the same tube by brief vortexing and the tube was again placed on ice for 2 min. After centrifuging another 5 min at 13000 rpm, the supernatant was removed to a new tube containing 0.6 ml of 100% (v/v) ethanol, mixed by vortexing and placed on ice for 5 min. The tube was centrifuged for 5 min, the supernatant removed by aspiration, and 0.2 ml of 70% (v/v) ethanol was added, centrifuged for 5 min and removed by aspiration. Residual solvent was removed under vacuum for 5 to 10 min and the pellet dissolved in a 0.2mg/ml of an RNAse solution.

M.4.2 Plasmid DNA extraction by alkaline lysis

This method was routinely used for plasmid DNA extraction from S.meliloti. It is based on the modified procedure of Sambrook (Birnboim and Doly, 1979; Ish-Horowicz and Burke, 1981; Sambrook et al., 1989). 3 ml of a saturated culture of S.meliloti was harvested in 1.5 ml microcentrifuge tubes and washed with 200 μl of 0.1% sarkosyl (made
in TE). Cell were resuspended in 100 μl of 50mM glucose, 25mM Tris pH 8.0 and 10mM EDTA and incubated for 5 min at room temperature. 200 μl of 0.2N NaOH and 1% SDS were added, mixed gently and put on ice for 5 min. 150 μl of precooled 5M KOAc pH 4.8 was added, mixed gently, and after 5 min on ice, samples were centrifuged for 1 min. The supernatant was then extracted with phenol-chloroform and reconstituted in a 0.2mg/ml of an RNase solution.

M.4.3 Total RNA extraction

This procedure, used for isolation of RNA from both *S. meliloti* and *E. coli* cultures, is based on the method described by (Cabanes et al., 2000). The bacterial pellet from a culture from *S. meliloti* or *E. coli* (optical density at 600 nm = 0.6) was resuspended and incubated for 10 min at 65°C in 2 ml of prewarmed lysis solution (1.4% SDS, 4 mM EDTA and 0.4 mg/ml Proteinase K solution). Proteins were removed by adding 150 μl of NaCl (5 M) at 4 °C for 10 min. After centrifugation, the nucleic acids present in the supernatant were precipitated by adding 1 ml of ethanol (100%). The pellet was resuspended in 85 μl of MQ water and digested with 50 units of RNase-free DNase I. The RNA was first extracted with 1 volume of phenol (pH 4.5)/chloroform/isoamyl alcohol (25:24:1) and then with 1 volume of chloroform/isoamyl alcohol (24:1). Finally, the RNA was precipitated with 3 volumes of ethanol (100%) and 75 mM NaOAc (pH 5.2). The RNA pellet was washed with 70% ethanol, dried, and resuspended in 20 μl of nuclease-free water.

M.5 CLONING AND ENZYMATIC MANIPULATION OF DNA

M.5.1 DNA digestion with restriction endonucleases

DNA was digested directly with commercially available restriction endonucleases according to the manufacturer’s recommendations (New England Biolabs). Typically, each microgram (μg) of DNA was digested in the presence of 5 units of enzymes and 1X reaction buffer, in a volume ranged from 10 μl (analytical digestion) to 50 μl (preparative digestion). When a double digestion was performed, reaction conditions were selected to ensure being amenable to any two restriction enzymes.
M.5.2 Conversion of protruding to blunt-ended restriction fragments

For converting a 5’ overhang to blunt end, a method based on (Wartell and Reznikoff, 1980) and modified by (Sambrook et al., 1989) was employed using the klenow fragment of *E.coli* DNA polymerase I (Roche). 0.5-1 μg of DNA restriction fragments were incubated for 15-30 min at 22°C with a mix of 1X filling buffer (500 mM Tris pH 7.5, 100 mM MgCl₂, 10 mM DTT, 5 00 μg/ml BSA), 1mM of each dNTP and 1U of Klenow Enzyme (Roche) in a volume of 10μl. Then, the enzyme was inactivated by heating at 70°C for 5 min.

DNA fragments with protruding 3’ ends were blunt-ended using the 3’-5’ exonuclease activity of bacteriophage T4 DNA polymerase (Sambrook et al., 1989). A mix of 1X incubation buffer (10 mM Tris-HCl 50 mM NaCl, 10 mM MgCl₂, 1 mM Dithiothreitol) supplemented with 100 μg/ml Bovine Serum Albumin and 100 μM dNTPs and 1U of T4 DNA Polymerase I (New England Biolabs) were added to 0.5-1 μg of DNA fragments, and incubated 5 min at 37°C. The reaction was stopped by adding EDTA to a final concentration of 10 mM.

M.5.3 Dephosphorilation of DNA fragments

Terminal phosphate groups from restriction endonuclease digested DNA were removed by incubation in the presence of calf intestinal alkaline phosphatase (CIP). Typically, 5-10 μg of DNA from a restriction digest was added directly to a reaction containing 1X capping buffer and 2 units of CIP (Roche) in a total volume of 50 μl. This reaction was incubated at 37 °C for 30 min, supplemented with a further 2 units CIP, and incubated at 37 °C for an additional 30 min. The reaction was terminated by adding a mix of 1% SDS in STE 1X buffer (100 mM NaCl, 1mM Tris-Cl pH 8 and 1 mM EDTA ) and heating at 75°C for 15 min. Then, the DNA was extracted with phenol/chloroform.

M.5.4 Nucleic Acids Purification

The cleanup of DNA fragments from enzymatic reactions was carried out by the commercial kits QIAquick PCR Purification Kit (Qiagen) and Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) or by phenol-chloroform extraction.

In the phenol-chloroform extraction, the DNA to be purified was extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) and centrifuged for
minutes. The aqueous phase was transferred to a fresh tube and extracted with chloroform/isoamyl alcohol (24:1), then centrifuged for 5 minutes. The supernatant was transferred to a fresh tube and precipitated with 3 volumes of 100% ethanol and 0.1X volume of 3M sodium acetate pH 5.2 for 5 minutes. The DNA was precipitated at -80°C for at least one hour and then pelleted by centrifugation for 15 minutes. The pellet was washed with 70% EtOH, dried, and resuspended in an appropriate solvent.

For extraction of DNA fragments from agarose gels in TAE (Tris acetate/EDTA) buffer, the commercial kits Perfectprep Gel Cleanup (Eppendorf) and Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) were used.

**M.5.5 Ligation of DNA fragments**

Restriction endonuclease digested or blunt end DNA was typically ligated into a vector backbone using the T4 DNA ligase (Engler and Richardson, 1983). Ligations were performed in a 10 μl reaction volume at 14-16°C for up to 24 h in the presence of 1 x ligase buffer and 1 U T4 DNA ligase (Roche). To calculate the appropriate amount of insert to include in the ligation reaction, the following equation was used:

\[
\frac{(\text{ng of vector}) \cdot (Kb \text{ size of insert})}{Kb \text{ size of vector}} \cdot (\text{insert: vector molar ratio}) = \text{ng of insert}
\]

**M.6 BACTERIAL TRANSFORMATION AND CONJUGATION**

**M.6.1 Competent *E.coli* cells and transformation**

Bacterial cells were made competent according to the rubidium chloride method of Rodríguez and Tait (1983). A culture of *E.coli* was grown until the OD₆₀₀ reached 0.4. Cells were pelleted by centrifugation at 6000 rpm for 10 min at 4°C. The cell pellet was resuspended in 30 ml of pre-chilled TFB1 solution [Composition: 30mM potassium acetate, 10mM CaCl₂, 50mM MnCl₂, 100mM RbCl, 15% glycerol, pH adjusted to 5.8 with 1M acetic acid. Filter sterilized and stored at 4°C]. The resuspended cells were incubated on ice for 15 minutes at 4°C and centrifuged again at 6000 rpm for 10 min at 4°C. The cells was resuspended in 4 ml of ice-cold TFBII solution [Composition: 100mM MOPS (pH 6.5), 5mM CaCl₂, 10mM RbCl, 15% glycerol, pH adjusted to 6.5 with 1M KOH. Filter
sterilized and stored at 4°C]. Cells were dispensed in aliquots, quick frozen in liquid nitrogen and stored at – 80 °C.

For transforming reaction, 50-500 ng of plasmid DNA or ligated DNA fragments were added to a 100 μl aliquot of competent cells and incubated for 30 min. Tubes were transferred from ice to a 42°C water bath, heat-shocked for 2 min and place on ice immediately to cool for 5 minutes. 1ml of LB medium was added and incubated for 1 hour at 37°C with shaking (~150 rpm). 100-200 μl of the transformation mix was plated and onto selection plates and incubated overnight at 37°C.

M.6.2 Plasmid conjugation between *E.coli* and *S.meliloti* strains

Plasmids were mobilized from *E.coli* to *S.meliloti* strain by tri-parental conjugation, using the helper plasmid pRK2013 (Ditta et al., 1980) for trans-complementation. A loopful of bacteria from each fresh plate was mixed together on a TY plaque without antibiotics and incubated at 28°C for 12-16 hours. The mix was then streaked onto MM medium with the appropriate antibiotics and incubated until the appearance of *S.meliloti* trans-conjugants colonies (2-3 days). A test of coliform-contaminant colonies were carried out in parallel onto Endo-Agar medium.

M.7 DNA AMPLIFICATION

M.7.1 The Polymerase Chain Reaction (PCR)

Typically, PCR reactions were performed in a final volume of 25 μl reaction including 10-100 ng of template DNA, 10 pmol of each specific primer, 0.1 mM of each dNTP, 1X reaction buffer and 2 U of DNA polymerase. The DNA polymerases used were AccuPrime Taq DNA Polymerase High Fidelity (Invitrogen), Pfu DNA Polimerasa (Biotools) and Phusion High-Fidelity DNA Polymerase (Finnzymes). Each set of reactions included a negative control consisting of all the reagents but the DNA was replaced with MQ water.

PCR amplifications were performed with an Eppendorf Mastercycler termal cycler. PCR conditions varied depending on the annealing primer temperature. A typical reaction consisted of a denaturing cycle at 94°C for 5 min, followed by 30 cycles: 30 s at 94 °C for
denaturation, 30 s at 60 °C for primer annealing and 30 s at 72°C for the extension of the polymerase. A final extension step at 72 °C for 10 min was also performed.

Amplified products were resolved on a 0.8 % agarose gel in TAE buffer and stained with ethidium bromide or RedGel (Biotium, Inc) or purified for later use.

**M.7.2 Quantitative real-time RT-PCR**

Real-time PCR is based on the procedure described by (Higuchi et al., 1993). Primers were designed with Primer 3 software (http://fokker.wi.mit.edu/primer3) (Rozen and Skaletsky, 2000) and examined for possible secondary structures with the mfold program (http://www.bioinfo.rpi.edu/applications/mfold) (Zuker, 2003). lacZ1 region, RmInt1 intron and ligated β-galactosidase gene lacZ were amplified from the pICG-WT plasmid. The sequences of the primers used are summarized in Table M5.

**TABLE M5.** Primers used in real-time qRT-PCR

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Q_CAT</td>
<td>CAGGAGCTAAGGAAGCTAAATGG</td>
</tr>
<tr>
<td>2QR_CAT</td>
<td>AGGCCGTAATATCCAGCTGAAC</td>
</tr>
<tr>
<td>3Q_RIB1</td>
<td>GCGTAAAGCTGCAGTGAT</td>
</tr>
<tr>
<td>4QR_RIB1</td>
<td>TCCTCGGAGGGTTCACCTT</td>
</tr>
<tr>
<td>5Q_CAT</td>
<td>TGTACTTATAACACAGGCGGTTA</td>
</tr>
<tr>
<td>6QR_LACZ</td>
<td>GATGTGCTGCAAGGGCAGT</td>
</tr>
</tbody>
</table>

The amplification efficiencies of all the primers were determined with serial ten-fold dilutions of the recombinant pICG-E1E2 and pICG-WT plasmids (Figure M4).
Figure M4. Generation of a standard curve to assess reaction optimization. A The standard curve was generated from plasmid DNA of pICG-E1E2 by using primers 1 and 2 and a 10-fold dilution on the iCycler iQ® real-time system. Each dilution was assayed in triplicate. B The standard curve generated from plasmid DNA using primers 1 and 2. The CT is plotted against the log of the starting quantity of template for each dilution. The equation for the regression line and the r value are shown below the graph. The calculated amplification efficiency was 96.8%.

PCR efficiency was high, at 94 to 97.7%, with a high degree of linearity (Pearson correlation coefficient r) (Table M6).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Assay value for indicated region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>lacZ1</td>
</tr>
<tr>
<td>PCR efficiency (%)</td>
<td>96.8</td>
</tr>
<tr>
<td>Melting temp (°C)</td>
<td>63.5</td>
</tr>
<tr>
<td>Amplicon length (pb)</td>
<td>149</td>
</tr>
</tbody>
</table>

The optimal annealing temperature was assessed by testing a range of annealing temperatures above and below the calculated Tm of the primers (Figure M5).

Figure M5. Annealing temperature optimization. An annealing temperature gradient of 55°C to 69°C was performed on the the iCycler iQ® real-time system. The 67°C gave a low CT value above the primer melting temperature and was selected as the optimal annealing temperature for this assay.
We carried out qRT-PCR on the iCycler System (Bio-Rad). Each reaction was run in triplicate and contained 5 µl of cDNA template (equivalent to 500 ng total RNA), 0.5 mM dNTP, 5 pmol of each oligonucleotide primer, 7.5 mM MgCl₂, 2.5 µl of a 1:10,000 dilution of SYBR Green, 2.3 µl of reaction buffer and 0.5 U of Platinum Taq DNA Polymerase (Invitrogen), with the final volume made up to 23 µl with DEPC water. The PCR cycling conditions were as follows: hot start, with heating at 94°C for 3 min, followed by 40 cycles of denaturation at 94°C for 30 s, annealing at 67°C for 30 s, and extension at 72°C for 15 s. A melting curve analysis was performed in all PCR runs, to check the identity of the PCR product obtained. Controls without a template were also included. Three independent RNA preparations from each strain were tested.

Absolute mRNA levels were determined, whereas relative quantification was carried out for splicing, based on the $2^{\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). Differences in relative expression were identified by carrying out Student’s t test. Levene’s test was used to check that the variances were equal.

M.8 ELECTROFORESIS

M.8.1 Non-denaturing agarose gel electroforesis

Plasmid DNA, PCR products or restriction DNA fragments were resolved by 0.8-2 % agarose gel electrophoresis in TAE buffer (Tris-HCl 40 mM, EDTA Na₂ 2 mM, glacial acetic acid 0,1142%(v/v)). 5 µl of PCR product or 2 µl of plasmid DNA or restriction DNA fragments were mixed with 2 µl of loading dye (Orange G 0.50%(w/v), EDTA Na₂ 0.01M and 50% Glycerol) and pipetted into the wells. 300 ng of DNA ladder (HindIII digested P29 phago or HindIII and EcoRI digested phago λ) was used to determine the size of the PCR products. Gels were stained with ethidium bromide (1 mg/ml) or 2% RedGel (Biotium, Inc). The stained gel was visualised under UV light in a Bio-Rad Gel Doc 2000 transilluminator. Images were captured, cropped, and printed with Gel Doc software version 4.3.1.

M.8.2 Denaturing agarose gel electroforesis

Total RNA was resolved by 1.4 % agarose gel electrophoresis in MOPS buffer (3-(N-morpholino)propanesulfonic acid) and 0.05 % (v/v) formaldehyde. 1 µl of total RNA
was mixed with 4 μl of MQ water and 2 μl of a solution of 1% RedGel (Biotium, Inc) in loading dye [1.8% sucrose, 1X TBE (Tris-borate EDTA, pH 8.3), 0.018% xylene cyanol, 36% (v/v) formamide and 25 mM EDTA]. The gel was visualised under UV light in a Bio-Rad Gel Doc 2000 transilluminator. Images were captured, cropped, and printed with Gel Doc software version 4.3.1.

**M.8.3 Denaturing polyacrylamide gel electrophoresis**

Single stranded nucleic acids were resolved by 8 M urea 5-6 % bisacrylamide (Acrisib 40%, acrilamidabisacrylamide (29:1), Amresco) gels in TBE buffer (Tris-HCl 0,089 M pH 8, Boric Acid 0,089 M and EDTA 0,002 M). For every 100 ml of gel casting solution, 40 μl of TEMED and 0.6 ml of freshly prepared 10 % Ammonium Persulfate were added. Samples were mixed with 2X loading dye [1.8% sucrose, 1X TBE (Tris-borate EDTA, pH 8.3), 0.018% xylene cyanol, 36% (v/v) formamide and 25 mM EDTA]. The gel was run in a vertical electrophoresis system in TBE buffer at 20-45 W. Gels were dried in a Gel Dryer® 583 (Bio-Rad) and exposed in phosphorus plates (Imagin Plate 2040, Fujifilm). The image was captured after 24-48 hours with the laser scanning system Personal Molecular Imager® FX and edited with the software Quantity One® Bio-Rad).

**M.9 NUCLEIC ACIDS HYBRIDIZATION**

**M.9.1 DNA-DNA Hybridization (Southern Blotting)**

This procedure, based on that of (Southern, 1975; Southern, 1979), was used in mobility assays of RmInt1 intron. DNA restriction fragments were separated in a 0.8 % agarose gel at 20-35 V for 17-20 hours. For partial fragmentation of large DNA fragment prior to transfer, the gel was left under UV light for 30 min. DNA was blotted onto a 20X SSC (1xSSC: NaCl 150 mM, sodium citrate 15 mM, pH 7.0) wet nylon filter (Biodine) by means of a denaturing (NaOH) vacuum system (VacuGene® de Pharmacia).

Membranes were hybridized with a digoxigenin-11-dUTP-labelled probe (0.1M) that was diluted down to 60 ng DNA/ml in prehybridization buffer [5xSSC, 0.1% sarkosyl, 0.02%SDS, 50% formamide and 2% blocking reagent (Roche)]. The membrane was incubated in a hybridization oven Hybridiser HB-1D (Techne) at 42°C for 20-24 h.
The membrane was soaked twice for 5 minutes each with 50 ml of 0.1 % (w/v) 2X SSC/SDS at room temperature, followed by soaking twice for 15 min each with 0.1 % (w/v) 0.1X SSC/SDS at 68°C and finally, soaking with washing buffer for 5 min at room temperature [0.3% Tween-20 in buffer 1 (0.1 M malic acid, 0.15 M NaCl, pH 7.5)].

After washing, the membrane was blocked with buffer 2 [1% (p/v) Blocking Reagent in buffer 1] for 30 min at room temperature and incubated with a 1:10000 dilution of an anti-digoxigenin-AP (Roche) antibody-conjugate for 30 min. To remove unbound antibody-conjugate, the membrane was washed twice for 15 min with washing buffer and equilibrated with buffer 3 [0.1 M Tris-HCl, 0.1 M NaCl, 50 mM MgCl₂, pH 9.5]. The chemiluminescent CSPD substrate (Roche) was diluted in buffer 3 and incubated with the blot for 5 min at room temperature. The blot was transferred onto a plastic report cover and incubated for 15 min at 37°C. Finally, the membrane was exposed to film for 4-5 hours (short exposition) or 18-20 (long exposition).

M.9.2 RNA-DNA hibridization (Northern Blotting)

This technique, described sby (Alwine et al., 1979), was used for the in vitro analysis of splicing products. 5 µg of S. meliloti total cellular RNA was subjected to electrophoresis in a denaturing 6% polyacrylamide gel (8M urea) and blotted onto a nylon filter (Biodine) in a TE 77 PWR Semi-Dry (GE Healthcare) transfer unit according to the manufacturer’s recommendations. The membrane was baked at 80°C for 1 hour.

Probes used in Northern hybridization (Table M7) were radiolabeled in a final volume of 10 µl containing 75 pmol of the oligonucleotide, 20 µCi of [γ-32P]ATP (3000 Ci/mmole; 10 µCi/µl; PerkinElmer) and 10 U of T4 polynucleotide (New England Biolabs), at 37°C, for 2 h, and the labeling mixture was then passed through Sephadex G-25 columns (GE Healthcare) for probe purification. The membrane was hybridized with about 20,000 cpm of probe/ml overnight at 42°C.

**TABLE M7.** Probes used in Northern Blotting assay

<table>
<thead>
<tr>
<th>Probe</th>
<th>Localization</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Q_CAT</td>
<td>5’ exon</td>
<td>CAGGAGCTAAGGAAGCCTAAAATGG</td>
</tr>
<tr>
<td>3Q_RIB1</td>
<td>Intron</td>
<td>GCGTAAAGCTGCGTGAATGAT</td>
</tr>
</tbody>
</table>
The blot was washed by soaking three-fold for 15 min each with 100 ml of 2X SSC/0.1% SDS, 1X SSC/0.1% SDS and 0.1X SSC/0.1% SDS at 42°C. Finally, the blot was placed against a phosphor screen, for visualization of the bands with the Personal Molecular Imager FX scanner and Quantity One software (BioRad).

M.10 MOBILITY ASSAY (HOMING) OF RmInt1 INTRON

*In vivo* homing efficiency of RmInt1 intron from pICG plasmids was determined by the double-plasmid assay in *S. meliloti* strain RMO17 using pICG-WT as donor plasmids and pBB0.6 (bearing the intron target sequence) or pBBA129 (negative control lacking the intron target sequence) (Martínez-Abarca et al., 2000) as recipient plasmids. Trans-conjugants were selected in minimal medium supplemented with antibiotics and the plasmid DNA extracted, digested with *XbaI* restriction enzyme and resolved in a non-denaturing 0.8 % agarose gel. DNA fragments were then transferred to a nylon membrane.

Mobility events were assessed by DNA hybridization analysis carried out using PCR-amplified probes. Primers used in the PCR reaction are listed in Table M7.

**TABLE M8.** Primers used to amplify probe oligonucleotides

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Probe</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011-b1</td>
<td>TGGACGAAGACGAACATGG</td>
<td>3’ exon of <em>I5</em>2011-2 (+3,+466 positions)</td>
<td>(Martínez-Abarca et al., 2000)</td>
</tr>
<tr>
<td>2011-b2</td>
<td>TTGAAGTGAGGCTGGCATT</td>
<td>RmInt1 intron (147-594 positions)</td>
<td></td>
</tr>
<tr>
<td>ε1</td>
<td>GTGAGCGTCGGATGAAAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18R.0</td>
<td>ACGTTTCTCAATTGAAACG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The homing efficiency was calculated as the percentage of the ratio of homing product to the addition of homing product and non-invaded recipient plasmid, following formulas:
Homing efficiency = \[
\frac{\text{Invaded recipient plasmid}}{\text{Invaded recipient p. + non – invaded recipient p.}} \times 100
\]

Images were edited with the software Quantity One Bio-Rad).

**M.11 β-GALACTOSIDASE ASSAY**

This assay was used for the determination of the expression level of the ligated β-galactosidase gene lacZ in pICG constructs. β-galactosidase assays were performed as described by Miller (Miller, 1972). Briefly, overnight cultures of *E. coli* grown at 28°C and 37°C and of *S. meliloti* grown at 28°C were diluted 1:50 in fresh LB and TY medium, respectively, and incubated until the cultures reached the exponential growth phase (OD\(_{600}\) = 0.6). Cultures were then cooled on ice for 20 minutes and bacterial density was recorded by measuring optical density at 600 nm. We then added aliquots of 10 µl to 300 µl of bacteria to the assay Z buffer containing 60 mM Na\(_2\)HPO\(_4\), 40 mM NaH\(_2\)PO\(_4\), 10 mM KCl, 1 mM MgSO\(_4\) and 50 mM β-mercaptoethanol and adjusted the pH to 7. We added 50 µl of chloroform and 25 µl of 0.1% SDS to the sample-buffer mixture, which we then vortexed for 30 seconds. The samples were incubated at 28°C for 5 minutes and the reaction was then started by adding 0.2 ml of o-nitrophenyl-β-D-galactopyranoside (ONPG) (4 mg/ml) (Fluka). The reaction was allowed to proceed for 10 minutes and was then stopped, by adding 0.5ml of a 1 M Na\(_2\)CO\(_3\) solution. Samples were centrifuged for two minutes to eliminate cell debris and absorbance was recorded at 420 nm. The results are expressed in Miller units, following formulas:

\[
β - \text{Galactosidase Activity (U)} = \frac{1000 \times DO_{420}}{t \times V \times OD_{600}}
\]

where \(t\) is the reaction time expressed in minutes and \(V\) is the volume of reaction expressed in milliliters.

**M.12 PRIMER EXTENSION ASSAY**

This procedure was initially described by (Boorstein and Craig, 1989) and later modified by (Marques et al., 1993) was used for the detection of the excised intron products derived of the splicing reaction. The primer annealing mixture contained 20 mg
of total RNA in 10 mM Pipes (pH 7.5), 400 mM NaCl and 0.2 pmol (300,000 cpm) of (5'-
32P)-labeled primer P, complementary to nucleotides 90 to 87 of the RmInt1 intron
sequence (TGA AAG CCG ATC CCG GAG). The mixture (10 μl) was first heated at 85
°C for five minutes, then incubated at 60 °C for three minutes and cooled slowly to 44 °C.
Extension reactions were started by adding 40 ml of 50 mM Tris–HCl (pH 8.0), 60 mM
NaCl, 10 mM DTT, 6 mM magnesium acetate, 1 mM each of all four dNTPs, 60 mg/ml
of actinomycin D (Sigma), two units of RNaseOUT (Invitrogen) and seven units of AMV
RT (Roche). Reaction mixtures were incubated at 44 °C for 60 minutes. The reaction was
stopped by adding 5 ml of 3 M sodium acetate (pH 5.2) and 150 ml of cold ethanol. Half
of each sample was analyzed on a denaturing (8 M urea) 6% (w/v) polyacrylamide gel.

M13. IN VITRO SPLICING ASSAYS

M.13.1 RNA synthesis and purification

The LT1 transcript (911 nt)(see Figure R3.3) was synthesized from plasmid pLM1
(see Section M3.3 and Figure R3.3), after digesting with BamHI restriction enzyme. The
LT2 transcript (908 nt)(see Figure R3.3) was synthesized from pLM2 (see Section M3.3
and Figure R3.3) linearized with Ndel restriction enzyme, while the ST2 transcript (775 nt)
(see Figure R3.3) was synthesized from the same plasmid linearized with BamHI enzyme
(see Section M3.3 and Figure R3.3).

T7 RNA polymerase transcriptions were performed in presence of 4μg of
linearized plasmid, 1X transcription buffer [10X transcription buffer: 1M MgCl2, 1M Tris-
HCl pH7.5, 2M spermidine, 1M DTT], 0.96mM ATP, 0.96mM CTP, 0.96mM UTP,
0.064mM GTP, 50 μCi of [x-32P]GTP (3000 Ci/mm; 10 mCi/ml; PerkinElmer), 10mM
DTT, 80 U of RNaseOUT (Invitrogen) and 3 μl of T7 RNA polymerase in a final volume
of 50 μl (Pyle and Green, 1994) (See Section M.X). Transcription was stopped after 1.5
hours at 37°C by adding an equal volume of gel loading dye (10 M urea, 0.1% (w/v) xylene
cyanol and bromophenol blue dyes, 40 mM Tris (pH 7.5), 8.3% (w/v) sucrose, and 0.83
mM EDTA, final concentrations) and then loading directly on a 5% (w/v) denaturing (7M
urea) polyacrylamide gel.

Following electrophoresis, the band was visualized by autoradiography for ten
minutes, excised from the gel, and eluted for 3 hours in elution buffer (0.3M NaCl, 10mM
MOPS pH6, 1Mm EDTA, final concentrations ). RNA was then isolated by ethanol
precipitation of the eluate (three volumes of 100% ethanol and 40μg of Glycogen
58
The RNA was dried and resuspended in 30 μl of RNA storage buffer (10 mM MOPS pH 6, 1 mM EDTA, final concentrations).

**M.13.2 Self-splicing reaction**

The splicing reactions were performed using a total volume of 40 μl. The precursor RNA (150000 to 300000 cpm) was preincubated in 80 mM MOPS buffer pH 7.5 at 95 °C for one minute. The RNA sample was then allowed to cool to reaction temperature before combining with 500 mM KCl or 500 mM (NH₄)₂SO₄ (final concentrations) monovalent salts. Reactions were initiated by adding 100 mM MgCl₂ (final concentration) and incubated at 42 °C (for KCl-containing reactions) or 50 °C (for (NH₄)₂SO₄-containing reactions).

To perform timecourses, aliquots (1–2 μl) were removed at the indicated timepoints, added to quench buffer (1.8% sucrose, 1X TBE (Tris-borate EDTA, pH 8.3), 0.018% xylene cyanol dye, 36% (v/v) formamide, and 25 mM EDTA) and placed on ice to stop the reaction. Samples were then loaded onto a 5% denaturing polyacrylamide gel and products were resolved by electrophoresis. The gels were dried and exposed to phosphorus plates (Imaging Plate 2040, FujiFilm). The image was captured after 24–48 hours with the laser scanning system Personal Molecular Imager FX and quantified with the software Quantity One (Bio-Rad).

In preparative-scale self-splicing reactions for product isolation, the total volume of transcript synthesized were loaded onto a preparative (0.750 mm thick) 5% denaturing polyacrylamide gel and products were resolved by electrophoresis. The wet gel was exposed in phosphorus plates for 10-15 minutes, after which they were scanned. The captured image was used as a template for cutting bands which were eluted for 3 hours in elution buffer (0.3M NaCl, 10mM MOPS pH6, 1mM EDTA) and ethanol precipitated (three volumes of 100% ethanol and 40 μg of Glycogen (Roche)). The RNA was dried and resuspended in an appropriate volume of RNA storage buffer (10mM MOPS pH6, 1mM EDTA).

**M.13.3 Identification of splicing products by DNAzyme treatment**

Purified self-splicing product RNA was combined with DNAzyme (about 40 μM final concentration) (Table M8), heated at 95 °C for 2 min for denaturation and then placed on ice for 5 min. The reaction was initiated by adding the reaction buffer [5X: 250 mM MgCl₂, 750 mM NaCl, 200 mM Tris-HCl pH8] and incubating at 37 °C for 2 hours.
The reaction stopped by adding an equal volume of quench buffer (1.8% sucrose, 1X TBE (Tris-borate EDTA, pH 8.3), 0.018% xylene cyanol dye, 36% (v/v) formamide, and 25 mM EDTA). Samples were resolved by PAGE and the gel dried and exposed for 24-48 hours in phosphorus plates (Imaging Plate 2040, Fujifilm). The image was revealed by the laser scanning system Personal Molecular Imager FX and quantified with the software Quantity One (Bio-Rad).

**TABLE M9.** Sequences of DNAzymes used in this work

<table>
<thead>
<tr>
<th>DNAzyme</th>
<th>Sequence¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>DZ1</td>
<td>5’ – gtcatgttgaa GGCTAGCTACAACGA tcttccgtgc</td>
</tr>
<tr>
<td>DZ2</td>
<td>5’ – cgttcataaagc GGCTAGCTACAACGA gcttctctcg</td>
</tr>
<tr>
<td>DZ3</td>
<td>5’ – gtaagcacta GGCTAGCTACAACGA ttcaacctac</td>
</tr>
<tr>
<td>DZ4</td>
<td>5’ – ctgcactcgta GGCTAGCTACAACGA tctgcttctg</td>
</tr>
<tr>
<td>DZ5</td>
<td>5’ – ctcctccagaa GGCTAGCTACAACGA gtgtcgcttga</td>
</tr>
<tr>
<td>DZ6</td>
<td>5’ – gaggagttcga GGCTAGCTACAACGA cagcgggcc</td>
</tr>
</tbody>
</table>

¹ Upper-case letters indicate the catalytic core; lower-case letters indicate left and right base-pairing arms.

**M.13.4 Quantitative analysis of splicing reactions**

Individual bands on autoradiographic images were quantified with the software Quantity One (Bio-Rad). Counts from a nearby blank position on the same lane were substracted as an internal measure of background RNA degradation at each timepoint. The resulted adjusted bands were normalized for decreasing counts with respect to precursor at time t=0. Finally, the relative fractions of precursor and products were calculated and corrected for the guanidine-content of each species.

Fractions of precursor and products were plotted against time. The rate constant of the overall reaction (kₚ) was calculated from the decrease of precursor, which follows a single exponential behavior, incorporating the next expression into the Kaleidagraph program (Abelbeck Software):

\[
[Pr]_t = [Pr]_0 \cdot e^{-k_pr t}
\]
where: \([Pr]_t\) is the amount of precursor molecules at any time \((t)\); \([Pr]_0\) is the amount of precursor molecules at time \(t=0\) and; \(k_r\) is the total rate of reaction.

In time courses carried out in KCl, where a unique population of molecules reacted through hydrolytic pathway (see Section R3.5, Scheme 4), equation (1) was enough for deciphering the hydrolytic rate constant. However, in time courses carried out in \((\text{NH}_4)_2\text{SO}_4\) buffer, where both branching and hydrolytic pathways are observed, branching \((k_{ba})\) and hydrolytic \((k_{hy})\) rates were extracted using a parallel kinetic model for simultaneous branching and hydrolysis reactions. First, fractions of lariat and linear products were determined from the fit of product evolution, following the equations (2) and (3):

\[
\begin{align*}
(2) \quad \text{[PLin]}_t &= [Pr]_0 \cdot f_{hy} (1 - e^{-k_t t}) \\
(3) \quad \text{[PLar]}_t &= [Pr]_0 \cdot f_{br} (1 - e^{-k_t t})
\end{align*}
\]

where: \([\text{PLin}]\) is the total amount of molecules that proceeds from hydrolytic pathway at any time \((t)\); \([\text{PLar}]_t\) is the total amount of molecules that proceeds from branching pathway at any time \((t)\); \(f_{hy}\) is the fraction of linear products and; \(f_{br}\) is the fraction of lariat products. Second, the hydrolytic \((k_{hy})\) and branching \((k_{ba})\) rates were solved from equations (4) and (5):

\[
\begin{align*}
(4) \quad k_{hy} &= \frac{f_{hy}}{f_{hy} + f_{br}} \cdot k_T \\
(5) \quad k_{br} &= \frac{f_{br}}{f_{hy} + f_{br}} \cdot k_T
\end{align*}
\]

where: \(k_{hy}\) is the individual rate of hydrolytic pathway; \(k_{ba}\) is the individual rate of branching pathway and where \(k_T = k_{hy} + k_{br}\). It is also correct that \(f_{hy} + f_{br} = 1\), but it is more accurate to use individual values for \(f\) in the denominator of equation (4) and (5) and not assume that reactions go to completion. The parallel model was verified by the fact that \(k_r\) value obtained from either product evolution or substrate decay were nearby the same, and that the sum of \(f_{hy}\) and \(f_{br}\) approached 1.

In order to dissect the branching and hydrolytic pathways under \((\text{NH}_4)_2\text{SO}_4\) buffer, and calculate all the rate constants, data were fit to equations, which varied depending on the transcript used. In LT1 transcripts (see Section R3.5, Scheme 1), which all products emerged independently and simultaneously from the precursor, following parallel first-order kinetics, the equations that described this reaction were: 

61
\[
[Pr]_t = [Pr]_0 \cdot e^{-(k_1+k_2+k_3)t}
\]

\[
[ILar - E2*]_t = [Pr]_0 \cdot \frac{k_1}{k_1+k_2+k_3} \cdot [1 - e^{-(k_1+k_2+k_3)}] =
= [Pr]_0 \cdot f_{ILar - E2*} \cdot [1 - e^{-(k_1+k_2+k_3)}]
\]

\[
[ILar]_t = [Pr]_0 \cdot \frac{k_2}{k_1+k_2+k_3} \cdot [1 - e^{-(k_1+k_2+k_3)}] =
= [Pr]_0 \cdot f_{ILar} \cdot [1 - e^{-(k_1+k_2+k_3)}]
\]

\[
[PLin]_t = [Pr]_0 \cdot \frac{k_3}{k_1+k_2+k_3} \cdot [1 - e^{-(k_1+k_2+k_3)}] =
= [Pr]_0 \cdot f_{PLin} \cdot [1 - e^{-(k_1+k_2+k_3)}]
\]

where \([Pr]_0\), \([ILar-E2*]_0\), \([ILar]\), and \([PLin]\), represent the amount of molecules of the precursor, lariat/3’-exon intermediate, intron lariat and linear products at any time, respectively; \(k_\nu\), each rate constant indicated in Scheme 1 and \(f_\nu\) the fraction of molecules reacting in each pathway.

In LT2 transcripts (see Section R3.5, Scheme 2), which followed a parallel-sequential kinetic model, the equations that described this reaction were:

\[
[Pr]_t = [Pr]_0 \cdot e^{-(k_1+k_3)t}
\]

\[
[ILar - E2]_t = [Pr]_0 \cdot \frac{k_1}{k_2-(k_1+k_3)} \cdot [e^{-(k_1+k_3)t} - e^{-k_2t}]
\]

\[
[ILin]_t = [Pr]_0 \cdot \frac{k_3}{k_1+k_3} \cdot [1 - e^{-(k_1+k_3)}]
\]

\[
[ILar]_t = [Pr]_0 \left[ 1 - [Pr]_t - [ILar - E2]_t - [ILar]_t \right]
\]

where \([Pr]_0\), \([ILar-E2]_0\), \([ILar]\), and \([ILin]\), represent the amount of molecules of the precursor, lariat/3’-exon intermediate, lariat intron and linear intron at any time, respectively and; \(k_\nu\), each rate constant indicated in Scheme 2.

In ST2 transcripts (see Section R3.5, Scheme 3), which followed a parallel-sequential kinetic model, the equations that described this reaction were:

\[
[Pr]_t = [Pr]_0 \cdot e^{-(k_1+k_4)t}
\]

\[
[ILar - E2]_t = [Pr]_0 \cdot \frac{k_1}{(k_2+k_3) - (k_1+k_4)} \cdot [e^{-(k_1+k_4)t} - e^{-(k_2+k_3)t}]
\]

\[
[ILar]_t = [Pr]_0 \cdot \frac{k_1}{k_3-(k_1+k_4)} \cdot [e^{-(k_1+k_4)t} - e^{-k_3t}] - [ILar - E2]_t
\]

\[
[Unk]_t = [Pr]_0 \cdot \frac{k_1}{k_2-(k_1+k_4)} \cdot [e^{-(k_1+k_4)t} - e^{-k_2t}] - [ILar - E2]_t
\]
\[
(18) \quad [ILin - E2]_t = [Pr]_0 \cdot \frac{k_4}{k_5 - (k_3 + k_4)} \cdot [e^{-(k_1 + k_4)t} - e^{-k_3t}]
\]
\[
(19) \quad [ILin]_t = [Pr]_0 [1 - [Pr]_t - [ILar - E2]_t - [ILar]_t] - [Unk]_t - [ILin - E2]_t
\]

where \([Pr]_0\), \([ILar - E2]_0\), \([ILar]_0\), \([Ilin - E2]_0\), \([Ilin]_0\), and \([Unk]_0\), represent the amount of molecules of the precursor, lariat/3’-exon intermediate, lariat intron, linear/3’-exon intermediate, linear intron and an unknown covalently circular RNA (see section R3.3 for a description) at any time, respectively and; \(k_o\), each rate constant indicated in Scheme 3.

In order to dissect the hydrolytic pathway under KCl buffer, and calculate all the rate constants, data from LT1 (see Section R3.5, Scheme 4) and LT2 (see Section R3.5, Scheme 5) were fit to first order single exponential equations:

\[
(20) \quad [Pr]_t = [Pr]_0 \cdot e^{-k_4t}
\]
\[
(21) \quad [PLin]_t/[ILin]_t = [Pr]_0 \cdot (1 - e^{-k_4t})
\]

where \([Pr]_0\), \([PLin]_0\), and \([ILin]_0\) represent the amount of molecules of the precursor, total products reacting through the hydrolytic pathway and the intron linear at any time, respectively and; \(k_o\), the global rate constant of the hydrolytic reaction, as shown in Schemes 4 and 5.

In ST2 transcripts (see Section R3.5, Scheme 6), which followed a parallel-sequential kinetic model, the equations that described this reaction were:

\[
(22) \quad [Pr]_t = [Pr]_0 \cdot e^{-(k_4 + k_6)t}
\]
\[
(23) \quad [ILin - E2]_t = [Pr]_0 \cdot \frac{k_4}{k_5 - (k_3 + k_6)} \cdot [e^{-(k_4 + k_6)t} - e^{-k_3t}]
\]
\[
(24) \quad [ILin]_t = [Pr]_0 [1 - [Pr]_t - [ILar - E2]_t]
\]

where \([Pr]_0\), \([Ilin - E2]_0\), and \([Ilin]_0\), represent the amount of molecules of the precursor, linear/3’-exon intermediate and linear intron at any time, respectively and; \(k_o\), each rate constant indicated in Scheme 6.

All the rate constants calculated were also verified by an independent fitting by nonlinear least-squares regression using the Dynafit software (Biokin). The variation coefficients associated to rate constants were less than 10 % in all cases, and the values differed in less than 1.5-fold from those calculated from equations above.
RESULTS
CHAPTER 1

CHAPTER 1: DESIGN OF A REPORTER SYSTEM FOR THE QUANTIFICATION OF EXON LIGATION DURING THE IN VIVO SPICING OF RMI INT1 INTRON
R1.1 BACKGROUND

Traditionally, the study of the in vivo splicing of RmInt1 has been focused on the
detection of the excised form of the intron (Muñoz-Adelantado et al., 2003; Molina-
Sánchez et al., 2006; Nisa-Martínez et al., 2007; Molina-Sánchez et al., 2010). This fact has
led up to a major development of tools aimed to fulfill this goal while resting attention to
approaches pointing to the detection of ligated exons as a measure of splicing efficiency.

Until date, plasmid systems used in the study of the in vivo splicing of RmInt1
intron were derivatives of the broad-host-range vector pBBR1MCS, a vector compatible
with IncP, IncQ and IncW group plasmids, as well as with ColE1- and P15a-based
replicons (Kovach et al., 1994; Kovach et al., 1995). In order to obtain plasmids that
constitutively expressed a functional RmInt1, intron derivatives containing the kanamycin
promoter from pKG10 plasmid (M. J. Soto, unpublished results), were constructed:
pKG2.5 and its derivatives (Martínez-Abarca et al., 2000), containing the full-length intron
and; pKGEMA4 and its derivatives, based on ΔORF constructs complemented with the
Intron Encoded Protein (IEP) located upstream or downstream of its ribozyme
counterpart (Nisa-Martínez et al., 2007). Furthermore, in the construction of pKGEMA4,
the context of long exons flanking the RmInt1 intron (-175/+466) from the insertion
sequence IS2011-2 was replaced with a variant of short exons (-20/+5), shown to allow
efficient homing of RmInt1 into DNA in vivo (Jiménez-Zurdo et al., 2003; Nisa-Martínez et
al., 2007) and because long exons has been postulated to promote inactivation of group II
ribozyme constructs (Nolte et al., 1998; Costa et al., 2006a).

To evaluate the exon ligation extent during the in vivo splicing of RmInt1 intron, a
new plasmid vector carrying a reporter system to monitor the splicing events was needed.
For that reason, the pICG series of plasmids were created.

R1.2 DESIGN OF A GENETIC ASSAY BASED ON A lacZ GENE
REPORTER SYSTEM FOR THE EVALUATION OF THE RmInt1
GROUP II INTRON SPICING EFFICIENCY

We developed a splicing reporter system by linking the splicing of RmInt1 to the
constitutive expression of the lacZ gene in pICG plasmids. Previous studies with reporter
genes have been carried out for other group II introns, like the L.lactis II.ltrB expressed in
E.coli (Singh et al., 2002) and group I introns (Mercure et al., 1997; Hong et al., 2008). Other
Figure R1.1. β-Galactosidase activity of *lacZ* gene in pICG constructs carrying the pSyn promoter. A On-plate blue-white discrimination of constructions expressing β-galactosidase protein. Cells from saturated cultures of *E. coli* and *S. meliloti* strains were plated on LB and TY medium, respectively, containing X-Gal (40 mg/l). *E. coli* DH5α and *S. meliloti* RMO17 cultures harboring no plasmid were used as *lac*- controls. B Miller assay of the intronless construct pICG-E1E2 and its precursors. Cultures were grown until exponential-phase and assayed for β-galactosidase activity by adding ONPG (ortho-Nitrophenyl-β-galactoside) (see Material and Methods). *E. coli* DH5α and *S. meliloti* RMO17 cultures were used as background measures of β-galactosidase activity. Error bars represent the standard error of the mean.

Screenable markers, such as the GFP gene, have also been used in other group II introns (Cui et al., 2004; Rawsthorne et al., 2006; Gu et al., 2010).

pICG plasmids are derivative of pMP220, an Inc-P broad host range transcriptional fusion vector, designed to measure transcriptional activity as β-galactosidase activity starting from a promoter cloned in the multicleoning site (Spaink et al., 1987) (See Figure M1, Material and Methods). pMP220 contains the *E. coli lacZ* gene, which constitutes the reporter marker, lacking the *lac* promoter and operator. Moreover, the *lacZ*
gene derived of the pMP220 vector was devoid of the first eight codons, which are not essential for enzymatic activity (Casadaban et al., 1980); instead, 57 amino-acid residues of the fragment containing the ribosome binding site derived of the chloramphenicol acetyl transferase gene (CAT) were added in the 5’ end of the lacZ gene to ensure a efficient translation of transcripts (Spaink et al., 1987). The use of pMP220 as the basis for the pICG system allowed us, on the one hand, the analysis of the splicing of RmInt1 in its natural host, S. meliloti, as well as in E.coli, due to its wide-ranging specificity of DNA replication initiation at origin and; on the other hand, because it could coexist with pBBR1MCS-derived plasmids (Kovach et al., 1994; Kovach et al., 1995) in E.coli and S. meliloti or with pBR322-derived plasmids (Bolivar, 1978) in E.coli, allowing assays of complementation in trans.

pICG1 was constructed by cloning the synthetic promoter pSyn (Giacomini et al., 1994) in the polylinker of pMP220. This promoter presented a consensus in the region from -50 to +5 with respect to the putative tsp (transcription start point) (+1) from known E.coli promoter sequences. Besides, the -35 (TTGACA) and -10 (TATAAT) regions were

![Diagram](image-url)

**Figure R1.2.** Strategy used in the construction of a genetic assay for the splicing of RmInt1 linked to the expression of β-galactosidase protein. **A** Schematic diagram of the nucleotide sequence of the lacZ mRNA in pICG-E1E2 and the ligated exons resulting from RmInt1 splicing, presented with the corresponding amino-acid sequence. This region consists of 42 nucleotides, including 25 nucleotides corresponding to the ISrwm2011-2 flanking exons and 17 corresponding to the additional PmA, Bln1 and BamHI cloning sites linked in-frame to the lacZ gene. Upper case letters indicate the lacZ sequence, whereas lower case letter indicate the exon flanking sequences and restriction sites added. Additional amino acids due to the addition of nucleotides to the residual lacZ remaining after intron excision are identified by asterisks. A vertical arrow indicates the site of intron insertion in the intron-containing constructs. **B** A diagram showing splicing of RmInt1 intron linked to the expression of β-galactosidase. Intronsplicing leads to the synthesis of a β-galactosidase protein with a short extension (14 amino-acid residues) corresponding to the ligated exons, whereas, in the absence of splicing, translation is terminated due to the presence of multiple stop codons within the intron, giving a lac phenotype.
separated by 17 nt, reported to be the optimal distance (de Boer et al., 1983; Giacomini et al., 1994). pICG1 displayed a constitutive and strong β-galactosidase activity in E. coli strain (10680 ± 60 Miller Units), in agreement with previous assays with this promoter (Giacomini et al., 1994), which was 180-fold the activity shown by pMP220 devoid of promoter (Figure R1.1). However, pICG1 expressed in S. meliloti host presented an 84% reduction in β-galactosidase activity with respect to the same construction expressed in E. coli, which is clearly perceptible in the XGal-based colorimetric assay on plate (Figure R1.1 A). This decrease is not well understood, since other related constructions (pICG11 and pICG12, see below), also expressed in S. meliloti, showed a similar level of activity than pICG1 expressed in E. coli.

pICG11 and pICG12 were removed of repeated BlnI and BamHI restriction sites located in the polylinker and after lacZ gene, respectively, aimed to keep only one copy of each for cloning purposes. These modifications resulted in an increase in the β-galactosidase activity both in E. coli and in S. meliloti host (Figure R1.1 B), although no changes were introduced in the lacZ gene. However, it is probably that length or sequence modifications in the polylinker could provoke an enhancement of the transcription or translation rates. Again, the β-galactosidase activity was about 1.4-fold greater in E. coli than in S. meliloti, in both cases.

pICG-E1E2 was the result of the inclusion of the minimal flanking exon sequences (-20/+5 nt length) from the ISRm2011-2 insertion sequence in the unique BamHI site of pICG12. This intron context has been shown to improve the efficiency of RmInt1 splicing both in vivo (Nisa-Martínez et al., 2007) and in vitro (Costa et al., 2006a; Costa et al., 2006b). It is not easy to insert intron RNA into reporter genes, because group II intron splicing requires base-pairing between the intron sequences EBS1, EBS2 and EBS3 located in DI and the 5’ and 3’-exon sequences IBS1, IBS2 and IBS3 (EBS and IBS indicate exon-binding and intron-binding sites, respectively). We solved this problem by introducing the flanking exon sequences from the ISRm2011-2 insertion sequence in frame with codon 58 of the lacZ gene, making a transcriptional and translational lacZ fusion (Figure R1.2 A). The resulting lacZ gene was increased by 42 base pairs which corresponded with a short extension of 14 amino-acid residues when translated (Figure R1.2 B), giving a fusion β-galactosidase protein with high hydrolase activity (Fig R1.1 B). Although a significant reduction in activity was revealed from the Miller assay in pICG-E1E2 construction with respect to pICG12 (about 20 % less β-galactosidase activity both in E. coli and S. meliloti, likely due to the elongation of the N-terminal end of the protein), the level of expression was about 8000 Miller Units, so it was considered to be active enough for being the basis for the rest of pICG constructs (Fig R1.1 B).
### A

<table>
<thead>
<tr>
<th>RmInt1 intron construction</th>
<th>Complementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-length intron</strong></td>
<td></td>
</tr>
<tr>
<td>pSyn E1 E2 lacZ</td>
<td>pICG-E1E2</td>
</tr>
<tr>
<td>pSyn E1 RmInt1 intron E2</td>
<td>pICG-WT</td>
</tr>
<tr>
<td>pSyn E1 RmInt1 intron E2</td>
<td>pICG-DV</td>
</tr>
</tbody>
</table>

**IEP-ΔORF**

| pSyn IEP E1 E2 lacZ       | pICG-IEP-E1E2   |
| pSyn IEP ΔORF E2 lacZ    | pICG-IEP-ΔORF  |
| pSyn IEP ΔORFΔDV E2 lacZ | pICG-IEP-ΔORFΔDV |

**ΔORF**

| pSyn E1 E2 lacZ           | pICG-E1E2       |
| pSyn E1 ΔORF E2 lacZ     | pICG-ΔORF       |
| pSyn E1 ΔORFΔDV E2 lacZ  | pICG-ΔORF       |

**Ter**

| pSyn ΔORF                 | pICG_{ter}ΔORF |
| pSyn ΔORF                 | pICG_{ter}ΔORF |

---

pKm 1EP

pKG0IEP 6.5 Kb

(Muñoz-Adelantado et al., 2003)

---

pKm 1EP

pKG0IEP 6.5 Kb

(Muñoz-Adelantado et al., 2003)
Figure R1.3. Constructions used in the in vivo splicing analysis of *S. meliloti* RmInt1 and *L. lactis* LLlTrB group II introns. **A** Schematic diagram of RmInt1 intron-based constructions in the pICG system. Four intron constructions were created, each belonging to a series of plasmids: the wild-type construct, pICG-WT, in the full-length intron series, contains the full-length RmInt1 intron (1.9 kb) with the protein encoded by a sequence in DIV; pICG-IEP-ΔORF, in the IEP-ΔORF series, and pICG-ΔORF, in the ΔORF series, that contain a 0.74 kb intron in which the IEP is either encoded by a sequence just upstream from the 5′-exon/*lacZ* sequence or constitutively produced from a compatible pKGIIEP plasmid, respectively. As a control for the ligated exons expected from the constructs described above, an intron-less construct, pICG-E1E2, consisting of the RmInt1 short flanking exons fused in frame with the *lacZ* gene, was used in which the IEP was either encoded by a sequence in the cis position (pICG-IEP-E1E2) or by a sequence in trans (pICG-E1E2+IEP), from compatible pKGIIEP plasmids. As a negative control in these splicing assays, plasmids containing splicing-defective mutants, with a mutation in the critical conserved AGC-GUU pairing of intron RNA domain V (GUU converted to CGA) were generated. These constructs were named pICGDV, pICG-IEP-ΔORFDV and pICG-ΔORFDV+IEP, for the corresponding full-length intron, and for the 0.74 kb intron with the IEP sequence expressed in cis and in trans, respectively. **B** Schematic diagram of LLlTrB intron-based constructions. All LLlTrB constructions are based in the full-length intron, with the LtrA protein encoded by a sequence in the RNA DIV. pICG-LLlTrB is the wild-type intron in the pICG system, with the LLlTrB intron cloned in sense with respect to the promoter orientation, allowing RNA and LtrA protein to express. As a control for the ligated exons of the above construction, the intronless construct, pICG-llTrB, where the intron flanking exons fused in-phase with the *lacZ* gene, was used. As a negative control of splicing, pICG-LLlTrBA, harboring a wild-type LLlTrB intron cloned in antisense with respect to the promoter pSyn, was used. The pLI1 construction (Matsuura *et al.*, 1997), carrying the wild-type LLlTrB intron under the inducible T7 promoter, was used as a control of expression and as a PCR template for cloning. The red and blue arrow-head represents the pSyn and T7 promoters, respectively. The two flanking exons (E1 and E2), the RmInt1 or LLlTrB introns, and the intron-encoded protein genes are indicated. The *lacZ* gene is indicated as a blue arrow. The transcription terminator is represented by a stem-loop hairpin in thick black lines.
The first RmInt1 intron construction in the pICG reporter system, the wild-type construct, pICG-WT, contained the full-length RmInt1 intron (1.9 kb) with the protein encoded by a sequence in DIV (Figure R1.3). This construction was the result of the invasion of the intronless construct, pICG-E1E2, with the RmInt1 intron expressed from the donor intron construction pKG-2.5 (Figure M2, See Materials and Methods). pICG-WT could potentially express both the intron RNA and the IEP protein, producing active RNP particles.

We carried out double-plasmid mobility assays to check that the pICG-WT plasmid was fully active. The pICG-WT plasmid harboring the RmInt1 intron was transferred to the RMO17 strain (an intron-less S. meliloti strain) by conjugation, together with the acceptor construct pBB0.6+ containing the RmInt1 homing site (-175/+466) (Figure R1.4, lane 3). As negative control, pBBΔ129, a recipient construct lacking the homing site, was used (lanes 2, 4) (Martínez-Abarca et al., 2000). The recipient plasmids pBB0.6 and pBBΔ129, not conjugated with the donor plasmid pICG-WT, were also assayed (lanes 1, 2). We analyzed the invasion of the recipient plasmids, pBB0.6 and pBBΔ129, by Southern Blotting Hybridizations by using an ISRm2011-2 exon 2 and an intron probes (Figure R1.4). Plasmid analysis for the transconjugants showed that almost 100% of the recipient plasmids were invaded by the RmInt1 intron (lane 3, exon 2 and intron probes). In contrast, the negative control (lane 4) and the recipient plasmids not complemented with the donor constructions (lanes 1 and 2) only hybridized with the exon 2 specific probe, and did not exhibit any hybridization signal with the intron specific probe.

![Figure R1.4](image_url)

*Figure R1.4. Homing assay of wild-type RmInt1 in the pICG reporter system. For homing assays, plasmid pools from RMO17 cells harbouring donor (pICG-WT) and recipient (pBB0.6) plasmids were blotted onto a nylon membrane and analyzed by Southern hybridization with DNA probes specific to the insertion sequence ISRm2011-2 (Exon 2 Probe, middle panel) and RmInt1 intron (Intron Probe, right panel). Recipient plasmid pBBΔ129 was used as negative control in the assays (lane 5) along with recipient plasmids pBB0.6 and pBBΔ129 not complemented with the donor construction (lanes 2 and 3).
A splicing defective construction, pICG-DV, was used as a negative control of the full-length wild-type pICG-WT plasmid in splicing analysis. This construction derived from pKG2.5-DV, which has a mutation in the critical conserved pairing AGC-GUU of intron RNA domain V (GUU → CGA) (Muñoz-Adelantado et al., 2003) (Figure R1.3, Figure M2, see Materials and Methods).
Figure R1.5. Characteristics and effect of the transcription terminator in pICG constructs. A Schematic of the transcription motif added to the pICG-ΔORF constructs in the *ter* series of plasmids. The terminator consists of a short stem-loop hairpin surrounded by 15 bases rich in adenine nucleotides on the 5′ end of the motif, called the “A-tail” and 15 bases rich in thymine nucleotides on the 3′ side called the “T-tail” (Adapted from Kingsford *et al* (2007). Below, the DNA sequence of the synthetic transcription terminator. Restriction sites added to the ends of the oligonucleotide are indicated. The motif was introduced in the unique *Ban*HI restriction site in pICG12. Nucleotides of the stem are highlighted in red. B Quantitative analysis of the effect of the transcription terminator in the expression of *laiZ* gene in pICG-ΔORF and pICG<sub>TER</sub>-ΔORF constructs in *S. meliloti* by real-time RT-PCR. The map of the *laiZ* region of pICG-ΔORF and pICG<sub>TER</sub>-ΔORF are depicted on the top. Green arrows under the map indicate the primers used in the real-time RT-PCR assay to amplify the *laiZ* gene (pICG-ΔORF construct, left) or the region located after the transcription terminator (pICG<sub>TER</sub>-ΔORF construct, right). Error bars represent the standard error of the mean.

pICG-ΔORF contained a 0.74 kb intron whereas the IEP protein was constitutively produced from a compatible pKGIEP plasmid (the *trans* position, as described in Muñoz-Adelantado *et al* (2003)) (Figure R1.3) as previously carried out for LL.ltrB group II intron *in vivo* (Matsuura *et al*., 1997; Cousineau *et al*., 1998; Mohr *et al*., 2000; Cui *et al*., 2004; Watanabe and Lambowitz, 2004). The IEP produced in trans could potentially bind to the intron RNA domain IV, similarly as in LL.ltrB intron (Wank *et al*., 1999; Matsuura *et al*., 2001; Singh *et al*., 2002; Noah and Lambowitz, 2003), promoting splicing of the RNA in the ΔORF construct. The negative control was comprised of a splicing defective construction, pICG-ΔORFDV, derived of pKGEMA4-DV.

pICG-IEP-ΔORF also contained a 0.74 kb intron in which the IEP was encoded by a sequence just upstream from the 5′-exon/*laiZ* sequence (referred to as the *cis* position; Figure R1.3) (Nisa-Martínez *et al*., 2007). This intron construct was derived of pKGMA4, along with its splicing defective counterpart, pICG-IEP-ΔORF, derived from pKGEMA4-DV, also impaired in the catalytic triad AGC of intron RNA domain V.

As a control for the ligated exons expected from the constructs described above, an intron-less construct, pICG-E1E2, consisting of the RmInt1 short flanking exons fused in frame with the *laiZ* gene, was used in which the IEP was either encoded by a sequence in the *cis* position (pICG-IEP-E1E2) or by a sequence in trans (pICG-E1E2+IEP), from compatible pKGIEP plasmids.

In order to get transcripts of a defined size, with homogenized 3′ ends of RNA precursor, for being analyzed on Northern Blotting assays as discrete bands, an intron ΔORF derivative construct was created in which the expression of the *laiZ* gene was interrupted by the introduction of a transcription terminator (Figure R1.5, Figure R1.3). The rho-independent synthetic transcription terminator was based on that designed by
Kigsford et al. (2007), which accomplished the requirements for intrinsic termination (Rosenberg and Court, 1979). It was constituted by a short, low-energy hairpin followed downstream by a stretch of thymine nucleotides (which are transcribed into uracils). On the 5’ side there was another stretch of adenine nucleotides, because the same sequence can function as a terminator on both strands (Kingsford et al., 2007) (Figure R1.5 A and B). No other recently described parameters were analyzed, such as the effect of the DNA downstream sequences (Nudler and Gottesman, 2002; Martinez-Trujillo et al., 2010).

To test the termination efficiency in vivo, the level of mRNA amplified in ΔORF constructions with and without terminator was compared (Figure R1.5 C). The termination efficiency varied from about 83 % in the intron construct to about 99.5 % in the splicing defective construction. This meant that RNA polymerase significantly paused at the hairpin, giving the chance to the destabilization of the rU-dA RNA-DNA hybrid, which has an unusually weak base-paired structure, and so terminating the transcription.

**R1.4 INTRODUCTION OF THE Lactococcus lactis L.ltrB INTRON IN THE REPORTER SYSTEM OF pICG PLASMIDS**

Since it has been reported that the comparison of values obtained from independent Miller assays may lead to inaccurate results, unless all experimental parameters are standardized (Giacomini et al., 1992), we performed a direct comparison of L. lactis L.ltrB intron and S. meliloti RmInt1 intron by introducing the L.ltrB intron in the reporter lacZ gene of pICG plasmids, as previously performed (Giacomini et al., 1994). Since L.ltrB intron is the most studied bacterial group II intron (Dunny and McKay, 1999; Klein and Dunny, 2002), many aspects of its molecular biology and in vivo behavior have been elucidated, enabling us to determine key aspects of the RmInt1 behavior.

Traditionally, L.ltrB intron has been studied in L. lactis within the ltrB relaxase gene in the natural context of the 48-kb conjugative element pRS01 (Mills et al., 1996; Chen et al., 2005). The copy number of pRS01 is less than 1 molecule per cell, and it likely integrates reversibly into the chromosome (Chen et al., 2005). In E. coli, several L.ltrB-containing plasmids were used for the splicing analysis of L.ltrB. Some of them are pLI1, a pET-11a derivative which expressed the wild-type full-length intron with long flanking exons (-176/+209) under the T7 promoter (Matsuura et al., 1997); pLE12-ΔORF, a pLE12-derivative (Mills et al., 1996), deleted of the intron-encoded LtrA protein coding region (Matsuura et al., 1997) or; the wild-type intron in the pALG1 series with minimal
flanking exons (−58/+38), the full-length construct pALG1 and the ΔORF-intron construct pALG2, both containing \textit{lreB}/GFP fusions (Cui \textit{et al.}, 2004).

pICG-LlLtrB contained the wild-type L.\textit{lreB} intron and short flanking exons (−58/+38) (Cui \textit{et al.}, 2004), cloned in sense orientation with respect to the promoter, so that this construct could potentially express both the intron RNA and the \textit{LtrA} ORF, the latter with its own promoter sequence within the intron (Zhou \textit{et al.}, 2000) and Shine–Dalgarno-like sequence for translation (Matsuura \textit{et al.}, 1997). For comparing, pL11 plasmid, in which L.LtrB intron is expressed under the inducible T7 promoter in \textit{E.coli}, was used in the splicing assays (see Chapter 2).

For negative control of splicing, an L.LtrB construction in antisense orientation with respect to the direction of DNA replication was used. As not being expressed the \textit{LtrA} protein, no splicing events were predicted to occur. As a control for the ligated exons, the construction pICG-\textit{lreB}, which contained the sequence of ligated \textit{lreB} flanking exons 1 and 2 (−58/+38), was used.

R1.5 POTENTIAL OF THE pICG REPORTER SYSTEM IN THE STUDY OF THE BIOLOGY OF \textit{S.meliloti}

In this work, pICG reporter system will be use in the analysis of splicing of the RmInt1 intron. Nevertheless, due to its characteristics, it could be useful for many other purposes:

- The intronless construct pICG-E1E2 contains the minimal target site for RmInt1 insertion. Conjugated along with a compatible intron donor construct, pICG-E1E2 can participate as a recipient plasmid in homing assays. Moreover, by using \textit{XbaI} restriction sites, located on both sides of the \textit{lacZ} gene, a 3.5 kb band harboring the inserted intron is released, suitable for Southern Blotting hybridizations with an exon 2 probe.
- It could be used in plant-based infection experiments for the localization of \textit{S.meliloti} bacteria on nodules due to the constitutive expression of the reporter gene \textit{lacZ}.
- Under the pSyn promoter, the \textit{lacZ} cassette can be replaced by digestion with \textit{BamHI} and \textit{NheI} restriction enzymes with another reporter gene. The RmInt1
target site could be introduced later in the BamHI site, bearing in mind not to impair the phase.

- Furthermore, other intronic elements could be potentially introduced in the BamHI site of lacZ gene, as the rybozime DIV for studying its interaction with the IEP protein as previously reported (Singh et al., 2002; Gu et al., 2010).
CHAPTER 2

*In vivo* quantification of the *S. meliloti* RmInt1 group II intron splicing. Comparison with the *L. lactis* Ll.ltrB group IIA intron expressed in *E. coli*
R.2.1 BACKGROUND

Since the discover of the first bacterial group II intron (Ferat and Michel, 1993), more than 30 full-length group II introns 9 partial group II introns had been identified in Eubacteria by 2002 (Dai et al., 2003) and about 20 more in Archaeabacteria (Dai and Zimmerly, 2003). Most of these introns were identified through sequence analysis in microbial genome sequencing projects or directly through genome database searches (Dai and Zimmerly, 2002). However, only a minority of bacterial group II introns have been unequivocally shown to be functional or to undergo splicing in vivo. LLtrB from Lactococcus lactis ML3 (Mills et al., 1996) and a group II intron similar to LLtrB identified in the sex factor from L. lactis 712 (Shearman et al., 1996); the Sinorhizobium meliloti RmInt1 (Martínez-Abarca et al., 1998); B.a.II and B.a.I2 from Bacillus anthracis (Robart et al., 2004); a Clostridium difficile intron (Roberts et al., 2001) and the Avi.groEL intron from Azotobacter vinelandii, which inserts within the termination codon of the essential groEL gene (Adamidi et al., 2003; Ferat et al., 2003).

While organellar group II introns are frequently located in essential genes (e.g. NAD dehydrogenase, ribulose-bisphosphate carboxylase, cytochrome oxidase), which require efficient transcription and splicing (Dai et al., 2003), nearly all bacterial group II introns are associated with mobile DNAs or even they are located outside genes, as those of bacterial class C, which insert exclusively after terminator structures. As far, no bacterial group II introns have yet been found in highly conserved or housekeeping genes. It is therefore possible that many of these bacterial group II introns are transcribed only weakly, if at all, in the cell.

RmInt1 splicing has been investigated principally by analyzing excised intron products. These analyses have shown that the maturase (X domain), including the conserved C-terminal tail of the IEP, is required for splicing (Molina-Sanchez et al., 2010). The excised intron products have been characterized in detail and shown to consist mostly of lariat forms, but with some putative circular forms (Molina-Sánchez et al., 2006). Exon junction products have been detected in vivo only in RT-PCR assays (Martínez-Abarca et al., 1998; Martínez-Abarca and Toro, 2000). Moreover, in vitro, it has been shown that exon ligation is unusually inefficient for RmInt1 (Costa et al., 2006a), partly because nucleotides 5 to 11 of the 3’ exon match the EBS1 5'-exon-binding site much better than the authentic IBS1 sequence in the 5’ exon (Costa et al., 2006b). For the reasons exposed above, we rationalized to develop an accurate study of the whole splicing reaction of RmInt1 group II intron. To achieve this purpose, we made use of the new context of pICG plasmid
system for analyzing all the products obtained in the splicing reaction, including excised intron and ligated exons, in vivo.

### R.2.2 THE *IN VIVO* EXCISION OF RmInt1

RmInt1 excision products were assessed by carrying out primer extension analysis on pICG constructs (Figure R2.1), as previously described for other plasmid constructs (Muñoz-Adelantado et al., 2003; Molina-Sánchez et al., 2006; Nisa-Martínez et al., 2007). In these assays, the excised lariat intron RNA (L) is detected as an extension product of 97 nt, together with larger bands derived from unspliced precursor RNA molecules (Pr) (Figure R2.1 B). In *S. meliloti*, the expected 97 nt product was obtained from constructs containing the RmInt1 intron (lanes 2, 5 and 8), whereas no band was obtained from intron-less constructs (lanes 1, 4 and 7) or from splicing-defective constructs with a mutation in DV of the intron (lanes 3, 6 and 9). RmInt1-ΔORF intron excision was three times more efficient when IEP was expressed in *trans* (lane 8) than when expressed in *cis* (lane 5), being almost as efficient as that observed with the wild-type intron construct (lane 2). This result contrasts with those for the LL.ltrB intron, for which ΔORF intron splicing occurred at wild-type levels when the IEP sequence was expressed in *cis*, but at much lower levels expressed in *trans* from a separate plasmid (Cui et al., 2004). We analyzed the pICG constructs in *E. coli*, and found that the amount of excision product obtained was 95% lower than that obtained in *S. meliloti* for constructs containing RmInt1 intron (lanes 11, 14 and 17). No band was observed for intron-less (lanes 10, 13 and 16) and splicing-defective (lanes 12, 15 and 18) constructs, as expected. Thus, the efficiency of RmInt1 excision depends on the bacterial host.

*Figure R2.1* In vivo excision of the RmInt1 intron in *S. meliloti* and *E. coli*. A Primer extension assays were performed on total RNA extracted from *S. meliloti* (lanes 1-9) and *E. coli* (lanes 10-18) harboring the indicated constructs. RNA was reverse-transcribed with a 32P [γ-ATP] 5'-labeled primer complementary to the first 97 nt of RmInt1. The products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel, which was dried and quantified with a PhosphorImager. The major cDNA product (97 nt, indicated with a dot) for the wild-type and ΔORF constructs (lanes 2, 5, 8, 11, 14 and 17) corresponds to the excised intron RNA. Larger products derived from unspliced precursor RNAs were also detected (110 and 126 nt). As negative controls, we used DV mutant intron constructs (lanes 3, 6, 9, 12, 15 and 18) and intron-less constructs (E1F2, Lanes 1, 4, 7, 10, 13 and 16). B Schematic diagram of primer extension on linear unspliced precursors and excised intron lariats, showing the expected extension products. The bulged A in DVI is also indicated.
### A

Unspliced Precursors (Pr) → Excised Intron Lariat (L)

**B**

**S. meliloti**

<table>
<thead>
<tr>
<th>Lane</th>
<th>126nt</th>
<th>110nt</th>
<th>97nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**E. coli**

<table>
<thead>
<tr>
<th>Lane</th>
<th>126nt</th>
<th>110nt</th>
<th>97nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

**% excision**

<table>
<thead>
<tr>
<th>Construct</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1E2 WT</td>
<td>100</td>
<td>34</td>
<td>92</td>
</tr>
<tr>
<td>ΔE1E2 DVI</td>
<td>3</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>ΔIEP&lt;sub&gt;es&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔEORF&lt;sub&gt;1&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔORF&lt;sub&gt;DVY&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**R.2.3 RmInt1 SPlicing IN S. meliloti**

The splicing products of the RmInt1 intron were analyzed in *S. meliloti*-host RNA samples, in which *in vivo* excision was the most efficient (Figure R.2.1).

First, we attempted to detect the cDNA from ligated exon RNAs with a primer extension assay from an exon 2 primer (primer P2; Figure R.2.2 A, B). The positive control, the intronless construct pICG-E1E2 (Figure R.2.2 B; lane 1), showed a cDNA triplet when extended from the primer P2. The longest 5' exon end corresponds to the predicted start site at the pSyn promoter of pICG plasmid (Giacomini *et al.*, 1994), whereas the other 5' ends were empirically determined to be 19 nt and 29 nt shorter than the longest one, respectively, and may reflect the creation of a new transcription start site (Figure R.2.2 B). However, neither in the full-length wild-type construct, pICG-WT (lane
Figure R2.2 Attempt of detection of ligated exons in the RmInt1 splicing. A RNAs from pICG plasmids expressed in S. meliloti was extracted and reverse transcribed with exon 2-specific primer P2. The cDNA bands, of lengths indicated to the right of each panel, correspond to ligated exons (E1E2), which are shown as a triplet in the intronless construct pICG-E1E2 (lane 1). No bands were detected in the intron-containing constructs, neither for the wild-type (lane 2) nor for the RmInt1-ΔORF intron construct (lane 3). The splicing defective construction pICG-ΔORFDV was used as negative control. B Schematic of RNAs and cDNAs products. RNA is depicted as a solid line and cDNAs as broken lines. Exons are represented as green (E1) and yellow (E2) boxes.

2), nor in the ΔORF construct, pICG-ΔORF + IEP (lane 3), ligated exons could be detected. As a negative control, the splicing defective construction pICG-ΔORFDV + IEP, was used (lane 4).

We also analyzed the splicing products of the RmInt1 intron by carrying out northern blotting (Figure R2.3). In this approach, the RmInt1-ΔORF intron was used, due to its shorter size respect to full-length constructions. In addition, we introduced an intrinsic transcription terminator (pICGter constructions, see Materials and Methods and Chapter 1) into the constructs, for the identification of RNA precursors of a precise size (Figure R1.5, Figure R2.3 B, lanes 1-6). As control for this element, we included constructs without the Rho-independent transcription terminator (lanes 7-9). A set of negative control samples, in which the IEP was absent, was used to identify the precursors and transcript forms. Precursor accumulated only for constructs containing the transcription terminator (lanes 2, 3, 5 and 6). The precursor was resolved as a triplet, because transcription is initiated by the Syn promoter at three different positions, as explained above. The lower intensity of the precursor band for intron-bearing constructs (lanes 2, 3, 5 and 6) than for intron-less constructs (lanes 1 and 4, left and right panel) probably reflects a lower efficiency of the RNA polymerase with longer templates and a greater susceptibility of the larger precursor RNA to nucleolytic degradation. Nevertheless, no traces of ligated exons were observed with the ΔORF construct complemented with the
Figure R2.3 RmInt1 splicing products in S. meliloti. A Schematic diagram showing the splicing products identified with the various probes used. The 5’ and 3’ exon probes hybridize to the precursor RNA and ligated exons, whereas the intron probe hybridizes to the precursor RNA, excised lariat intron and broken lariat intron. Exons are depicted in green (E1) and yellow (E2), the intron is shown as a solid black line with a circle indicating the bulging adenosine in DVI, the transcription terminator is a stem-loop hairpin and the probes are represented by a bold red line. The pSyn promoter includes three origins of transcription, generating three bands for the precursor or ligated exons, the size of which is indicated. B Northern blot hybridizations were performed with 5 μg of total RNA extracted from S. meliloti harboring the indicated pICG plasmids with (lanes 1-6; ter series) and without (lanes 7-9) the Rho-independent transcription terminator and complemented (lanes 4-9) or not (lanes 1-3) with the IEP in trans. The splicing products detected are indicated to the right of the panels. Three bands corresponding to the expected ligated exons are visible only in bacteria containing intron-less constructs of the ter series (Ter-E1E2; lanes 1, 4). Excised and broken lariat bands are detectable only in samples containing ΔORF constructs complemented with IEP (lanes 5 and 8). Other bands corresponding to nonspecific cleavage or cross-hybridization products of about 600 nt are indicated (x).

IEP (lane 5, left and right panel). Lariat intron forms were also observed for RmInt1 constructs (lanes 5 and 8, middle panel). Their electrophoretic mobility was low due to their three-dimensional configuration. The absence of ligated exons cannot be accounted for by the accumulation of intermediates of the splicing reaction steps, as no lariat-3’ exon intermediates were observed, as shown by the lack of bands on hybridization with 3’ exon probe (lanes 5 and 8, right panel). Further, a 740 nt band was shown, probably corresponding to a broken lariat form because it was detected only with the intron probe for constructs containing ΔORF complemented with the IEP in trans (lanes 5 and 8, middle panel), with no bands detected for the intron-less (lanes 1, 4 and 7, middle panel) and DV mutant (lanes 3, 6 and 9) constructs. The other bands observed on hybridization with the 3’ exon probe correspond to nonspecific cleavage products (lanes 2, 3, 5 and 6, right panel, x products). Our results provide direct evidence that RmInt1 splicing in S. meliloti is inefficient, with the accumulation of the intron lariat not accompanied by the accumulation of ligated exons.

R.2.4 QUANTITATIVE ANALYSIS OF RmInt1 EXPRESSION AND SPLICING

No ligated exons were detected on northern blotting. We therefore used a reporter gene system to determine the extent to which exon ligation occurred during RmInt1 splicing in vivo. In this system, intron excision is linked to lacZ gene expression. Quantification was carried out in S. meliloti at its growth temperature (28°C) and in E. coli at
37°C and 28°C (Table R2.1). The intron-less construct of each plasmid series was used to test the system and the reaction conditions. With these constructs, β-galactosidase activity ranged from about 4000 Miller units in the cis series in S. meliloti to 15000 Miller units in the wild-type series in E. coli grown at 37°C, demonstrating an effect of temperature on the enzymatic activity of β-galactosidase in E. coli. However, constructs harboring introns had very low levels of activity (from 5 to 20 Miller units) in both S. meliloti and E. coli and, in most cases, activity levels were similar to those obtained with the splicing-defective DV mutant. Thus, there are presumably too few ligated exons for quantification by this enzymatic method.

**TABLE R2.1. RmInt1 intron splicing measured as β-galactosidase activity**

<table>
<thead>
<tr>
<th>Construction</th>
<th>E. coli 37°C</th>
<th>E. coli 28°C</th>
<th>S. meliloti 28°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1E2</td>
<td>15000 ± 300</td>
<td>10300 ± 300</td>
<td>6580 ± 20</td>
</tr>
<tr>
<td>WT</td>
<td>18 ± 1</td>
<td>14.1 ± 0.8</td>
<td>5 ± 1</td>
</tr>
<tr>
<td>DV</td>
<td>16.9 ± 0.3</td>
<td>12 ± 1</td>
<td>7.5 ± 0.3</td>
</tr>
<tr>
<td>IEP <em>trans</em> series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1E2</td>
<td>13600 ± 400</td>
<td>10300 ± 300</td>
<td>8400 ± 300</td>
</tr>
<tr>
<td>ΔORF</td>
<td>20.4 ± 0.6</td>
<td>19 ± 1</td>
<td>14.4 ± 0.3</td>
</tr>
<tr>
<td>ΔORF DV</td>
<td>24 ± 2</td>
<td>22 ± 1</td>
<td>21.3 ± 1.3</td>
</tr>
<tr>
<td>IEP <em>cis</em> series</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E1E2</td>
<td>9900 ± 200</td>
<td>5700 ± 200</td>
<td>3980 ± 70</td>
</tr>
<tr>
<td>ΔORF</td>
<td>18.4 ± 0.6</td>
<td>15.3 ± 0.9</td>
<td>12 ± 1</td>
</tr>
<tr>
<td>ΔORF DV</td>
<td>18.6 ± 0.3</td>
<td>17.9 ± 1.1</td>
<td>15.1 ± 0.9</td>
</tr>
</tbody>
</table>

* β-galactosidase activity is expressed in Miller Units as described in the Materials and Methods

We developed a quantitative real-time reverse transcription-PCR (qRT-PCR) assay, to quantify the ligated exons generated by intron splicing and to analyze RmInt1 and laiZ transcript levels in the context of pICG plasmids, as previously carried out for L.LitrB in L. lactis (Chen et al., 2005). Absolute quantification was used to determine mRNA levels in the wild-type series of plasmids. Three primer pairs were used to amplify laiZ1, RmInt1 intron and ligated β-galactosidase gene laiZ produced as a consequence of the splicing of RmInt1 (Figure R2.4 and Table M.5, see Materials and Methods). All transcripts displayed constitutively high levels of expression (Figure R2.4 B). The amount of laiZ1 mRNA...
(primer pair 1-2), which represents total mRNA transcribed, was almost constant in all samples. The RmInt1 intron mRNA (primer pair 3-4) is produced only from constructs harboring the intron (pICG-WT and DV), both excised and being part of the precursor mRNA. RmInt1 mRNA levels were highest, due to accumulation of the more stable lariat form of the intron. The ligated β-galactosidase gene lacZ transcripts (primer pair 5-6) are produced by intron excision and ligation of the 5’ and 3’exons containing lacZ1 and lacZ2 fragments. To avoid amplification of precursor molecules by primers 5-6, the extension step in the qRT-PCR was reduced until ligated exons were the unique product detected (see Material and Methods). The mRNA for the intron-less construct pICG-E1E2 amplified by primers 5-6 was produced at almost the same level as that for the lacZ1, and it constituted the positive control for exon ligation. In contrast, the intron-containing construct pICG-WT generated three orders of magnitude less mRNA than pICG-E1E2 and 15 times as much mRNA as the splicing defective pICG-DV construct, which gave no distinct amplicon even if PCR was carried out over 40 cycles (Figure R2.4 B).

We investigated the splicing efficiency of RmInt1 in S. meliloti, by calculating the ratio of the ligated lacZ gene mRNA to lacZ1 mRNA for pICG-WT and pICG-DV, normalizing the value obtained with respect to the intron-less exon-ligated control and
Figure R2.4. Quantitative analysis of RmInt1 in *S. meliloti*. A Map of the *laiz* region of pICG-WT. The RmInt1 intron with the IEP (pink arrow) encoded in DIV and minimal flanking sequences from the *ISRm2011-2* insertion sequence (E1 and E2) (green and yellow boxes, respectively) are located in the *laiz* gene (blue arrow). The interrupted *laiz* gene is shown as *laiz*1 and *laiz*2 exons. Arrows under the map indicate the primers used in the real-time RT-PCR assay. Primer pair 1-2 was used to amplify the *laiz*1; primers 3-4 were used to amplify the RmInt1 intron and primers 5-6 were used to amplify the ligated β-galactosidase gene *laiz*. B Total RNA was isolated from *S. meliloti* containing the wild-type intron pICG-WT, the intron-less construct pICG-E1E2 and the splicing-defective mutant pICG-DV. The levels of *laiz*1 (1-2), RmInt1 intron (3-4) and ligated β-galactosidase gene *laiz* (5-6) mRNA were determined by an absolute quantification method. Each measurement is the mean of at least three independent RNA preparations. A 40 cycle end-point RT-PCR was carried out, analyzed by an electrophoresis in a 0.8% agarose gel and observed under UV light. A photograph of the revealed cDNA bands corresponding to each sample are depicted under the graphic bars. Note that no distinct bands are detected in lanes with an expression lower than the background level in qRT-PCR.

expressing the result as a percentage. We compared the splicing efficiency of RmInt1 in *S. meliloti* with that in *E. coli* (Table R2.2). The splicing efficiency of RmInt1 was very low in *S. meliloti*, at 0.07 ± 0.02%, but was nonetheless 4.5 times higher than that in *E. coli*, confirming that the intron was more functional in its natural host.

<table>
<thead>
<tr>
<th>Strain and construction</th>
<th>% splicing ± error*</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. meliloti</em> RmInt1</td>
<td></td>
</tr>
<tr>
<td>E1E2</td>
<td>100</td>
</tr>
<tr>
<td>WT</td>
<td>0.07 ± 0.02</td>
</tr>
<tr>
<td>DV</td>
<td>0.002 ± 0.002</td>
</tr>
<tr>
<td><em>E. coli</em> RmInt1</td>
<td></td>
</tr>
<tr>
<td>E1E2</td>
<td>100</td>
</tr>
<tr>
<td>WT</td>
<td>0.016 ± 0.006</td>
</tr>
<tr>
<td>DV</td>
<td>0.004 ± 0.002</td>
</tr>
<tr>
<td><em>E. coli</em> Ll.ltrB</td>
<td></td>
</tr>
<tr>
<td>ltrB</td>
<td>100</td>
</tr>
<tr>
<td>Ll.ltrB</td>
<td>0.23 ± 0.05</td>
</tr>
<tr>
<td>Ll.ltrBA</td>
<td>0.001 ± 0.001</td>
</tr>
</tbody>
</table>

* The percentage of RmInt1 splicing is defined by the following equation: $2^{\Delta C_F}$ where $\Delta C_F = \Delta C_T$ (mRNA amplified by 5–6 primer pair) - $\Delta C_T$ (mRNA amplified by 1–2 primer pair) expressed as a percentage related to each ligated-exon construction. The error is calculated applying the formula: $2^{\Delta C_F} \cdot \ln 2 \cdot \Sigma SE$. 

91
R.2.5 ANALYSIS OF THE \textit{L.lactis} LL.ltrB INTRON SPlicing IN \textit{E.coli}I AND \textit{S}. \textit{meliloti}

We made use of the LL.ltrB-containing pICG constructs to analyze the splicing level of the LL.ltrB intron in the context of the \textit{lacZ} reporter system and to compare it with that of RmInt1, carried out in \textit{E.coli} and \textit{S}. \textit{meliloti} hosts and under different temperature conditions.

A primer extension assay was carried out to determine whether the LL.ltrB intron, expressed in \textit{E.coli}, was excised in the pICG system (Figure R2.5). We performed cDNA synthesis by using a primer complementary to the LL.ltrB intron. A cDNA band, corresponding to the excised product (L) was detected as an extension product of 92 nt in the intron-harboring construct, pICG-LL.ltrB (lane 2), while no band was detected in the construction in which the intron was inserted in antisense orientation with respect to the promoter (pICG-LL.ltrBA, lane 1). For comparison, it was included the pLI1 construct (Matsuura \textit{et al.}, 1997), expressed in \textit{E.coli} under the T7 promoter and analyzed after 3 hours of induction with IPTG (lane 3). Compared to pICG-LL.ltrB, whose excised lariat band was barely detectable, a strong signal, belonging to pLI1, was identify. The contrast in the detection level of excised intron lariat forms when the LL.ltrB intron is expressed either in the pICG vector, under the constitutive promoter pSyn or in the pLI1 plasmid, under the inducible promoter T7, may represent a key factor for the detection of the spliced forms produced during the splicing process.

A quantitative evaluation of the splicing level in the LL.ltrB constructs was carried out by the determination of the \(\beta\)-galactosidase activity in strains of \textit{E.coli} and \textit{S}. \textit{meliloti} carrying the intronless construct pICG-\textit{ltrB}, the full-length wild-type intron construction pICG-LL.ltrB and the negative control, pICG-LL.ltrBA.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure_r2_5.png}
\caption{In vivo excision of LL.ltrB intron in \textit{E.coli}. Primer extension analysis was performed on total RNA extracted from pICG plasmids pICG-LL.ltrBA (lane 1) and pICG-LL.ltrB (lane 2) expressed in \textit{E.coli} cultures until exponential phase of growth and from pLI1 plasmid (Matsuura \textit{et al.}, 1997), after 3 hours of induction with IPTG (lane 3). RNA was reverse-transcribed from a primer complementary to the intron. cDNA bands, of 92 nt length, correspond to the excised intron lariat. No bands were detected in pICG-LL.ltrBA, in which the intron was inserted in antisense orientation with respect to the promoter.}
\end{figure}
We evaluated the impact of temperature in the splicing of L1.ltrB in E.coli (Figure R2.6, left panel). Three different temperatures: 37°C, 28°C and 24°C was assayed. In its natural host, L.lactis, the L1.ltrB intron encounters various temperatures, from low temperatures as 24°C to higher ones as 37°C (Chen et al., 2005), which may influence the ability of L1.ltrB to splice in a variable environment. In the heterologous host E.coli, the optimal growing temperature is 37°C, although it was forced to grow as well at 28°C and 24°C. At 37°C, the β-galactosidase activity measured in E.coli cultures in pICG-ltrB was about 12000 U Miller, in the order of magnitude as pICG-E1E2 (Table R2.1). Nonetheless, the level of activity in the intron-carrying construct pICG-L1.ltrB (4 ± 1) was under the background, as determined by the β-galactosidase activity of pICG-L1.ltrBA (14 ± 2), in the same way as observed for RmInt1. When the activity was determined from a culture of E.coli grown at 28°C, the level of expression of the lacZ gene in the positive control, pICG-ltrB, dropped until ~5000 U Miller although, surprisingly, the activity in pICG-L1.ltrB increased 4.5-fold respect to that observed at 37°C and above 3.5-fold above the background level (Figure R2.6, left panel). The splicing efficiency calculated for L1.ltrB expressed at 28°C was (0.39 ± 0.05) % respect to that of ligated-exon construction (100 %). The β-galactosidase activity was also determined in cultures grown at 24°C, showing again a decrease in ltrB expression accompanied by a 2-fold increase in the activity of the L1.ltrB intron construct with respect to that assayed at 28°C. The result was a net raise of splicing efficiency, which was calculated to be (0.89 ± 0.04) %. The expression level was also assayed in S.meliloti cultures harboring the L1.ltrB intron grown at 28°C, the optimal temperature for the growth of this strain (Figure R2.6, right panel). As previously observed for RmInt1, the expression of the intron construction was similar to the negative control, not allowing any quantification of the splicing activity.

We performed real-time quantitative RT-PCR on samples of E.coli carrying L1.ltrB constructs grown at 28°C as described above (also see Materials and Methods). The splicing efficiency was calculated normalizing with respect to the intron-less exon-ligated control and expressing the result as a percentage (Table R2.2). The splicing efficiency of L1.ltrB in the pICG context was quantified at (0.23 ± 0.05) %, significantly higher than the negative control (0.001 ± 0.001) %. This value was in agreement with that obtained from

Figure R2.6. Quantitative analysis of L1.ltrB intron in E.coli and S.meliloti hosts. Expression of the lacZ gene in pICG constructs harboring the L1.ltrB intron was measured by a β-galactosidase assay (Miller, 1972) both in E.coli (left panel) and S.meliloti (right panel) cultures, grown until exponential phase. Expression in E.coli was assessed at three different temperatures: 37°C, 28°C and 24°C. The expression of the negative control construction, pICG-L1.ltrBA, determines the background level for each temperature assayed. The evaluation of the β-galactosidase activity in S.meliloti was only performed at 28°C. Miller Units are indicated in the graphic in logarithmic scale.
the Miller assay and it was about 3-fold higher than RmInt1 splicing efficiency in *S. meliloti* and about 14-fold higher if measured in an *E. coli* culture.

**R.2.6 DISCUSSION**

The *in vivo* splicing of group II introns has been studied in detail for a small number of well known group II introns, such as Ll.ltrB from *L. lactis* and a15γ from *S. cerevisiae* (Arnberg *et al.*, 1980; Hensgens *et al.*, 1983; Domdey *et al.*, 1984; Peebles *et al.*, 1993; Boulanger *et al.*, 1996; Mills *et al.*, 1996; Podar *et al.*, 1998; Cui *et al.*, 2004; Yao *et al.*, 2006; Belhocine *et al.*, 2007). In most cases, these analyses were limited to the detection of the spliced form of the intron, with very few studies considering the detection (Matsura *et al.*, 1997) or quantification (Chen *et al.*, 2005) of ligated exon forms. In this study, we analyzed the splicing of the bacterial group II intron RmInt1, including intron excision and exon ligation, in *S. meliloti* and in *E. coli*, by northern blotting, β-galactosidase assays and qRT-PCR approaches, and compared it with the *L. lactis* Ll.ltrB intron in *E. coli*. We show that the splicing of a group II intron naturally located into an insertion sequence (ISRm2011-2) interrupting the transposase gene is almost completely abolished, but this intron retains its invasion capacity.

Assays of RmInt1 excision efficiency in the pICG system showed similar rates of excision for the wild-type construct and the construct in which the IEP sequence was
present in trans on a separate plasmid. These findings contrast with those for LtrA in E. coli, for which lower levels of excision were reported in trans (Cui et al., 2004). As the plasmid carrying the IEP sequence in trans is present at a higher copy number than the IEP sequence in cis, the larger amount of protein produced in the trans system may rapidly become available for the binding of intron RNA, preventing degradation of the precursor RNA. RmInt1 in a cis position upstream from the 5' exon was excised less efficiently, probably due to the larger distance to the promoter and the higher probability of the RNA polymerase releasing the RNA chain or dissociating from the DNA template. By contrast, RmInt1 excision in E. coli was 95% less efficient than excision of the intron from the wild-type construct in S. meliloti, suggesting that host factors may play an important role in the splicing reaction, probably during ribozyme folding, as previously reported for other group II introns (Smith et al., 2005; Beauregard et al., 2008).

The analysis of RmInt1 splicing products in vivo in S. meliloti confirmed that lariat form accumulates in the splicing process and that this product is free of lariat-3’ exon reaction intermediates. However, ligated exons forms, more susceptible to cell degradation and thus less prone to accumulate, were not detected. The absence of ligated exon forms contrasts with the in vivo detection of exon ligation during splicing of the lactococcal group II intron Ll.ltrB in E. coli (Cui et al., 2004), which suggests that splicing of RmInt1 may be very inefficient. The lower level of accumulation of splicing products from constructs harboring a transcription terminator may be due to higher rates of RNA polymerase dissociation from short templates with a high degree of secondary structure than from templates lacking terminator sequences. We also detected a 740 nt intron band, probably corresponding to a broken lariat form. Linear group II introns have been reported to occur in vivo (Vogel and Borner, 2002), but this form is produced predominantly during in vitro splicing reactions (Daniels et al., 1996) or as a consequence of branch-site mutations in vivo (Podar et al., 1998).

The absence of ligated exons on northern blots led us to use more sensitive approaches for monitoring this product of intron RNA splicing. Nevertheless, the level of exon ligation observed for RmInt1 intron, in Miller units, was below background levels. Quantitative real-time qRT-PCR assays enabled us to analyze RmInt1 and lacZ mRNA levels and the intron splicing efficiency. Large amounts of lacZ1 and RmInt1 RNA were produced, but ligated lacZ gene produced three orders of magnitude less RNA, demonstrating the strong effect of RmInt1 on the genes it interrupts. We found that RmInt1 was very inefficiently spliced in vivo in S. meliloti (0.07 ± 0.02%) and in E. coli (0.016 ± 0.006%).

95
The *L. lactis* Gram-positive bacterium LL.ltrB group II intron has been extensively studied for its splicing activity. Since it was initially reported in its natural host *L. lactis* (Mills et al., 1996; Shearman et al., 1996), diverse methods have been used for the detection of the excised form as well as the ligated exons of LL.ltrB in *vivo* (Matsuura et al., 1997; Cui et al., 2004; Klein et al., 2004; Chen et al., 2005) both in *L. lactis* and the heterologous host *E. coli*. Here, we inserted the LL.ltrB intron in the reported system pICG for comparing the splicing abilities of both introns expressed under the same conditions.

First, we confirmed that LL.ltrB was efficiently excised from pICG plasmids and compared it with pLI1, previously used by Matsuura *et al* (1997), observing that the expression under the inducible promoter T7 in pLI1 was several orders of magnitude higher. This finding may be the responsible for the accumulation of a big amount of intron mRNAs in the cell, allowing the detection of splicing reaction products which are in low numbers.

The study of LL.ltrB intron in the *lacZ* context in *E. coli* revealed an splicing efficiency of LL.ltrB ranged from (0.23 ± 0.05) %, as measured by qRT-PCR at 28°C to (0.89 ± 0.04) %, as reported by β-galactosidase activity analysis in a culture grown at 24°C. These values were 3.2 to 13-fold higher than those of *S. meliloti*. In addition, a previous work on LL.ltrB intron expressed in its natural host *L. lactis* (Chen *et al*., 2005) reported splicing efficiencies 85 to 315 times more efficient than that for RmInt1 in *S. meliloti*, probably due to the assistant of specific host-encoded proteins of *L. lactis* recruited to ensure the intron function. In that work, it was also postulated that temperature stress had no significant effects on expression and splicing of LL.ltrB. Consistent with Yao *et al* (2006), who reported an inherent temperature sensitivity of the splicing reaction of LL.ltrB expressed in *E. coli*, we have observed a decrease in the expression of LL.ltrB when the temperature rises from 24°C or 28°C to 37°C. Probably, this controversy could be the result of the influence of cell physiology and environmental conditions on the intron splicing. On the one hand, the changing dairy environment of *L. lactis* may influence that the host better resists temperature and other stresses while, in the other hand, a constant environment as that of *E. coli* could explain why this host is more sensitive to temperature variations.

Considered together, our results suggest that the splicing of RmInt1 in its natural host *S. meliloti*, where it interrupts the transposase of ISRm2011-2, is much less efficient than that of other described bacterial group II introns, as LL.ltrB. However, this inefficiency does not lead to defects in intron mobility, since a small amount of excised intron remains stable and accumulates, likely because of its association with the IEP,
forming ribonucleoprotein particles. In fact, this intron has been shown to spread well in *S. meliloti* (Martínez-Abarca *et al.*, 2000; Martínez-Abarca and Toro, 2000; Nisa-Martínez *et al.*, 2007). In contrast, the minimal quantities of ligated exons produced during the splicing process are more likely to be susceptible to cell degradation, giving little chance of transposase translation. According to this, it could be of interest the use of strong inducible promoters as that of T7, using T7 RNA polymerase introduced via a λDE3 lysogen or from a separate plasmid (Guo *et al.*, 2000; Karberg *et al.*, 2001) to increase the steady-state level of a significant cross section of transcripts (Carpousis, 2007). Another possibility it could be the use of an overexpressed RraA protein (*regulator of ribonuclease activity A*), that binds to RNase E and inhibits RNase E endonucleolytic cleavages without interacting with substrate RNAs (Lee *et al.*, 2003). Nevertheless, this option is only available for strains containing a functional ortholog of *E. coli* RraA, not found in *S. meliloti*. 

Finally, our findings support the postulated differentiation between bacterial and organellar group II introns (Dai and Zimmerly, 2002). In agreement with that, RmInt1 have many characteristic features attributed to bacterial group II introns. Its low splicing efficiency, as described above, occurs along with its localization in the insertion sequence ISRm2011-2 which, in turn, is located outside of any gene. In addition, although RmInt1 splicing efficiency is very low, its invasion capacity is high, helped by the presence of a reverse-transcriptase codified in the ribozyme DIV, because survival of group II introns in bacteria appears to rely on constant movement (Dai and Zimmerly, 2002). According with that, RmInt1 provides evidence of retroelement behavior, while its intronic character is less prominent.
CHAPTER 3

Effect of point mutations in conserved positions of the RmInt1 group II intron on the splicing reaction in vivo
R.3.1 BACKGROUND

Splicing efficiency of group II introns located in bacterial genomes has been reported to be less efficient than that of those placed in organellar genomes of eukaryotic organisms (Dai and Zimmerly, 2002). The in vivo splicing of RmInt1 intron even shows a lower efficiency than that of other bacterial introns (See Chapter 1). However, it seems that there is no uncoupling between the two steps of splicing, since no traces of accumulation of intron intermediates of reaction has been observed (See Chapter 2, Figure R2.3).

The six domains of group II introns fold into a catalytically active structure following well-known conserved long-range tertiary interactions identified by phylogenetic covariation and biochemical analyses (Michel et al., 1989; Lehmann and Schmidt, 2003; Fedorova and Zingler, 2007; Dai et al., 2008). These include those involving Watson-Crick base pairs (α-α', β-β', γ-γ', δ-δ', ε-ε', IBS1-IBS1, IBS2-IBS2, and IBS3-IBS3), tetraloop-receptor interactions of known geometries (ζ-ζ', η-η', and θ-θ'), and other of less defined geometries (λ-λ', κ-κ', and μ-μ'). Recently, the use of site-specific cross-linkers to the establishment of distance constraints and the evolution of crystallography techniques have allow a more complex panorama of tertiary interactions (Toor et al., 2009; Keating et al., 2010; Pyle, 2010). Of these tertiary interactions, those involving 1) the conserved linker sequence between domain II and III (J2/3), which have been previously reported to interact with the catalytic domain V and to affect one or both splicing steps (Jacquier and Michel, 1990; Bar-Shalom and Moore, 2000; Mikheeva et al., 2000; de Lencastre et al., 2005; Keating et al., 2010) and; 2) the 3'-intron boundary and the intron/3'-exon EBS3-IBS3 interaction, previously reported to inhibit the second step of the splicing reaction (Chanfreau and Jacquier, 1993; Costa et al., 2000; Molina Sanchez et al., 2011) were of our interest.

Compared to in vitro investigations, only few reports show data from in vivo studies using site-directed intron mutations. In the mitochondrial yeast a15γ intron, in vivo splicing has been analyzed using ribozyme mutants affected in the conserved first nucleotide G1 (Peebles et al., 1993), in selected positions of the catalytic DV (Boulanger et al., 1995; Schmidt et al., 1996) and in the length of the joining segment between domains 5 and 6 (Boulanger et al., 1996). The in vivo studies were also extended to the Scenedesmus obliquus algae mitochondrial rH1 group II intron, where complete deletion of domains V and VI and substitutions of the first intron nucleotide and γ-γ' base pair were analyzed (Hollander and Kuck, 1999a; 1999b). A comprehensive study of mutations retaining mobility in vivo
were also developed for the bacterial L.ltrB intron from *Lactococcus lactis* (D’Souza and Zhong, 2002).

In this chapter, we studied the effect on the splicing of the RmInt1 group II intron of point mutations in positions in the intron ribozyme previously reported to be part of conserved tertiary interactions in α5γ, L.ltrB and *Oceanobacillus iheyensis* group II introns (Dai *et al.*, 2008; de Lencastre and Pyle, 2008; Toor *et al.*, 2009; Toor *et al.*, 2010). Through the analysis of the effect of these mutations on the first and second splicing steps, we identified long-range interactions occurring in the RmInt1 tertiary structure which are common to other group II introns and deepened the understanding of the structural constraints involved in the intrinsic low efficiency displayed by RmInt1 intron.

### R.3.2 MAP OF POINT MUTATIONS INTRODUCED IN THE RIBOZYME OF RmInt1 INTRON

Using site-directed mutagenesis (See section M.X, Materials and Methods), point mutations (substitutions and deletions), were introduced in the ribozyme of RmInt1 intron (Figure R3.1). All mutated intron positions have been shown previously to be part of conserved tertiary interactions during the *in vivo* and/or *in vitro* splicing of yeast and bacterial group II introns. The wild-type intron and its mutant derivatives were inserted in the pICG system for the *in vivo* splicing analysis.

We introduced a triple mutation in positions G1840, U1841 and U1842 which base pair with the conserved catalytic triad A1811, G1812 and C1843, located in the stem 1 of domain V and disrupting the Watson-Crick base pair (DV-CGA mutant) (Figure R3.1, violet box). Related mutations affecting the catalytic domain V has also been reported *in vitro* and *in vivo* (Jarrell *et al.*, 1988; Koch *et al.*, 1992; Dib-Hajj *et al.*, 1993; Boulanger *et al.*, 1995; Peebles *et al.*, 1995). Three mutations where introduced in J2/3 region. We deleted the conserved nucleotides G453 and A454 as previously reported by Mikheeva *et al* (2000) (J2/3-ΔGA mutant) (Figure R3.1, orange box). In addition, we assayed two mutations in the nucleotide G452, which is part of the conserved tertiary interaction γ-γ’ with the C1884 position: its deletion (Δγ mutant) and its substitution by a C (γC mutant) (Molina Sanchez *et al.*, 2011), disrupting the long-range interaction (Figure R3.1, blue box). On the other hand, the C1884 mutated to a G (Cγ’G mutant) (Molina Sanchez *et al.*, 2011) was assayed, which also disrupted the Watson-Crick interaction. A double G452C C1884G, restoring the γ-γ’ base pair, was used as well (γC-γC’G mutant) (Figure R3.1, blue box).
Figure R3.1. Diagram of point mutations and deletions of the ΔORF RmInt1 group II intron flanked by short exons (-20/+5). The mutation in the C+1 position was introduced in the full-length intron flanked by long exons (-146/+466). The RmInt1 intron sequence is indicated by upper-case letters; the exonic sequences are indicated by lower-case letters. Point mutations are indicated by an arrow follow by the new nucleotide. Deletions are represented by a Δ symbol. Interactions in which mutated position participate include conserved described interactions (Michel et al., 1989; Jacquier and Michel, 1990; Chanfreau and Jacquier, 1993; Boulanger et al., 1995; Costa et al., 2000) and those obtained from crosslinks and/or crystal structure determinations described for α5γ (de Lencastre et al., 2005; de Lencastre and Pyle, 2008) O.ι. (Toor et al., 2008a; 2008b; Keating et al., 2010) and L.rtrB (Dai et al., 2008) introns.
(Molina Sanchez et al., 2011). Mutations of the γ-γ’ base pair have been previously described for the yeast a15γ and the ascomycete *Podospora anserina* group II intron *in vitro* (Jacquier and Michel, 1990; Schmidt et al., 1993) and for *S. melioli* RmInt1 and *Chlamydomonas reinhardtii* r11 *in vivo* (Hollander and Kuck, 1999a; Hollander and Kuck, 1999b; Molina Sanchez et al., 2011). In the 3’ boundary, the penultimate position C1883 was mutated to the other three bases (C to A, G and U) as previously assayed (Chanfreau and Jacquier, 1993) (C1883A, C1883G and C1883U mutants, respectively) (Figure R3.1, pink box). Finally, position G329 of the intron and C+1, located in the 3’-exon, were mutated to C and A, respectively (GEBS3C and C+1A mutants, respectively) (Figure R3.1, green box) (Molina Sanchez et al., 2011). These positions, which are part of the conserved tertiary interaction EBS3-IBS3 observed in group IIB introns, has been analyzed both *in vitro* (Costa et al., 2000) as *in vivo* (Robart et al., 2004; Molina Sanchez et al., 2011).

### R.3.3 Quantitative Analysis of Splicing Efficiency in Wild-Type and Mutant RmInt1 INtron

We use the pICG reported system to accurately and quantitatively evaluate the splicing efficiency through the measurement of ligated exons in the wild-type and mutant RmInt1 intron constructions.

A relative quantification was used to determine the splicing efficiency of mutant derivatives related to wild-type RmInt1 intron construct (Table R3.1). The majority of mutants were severely affected in the splicing efficiency *in vitro*, showing less than 5% splicing activity. These include DV-CGA, J2/3ΔGA, Δγ, Δγ+GEBS3C, γ’C, Cγ’G and GEBS3C. γ’C-Cγ’G and C1883G mutants showed between 15-20% splicing activity. These data shed light on the importance of such positions in the splicing *in vivo*. Mutants in the penultimate position of the intron C1883A and C1883U were moderately affected, displaying splicing activities of 70 and 40%, respectively. Finally, the relative activity of the C+1A mutant was similar to the wild-type allele, indicating that this position is not important for the *in vivo* splicing.

Note, however, that the splicing activity depends on the ability of the wild-type and mutant introns to efficiently and accurately go through the first and second steps of the splicing reaction and join the exons of the gene interrupted by the intron. Thus, effects observed in the splicing of mutant intron variants are the result of the addition of the defects added in each step of the reaction and, then, splicing efficiency values give no
information whether the first or second splicing steps were affected. As a consequence, it was necessary to compare with data of excision and accumulation of intron intermediates to figure out the effect of each mutation in the complete splicing reaction.

**TABLE R3.1. RmInt1 splicing activity** in ΔORF and ΔORF mutant constructs expressed in *S. meliloti*.

<table>
<thead>
<tr>
<th>Construction</th>
<th>% splicing ± error</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔORF</td>
<td>100</td>
</tr>
<tr>
<td>DV-CGA</td>
<td>1.2 ± 0.3</td>
</tr>
<tr>
<td>J2/3ΔGA</td>
<td>2 ± 2</td>
</tr>
<tr>
<td>Δγ</td>
<td>0.016 ± 0.007</td>
</tr>
<tr>
<td>Δγ+GEBS3C</td>
<td>1.3 ± 0.1</td>
</tr>
<tr>
<td>GγC</td>
<td>0.43 ± 0.09</td>
</tr>
<tr>
<td>GγC-Cγ′G</td>
<td>15 ± 3</td>
</tr>
<tr>
<td>Cγ′G</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>C1883A</td>
<td>70 ± 30</td>
</tr>
<tr>
<td>C1883G</td>
<td>19 ± 3</td>
</tr>
<tr>
<td>C1883U</td>
<td>40 ± 10</td>
</tr>
<tr>
<td>GEBS3C</td>
<td>4 ± 1</td>
</tr>
<tr>
<td>C+1A</td>
<td>98 ± 1</td>
</tr>
</tbody>
</table>

*a Activity measured by relative quantification of a qRT-PCR assay.

*b All mutants were introduced in the ΔORF construct and co-expressed in *trans* with the IEP protein, with the exception of the C+1A mutation, which was introduced in the full-length long-exon RmInt1 intron context.

*c RmInt1 splicing is defined by the following equation: \(2^\Delta\Delta CT\) where \(\Delta CT = CT\) (mutant ΔORF) - \(CT\) (ΔORF). \(\Delta CT\) (mutant ΔORF) = \(CT\) (mutant ΔORF mRNA amplified by 5-6 primers) – \(CT\) (mutant ΔORF mRNA amplified by 16S_Fw-16S_Rv2 primers). \(\Delta CT\) (ΔORF) = \(CT\) (ΔORF mRNA amplified by 5-6 primers) – \(CT\) (ΔORF mRNA amplified by 16S_Fw-16S_Rv2 primers). The result is expressed as a percentage related to ΔORF construct. The error is calculated applying the formula: \(2^\Delta\Delta CT \cdot \ln2 \cdot \Sigma SE\).
R.3.4 QUANTIFICATION OF THE FIRST STEP OF THE IN VIVO SPlicing IN WILD-TYPE AND MUTANT RmInt1 INTRON DERIVATIVES

The first step of the splicing reaction in vivo was determined by the efficiency at which the intron excises from the precursor RNA. In order to accurately quantify excised intron products during the in vivo splicing both wild-type and mutant RmInt1 introns were subjected to primer extension assays (Figure R3.2). As previously assessed in Chapter 2, the excised lariat intron RNA (L) was detected as an extension product of 97 nt together with larger bands (not shown) derived from the 330-nt unspliced precursor (Pr) (Figure R3.2 A). The wild-type intron (lane 1) produced a strong intron band, which was quantified (see Materials and Methods and Figure R3.2 legend). The extension products of mutant-intron derivatives were also quantified and normalized to that of wild-type, giving relative excision efficiency values (Figure R3.2 B).

The most affected mutants in the excision capacity were DV-CGA (lane 2) and Δγ+GEBS3C (lane 5), with relative excision efficiencies of 3 and 7 %, respectively. They
Figure R3.2. *In vivo* excision of the wild-type and mutant ΔORF intron in *S. meliloti*. A Schematic diagram of primer extension on linear unspliced precursor and excised and partial-excised intron lariats, showing the expected extension products derived from pICG plasmids. The bulged A in DVI is also indicated. B Primer extension assays were performed on total RNA extracted from *S. meliloti* harboring either the wild-type ΔORF (lane 1) or the mutant intron constructions (lanes 2-12) complemented with the IEP *in trans*. RNA was reverse-transcribed with a \(^{32}P \gamma\)-ATP \(^{5}\)-labeled primer complementary to the first 97 nt of RmInt1 ribozyme. The products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel, which was dried and quantified with a PhosphorImager. The major cDNA product of 97 nt corresponds to the excised intron RNA. Transcription was started from three different points, giving a strong unspliced precursor of 330 nt, characteristic of pICG plasmid, which were used as a loading control. Excision efficiency was calculated as the ratio of the intensity of the 97 nt product to that of the transcription start point, related to wild-type ΔORF construct (lane 1) and expressed as a percentage.

were followed by Δγ (lane 4) and GγC-GγG (lane 7), with excision efficiencies of 23 and 20 % respectively. Mutants J2/3ΔGA (lane 3) and GEBS3C (lane 46) showing 46 % excision efficiency and GγC (lane 6) with 43 %, were moderately affected. By contrast, C1883G (lane 10) presented a low decrease in the excision efficiency, whereas Gγ'G (lane 8), C1883A (lane 9) and C1833U (lane 11) showed a wilt-type phenotype.

Considered together, it was observed that mutations primarily affecting the first step of the RmInt1 splicing reaction *in vivo* are located in the catalytic domain V, in the J2/3 region and in the EBS3 position in the coordination loop of the ribozyme. By contrast, mutations located in the 3'-intron boundary produced no decrease in the intron excision capacity. Note, however, that the measure of the first step of splicing makes no difference between partial and full-excised intron forms, then requiring the determination of the extent at which the intron is completely excised from the flanking exons (*vide infra*).

R3.5 OVERVIEW OF THE COMPLETE SPLICING REACTION OF MUTANT-DERIVATIVES AND WILD-TYPE RmInt1 ΔORF INTRON IN *S.meliloti*

*In vivo* splicing products from *S.meliloti* cultures harboring mutant-derivatives and wild-type ΔORF RmInt1 intron were tested by RNA gel blot analysis. Figure R3.3 shows northern hybridizations of various intron mutants with the intron-specific probe (left panel). For comparison, all filters were rehybridized with a probe specific for the 3'-exon of pICG plasmids (right panel). The ΔORF wild-type strain (lane 1) was used as a control.
Figure R3.2. In vivo excision of the wild-type and mutant ΔORF intron in *S. meliloti*. A Schematic diagram of primer extension on linear unspliced precursor and excised and partial-excision intron lariats, showing the expected extension products derived from pICG plasmids. The bulged A in DVI is also indicated. B Primer extension assays were performed on total RNA extracted from *S. meliloti* harboring either the wild-type ΔORF (lane 1) or the mutant intron constructions (lanes 2-12) complemented with the IEP in trans. RNA was reverse-transcribed with a 32P [γ-ATP] 5'-labeled primer complementary to the first 97 nt of RmInt1 ribozyme. The products were analyzed by electrophoresis in a denaturing 6% polyacrylamide gel, which was dried and quantified with a PhosphorImager. The major cDNA product of 97 nt corresponds to the excised intron RNA. Transcription was started from three different points, giving a strong unspliced precursor of 330 nt, characteristic of pICG plasmid, which were used as a loading control. Excision efficiency was calculated as the ratio of the intensity of the 97 nt product to that of the transcription start point, related to wild-type ΔORF construct (lane 1) and expressed as a percentage.
Intron mutations resulted in alterations of splicing activity in vivo in a variable extent. Intron mutants with altered positions located in DV, the J2/3 region and the EBS3 position in the coordination loop (lanes 2-7 and 13) showed intron probe hybridization signals lesser than 35 % related to wild-type ΔORF construct (lane 1) (left panel). Of them, the DV-CGA (lane 2) was the most affected, with barely detectable hybridization signal. It was followed by the double mutant Δγ+GEB3S3C (lane 5), Δγ (lane 4) and the double mutant GγC-Gγ’G (lane 7), restoring the γ-γ’ base pair, which showed intron hybridization signals lesser than 15 %. In the contrary, mutants affected in the 3’-intron boundary (lanes 8-11) showed intron signals higher than 50 % related to wild-type (lane 1). Of those, the least affected mutant was C1883A (lane 9), which displayed a strong intron signal, higher of that of wild-type intron (lane 1).

The accumulation of intermediate products of reaction was used as a rough measurement of the second splicing step efficiency, as observed when the filter was hybridized with a 3’-exon-specific probe (right panel). As expected, the wild-type intron (lane 1) did not show any 3’-exon signal and then, no reaction intermediate accumulation. Alterations in the G453 position, the J2/3ΔGA (lane 3) and GγC (lane 6) mutants, displayed a marked 3’-exon probe signal compared with that of intron probe, indicating an important accumulation of lariat/3’-exon. Particularly striking was the latter, which presented a 95 % decrease in the second splicing step with respect to the wild-type intron. The rest of mutant-intron derivatives resulted in a second splicing step efficiency lesser than 20 %, with the exception of those with altered penultimate position of the intron C1883A and C1883 U (lanes 9 and 11, respectively) were moderately affected in the second splicing step, as indicated the intensity of the 3’-signal probe. Note that both lariat intron and lariat/3’-exon intermediate products co-migrated, so just one lower-migrating band could be observed in samples accumulating reaction intermediates when hybridized with the intron probe.

From the analysis of data from Northern Blotting hybridization, we can conclude that mutations located in DV, the J2/3 region and the EBS3 position in the coordination loop, all affected the first splicing step in some extent. By contrast, mutations altering positions involved in the establishment of the conserved interaction γ-γ’, the EBS3 nucleotide and the penultimate nucleotide of the intron, was all affected in the second step of the splicing reaction.
R.3.6 DISCUSSION

Here, we have elaborated a map of RmInt1 intron mutations affecting the first, second or both splicing steps based on the disruption of tertiary interactions established between conserved positions on the ribozyme.

We introduced a triple mutation in positions G1840, U1841 and U1842 which base pair with the conserved catalytic triad A1811, G1812 and C1843, located in the stem 1 of domain V. The mutation, disrupting the Watson-Crick base pair, strongly reduced the splicing efficiency \textit{in vivo} (of which, at least the first step was affected, although we were not able to determine the effect of the second step). Boulanger \textit{et al} (1995) previously assayed these mutations in the \textit{in vivo} splicing of the ai5\gamma intron, showing an asymmetry of mutational effects in DV stem 1. Particularly, they found that A2, G3 and C4 (referred to the first position of DV) were essential for splicing \textit{in vivo}, whereas U33, U32 and G31, base pairing with them, tolerated well the introduction of mutations. These results were also observed for the ai5\gamma intron \textit{in vitro} (Peebles \textit{et al}., 1995). This made sense in the perspective of structure, since it has been shown that only the conserved 5'-AGC-3' nucleotides (and not their pairing partners) form hydrogen bonds with nucleotides of the J2/3 region and the AC bulge in DV, constituting a triple helix (triplex) which is essential for splicing catalysis (Toor \textit{et al}., 2008a; Toor \textit{et al}., 2009; Keating \textit{et al}., 2010). However, the simultaneous introduction of mutations in the RmInt1 DV and the loose of three base pairs could have an additive and deleterious effect on the splicing reaction which could explain how mutations in the opposite strand of that containing the conserved 5'-AGC-3' nucleotides, also produces a strong reduction in splicing efficiency.

The deletion of the dinucleotide G453 A454, located in the J2/3 region, produced an effect in both first and second steps of the splicing reaction, being the latter the most accused. Previously, Mikheeva \textit{et al} (2000) reported a profound negative effect on the second step of splicing \textit{in vitro}, but little effect on the first step, according with Chanfreau and Jacquier (1996), which observed a strong and specific effect on the second splicing step when they modified the adenosine residue. However, mutations in the same position resulted in affection of both steps of splicing (Chanfreau and Jacquier, 1996). Despite its importance in the second splicing step repeatedly observed, the effect on the first step of splicing is supported by the crystallographic results recently obtained. In these, the position G288 (G453 for RmInt1 intron) of the J2/3 region has been shown to take part of the triple interaction with the G359 U384 base pair (G1812 U1841 in RmInt1 intron) and the C289 (equivalent to the A454) in the triple interaction with the C358 G385 (equivalent to the A1811 U1842) of the catalytic triad 5'-AGC-3' in DV (Keating \textit{et al}.,
2010). In addition, the G588 (in a15γ intron) has been previously observed to crosslink with the first nucleotide of the intron, which indicates its role in the approximation of the catalytic core to the 5'-splice site (Hamill and Pyle, 2006; de Lencastre and Pyle, 2008). Moreover, since the fact that a mutation located in the catalytic core affect both steps of splicing, is in accordance with results suggesting the occurrence of an unique active-site catalyzing both first and second steps of the splicing reaction (Chanfreau and Jacquier, 1994; de Lencastre et al., 2005).

We also observed that mutations or deletions in the G452 (γ position) and mutations in the C1884 (γ' position) disrupting the conserved γ-γ' interaction resulted in a pronounced alteration of the second splicing step, accordingly with previous in vivo and in vitro studies. Thus, a complete loss of exon-exon ligation was observed for the r11 intron from S. obliquus algae after disruption of the wild-type AyUγ' base pair when expressed in vivo in E.coli and in chloroplasts from Chlamydomonas reinhardtii (Hollander and Kuck, 1999a; 1999b). In vitro, it has also been shown that the γ·γ' base pairing is essential for an efficient second splicing step (Jacquier and Michel, 1990; Jacquier and Jacquesson-Breuleux, 1991). However, unlike the studies above, the double mutation restoring the γ·γ' base-pair did not eliminated defects on the second splicing step. Moreover, a decrease in the efficiency for 5' cleavage when the G452 was mutated, independent of the γ·γ' interaction, was observed both in single and double mutants restoring the base pair, whereas mutation in the C1884 showed a wild-type phenotype for the first step of splicing, in agreement with previous assays in vivo on the RmInt1 intron (Molina Sanchez et al., 2011). These results can be explained on the base of crystallographic experiments on the O.i group II intron. In those, it was showed that the γ nucleotide is located near the domain V bulge, positioning the 3' splice site for catalysis of the exon junction (Toor et al., 2008c). In addition, as mentioned above, it has been establish the existence of a single active-site catalyzing both steps of splicing (Chanfreau and Jacquier, 1994; de Lencastre et al., 2005). We proposed that the γ·γ' interaction is important for the second splicing step, as shown by the fact that mutations in the γ' position disrupting the γ·γ' base pair specifically affect the second step of splicing reaction. However, we think that mutations in the catalytic core have a predominant character in the RmInt1 intron, and so, when the γ position is mutated, both steps of splicing is altered, independently of the integrity of the γ·γ' interaction.

Mutations in the penultimate position of the intron resulted in a specific but slight affectation of the second step of the splicing reaction. The less tolerated base change was the C to G, which resulted in a 50% reduction in the first step and an 80% decrease in the second step of splicing. Mutations of the C1884 position to A and U resulted in a wild-
type phenotype. This result partially agrees with previous data obtained from the ai5γ intron (Chanfreau and Jacquier, 1993), in which the penultimate position is an A. According with this study, the substitution to a G strongly reduced the second splicing step as well as to a U, unlike that observed for the RmInt1 intron. Interestingly, we have found that the wild-type phenotype is kept both A and C nucleotides in the penultimate position of the intron. This position has been shown to interact with the first nucleotide of the intron, approximating the exon ends, which occurs after the lariat formation, through a non Watson-Crick interaction, as confirmed by the absence of compensatory substitutions (Chanfreau and Jacquier, 1993).

The substitution of the EBS3 position (G329) in the coordination loop resulted in a moderate effect in the first and second splicing steps, whereas the mutation of the first nucleotide of the 3'-exon showed a wild-type splicing efficiency level. This effect on the first step splicing has been recently reported by Molina-Sánchez and colleges (Molina Sanchez et al., 2011), who proposed that this observation could be the result of the increase in the reversibility of the first step (and so the reduction of excision product) due to the impairment in the second step, thus preventing the reaction to proceeds until completion. On the other hand, the wild-type splicing efficiency produced by the mutation in the first nucleotide of the 3’ exon is in agreement with the observation that the 3’ exon could be minimized to a single monophosphate group, not being determinant in the second transesterification reaction during splicing neither the nature nor the sequence composition of that (Müller et al., 1991). Moreover, in agreement with our results, it has been previously reported that the coordination loop spatially organizes all reactants involved in both steps of group II intron self-splicing in vitro (Costa et al., 2000; Hamill and Pyle, 2006). The role of the EBS3 in the coordination loop on the approximation of exon ends and the interaction with the DV has been elucidated through the analysis of the structure. Thus, the 5’ and 3’ exons are brought together and recognized by EBS1 and EBS3 motifs (interacting with the IBS1 motif in 5’ exon and IBS3 nucleotide in 3’ exon, respectively), which are joined through the δ-δ’ interaction (Toor et al., 2008b); on the other hand, the coordination loop interacts with the ζ motif, which in turn, forms a binding interface within the noncatalytic face of DV (Toor et al., 2010). This arrangement presents then the scissile phosphate to the active site in DV, carrying on both steps of splicing, which are coordinated within the active site.

Taking all together, our data show that disruption of interactions in which components of the catalytic core are included affected both steps of splicing, supporting that a unique catalytic site carries out both splicing steps. In the contrary, interactions
promoting the approximation of exons but not included in the catalytic core exclusively affected the second splicing step.

Moreover, the analysis of mutations affecting conserved positions in the ribozyme has allowed us a deeper understanding of tertiary interactions in the RmInt1 group II intron, which in turn could give us the clues needed for the improvement of the splicing efficiency and the use of RmInt1 as a biotechnological tool.
CHAPTER 4

SELF-SPICING OF RmINT1 GROUP II INTRON: A QUALITATIVE, QUANTITATIVE AND MUTATIONAL ANALYSIS
R.4.1 BACKGROUND

Some group II introns can catalyze their own splicing in vitro (self-splicing reaction in vitro), in a manner independent of protein assistance, though under non-physiological conditions such as high concentrations of monovalent salts (Lambowitz and Zimmerly, 2010). As occurring in vivo, self-splicing can occur via a branching pathway by means of two sequential transesterification reactions. In the first step, the 2’ OH of the branchpoint adenosine in DVI acts as the nucleophile to attack the 5’-splice site, producing an intron lariat-3’ exon intermediate. In the second step, the 3’ OH of the cleaved 5’ exon is the nucleophile and attacks the 3’-splice site, resulting in exon ligation and excision of an intron lariat RNA (Figure R4.1 A) (van der Veen et al., 1986). A hydrolytic pathway, in which the first step of splicing involves the use of a water molecule or a hydroxyl ion as the nucleophile, may also occur both in vitro and in vivo (Figure R4.1 B) (Jarrell et al., 1988b).

**Figure R4.1.** Schematic representation of the observed reactions during in vitro self-splicing of group II introns. All reaction products can be observed in high resolution polyacrylamide gels under a particular set of reaction conditions. **A** Schematic of the branching pathway. It is the predominant pathway observed in the presence of high concentrations of ammonium salts. The first step of splicing proceeds through a branch-point attack on the 5’-splice site, followed by a second attack of the 5’-exon on the 3’-splice site. The reaction leads to the formation of a lariat intron molecule, along with ligated exons. **B** Schematic of the hydrolytic pathway. It is the predominant pathway observed in the presence of high concentrations of KCl. In this route, the first step of splicing is mediated by a water molecule or a hydroxide ion on the 5’-splice site. After the second step of splicing, it results in the formation of a linear intron molecule and ligated exons. Adapted from Daniels et al. (1996).
These two pathways are sensitive to ionic conditions, in such way that, by varying the monovalent cation in the reaction mixture, one can force the ribozyme to prefer one reaction pathway over the other (Fedorova et al., 2002). In the presence of ammonium salts, the branching pathway is predominant, leading to an intron excision as lariats (Peebles et al., 1986; van der Veen et al., 1986; van der Veen et al., 1987; Jarrell et al., 1988b; Bachl and Schmelzer, 1990; Daniels et al., 1996; Zingler et al., 2010). In this case, \((\text{NH}_4)_2\text{SO}_4\) at high concentration causes the highest increment of lariat with respect linear intron whereas in other monovalent salts as NH\(_4\)Cl, LiCl or NaCl is less prominent (Jarrell et al., 1988b; Peebles et al., 1993; Daniels et al., 1996). In the presence of KCl buffer, the predominant route is the hydrolytic one, with linear intron as the main excised intron form (Jarrell et al., 1988b; Peebles et al., 1993; Daniels et al., 1996; Zingler et al., 2010).

To date a myriad of self-splicing assays have been performed in the well-studied a15γ intron from *S.cerevisiae* mitochondria (Kwakman et al., 1989; Koch et al., 1992; Su et al., 2001) and in other group II introns such as the first intron of the mitochondrial gene coding for cytochrome oxidase subunit I (COI II) of *Podospora anserina* (Schmidt et al., 1990), the a11 intron of the mitochondrial *cox1* gene of *S.cerevisiae* (Hebbar et al., 1992), the bacterial Ll.1tB intron from *Lactococcus lactis* (Matsuura et al., 1997) or that of the mitochondrial *rns* gene of the filamentous fungus *Leptographium truncatum* (Mullineux et al., 2010). In general, these studies can be carried out through two major approaches: either studying self-splicing in an intron containing all intronic domains and flanking exons or dissecting the intron into several individual constructs (Fedorova et al., 2002). Advantage of the first is the possibility of getting a global understanding of the overall reaction, observing the accumulation of each product and studying the effect of varying reaction conditions (Jarrell et al., 1988b; Daniels et al., 1996; Fedorova et al., 2002). The second approach offers the means for dissecting the splicing process in individual reactions which can be readily characterized (Chin and Pyle, 1995; Fedorova et al., 2002).

The bacterial group II intron RmInt1 has been also subjected to *in vitro* self-splicing analysis (Costa et al., 2006a; 2006b). In these studies, constructs containing the ΔORF RmInt1 intron along with its flanked exons were assayed in \((\text{NH}_4)_2\text{SO}_4\) as well as in KCl buffers, showing a predominant lariat and linear intron formation, respectively (Figure R4.2 A). The 5’ exons in these constructs were trimmed to its last 20 and 15 nucleotides, although the latter (pLM1) was shown to react better, with a first-order rate constant for conversion of precursor RNA into products of about 0.1 min\(^{-1}\) at 45 °C and a fraction of reactive precursor molecules of 0.30 (Costa et al., 2006a). Unlike other group II introns studied, self-splicing in RmInt1 led to the accumulation of lariat-3’ exon intermediate,
Figure R4.2. Self-splicing of pLM1 (ΔORF RmInt1 construct) at 45°C (Costa et al., 2006). A 32P[α-UTP] radiolabelled precursor molecules were subjected to a self-splicing reaction in the presence of KCl (left side of the panel) or (NH₄)₂SO₄ (right side of the panel) buffer salts, respectively. Products were analyzed after separation by denaturing PAGE. On the far left and right, representations of products identified are shown. Products generated by cleavage at positions +10 or +11 of the 3′ exons (IBS1* sequence; see the text for explanation) are marked with an asterisk. KCl buffer is 1 M KCl, 40 mM Tris-Cl pH 7.5, 100 mM MgCl₂, 1 mM Na₂EDTA. (NH₄)₂SO₄ buffer is 1 M (NH₄)₂SO₄, 40 mM Tris-Cl pH 7.5, 50 mM MgCl₂. B Proposed pathway framework for reactions of the ΔORF RmInt1 intron expressed from pLM1 plasmid under ammonium salt buffer. The intron is depicted as a black line; the branchpoint adenosine is marked as a white A into a blue circle and; the exons are green (5′) and yellow (3′) boxes.

which predominated for at least 2 first hours of reaction, resulting in an unusually inefficient exon ligation. Moreover, an unexpected truncated lariat/3′-exon intermediate, with only 10 or 11 nucleotides in its 3′-exon, was found (Figure R4.2 A). This aberrant product was suggested to be a result of the ribozyme-promoted cleavage 3′ of the sequence GACGAA or GACGAAG (hereafter designated as IBS1*), located in the 3′-exon, which matched better the EBS1 (UUUCGUC) sequence sited in the DI of the ribozyme than the authentic IBS1 (GAUGAGA) sequence in the 5′ exon (Costa et al., 2006a). This situation was explained as two parallel pathways competing for the lariat/3′-exon intermediate to yield the canonical lariat and ligated exons as well as the lariat/3′-truncated-exon plus a smaller 3′ exon not ligated to 5′ exon (Figure R4.2 B).

Based on the previous studies, our research had the following goals: (1) to improve the efficiency of the in vitro self-splicing reaction of RmInt1, and thus prevent the formation of unexpected products by testing different RmInt1-transcript contexts and reaction conditions (2) to elucidate the molecular mechanisms involved in the low splicing
efficiency displayed by RmIn1 intron in vitro, for which we introduced point mutations at conserved positions of the ribozyme in order to analyze their influence on the splicing reaction.

**R.4.2 CONSTRUCTS FOR SELF-SPLICING ASSAYS**

The pLM1 construct used in the ΔORF RmInt1 self-splicing assays by Costa et al (2006a; 2006b) is shown in Figure R4.3. In this RNA, RmInt1 5’ exon was trimmed to its last 15 nucleotides, because it has been previously shown that long 5’ exons can inhibit group II intron ribozyme self-splicing reactions (Nolte et al., 1998). pLM1 contained a long 3’ exon of 146 bp, similar to the previously studied ai5γ intron constructs (Daniels et al., 1996), resulting in a 911 nt LT1 transcript synthesized from the T7 promoter (Figure R4.3 A). Interestingly, the 3’ exon of pLM1, which was derived from the natural RmInt1 flanking exon of the ISRm2011-2, carried an IBS1* sequence in its 5 to 10 or 11 first nucleotides. As mentioned above, it competes for the intron binding with the authentic IBS1 sequence (Figure R4.3 B) (Costa et al., 2006a). As a result, a hydrolytic cleavage at this position occurred and truncated products appeared during self-splicing (Figure R4.2).

A pLM1-derivative pLM2 construct was created to circumvent problems associated with cross-interactions produced between the intron and the 3’ exon in pLM1 plasmid. Similar to constructs used in vivo, we trimmed the RmInt1 3’ exon of the ISRm2011-2 to its first 5 nucleotides CCTGG to which a 128 bp fragment from the E.coli lacZ gene was added (Figure R4.3 B). The pLM2 plasmid was linearized by either NdeI or BamHI restriction enzymes, which allowed us to synthesize two alternative transcripts. NdeI digest resulted in a 908 nt LT2 transcript (LT2), which no longer contained the IBS1* sequence. We predicted that the new 3’-exon would not be involved in novel tertiary contacts and it would potentially eliminate the kinetic trap caused by IBS1* sequence. BamHI digest resulted in a smaller 775 nt ST2 transcript. This transcript keeps the EBS3-IBS3 tertiary interaction, while exons are long enough to resolve the precursor from the excised linear intron on a gel. Thus we could assay a transcript similar in length to RNAs produced for the in vivo splicing analysis.
Figure R4.3. Constructs for RmInt1 self-splicing analysis. A Schematic diagram of RmInt1 intron-based constructions shown along with the RNA transcripts generated (below, wavy lines) after treatment with the restriction enzymes indicated. All the pLM constructs, derivative of pUC19 plasmid, contain the last 15 nucleotides of the RmInt1 5‘ exon (E1, green box) and the ΔORF RmInt1 intron (purple heave line), following the T7 promoter (blue arrow head) and a short transcription start sequence recognized by T7 polymerase (in a pink box). The 3’ region in the pLM1 plasmid (Costa et al., 2006) (yellow box) is comprised of 146 bp of the RmInt1 3‘ exon corresponding to the sequence of the ISRm2011-2 (E2, yellow box) and the BamHI (B) restriction site, which allows the generation of 911 bp LT1 transcript by the T7 RNA polymerase from the T7 promoter. The 3’ region in the pLM2 plasmid (yellow and blue boxes) is comprised of 5 bp of the RmInt1 3‘ exon (E2, yellow box) and 145 bp of the lacZ gene (positions 42-169) flanked by Bsh (B) and BamHI (B) sites on the left and Ndel (N) site on the right (blue box). Two RNA transcripts can be synthesized from the T7 promoter in pLM2 plasmid: the 908 nt LT2, after Ndel digestion and the 775 nt ST2, after BamHI digestion. B Detail of the pLM plasmids highlighting relevant sequences of the intron flanking exons. Different regions are colored according to elements in A. IBS elements and restriction sites are indicated.

R.4.3 OPTIMIZATION OF REACTION CONDITIONS

Prior to qualitatively and quantitatively study the kinetic behavior of RmInt1 intron splicing in these three constructs, we tested some reaction conditions previously used for \textit{in vitro} α15γ intron self-splicing (Daniels et al., 1996; Fedorova et al., 2007). We chose ST2 transcript because short exons transcripts has been reported to react with a higher rate
Figure 4.4. Optimization of RmInt1 self-splicing reaction conditions. Timecourses with ST2 transcript at different reaction conditions were carried out both in (NH₄)₂SO₄ buffer, predominantly promoting branched molecules and KCl buffer, mainly giving rise to linear molecules. Representative datasets of precursor disappearance were plotted against time. The data were fit to a single exponential. **A** Optimization of self-splicing of ST2 in (NH₄)₂SO₄ buffer varying the temperature reaction and the composition of salts. In all cases, (NH₄)₂SO₄ buffer contained 500 mM (NH₄)₂SO₄, 100 mM MgCl₂ and 80 mM MOPS (pH 7.5). Rate constants obtained from at least two independent experiments are (0.0027 ± 0.0003) min⁻¹ for ST2 splicing adding 500 mM urea at 42 °C; (0.0039 ± 0.0001) min⁻¹ for ST2 splicing at 42 °C; (0.006 ± 0.001) min⁻¹ for ST2 splicing adding 500 mM urea at 50 °C and; (0.0081 ± 0.0002) min⁻¹ for ST2 splicing at 50 °C. **B** Optimization of self-splicing of ST2 in KCl buffer varying the composition of salts. In all cases, KCl buffer contained 500 mM KCl, 100 mM MgCl₂ and 80 mM Mops (pH 7.5). Rate constants obtained from at least two independent experiments are (0.005 ± 0.001) min⁻¹ for ST2 splicing adding 500 mM urea at 42 °C and (0.0054 ± 0.0006) min⁻¹ for ST2 splicing at 42 °C.

constant (Zingler et al., 2010). All reaction were carried out at high monovalent and Mg²⁺ salt concentrations, since an unbalance in favor of the former has been observed to result in an inhibitory effect (Daniels et al., 1996).

In ammonium sulfate buffer (500 mM (NH₄)₂SO₄, 100 mM MgCl₂ and 80 mM MOPS pH 7.5), we first conducted self-splicing reactions at 42 °C, as described for the α₁γ intron (Daniels et al., 1996; Chu et al., 1998; Chu et al., 2001; Zingler et al., 2010) (Figure R4.4 A, green line). The apparent first order rate constant of this reaction was (0.00391 ± 0.00006) min⁻¹. One of possible reasons for such a slow reaction is misfolding of the precursor RNA. Therefore we attempted to alleviate this potential problem by using subdenaturing urea concentrations (0.5 M final concentration) in our splicing reactions (Fedorova et al., 2007). However, we observed a decrease in the global rate constant which
was (0.00249 ± 0.00006) min⁻¹ (Figure R4.4 A, pink line), which suggests partial unfolding of the RNA structure. We also observed that precursor did not react completely after long timepoints (not shown). In order to increase the amplitude of reaction, we raised the temperature from 42 °C to 50 °C, which resulted in almost complete disappearance of the precursor and two-fold increase of the reaction rate (0.0085 ± 0.0002) min⁻¹ (Figure R4.4 A, red line). Again, the addition of urea resulted in a reduction in the observed rate constant, which was (0.0049 ± 0.0002) min⁻¹ (Figure R4.4 A, blue line). We, therefore, chose to carry out the kinetic analysis of ammonium sulfate self-splicing reactions at 50 °C in the absence of urea.

In potassium chloride buffer (500 mM KCl, 100 mM MgCl₂ and 80 mM MOPS pH 7.5), self-splicing reactions were tested at 42 °C in the presence or in the absence of urea (Figure R4.4 B) (Jacquier and Rosbash, 1986b; Jarrell et al., 1988b; Daniels et al., 1996). A preliminary test performed at 50 °C (not shown) resulted in a vast degradation at long timepoints, therefore we chose not to use these conditions. At 42 °C, the apparent reaction rate was (0.0051 ± 0.00002) min⁻¹. Like in ammonium sulfate buffer, addition of subdenaturing urea concentrations resulted in a decrease in the rate constant, which was (0.0039 ± 0.0001) min⁻¹. Therefore, the kinetic analysis was performed at 42 °C without the addition of urea.

R.4.4 QUALITATIVE ANALYSIS OF BRANCHING AND HYDROLYTIC SELF-SPlicing PATHWAYS OF RmInt1 INTRON

We then analyzed products obtained in course of branching and hydrolytic self-splicing pathways for different RmInt1 intron transcripts. For this purpose, reaction timecourses of LT1, LT2 and ST2 transcripts were carried out in the presence of high concentration of monovalent salts and the products were examined by PAGE (Figure R4.5). The branching pathway was induced by adding a (NH₄)₂SO₄ buffer to the reaction mixture (Figure R4.5 A) while the hydrolytic route was forced by adding a KCl buffer (Figure R4.5 B). Although not being favored, the hydrolytic pathway also occurred in the presence of ammonium salts (Figure R4.5 A).

In the presence of (NH₄)₂SO₄ buffer, LT1 gave rise to a comparable pattern of products than those observed by Costa et al (2006a) (Figure R4.2), despite some changes in the composition of reaction buffer (see Materials and Methods and legends of Figures
Figure R4.5 Self-splicing of ΔORF RmIn1 from pLM constructs. Each sample, containing about 3 nM of 32P [α-GTP] labelled precursor was subjected to self-splicing reactions in the presence of high concentrations of monovalent salts prior to be analyzed in a high resolution polyacrylamide gel. On the far left and far right of autoradiograph panels are tentative interpretations of the various bands observed following color codes as in Fig R3.3. Upper panels represent self-splicing reactions from LT1 transcript; middle panels, those from LT2 transcript and; lower panels, those from ST2 transcript. **A** Self-splicing reactions carried out in the presence of (NH₄)₂SO₄ buffer at 50 °C. The reaction buffer contained 500 mM (NH₄)₂SO₄, 100 mM MgCl₂ and 80 mM MOPS (pH 7.5). **B** Self-splicing reactions carried out in the presence of KCl buffer at 42 °C. KCl buffer is 500 mM KCl, 100 mM MgCl₂ and 80 mM Mops (pH 7.5). Pr is precursor RNA; ILar-E2 is intron-lariat/3'-exon intermediate; ILar-E2* is the intron-lariat/3'-exon intermediate cleaved at the IBS1* sequence; ILar is the intron lariat; PLin is the complete set of linear products; ILin is the linear intron; Hyp is an hypothetical intron-derived product.

R4.2 A and R4.4 A). Low-migrating (presumably) branched molecules were predominant, especially those assumed to be lariat intron and lariat-intron/IBS1*-truncated-3'-exon (Figure R4.5 A, upper panel). The band corresponding to a lariat-intron/complete-3'-exon was barely detectable, disappearing completely at 150 minutes of reaction. Linear products were also observed. Of these, three were predominant over a background of linear products probably caused by RNA degradation. We hypothesized that these linear products are an IBS1*-truncated precursor, an IBS1*-truncated linear-intron intermediate and a linear intron form, as previously reported by Costa et al (2006a). However, unlike in previously published observations, precursor reacted to completion. In LT2 transcript, self-splicing timecourse revealed a smaller number of discrete product bands (Figure R4.5 A, middle panel). Early in time, a low-migrating band, probably a branched intermediate was observed, which then gave rise to a branched smaller-sized lariat molecule. A unique linear band, which accumulated over time, was also detected (Figure R4.5 A). Precursor, which reacted completely, was accompanied by a less intense lower-migrating linear band, probably because the transcript is a mixture of two species. In timecourses conducted with ST2 transcript (Figure R4.5 A, lower panel), the precursor reacted almost to completion, giving rise, via the branching pathway, to two low-migrating bands assumed to be lariat/3’-exon intermediate and lariat intron, respectively. Along with these, a third product, deriving apparently from the intermediate/3'-lariat form, was observed. Alternatively, by the hydrolytic pathway, two linear bands were observed, the first one originated from the precursor and interpreted to be a linear intermediate, and the second, a likely linear intron.

Consistent with previously published data (Costa 2006a), in KCl buffer the LT1 transcript gave rise to three linear bands, previously identified as an IBS1*-truncated
precursor, a linear-intron/IBS1*-truncated-3’-exon intermediate and a linear intron form (Figure R4.5 B, upper panel). In timecourses from the LT2 transcript, a prominent product band, assumed to be a linear intron, accumulated over time as precursor reacted; however, no other linear intermediate product was observed. Finally, the reaction of the ST2 transcript resulted, first, in a transient linear band of a slightly smaller size than the precursor that, in turn, produced two smaller linear bands interpreted to be a linear intermediate and a linear intron form, respectively.

**R.4.5 CHARACTERIZATION OF SELF-SPlicing PRODUCTS BY DNAZyme CLEAVAGE**

DNAzymes, also known as deoxyribozymes, are catalytically active DNA molecules that cleave RNA (Silverman, 2005). DNAzymes were first identified through in vitro selection in 1994 (Breaker and Joyce, 1994). The mechanism of RNA cleavage by DNAzymes involves the attack of an RNA 2’-hydroxyl group on the adjacent phosphodiester linkage in the presence of Mg^{2+} or other divalent ions, resulting in 2’,3’-cyclic phosphate and 5’-hydroxyl RNA termini (Figure R4.6 A) (Pyle et al., 2000; Silverman, 2005). The DNAzyme in the 10-23 family, which has been used in this work, is comprised of a small catalytic core flanked by two arms that base pair with RNA (9-14 nucleotides depending on the sequence) (Figure R4.6 A) (Santoro and Joyce, 1997) [see Silverman (2005) for a comprehensive study of different DNAzyme families]. They can cleave any 5’-purine-pyrimidine-3’ linkage (5’-RY-3’) in the RNA sequence, although a uridine is preferred rather than a cytidine (Pyle et al., 2000). DNAzymes can be used for excision of small fragments of RNA from a larger parent RNA in order to study the fragment in detail (Pyle et al., 2000) or for linearization of circular molecules.

In this study, DNAzymes were used for the generation of a pattern of fragments characteristic of each RNA molecule, as in a restriction analysis. Through the comparison of motifs we carried out a preliminary identification of forms produced during self-splicing, which was the previous step to the quantitative analysis of reaction rates.

**R.4.5.1 Design of DNAzymes for splicing product analysis**

Six DNAzymes were designed taking into account cleavage site preferences. Cleavage sites were interspersed in the sequence, in order to get as many different cleavage
Results and Discussion – Chapter 4

A

RNA substrate DNAzyme

B

DZ1

DZ2

DZ3

DZ4

DZ5

DZ6

C

0 10 30 60 120 (min) / α

Pr

DZ1

Pr

DZ3

Pr

DZ4

Pr

DZ5

Pr

DZ6

0 10 30 60 120 (min) / α

0 10 30 60 120 (min) / α

0 10 30 60 120 (min) / α

775

525

250

407

368

449

226

775

716

775

38
Figure R4.6. DNAzymes for the analysis of self-splicing products. A Schematic representation of a DNAzyme with the 10-23 motif (Santoro and Joyce, 1997). It is comprised of two arms (light green) with a duplex stability of 10-12 kcal/mol (9-14 nt depending on the sequence) base paired to the RNA substrate (orange). The catalytic core is denoted by a green uppercase-letter loop. The arrow indicates the DNAzyme cleavage site produced between an unpaired purine nucleotide (R) and a pyrimidine (Y) in the RNA substrate in the presence of a divalent metal ion cofactor (M$^{2+}$). The cleavage results in 2',3'-cyclic phosphate and 5'-hydroxyl termini in the 5' and 3' RNA products, respectively. B Schematic of RmInt1 intron RNA showing the cleavage sites of designed DNAzymes on the molecule. DNAzymes DZ1, DZ2 and DZ3 were designed to cut the RNA domain I, whereas other three, DZ4, DZ5 and DZ6, recognized regions in domains II, V and VI, respectively. Base-paired regions of DNAzymes in the RNA substrate are denoted by bold lines in green. The cleavage site is indicated by a red arrow between the uppercase letter nucleotides (green) in the RNA substrate. C Cleavage efficiency of designed DNAzymes. ^{32}P [z-GTP] labelled precursor from the ST2 transcript (775 nt) was combined with an excess of DNAzyme, heated at 95 °C to denature RNA and DNA and added to a magnesium salt buffer (see Material and Methods). Aliquots were taken out at selected timepoints and cleavage products were resolved by denaturing polyacrylamide gel electrophoresis (PAGE). Precursor RNA is indicated by a green arrowhead on the left of each panel, while sizes of precursor and cleavage products are shown on the right of each panel.

patterns as possible. (Figure R4.6 B; see Table MX.X in Materials and Methods for DNAzyme sequences). Three DNAzymes (DZ) were located in the domain I of the ribozyme: DZ1, cleaving between nucleotides G25 and U26 of RmInt1 intron; DZ2, between G230 and U231 and; DZ3, between A348 and U349. Other three DNAzyme cleavage sites were located in domains II, V and VI, respectively: DZ4 cleaved between G429 and U430; DZ5 between G1840 and U1841 and; DZ6, between G1860 and U1861.

The cleavage efficiency of these DNAzymes was determined using ST2 transcript as a substrate, which was chosen because of its small size and the appropriate range of fragments produced (Figure R4.6 C). Cleavage efficiency of DZ1, calculated as the reaction extent after 2 hours of incubation at 37 °C in the presence of a magnesium salt (see Material and Methods for reaction conditions) was 98.5 %, which is the best among all DNAzymes tested. Two fragments of 730 nt (Figure R4.5 C, upper left panel) and 45 nt (not shown in the image) were obtained. DZ2 exhibited 89.2 % efficiency and produced two fragments of 525 nt and 250 nt (Figure R4.5 C, upper middle panel). DZ3, with a 45.5 % of cleavage efficiency, gave rise to two fragments of similar size, 407 and 368 nt. The DZ4 cut the ST2 transcript with an efficiency of 40.3 %, resulting in RNA fragments of 449 and 326 nt length. DZ5 showed a 20.2 % of cleavage efficiency and produced a large fragment of 716 nt and a smaller one of 59 nt in the ST2 molecule (not shown). Finally, DZ6 did not cleave the ST2 transcript at all, probably because the base pairing region is rich in GC and then hard to denature and access. Two DNAzymes, DZ1 and DZ2, were
classified as fast cutters (Pyle et al., 2000), because they were able to almost completely cleave the RNA substrate in the first 10 minutes of reaction. In contrast, DZ3, DZ4 and DZ5 were considered as slow cutters, producing less than 50% cleavage after 2 hours.

We chose to use DZ1 and DZ4 for the characterization of self-splicing products. DZ1, due to its ability to process the RNA substrate with a very high efficiency, was used to linearize circular forms (i.e. covalently closed molecules). DZ4, which yielded less than 50% cleavage, was chosen because, in combination with DZ1, the intron and other splicing forms were divided into two fragments in a range of lengths appropriate to be accurately resolved by PAGE.

**R.4.52 Self-splicing product characterization**

Self-splicing products generated by two-hour preparative reactions of LT1, LT2 and ST2 transcripts were gel-extracted and reacted with DZ1, DZ4 or a combination of both (Figure R4.7). Precursor, undigested and partially digested products were used as size controls. LT1 self-splicing bands cut with DNAzymes were identified taking into account self-splicing product analysis previously reported for RmInt1 intron (Costa et al., 2006a). According to this, probable unusual intermediate products were detected, in which the truncated 3’ exon was attached to lariat (Figure R4.6 A, IIar-E2*, as part of band labeled as a, or linear intron (IIin-E2*, as part of band b. Lariat (IIar) and linear (IIin) intron bands were co-purified with lariat/3’-truncated-exon (IIar-E2*) and linear/3’-truncated-exon (IIin-E2*) bands, respectively. We could not isolate the lariat/3’-full-exon described in Costa’s work, although we strongly believe it corresponds to the low-migrating band observed in time course gels (Figure R4.5 A, upper panel) but almost completely reacted after 2 hours.

In LT2 transcript, three self-splicing product bands, with the corresponding precursor, were isolated and preliminary characterized by DNAzyme digestion (Figure R4.7 B; bands a, b and c). The lariat intron (IIar) likely corresponded to band b, since it was linearized when treated with DZ1 (giving a 740 nt band) and divided into two fragments of about 404 and 336 nt when digested with a mix of DZ1 and DZ4, which were the sizes expected after this treatment. The lariat/3’-exon intermediate (IIar-E2) was isolated as band a. In this case, the linearized band had nearly the same mobility as the precursor while, as a result of the treatment with a DZ1/DZ4 mix, the band released a 404 nt fragment common with lariat intron plus an additional fragment of lower mobility,
Figure 4.7. Identification of RmInt1 self-splicing products by DNAzyme treatment. ³²P [α-GTP] labelled LT1 (A), LT2 (B) and ST2 (C) transcripts were subjected to a preparative-scale self-splicing reaction for 2 hours in (NH₄)₂SO₄ buffer at 50 °C (see Materials and Methods). Products of self-splicing were purified by PAGE, eluted from the gel and cleaved with the DNAzymes indicated. In upper left panels, details of the self-splicing products at 2 hours of reaction are shown. The precursor and main products are indicated on the left while the corresponding isolated band are indicated with lower-case letters on the right. The cleavage products after DNAzyme treatment are shown in the upper right panels, along with known molecular sizes indicated on the left. Arrows indicate undigested remaining products. Below, schematic interpretations of the self-splicing products are displayed, indicating DNAzyme cleavage sites and the various fragments produced. Asterisks indicate branched products which can have a retarded migration. The color code and abbreviations are those specified in Fig R3.3 and R3.4, respectively.

which could be that expected of 484 nt. The band ε was preliminary identified as the linear intron (IIIn), since it gave rise to two products of sizes similar to the expected 429 and 311 nt after DZ4 cleavage. Besides, we noted that two bands close to the previous ones were detected, giving two doublets, which probably were due to degradation occurred after product isolation. There was no evidence of the presence of a linear/3′-exon intermediate (IIIn-E2).
The ST2 transcript yielded self-splicing products of closed related sizes more difficult to resolve in a preparative gel. For that reason, branched and linear products were extracted as two single bands \(a\) and \(b\), respectively and subjected to DNAzyme cleavage (Figure R4.7 C). Branched molecules included three products observed in time-course gels, which probably were the lariat/3’exon intermediate (ILar-E2), lariat intron (ILar) and a hypothetical intron-derived band (Hyp). After DZ1 and DZ4 digestion, a 404 nt fragment common to the three products was produced, along with three smaller bands, which could be those of 351, 336 and 330 nt, respectively, expected to obtain after digestion of the branching products above. Since linear products were partly contaminated by the slower migrating precursor, these were more difficult to characterize. Two cleavage fragments of about 326 and 311 nt were obtained, which could correspond to the expected linear intron (ILin) and linear/3’-exon intermediate (ILin-E2) forms, along with a fragment of about 429 nt, probably derived from both products after DZ4 cleavage.

**R4.6 KINETIC ANALYSIS OF BRANCHING AND HYDROLYTIC SELF-SPLICING PATHWAYS WITH HIGH CONCENTRATIONS OF MONOVALENT ION**

Once the self-splicing products formed via branching and hydrolytic pathways were identified, we analyzed their evolution over time under the best reaction conditions, determining their rate constants and amplitudes for the three classes of transcripts (see above). With the data collected, we put them all together to establish the best context of expression and reaction conditions for further *in vitro* studies of RmInt1 intron.

In order to determine the reaction rates, we first described each model by a differential equation and found an exact analytical solution for each equation (not shown). Then, the models (equations) were fitted to the experimental data using nonlinear regression to obtain estimates of parameters using the software packages Kaleidagraph (Abelbeck Software) and DynaFit (Kiokin Ltd). In complex reaction pathways, a simplification was introduced for the estimation of \(k_{br}\) and \(k_{hr}\), in which all products proceeding from the branching and the hydrolytic pathways, respectively, were added together. For a detailed description of this approach and the equations used, refer to Materials and Methods (Section M.13.4).
R.4.6.1 Kinetics of self-splicing in the LT1 transcript

Self-splicing reactions of LT1 transcript under (NH₄)₂SO₄ buffer at 50 °C (Figure R4.8 A, left panel) are described by the model shown in Figure R4.2 B. Notably, we do not observe the accumulation of any products containing intron and full-length 3’-exon, which are the intermediates formed after the first step of each reaction. This indicates that for all three pathways (branching, hydrolysis and exon truncation) the second steps are much faster than the first steps. Thus, the first steps are rate-limiting for all three reactions. Therefore we are determining rate constants only for the first steps of all respective reactions and the kinetic model can be simplified as illustrated in Scheme 1:

\[ \text{Lariat-3’-exon}^* \text{ intermediate [ILar-E2*]} \]
\[ \begin{array}{c}
\text{RmInt1} \\
\text{L11 precursor [Pr]} \\
\mid \overrightarrow{\text{k}_2} \\
\mid \overrightarrow{\text{k}_{hv=3}} \\
\text{Lariat intron [ILar]} \\
\mid \overrightarrow{\text{Linear products [Pl.in]}}
\end{array} \]

Scheme 1

where \( k_{br} = k_1 + k_2 \) and \( k_T = k_1 + k_2 + k_3 \). All products are considered to evolve simultaneously and independently from precursor, so the sum of product fractions \( f_{nu} \) and \( f_{hv} \) will equal 1.

The overall \( k_r \) value of the reaction was 0.020 min⁻¹ (note that standard errors for all rates are provided in Table R4.1), which was two-fold higher than that previously reported for RmInt1 intron (0.01min⁻¹) (Costa et al., 2006a). The fraction of reactive

Figure R4.8. Kinetic analysis of RmInt1 self-splicing under high concentrations of monovalent ions. Splicing reactions were performed in the presence of either 500 mM (NH₄)₂SO₄ buffer (A) or 500 mM KCl buffer (B). Experimental data for the decrease in precursor RNA and the evolution of products from Figure R3.4 were plotted and fit using DinaFit (BioKin Ltd). Models describing reactions in (NH₄)₂SO₄ buffer were fitted to equation (6) to (9) for LT1, (10) to (13) for LT2 and (14) to (19) for ST2 (see Materials and Methods). Models describing reactions in KCl buffer were fitted to equation (20) and (21) for LT1 and LT2, and to equations (22) to (24) for ST2 (see Materials and Methods). All reactions proceeded to completion at long times (8–9 hours), but the significant amount of RNA degradation that occurred after four hours of reaction prevented quantitative analysis after that time in the majority of samples. Reaction rates and product constants are summarized in Table R3.1. Pr is precursor RNA; ILar-E2 is intron-lariat/3’-exon intermediate; ILar-E2* is the intron-lariat/3’-exon intermediate cleaved at the IBS1* sequence; ILar is the intron lariat; Pl.in is the complete set of linear products; ILin is the linear intron; Hyp is an hypothetical intron-derived product.
precursor observed was also increased [about 70% versus 30% (Costa et al., 2006a) after 7 hours of reaction]. The plot of the overall reaction showed a monophasic behaviour, although the semi-log plot could only be adjusted to a straight line as far as 60 minutes of reaction (Figure R4.9 A, green line), which may be caused by more noticeable degradation of the precursor RNA after this time point. Comparing with a15γ intron, all precursor molecules of RmInt1 intron fell into the “fast” population category (Daniels et al., 1996), although the RmInt1 reaction rate was 2.5-fold lower than that of a15γ.

In this buffer, the branching pathway predominated over the hydrolytic one ($k_{br} = 0.012 \text{ min}^{-1}$ versus $k_{hy} = 0.008 \text{ min}^{-1}$). Relative amounts of lariat and linear intron were 0.571 and 0.4, respectively (Table R4.1). The branching population was dissected as lariat intron and lariat/truncated-3'-exon (Figure R4.8 A, left panel; Scheme1), with relative fractions of (0.259 ± 0.008) and (0.36 ± 0.01), respectively. Partial reaction rates of lariat/truncated-3'-exon and lariat intron were $k_1 = (0.0081 \pm 0.0006) \text{ min}^{-1}$ and $k_2 = (0.0047 \pm 0.0005) \text{ min}^{-1}$, respectively. The truncated intermediate, therefore, predominated over the lariat intron.

Self-splicing timecourse of I.T1 transcript in KCl buffer at 42 °C (Figure R4.5 B, upper panel) is shown in Figure R4.8 A (right panel). We simplified the model by treating multiple linear products as a single species. As a result, linear products were fitted to a first order single exponential kinetics, in accordance with Scheme 2:

\[
\begin{align*}
\text{RmInt1} & \quad \text{I.T1 or I.T2 precursor} \quad k_{hy} = k_4 \quad \text{Linear intron products [PLin]} \\
& \quad \text{[Pr]} \quad \text{or Linear intron [ILin]} \\
\end{align*}
\]

Scheme 2

where $k_f = k_{hy} = k_{4, y}$ is the sum of all products evolving through the hydrolytic pathway.

The global reaction rate under KCl buffer at 42 °C was 0.0069 min’1, which is substantially slower than that observed under (NH₄)₂SO₄ buffer at 50 °C. The reaction showed a monophasic behavior, with a single slowly reacting population of molecules. Fractions of linear products at the reaction endpoint was 0.87 (Table R4.1).
<table>
<thead>
<tr>
<th>Transcript</th>
<th>Ionic reaction conditions</th>
<th>Temp (°C)</th>
<th>$k_T$ (min$^{-1}$)</th>
<th>$k_r$ (min$^{-1}$)</th>
<th>$f_r$</th>
<th>$k_{11}$ (min$^{-1}$)</th>
<th>$f_{11}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT1</td>
<td>500 mM (NH$_4$)$_2$SO$_4$</td>
<td>50</td>
<td>0.020 ± 0.002</td>
<td>0.012 ± 0.002</td>
<td>0.571 ± 0.009</td>
<td>0.008 ± 0.001</td>
<td>0.4 ± 0.03</td>
</tr>
<tr>
<td>LT2</td>
<td>500 mM (NH$_4$)$_2$SO$_4$</td>
<td>50</td>
<td>0.0115 ± 0.0005</td>
<td>0.0077 ± 0.0007</td>
<td>0.71 ± 0.02</td>
<td>0.0038 ± 0.0002</td>
<td>0.352$^1$</td>
</tr>
<tr>
<td>ST2</td>
<td>500 mM (NH$_4$)$_2$SO$_4$</td>
<td>50</td>
<td>0.0083 ± 0.0002</td>
<td>0.0054 ± 0.0004</td>
<td>0.641 ± 0.008</td>
<td>0.0029 ± 0.0004</td>
<td>0.35 ± 0.03</td>
</tr>
<tr>
<td>LT1</td>
<td>500 mM KCl</td>
<td>42</td>
<td>0.0069 ± 0.0005</td>
<td>0</td>
<td>0</td>
<td>$k_T$</td>
<td>0.87 ± 0.04</td>
</tr>
<tr>
<td>LT2</td>
<td>500 mM KCl</td>
<td>42</td>
<td>0.0045 ± 0.0002</td>
<td>0</td>
<td>0</td>
<td>$k_T$</td>
<td>0.9 ± 0.1</td>
</tr>
<tr>
<td>ST2</td>
<td>500 mM KCl</td>
<td>42</td>
<td>0.0060 ± 0.0002</td>
<td>0</td>
<td>0</td>
<td>$k_T$</td>
<td>0.83 ± 0.03</td>
</tr>
</tbody>
</table>

$^1$ Rates reported here are observed rates derived from the plots in Figure R3.6. $k_T$ is the total rate of splicing; $k_r$ is the rate of branching; $f_r$ is the fraction of molecules that splice through branching; $k_{11}$ is the rate of hydrolytic splicing and; $f_{11}$ is the fraction of molecules that splice through hydrolysis. $k_T$ value was determined from the decrease in precursor using Kaleidagraph (Abelbeck Software). The fractional coefficients, $f_r$ and $f_{11}$, were determined from the fit of product evolution data. From $k_T$, $f_r$ and $f_{11}$, $k_r$ and $k_{11}$ were calculated using equations (4) and (5) (see Materials and Methods). Rates and coefficients were also calculated using DynaFit (BioKin Ltd.) and equations (6) to (10), showing a high degree of correspondence. Experimental errors represent the standard error calculated from the fits to rate equations.

$^2$ In addition to the salts listed, each reaction contained 100 mM MgCl$_2$ and 80 mM Mops (pH 7.5).

$^3$ This value was fixed based on the linear population observed at long timepoints.
Figure R4.9. Semi-log plots of reaction in LT1, LT2 and ST2 transcripts. Logarithms to base ten of the fractions of precursor from timecourses were plotted against time. LT1 timecourse was reduced to its first 60 minutes of reaction since after that time dataset did not fit well to equation (1). The self-splicing reaction timecourses, carried out in A (NH₄)₂SO₄ buffer at 50 °C and B KCl buffer at 42 °C, showed a monophasic kinetic profile for which a single slope was observed.

### R.4.6.2 Kinetics of self-splicing in the LT2 transcript

Evolution of precursor and products of self-splicing reactions of LT2 transcript in (NH₄)₂SO₄ buffer at 50 °C conditions is shown in Figure R4.8 B (left panel). Interestingly, in this case we observed the formation of lariat-3’exon intermediate. This suggested that first step of splicing via branching pathway was not rate-limiting for this construct, and we could therefore determine rate constants for both steps of splicing. Surprisingly, this was not the case for the hydrolytic pathway, where the formation of the intermediate was not observed and the first step was still rate-limiting. This behavior can be described by a parallel-sequential kinetic model, as shown in Scheme 3:

![Scheme 3](image)

138
where \( k_T = k_{br} + k_{hy} \). Lariat/3'-exon intermediate and linear intron evolved simultaneously and independently from precursor, while the lariat intron was produced from the intermediate. The sum of product fractions \( f_{pr} \) and \( f_{hy} \) equals 1, being the \( f_{pr} \) the sum of all the products evolving through the branching pathway.

The global reaction rate \( k_1 \) was 0.0115 min\(^{-1}\) (Table R4.1), a value slightly lower but in the range of that observed in LT1 transcript. The fraction of reactive precursor (observed at 7 hours of reaction, for comparing) was 97.5%, which reacted completely at the end of the timecourse (8 hours). The reaction also showed that all the precursor molecules fell into one homogeneous population with a clear monophasic behavior (Figure R4.9 A, blue line). According to Daniels et al (1996), the value of the overall reaction of the LT2 precursor was classified as a fast population of molecules.

The rate constant for the first step of splicing via branching pathway in (NH\(_4\))\(_2\)SO\(_4\) buffer was 2-fold higher than that of the hydrolytic pathway: 0.0077 min\(^{-1}\) versus 0.0038 min\(^{-1}\), respectively (Table R4.1). Branched forms predominated over linear ones, with a relative population of 0.71 for the first and 0.352 for the second (Table R4.1). The rate of conversion of lariat/3'-exon intermediate to lariat intron was estimated in \((0.0122 \pm 0.0008)\) min\(^{-1}\), slightly higher than the rate for the production of intermediate, which led to the almost complete disappearance of this form at the end of the timecourse. In the hydrolytic pathway, the linear intron was the only product observed, with a relative fraction of 0.352 (Table R4.1). No linear intermediates were detected.

In KCl buffer (Figure R4.8 B, right panel), the only product formed was the linear intron. For that reasons, we considered that LT1 fitted well to a first order single exponential kinetics, in accordance with Scheme 2 (see above). The global reaction rate was 0.0045 min\(^{-1}\). The reaction showed a monophasic behavior, with only one slow population of molecules reacting. The fraction of linear products corresponding to the reaction endpoint was 0.9 (Table R4.1).

**R.4.6.3 Kinetics of self-splicing in the ST2 transcript**

Self-splicing reaction of ST2 transcript in (NH\(_4\))\(_2\)SO\(_4\) buffer at 50 °C (Figure R4.9 A, upper panel) is shown in Figure R4.8 C (left panel). In this case, the pattern of product evolution is more complex. Intron-3'-exon intermediates were observed for both branching and hydrolysis pathways, suggesting that the first step of splicing by either pathway was not rate-limiting. Thus, for each pathway, we could determine the rate
constants for both first and second steps of splicing. In addition, we observed that the lariat-3'-exon intermediate gave rise to a secondary product in addition to the lariat intron. This system can be described by the following parallel-sequential kinetic model (Scheme 4):

\[ \text{RmInt1} \xrightarrow{k_{br}} \text{Lariat-3'-exon intermediate} \xrightarrow{k_2} \text{Lariat intron [II.lar]} \]
\[ \xrightarrow{k_{hy}} \text{Putative intron-derived form (Put)} \]
\[ \text{ST2 precursor} \xrightarrow{k_{hy}} \text{Linear-3'-exon intermediate} \xrightarrow{k_5} \text{Linear intron [II.in]} \]

where \( k_T = k_{br} + k_{hy} \). Lariat/3'-exon and linear/3'-exon intermediate forms evolved simultaneously and independently from precursor, while the lariat intron and the putative intron-derived form (Put) were produced from the intermediate. The sum of product fractions \( f_{br} \) and \( f_{hy} \) equals 1, being the \( f_{br} \) the sum of all the products evolving through the branching pathway and \( f_{hy} \) the sum of all products evolving through the hydrolytic pathway.

The overall observed reaction rate \( (k_T) \) was 0.0083 min\(^{-1}\) (Table R4.1), the lowest value obtained among the three transcripts studied. The fraction of reactive precursor (observed at 7 hours of reaction, for comparing with LT1 and LT2) was 89 %, and it reacted almost completely at the end of the reaction. Plot of the overall reaction of ST2 showed a monophasic behavior (Figure R4.9 A, red line). According to Daniels et al (1996), the value of the reaction rate of the ST2 precursor was classified as belonging to a slow population of molecules.

The branching pathway in (NH\(_4\))\(_2\)SO\(_4\) buffer was predominant over the hydrolytic one. The relative amounts of lariat and linear forms were \( f_{br} = 0.641 \) and \( f_{hy} = 0.35 \), respectively (Table R4.1). The rate constants for the first step of branching and hydrolytic splicing were 0.0054 min\(^{-1}\) and 0.0029 min\(^{-1}\), respectively (Table R4.1). The rate at which lariat/3'-exon intermediate was converted to lariat intron was \( (0.0108 \pm 0.0004) \) min\(^{-1}\) whereas the rate at which the linear/3'-exon intermediate was converted to linear intron was \( (0.0065 \pm 0.0006) \) min\(^{-1}\).

In KCl buffer (Figure R4.8 C, right panel), the linear/3'-exon intermediate accumulated, indicating that the first step was not rate limiting. Thus, for each pathway, we
could determine the rate constants for both first and second steps of splicing. Product evolution followed a parallel-sequential kinetic model, according with Scheme 5:

\[
\begin{align*}
\text{ST2 precursor} & \quad [P_4] \quad \xrightarrow{k_4} \quad \text{Linear-3’ exon intermediate} \quad [\text{ILin-E2}] \quad \xrightarrow{k_5} \quad \text{Linear intron} \quad [\text{ILin}]
\end{align*}
\]

where \( k_4 = k_{h_0} = k_{v} \).

The global reaction rate under KCl buffer was 0.0048 min\(^{-1}\). The reaction showed a monophasic behavior, with only one slow population of molecules reacting. The reaction endpoint was 0.83 (Table R4.1).

R4.7 ANALYSIS OF POINT MUTANTS FOR THE STUDY OF SELF-SPLICING MECHANISM IN THE RmInt1 INTRON

In the previous chapter, we analyzed the effect of point mutations on the \textit{in vivo} splicing reaction. Aspects influencing the observed results included not only the secondary structure and tertiary interactions of the ribozyme but also the modulatory effect of the intron-encoded protein (IEP) acting as a maturase in the folding and maintenance of the intron structure as well as a splicing cofactor and; the possible role of host-encoded proteins in the splicing reaction. Of all this factors, the effect of point mutations at conserved positions in the ribozyme active site can be studied in isolation using the self-splicing reaction.

For this purpose, we introduced point mutations and deletions described in the previous chapter (Figure R3.1) in the context of the ST2 transcript (see Materials and Methods, section MX.X). As we recall from Chapter 3, we studied point mutations and deletions in conserved positions in the RmInt1 ribozyme which had been previously reported to affect the first, second or both steps of the splicing reaction, reading it in conjunction with the location of mutated positions in the three-dimensional structure of the intron and the tertiary interactions in which they are involved. Thus, we assayed the ability of each mutant transcript to catalyze the complete splicing reaction by estimating the overall reaction rate and all the reaction constants in each individual step in the process. Besides, in order to examine the two steps of splicing, we calculated relative
E  $\Delta\gamma$+GEBS3C

F  $\gamma C$

G  $\gamma C-C\gamma G$

H  $c\gamma G$
Results and Discussion – Chapter 4

**I** C1883A

**J** C1883G

**K** C1883U

**L** GEB50C
efficiencies of the first and second splicing steps of mutants with respect to WT, under (NH₄)₂SO₄ and KCl buffers conditions.

As seen in Section R4.5, the wild-type control RNA (WT-ST2) reacted with a rate constant of 0.0086 min⁻¹ in (NH₄)₂SO₄ buffer, yielding a predominant intron lariat form (k₂ = 0.0108 min⁻¹) produced through a lariat/3'‐exon intermediate via branching pathway (k₁ = 0.0063 min⁻¹). We also observed a putative intron‐derived product migrating near the lariat form (k₃ = 0.0028 min⁻¹). The linear forms produced through the hydrolytic pathway reacted with a rate constant of 0.0023 min⁻¹ in (NH₄)₂SO₄ buffer, and with a rate constant of 0.0044 min⁻¹ in KCl buffer. The rate of linear intron formation under the latter buffer conditions was 0.0063 min⁻¹.

Time courses of mutant transcripts carried out under high concentrations of (NH₄)₂SO₄ and KCl buffers are shown in figure R4.10. Reaction rates of self‐splicing are summarized in Table R4.2. Relative first and second splicing step efficiencies are summarized in Table R4.3.

The mutant DV-CGA, in which positions GUU (1840-2) of the domain V base-pairing with the catalytic triad AGC were substituted by CGA, was severely affected in the first step of splicing under (NH₄)₂SO₄ (kₐ = 17x10⁶ min⁻¹; kₚ = 183x10⁶ min⁻¹) and KCl buffers, in which the transcript remained unreactive. Notably, this mutant construct was no affected in the second splicing step of the branching pathway (Table R4.3), showing a tiny accumulation of lariat/3'‐exon intermediate which rapidly reacted to give lariat intron (k₃ = 0.02).

We observed that the deletion of the dinucleotide GA (453-4) (J2/3-ΔGA mutant), located in the linker region J2/3, dramatically inhibited the first step of the splicing reaction as well (Table R4.2). In addition, J2/3ΔGA did not react through the hydrolytic pathway, and only branching was contributing to the reaction rate (178·10⁶ min⁻¹). Moreover, the k₃ for lariat formation was not estimated, since there was no possibility of discerning if the accumulation of the unique branched product observed corresponded to
**TABLE R4.2.** Rate constants for self-splicing in wild-type and mutant transcripts\(^a\)

<table>
<thead>
<tr>
<th>Transcript(^b)</th>
<th>(k_T) (min(^{-1}))</th>
<th>(k_{0-1}) (min(^{-1}))</th>
<th>(k_2) (min(^{-1}))</th>
<th>(k_3) (min(^{-1}))</th>
<th>(k_{D}-k_s) (min(^{-1}))</th>
<th>(k_5) (min(^{-1}))</th>
<th>(k_6) (min(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(NH(_4))(_2)SO(_4) buffer(^c)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>0.0086 ± 0.0002</td>
<td>0.0063 ± 0.0001</td>
<td>0.0108 ± 0.0003</td>
<td>0.0038 ± 0.0003</td>
<td>0.0023 ± 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV-CGA</td>
<td>200 ± 10(10(^{-6}))</td>
<td>17 ± 5 (10(^{-6}))</td>
<td>0.02 ± 0.01 &amp;</td>
<td>–</td>
<td>183 ± 6 (10(^{-6}))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2/3-ΔGA</td>
<td>178 ± 4 (10(^{-6}))</td>
<td>178 ± 4 (10(^{-6}))</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δγ</td>
<td>29 ± 2 (10(^{-6}))</td>
<td>2 ± 1 (10(^{-6}))</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>27 ± 1 (10(^{-6}))</td>
<td></td>
</tr>
<tr>
<td>Δγ+GEBS3C</td>
<td>307 ± 9 (10(^{-6}))</td>
<td>35 ± 5 (10(^{-6}))</td>
<td>0.004 ± 0.002 &amp;</td>
<td>–</td>
<td>–</td>
<td>272 ± 4 (10(^{-6}))</td>
<td></td>
</tr>
<tr>
<td>GγC</td>
<td>0.0070 ± 0.0003</td>
<td>0.0049 ± 0.0002</td>
<td>0.015 ± 0.001</td>
<td>0.0050 ± 0.0008</td>
<td>0.0021 ± 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GγC-Cγ(_G)</td>
<td>0.0058 ± 0.0004</td>
<td>0.0024 ± 0.0002</td>
<td>0.028 ± 0.009 &amp;</td>
<td>0.014 ± 0.006 &amp;</td>
<td>0.0034 ± 0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cγ(_G)</td>
<td>0.0062 ± 0.0002</td>
<td>0.0027 ± 0.0001</td>
<td>0.023 ± 0.003</td>
<td>0.010 ± 0.001</td>
<td>0.0035 ± 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883A</td>
<td>0.0084 ± 0.0002</td>
<td>0.0059 ± 0.0001</td>
<td>0.0100 ± 0.0004</td>
<td>0.0027 ± 0.0003</td>
<td>0.0025 ± 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883G</td>
<td>0.0068 ± 0.0002</td>
<td>0.0045 ± 0.0001</td>
<td>0.0135 ± 0.0005</td>
<td>0.0042 ± 0.0003</td>
<td>0.0023 ± 0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883U</td>
<td>0.0098 ± 0.0005</td>
<td>0.0069 ± 0.0003</td>
<td>0.09 ± 0.02 &amp;</td>
<td>0.020 ± 0.006 &amp;</td>
<td>0.0029 ± 0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEBS3C</td>
<td>0.0095 ± 0.0004</td>
<td>0.0059 ± 0.0002</td>
<td>0.046 ± 0.08</td>
<td>0.012 ± 0.003</td>
<td>0.0036 ± 0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCl buffer(^d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>0.0033 ± 0.0003</td>
<td>0.0033 ± 0.0003</td>
<td>0.0038 ± 0.0005</td>
<td>0.0013 ± 0.0003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV-CGA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>J2/3-ΔGA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Δγ</td>
<td>0.0009 ± 0.0001</td>
<td>0.0009 ± 0.0001</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Δγ+GEBS3C</td>
<td>0.0008 ± 0.0001</td>
<td>0.0008 ± 0.0001</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GγC</td>
<td>0.0065 ± 0.0001</td>
<td>0.0065 ± 0.0001</td>
<td>0.0010 ± 0.0001</td>
<td>0.0012 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GγC-Cγ(_G)</td>
<td>0.0053 ± 0.0003</td>
<td>0.0053 ± 0.0003</td>
<td>0.0024 ± 0.0003</td>
<td>0.0016 ± 0.0003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cγ(_G)</td>
<td>0.0025 ± 0.0002</td>
<td>0.0025 ± 0.0002</td>
<td>0.0020 ± 0.0005 &amp;</td>
<td>0.0010 ± 0.0002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883A</td>
<td>0.0040 ± 0.0001</td>
<td>0.0040 ± 0.0001</td>
<td>0.0028 ± 0.0002</td>
<td>0.0010 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883G</td>
<td>0.0035 ± 0.0001</td>
<td>0.0035 ± 0.0001</td>
<td>0.0009 ± 0.0001</td>
<td>0.0015 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1883U</td>
<td>0.0038 ± 0.0001</td>
<td>0.0038 ± 0.0001</td>
<td>0.0026 ± 0.0002</td>
<td>0.0008 ± 0.0001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEBS3C</td>
<td>0.0054 ± 0.0004</td>
<td>0.0054 ± 0.0004</td>
<td>0.0022 ± 0.0003</td>
<td>0.0022 ± 0.0003</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
a Rates reported here are observed rates. $k_f$ is the total rate of splicing; $k_w$ is the rate of branching pathwa; $k_2$ is the rate of conversion of lariat/3'-exon intermediate to lariat intron; $k_f$ is the rate of conversion of lariat/3'-exon to the hypothetical intron-derived product; $k_h$ is the rate of hydrolysis; $k_3$ is the rate of conversion of linear/3'-exon intermediate to linear intron and; $k_3$ is the rate of linear intron formation directly from the precursor. Rates were estimated using DynaFit (BioKin Ltd). Experimental errors represent the standard error calculated from the fits to rate equations.

b All transcripts proceeded form pLM2 and mutant-derived pLM2 plasmids digested with BamHI restriction enzyme. WT = ST2 transcript.
c The reaction buffer contained 500 mM (NH₄)₂SO₄, 100 mM MgCl₂ and 80 mM MOPS (pH 7.5).
d The reaction buffer contained 500 mM KCl, 100 mM MgCl₂ and 80 mM MOPS (pH 7.5).
e Rates followed by an asterisk presented a coefficient of variation higher than 20 %.

lariat/3'-exon intermediate or to the lariat form, although we supposed there was no lariat production (not shown).

Positions involved in the conserved $\gamma$-$\gamma'$ interaction, that is, the G452 in the linker region J2/3 and the last nucleotide of the ribozyme C1884, were subjected to several mutations: deletion of the $\gamma$ position (Δ$\gamma$ and Δ$\gamma$-GEBS3C mutants); single substitutions of the G432 by a C (GγC) or the C1884 by a G (GγG mutant), therefore disrupting the $\gamma$-$\gamma'$ interaction and; double substitutions of G432 by a C and C1884 by a G (GγC-GγG), restoring the $\gamma$-$\gamma'$ interaction. The Δ$\gamma$ and Δ$\gamma$-GEBS3C mutants inhibited the first step of splicing both in (NH₄)₂SO₄ and KCl buffers, with rate constants of about $10^{5}$ to $10^{6}$ (min⁻¹). In addition, the $k_3$ for lariat formation in the Δ$\gamma$ mutant could not be calculated either, although it seems that all the branched products corresponded to lariat/3'-exon intermediate (Figure SR4.2).

Single GγG and double GγC-GγG mutants in the RmInt1 ribozyme were about 50 % affected in the first step of the branching reaction, whereas GγC was only 20 % reduced in the first step splicing efficiency (Table R4.3). In the GγC mutant, the reduction in the branching rate ($k_{br} = 0.0049$ min⁻¹) was offset by a 40 % increase in the $k_2$ rate for the formation of intron lariat (0.015 min⁻¹). Notably, single GγG and double GγC-GγG mutants caused a 2-fold increase in the conversion of lariat/3'-exon intermediate to lariat intron ($k_2 = 0.023$ min⁻¹ and $k_2 = 0.028$ min⁻¹, respectively). However, the amplitude of lariat formation was still low due to the impairment in the first step (Figure R3.10). In (NH₄)₂SO₄ buffer, a 1.5-fold increase in the hydrolytic pathway was observed for GγG and GγC-GγG mutants ($k_{hy} = 0.0034$ min⁻¹ and $k_{hy} = 0.0035$ min⁻¹, respectively). However, in KCl buffer, GγC and GγC-GγG showed 1.5-fold increase in the hydrolytic pathway, while all three were affected in the second step splicing from slight (GγC-GγG and GγG) to moderate (GγC) (Table R4.2 and R4.3).
### TABLE R4.3. Relative self-splicing efficiencies of mutants in the ribozyme of RmIntr1 intron

<table>
<thead>
<tr>
<th>Transcript</th>
<th>Relative active fraction&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Branching pathway</th>
<th>Hydrolytic pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1&lt;sup&gt;st&lt;/sup&gt; step&lt;sup&gt;c&lt;/sup&gt; (%)</td>
<td>2&lt;sup&gt;nd&lt;/sup&gt; step&lt;sup&gt;d&lt;/sup&gt; (%)</td>
<td>1&lt;sup&gt;st&lt;/sup&gt; step&lt;sup&gt;e&lt;/sup&gt; (%)</td>
</tr>
<tr>
<td>(NH&lt;sub&gt;4&lt;/sub&gt;)&lt;sub&gt;2&lt;/sub&gt;SO&lt;sub&gt;4&lt;/sub&gt; buffer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>DV-CGA</td>
<td>2</td>
<td>0.3</td>
<td>185</td>
</tr>
<tr>
<td>J2/3-ΔGA</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Δγ</td>
<td>3</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>Δγ+GEBS3C</td>
<td>4</td>
<td>0.6</td>
<td>37</td>
</tr>
<tr>
<td>GγC</td>
<td>81</td>
<td>78</td>
<td>139</td>
</tr>
<tr>
<td>GγC-Cγ’G</td>
<td>67</td>
<td>38</td>
<td>259</td>
</tr>
<tr>
<td>Cγ’G</td>
<td>72</td>
<td>43</td>
<td>213</td>
</tr>
<tr>
<td>C1883A</td>
<td>98</td>
<td>94</td>
<td>93</td>
</tr>
<tr>
<td>C1883G</td>
<td>79</td>
<td>71</td>
<td>125</td>
</tr>
<tr>
<td>C1883U</td>
<td>114</td>
<td>110</td>
<td>833</td>
</tr>
<tr>
<td>GEBS3C</td>
<td>110</td>
<td>94</td>
<td>426</td>
</tr>
<tr>
<td>KCl buffer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WT</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>DV-CGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>J2/3-ΔGA</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Δγ</td>
<td>27</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>Δγ+GEBS3C</td>
<td>24</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>GγC</td>
<td>197</td>
<td>197</td>
<td>43</td>
</tr>
<tr>
<td>GγC-Cγ’G</td>
<td>161</td>
<td>161</td>
<td>78</td>
</tr>
<tr>
<td>Cγ’G</td>
<td>76</td>
<td>76</td>
<td>59</td>
</tr>
<tr>
<td>C1883A</td>
<td>121</td>
<td>121</td>
<td>75</td>
</tr>
<tr>
<td>C1883G</td>
<td>106</td>
<td>106</td>
<td>47</td>
</tr>
<tr>
<td>C1883U</td>
<td>115</td>
<td>115</td>
<td>67</td>
</tr>
<tr>
<td>GEBS3C</td>
<td>164</td>
<td>164</td>
<td>86</td>
</tr>
</tbody>
</table>

<sup>a</sup> All values are relative to the WT (ST2) transcript.
<sup>b</sup> The fraction of reactive precursor molecules was determined by normalizing values of <i>k<sub>T</sub></i> from mutant samples with respect to WT. It represents the percentage of molecules reacting through branching and hydrolysis pathways.
<sup>c</sup> The relative first splicing step of the branching pathway was determined by normalizing values of <i>k<sub>B1</sub></i> under (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> buffer conditions from mutant samples with respect to that of WT.
<sup>d</sup> The relative second splicing step of the branching pathway was determined by normalizing values of <i>k<sub>B2</sub></i> under (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> buffer conditions from mutant samples with respect to that of WT.
<sup>e</sup> The relative first splicing step of the hydrolytic pathway was determined by normalizing values of <i>k<sub>H1</sub></i> under (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> or KCl buffer conditions from mutant samples with respect to that of WT.
<sup>f</sup> The relative second splicing step of the hydrolytic pathway was determined by normalizing values of <i>k<sub>H2</sub></i> under KCl buffer conditions from mutant samples with respect to that of WT.

Mutations in the penultimate position (C1883A, C1883G and C1883U), in which a non Watson-Crick interaction with the G1 position has been disrupted, did not affect the first step of splicing neither by branching nor by the hydrolytic pathways (Table R4.2).
Surprisingly, we observed an 8-fold increase in the second splicing step of the branching pathway in the C1883U mutant ($k_2 = 0.09$ min$^{-1}$) (Table R4.2 and R4.3). However, a reduction of about 40-60% in the second step of splicing for the hydrolytic pathway under KCl buffer was observed for C1883U and C1883G, respectively (Table R4.3).

Finally, the mutant GEBS3C, in which the contact between the G329 (EBS3 position) and the first nucleotide of the 3’-exon (IBS3 position) was impaired (EBS3-IBS3 interaction), displayed a rate constant for branching ($0.0095$ min$^{-1}$) close to that of a WT, while the rate constant for hydrolysis was about 1.5-fold higher in ($\text{NH}_4$)$_2\text{SO}_4$ ($0.0036$ min$^{-1}$) and KCl ($0.0069$ min$^{-1}$) buffers. An increase in the rate constant for the second step of splicing was only observed for the branching pathway ($0.046$ min$^{-1}$).

**R.3.7 DISCUSSION**

Here, we have designed the RmInt1 intron construct, which can self-splice in vitro much more efficiently than the previously described variants, and without the formation of unusual side products. Through the introduction of a set of mutations in the ribozyme of RmInt1 intron we demonstrated the existence of tertiary interaction common to other group IIB introns and identified positions, which affected rate constants of individual steps of the splicing reaction.

**The self-splicing reaction of RmInt1 can be improved by varying the composition of its flanking 3’-exon sequence derived of the ISRm2011-2**

The in vitro self-splicing of RmInt1 intron, as far as it had been analyzed to date (Costa et al., 2006a; Costa et al., 2006b), showed some unconventional characteristics when compared with that of other well-studied group II introns, particularly that of yeast *S.cerevisiae* al5γ intron (Jacquier and Rosbash, 1986a; Pyle and Green, 1994; Daniels et al., 1996; Chu et al., 1998; Su et al., 2001; Zingler et al., 2010). Unlike the final products commonly described under high salt conditions (Jacquier and Rosbash, 1986b; Peebles et et al., 1986; van der Veen et al., 1986; Jarrell et al., 1988b; Daniels et al., 1996), namely lariat and linear intron, ligated and reopened exons, RmInt1 intron showed unusual truncated-3’-exon products derived of the presence of an IBS1-like sequence in the first 10 or 11 nucleotides of the 3’ exon, named IBS1*, which simulated the authentic IBS1.
We eliminated the IBS1* sequence of the 3’-exon in the pLM1 construct used in Costa’s work which, potentially, was trapping the structure in a misfolded conformation, and subsequently producing the cleavage in the 3’ exon of the transcript. We also constructed a pLM1-derivative plasmid, pLM2, in which the RmInt1 3’ exon was trimmed to its first 5 nucleotides, while the rest of the 3’ exon was replaced by several restriction sites and part of the laiz gene, which were not expected to form long range tertiary interactions with the intron nor with other regions of the exon itself (Figure R3.3). Although it had been proposed that the length of the 3’-exon did not affect the splicing kinetic behavior (Nolte et al., 1998), recent studies showed that shortening both exons resulted in a single population of fast-reacting molecules (Zingler et al., 2010). For that reason, we designed two kinds of transcripts from pLM2 plasmid with different length that shared the same 5’ exon: LT2 (long transcript) and ST2 (short transcript).

The analysis of the self-splicing products from LT1, LT2 and ST2 revealed that transcripts obtained using pLM2 as a template (i.e. LT2 and ST2) did not form side products with the truncated 3’-exon neither in (NH₄)₂SO₄ nor in KCl buffers (Figures R4.5). By contrast, in ammonium sulfate buffer ST2 transcript produced a stable covalently closed RNA form, which co-migrated with lariat intron on preparative gels. Through DNAzyme and kinetic analysis, we determined that this form, of about 10 nucleotides smaller than lariat intron, was produced from the lariat/3’-exon intermediate. We hypothesize that this intron-derived form could be a “pseudo-lariat form”, lacking the 6 nucleotides of its “tail”, or an unusual circularized fragment, as previously observed for RmInt1 intron (Costa et al., 2006a), consistent with its higher electrophoretic mobility.

The kinetic analysis indicated that reactions carried out in (NH₄)₂SO₄ buffer followed parallel and parallel-sequential kinetic models (Fersht, 1985; Marangoni, 2003). In ammonium sulfate we observed two parallel pathways for the first step of splicing, branching and hydrolysis, which is consistent with previously published works (Daniels et al., 1996; Roitzsch and Pyle, 2009; Stabell et al., 2009). In some cases we observed unusual side products, for example, species containing truncated 3’-exon formed by the LT1 construct, which is consistent with previous studies on this substrate (Costa et al., 2006a). In potassium chloride the reaction followed a sequential kinetic model in the ST2 substrate, in which a linear intron/3’-exon intermediate accumulated, as previously shown by the first group IIA intron form Podospora anserina (Schmidt et al., 1990), indicating that the second transesterification reaction of splicing was, in this case, the limiting step, contrary to that previously observed for the ai5γ intron (Chin and Pyle, 1995). By contrast, LT2 substrate in this buffer followed a simple first-order exponential kinetics, with no
intron/3’-exon intermediate accumulation detected, as reported for the aI5γ intron (Jarrell et al., 1988b; Daniels et al., 1996; Zingler et al., 2010).

We found that precursor molecules from the three transcripts were part of a homogeneous population of molecules reacting with a monophasic kinetic behavior. This fact was consistent with the existence of a short 5’-exon in pLM constructs, since it was stated that long 5’ exon sequences can induce unfavorable structures which, in turn, can introduce a folding step limiting the rate of in vitro self-splicing (Nolte et al., 1998). An example is the aI5γ intron long-exon construct used by Daniels et al (1996), in which two conformation of precursor molecules resulted in a slow and a fast population, giving a biphasic kinetic behavior.

From the analysis of reaction rates, it was shown that LT1 and LT2 reacted uniformly as a fast population, with rate constant of 0.020 and 0.0115 (min⁻¹), respectively. The reaction endpoint for the LT1 precursor was only 20%, whereas it reached more than 60% for the LT2 construct. This finding was consistent with the existence of a kinetic trap in the former (Costa et al., 2006b), which precluded a higher lariat intron population to form (Russell, 2008). ST2 reacted as a slow population (0.0083 min⁻¹), even when this transcript has been shortened in its 5’ and 3’ exon sequences. In addition, subdenaturing urea concentrations did not increase the rate constant or even diminished it, suggesting that misfolding was not the likely cause for the slow reactivity (Fedorova et al., 2007). It was proposed that the existence of a homogeneous population reacting with a lower rate when compared to a more optimal precursor transcript could be interpreted by the inhibitory structures being in fast equilibrium with the active conformation (Nolte et al., 1998). We hypothesize that any stable structure formed in the 3’-exon of ST2 or some atypical interaction between the 3’ exon and other structure of the intron or 5’ exon, not occurring in LT2, could be interfering with the splicing process, leading to a reduction in the reaction rate or even resulting in the production of the unconventional intron-derived form described above. In agreement with previous observations (Nolte et al., 1998), rates for hydrolysis were about 2-fold lower than those for branching pathway.

However, the most striking characteristic of RmInt1 intron is the persistence of a lag phase in all transcripts due to the formation of a transient intermediate prior to the excision of the intron form, both in the branching and hydrolytic pathways (Figure R3.5 and R3.8). Unlike aI5 γ intron, RmInt1 forms RNP particles in vivo, which could explain the need of protein assistance for a correct splicing of RmInt1. Nevertheless, the
inefficiency of RmInt1 intron \textit{in vitro}, under high-salt conditions, show how the constrains imposed by the RNA structure preclude a productive splicing reaction.

Considered together, our results show that the LT2 construct of RmInt1 intron is better behaved \textit{in vitro} than LT1 RNA, because it does not form unconventional side products and it produces lariat intron with at much higher amplitude. However, RmInt1 is still less efficient that other group II introns studied \textit{in vitro}. For that reason, a further analysis of the secondary and tertiary structures of the RmInt1 intron is necessary in order to elucidate the elements responsible for the defects in the splicing reaction.

**Point mutations in the RmInt1 intron affect branching and hydrolytic pathways**

Site-directed mutagenesis is a useful and extended tool in the study of self-splicing activity of group II introns. Through different approaches, it has been able to associate specific roles to conserved positions in the ribozyme affecting several aspects of catalysis (Jacquier and Michel, 1990; Jacquier and Jacquesson-Breuleux, 1991; Peebles \textit{et al}., 1995; Bar-Shalom and Moore, 2000; Costa \textit{et al}., 2000; Mikheeva \textit{et al}., 2000; Molina Sanchez \textit{et al}., 2011). Moreover, recent advances in NMR (Sigel \textit{et al}., 2000; 2004; Erat \textit{et al}., 2007) and crystal structure determinations (Zhang and Doudna, 2002; Toor \textit{et al}., 2008a), have allow the association of known functionalities of conserved positions to defined structural motifs.

Aimed to the determination of critical positions of the RmInt1 ribozyme affecting the first, second or both splicing steps \textit{in vitro}, here we introduced mutations at nucleotides known to participate in tertiary interactions in other group II introns, as previously carried out for the \textit{in vivo} splicing of RmInt1 intron (Figure R3.1, see Chapter 3). We assayed their ability to self-splice \textit{in vitro} and determined their relative efficiencies of first and second steps of the splicing reaction. The effects on the self-splicing reaction could be then explained from the perspective of structure.

Mutations in the conserved catalytic triad 5-AGC-3’ at the base of domain V that abolish splicing have been reported repeatedly for the \textit{in vitro} self-splicing of a15γ intron (Jarrell \textit{et al}., 1988a; Koch \textit{et al}., 1992; Dib-Hajj \textit{et al}., 1993; Peebles \textit{et al}., 1995). In RmInt1, mutations that disrupted base-pairing with the catalytic triad at positions A1811, G1812 and C1843 (G1840 U1841 U1842 to C1840 G1841 A1842) inhibited the splicing reaction. It has been reported that mutations at positions base pairing with those of the catalytic triad did not lead to severe self-splicing defects \textit{in vitro} (Boulanger \textit{et al}., 1995; Peebles \textit{et al}.,
1995). However, the same authors observed that base pairings of the A2 and C4 (referred to DV positions) were more important than the identities of the nucleotide bases at those positions and the maintenance of the wobble pair at position G3 was also the responsible of a local distortion of one helix functionally important, which could influence the loss of activity of the triple mutant (Boulanger et al., 1995; Peebles et al., 1995). Noticeable, according to Boulanger et al (1995), the triple mutation at G1840 U1841 U1842 did not lead to a detectable defect in the second step of splicing. As mentioned in Chapter 3, the crystal structure of the O.i. group II intron has revealed that only the conserved 5'-AGC-3' nucleotides (and not their pairing partners) form hydrogen bonds with nucleotides of the J2/3 region and the AC bulge in DV, constituting a triple helix (triplex) which is essential for splicing catalysis (Toor et al., 2008a; Toor et al., 2009; Keating et al., 2010).

Mikheeva et al (2000) previously reported that the deletion of the dinucleotide G588 A589 (G453 A454 in RmInt1 intron) affected the second splicing step in the yeast a15γ intron. However, in RmInt1, both the first and the second splicing steps were significantly impaired. This result better matched that of Chanfreau and Jacquier (1996), in which the mutation of the A589 of a15γ intron affected both steps of the splicing. Photocrosslinking studies initially confirmed these studies, since it was shown that the catalytic domain V interacted with the dinucleotide G588 A589 of a15γ intron (Podar et al., 1998; de Lencastre and Pyle, 2008), confirming the occurrence of a single active site catalyzing both transesterification reactions (Chanfreau and Jacquier, 1994; de Lencastre et al., 2005). As explained in Chapter 3, recent crystallographic studies have also shown that position G288 and C289 of the O.i. group II intron are part of a triple interaction with the G359 U384 and C358 G385 base pairs of DV, respectively, establishing a triple helix (triplex) essential for catalysis (Keating et al., 2010). Moreover, the high correlation observed for self-splicing in vitro and splicing in vivo of DV and J2/3 region mutant constructs reflects that splicing depends very much on secondary and tertiary structures rather than on an effect of protein co-factors.

The deletion of the G452 (γ position), which also inhibited splicing, is in agreement with the observation of a close proximity of this position with the bulge of the catalytic core through the interaction with the fourth (or fifth in O.i. intron) nucleotide of the intron, positioning the 3’ splice site for catalysis of the exon junction (de Lencastre and Pyle, 2008; Toor et al., 2008b). Although G452 is also involved in the conserved interaction γ-γ’ with the last nucleotide of the intron, it seems that its interactions with the catalytic core could be determinant for the appearance of this phenotype. However, the substitution of the G452 to a C, which is not represented on the nature (Jacquier and
Michel, 1990) did not altered significantly the splicing efficiencies shown for the wild-type strain. It is possible that *in vitro*, the presence of a C nucleotide in the γ position partially substitutes the role of the G in the network of tertiary interactions with the 5′ splice site and the active site. This result agrees with the previous observation that substitutions in the γ position did not grossly affect the first step of the splicing reaction (Jacquier and Michel, 1990). By contrast, the same author noted that mismatched combinations accumulated significantly more lariat/3′-exon intermediate (Jacquier and Michel, 1990) while, in RmInt1 intron, a slight increase in the second splicing step was observed. Another possibility could be that a cryptic interaction could occur between the γ′ position (C1884) and the G453, located next to the γ position in the wild-type intron, which would restore splicing. *In vivo*, the IEP protein could help to correctly position γ and γ′ nucleotides, preventing cryptic interactions to form and so explaining the complete block of splicing observed. On the other hand, simple and double mutations of the γ′ position, disrupting or restoring the Watson-Crick interaction, respectively, resulted in a moderate decrease in the first step of splicing and an increase in the second step. Although we did not find any good reason for these observations, one possibility is that the substitution of the C1884 to a G could lead to a cryptic interaction with the C451 in the J2/3 region, which distorted the complex interactions of that region with elements of the DV triplex (Keating et al., 2010) and the 5′-splice site (de Lencastre and Pyle, 2008), thus enabling the γ-γ′ to occur while partially impairing the first step of splicing. However, another possibility could be the formation of a cryptic splice site downstream in the 3′ exon, as previously suggested (Schmidt et al., 1996; Molina Sanchez et al., 2011), whereas a novel tertiary interaction in the ribozyme of RmInt1, involving the γ′ position, could be affecting the 5′-cleavage efficiency *in vitro*.

It has been previously reported that the penultimate position of group II introns establishes contacts with the first nucleotide of the intron by means of a non Watson-Crick interaction and that substitutions at the penultimate nucleotide of the ribozyme affected the efficiency of the second splicing step *in vitro* (Chanfreau and Jacquier, 1993). We assayed substitutions for the C1883 position to A, C and U. The mutation to adenosine, and in a less extent to guanidine, led to a wild-type phenotype, according with *in vivo* results and in agreement with the presence in the nature of the combination C(first nucleotide)-A(last nucleotide), (Michel et al., 1989; Chanfreau and Jacquier, 1993). The mutation of C to U also showed a wild-type first splicing step, through branching and hydrolysis pathways. However, in (NH₄)₂SO₄ buffer, this mutation produced a 9-fold increase of the second step of the splicing reaction. Since there is no available data of intronic 5′ and 3′ ends in the crystal structure, probably because these regions are
disordered or have been cleaved through secondary reactions (Toor et al., 2010), it is
difficult to figure out a structural interpretation for this finding. However, one explanation
might be that, in vitro, a novel interaction involving the penultimate nucleotide could
promote or stabilize the approximation of 5’ and 3’ exons, leading to a productive second
transesterification reaction.

Finally, the disruption of the EBS3-IBS3 interaction through the mutation of the
G329 to C caused no effect on the first splicing step of the branching pathway and a
moderate increase in that of the hydrolytic pathway. By contrast, a 4-fold increase in the
second step of the splicing reaction was observed, as occurred for the mutation of the
C1883 to U. The role of the EBS3-IBS3 interaction has been previously shown to be
essential for efficient exon ligation in vitro, by positioning the 3’ splice site in the active site
(Costa et al., 2000). In that work, the combination of C(EBS3)—C(+1) assayed here, led to
a low second splicing step efficiency, accordingly with the fact that this combination is
absence in the nature (Costa et al., 2000). Observations of the crystal structure also reveal
that EBS1 and EBS3 form a continuous binding interface which simultaneously recognize
both 5’ and 3’ exon sequences (Toor et al., 2008b). On the basis of structure and previous
observations we could not find an explanation which accommodates these findings.
However, for an unknown mechanism, this mutation could be stabilizing the interaction
between 5’ and 3’ exons.

Taking all together, we observe an agreement between in vivo and in vitro results
from mutations involved in tertiary interactions with the catalytic core, indicating that
splicing rely essentially on the secondary and tertiary intron structures, such as the triple
helix between DV and nucleotides within the J2/3 linker. However, some significant
differences are observed between in vitro and in vivo data involving tertiary interactions
aimed to approximate 5’ and 3’ exons (γ-γ’ and EBS3-IBS3 interactions) and intron
boundaries (G+1 – penultimate nucleotide of the intron), which suggest that trans-acting
factor, such as the IEP protein, play an important role in these processes.

This knowledge opens the possibility to go deep into the mechanisms involved in
the low splicing efficiency actually observed for RmInt1 intron, which, in turn, could be
modulated and used as a biotechnological tool.
CONCLUDING REMARKS
1. The reported system based on the lacZ gene, containing the minimal target site for RmInt1 intron insertion, links the splicing of the intron to the strong and constitutive expression of the β-galactosidase protein. The level of expression of β-galactosidase protein can be accurately measured by a Miller assay, which shows a range of detection varying from 60-14000 Miller Units.

2. The wild-type RmInt1 intron excises efficiently from the RNA precursor. The construction in which the intron-encoded protein (IEP) is present in trans in a separate plasmid excises at a level similar as that of the full-length intron, whereas the construction with the IEP expressed in cis presents a 70 % excision efficiency reduction. RmInt1 excision in E. coli was 95% less efficient than excision of the intron in S. meliloti, suggesting that host-specific factors may play an important role in the splicing reaction.

3. The lariat form accumulates in the in vivo splicing process and this product is free of lariat/3’-exon reaction intermediates. The level of exon ligation observed for RmInt1 intron, in Miller units, is below background levels. By quantitative real-time qRT-PCR the splicing efficiency of RmInt1 intron has been determined in vivo in S. meliloti (0.07 ± 0.02) % and in E. coli (0.016 ± 0.006) %. These findings provide evidence that RmInt1 intron functions more like a retroelement than a spliceosomal intron.

4. The self-splicing reaction efficiency in vitro of the RmInt1 intron has been improved by eliminating an IBS1-like sequence present in the first 10 or 11 nucleotides of the 3’ exon from the natural target IS2011-2 insertion sequence. The new construct prevents the formation of unusual truncated side products previously observed, producing an increment of the amplitude of the reaction.
5. The in vitro self-splicing reaction of the RmInt1 intron carried out in (NH₄)₂SO₄ buffer follows a parallel-sequential kinetic model. In this buffer, two parallel pathways for the first step of splicing, branching and hydrolysis, are observed. The final lariat and linear introns are the result of a sequential reaction in which a transient intron intermediate in formed. In KCl buffer, the RmInt1 intron follows a sequential kinetic model, in which the linear intron is produced from a transient linear-intron intermediate. The transcripts from the RmInt1 intron improved constructs react as a fast and homogeneous population of molecules, showing a monophasic kinetic behavior.

6. The persistence of a lag phase in all transcripts due to the formation of a transient intermediate prior to the excision of the intron form indicates an intrinsic low efficiency of the first-step of splicing of RmInt1 intron in vitro.

7. The catalytic core in RmInt1 intron is composed of the triple helix (triplex) between DV and nucleotides within the J2/3 linker common to other group II introns. The disruption of interactions involving components of the catalytic core results in the affectation of both steps of splicing in vivo and in vitro, indicating that splicing rely essentially on the secondary and tertiary intron structures. This observation also supports the occurrence of an unique active site catalyzing both steps of the splicing reaction.

8. The tertiary interactions aimed to approximate 5’ and 3’ exons (γ-γ’ and EBS3-IBS3 interactions) and intron boundaries (G+1 – penultimate nucleotide of the intron) have been shown to occur for the RmInt1 intron in vivo. Significant differences observed in vitro indicate that the IEP protein plays an important role helping to correctly position these elements in the RmInt1 ribozyme in vivo.
CONCLUSIONES
1. El sistema informador basado en el gen lacZ, que contiene la diana mínima de reconocimiento para la inserción del intrón RmInt1, une el proceso de splicing del intrón a una expresión fuerte y constitutiva de la proteína β-galactosidasa. El nivel de expresión de la proteína β-galactosidasa puede ser medido con precisión mediante un ensayo Miller, que posee un margen de detección que varía entre 60-14000 unidades Miller.

2. El intrón RmInt1 silvestre se escinde eficientemente de su precursor de ARN. La construcción en la cual la proteína codificada por el intrón (del inglés Intron-Encoded Protein, IEP) está expresada en trans desde un plásmido separado se escinde a un nivel similar al del intrón de longitud completa. La construcción con la IEP expresada en cis presenta un 70 % de reducción de la escisión. La escisión de RmInt1 en E.coli es un 95 % menos eficiente que la misma en S.meliloti, lo que sugiere que determinados factores específicos del huésped podrían jugar un papel importante en la reacción de splicing.

3. La forma en lazo o lariat se acumula durante el proceso de splicing in vivo libre de intermediarios de la reacción del tipo intrón en lazo/exón 3’. El nivel de unión de exones observado en el intrón RmInt1, en unidades Miller, se encuentra por debajo del nivel de detección. Mediante PCR cuantitativa (qRT-PCR) se ha determinado que la eficiencia de splicing in vivo del intrón RmInt1 es \(0.07 \pm 0.02\) % en S.meliloti y \(0.016 \pm 0.006\) % en E.coli. Estos hallazgos muestran la evidencia de que el intrón RmInt1 funciona de un modo más parecido a los retroelementos que a los intrones espliceosómicos.

4. La eficiencia de la reacción de self-splicing in vitro se ha mejorado mediante la eliminación de una secuencia parecida a IBS1, presente en los primeros 10 u 11 nucleótidos del exón 3’ de la diana de inserción natural procedente de la secuencia de inserción IS2011-2. La nueva construcción evita la formación de co-productos truncados bastante inusuales, que habían sido observados previamente, produciendo, de este modo, un aumento en la amplitud de la reacción.
5. La reacción de *self-splicing in vitro* del intrón RmInt1 llevada a cabo en tampón (NH₄)₂SO₄ sigue un modelo cinético del tipo paralelo-secuencial. En este tampón, se observan dos rutas paralelas durante el primer paso del *splicing* conocidas como *branching* e hidrolisis. Los intrones en lazo y lineal finales son consecuencia de una reacción secuencial a partir de un intermediario transitorio del intrón. En tampón KCl, el intrón RmInt1 sigue un modelo cinético secuencial, en el cual el intrón lineal se produce a partir de un intermediario lineal transitorio. Los transcriptos de las construcciones mejoradas del intrón RmInt1 reaccionan rápida y homogéneamente, y muestran un comportamiento cinético monofásico.

6. La persistencia de un retraso en la escisión del intrón, debida a la formación de un intermediario transitorio, indica una baja eficiencia en el primer paso de la reacción de *splicing in vivo* intrínseca del intrón RmInt1.

7. El centro catalítico del intrón RmInt1 se compone de una triple hélice (*tripléx*) que se establece entre el DV y los nuclétoidos de la región intermedia J2/3, común a otros intrones del grupo II. La interrupción de interacciones terciarias en las que participan componentes del centro catalítico afectan tanto al primer como al segundo paso de la reacción de *splicing*, tanto *in vivo* como *in vitro*, lo que indica que el *splicing* depende necesariamente de las estructuras secundaria y terciaria del intrón. Esta observación también apoya la existencia de un único sitio activo que catalice ambos pasos de la reacción de *splicing*.

8. Se ha observado que las interacciones terciarias que producen el acercamientos de los exones 5’ y 3’ (interacciones y-γ’ y EBS3-IBS3) y los extremos del intrón (G+1 - penúltimo nucléótido del intrón) tienen lugar *in vivo* para el intrón RmInt1. También se han encontrado diferencias significativas con los resultados *in vitro*, lo que indica que la proteína IEP juega un papel importante ayudando a disponer correctamente estos elementos en la ribozima de RmInt1 *in vivo*.
BIBLIOGRAPHY


147. Ostertag O, Cooke AM, Watkins KP and Barkan A (2005) CRS1, a chloroplast
group II intron splicing factor, promotes intron folding through specific interactions

148. Peebles CL, Perlman PS, Mecklenburg KL, Petrillo ML, Tabor JH, Jarrell KA

Mutation of the conserved first nucleotide of a group II intron from yeast mitochondrial DNA

nucleotide in domain 5 of a group II intron. Proc Natl Acad Sci USA 92(10):4422-4426.

151. Podar M, Chu VT, Pyle AM and Perlman PS (1998a) Group II intron splicing in

Domain 5 binds near a highly conserved dinucleotide in the joiner linking domains 2 and 3
of a group II intron. RNA 4(2):151-166.

ribozyme activity: quantitation of interdomain binding and reaction rate. Biochemistry
33(9):2716-2725.

cut, process, and map RNA molecules for structural studies or modification. Methods
Enzymol 317:140-146.

7(7-8):679-690.

and invade DNA. The RNA world: the nature of modern RNA suggests a prebiotic RNA world,
eds Gesteland RF, Cech TR, & Atkins JF (Cold Spring Harbor Laboratory Press, Cold


