



UNIVERSIDAD DE GRANADA  
FACULTAD DE CIENCIAS

**Intrones del grupo II en *Sinorhizobium meliloti*:  
Contribución de la Proteína y la Ribozima  
codificadas por RmInt1 a su escisión y movilidad**

*M<sup>a</sup> Dolores Molina Sánchez*

Tesis Doctoral  
2008

Editor: Editorial de la Universidad de Granada  
Autor: María Dolores Molina Sánchez  
D.L.: Gr. 962- 2008  
ISBN: 978-84-338-4968-7



**Intrones del grupo II en *Sinorhizobium meliloti*:  
Contribución de la Proteína y la Ribozima codificadas  
por RmInt1 a su escisión y movilidad**

Memoria que presenta la licenciada en Bioquímica  
Dña. M<sup>a</sup> Dolores Molina Sánchez  
como aspirante al grado de Doctor

Fdo: ***M<sup>a</sup> Dolores Molina Sánchez***

V<sup>o</sup> B<sup>o</sup>  
El Director

Fdo: Nicolás Toro García  
Doctor en Ciencias Biológicas  
Profesor de Investigación del C.S.I.C.



Esta Tesis Doctoral ha sido realizada en el Grupo de Ecología Genética de la Rizosfera, del Departamento del Microbiología del Suelo y Sistemas Simbióticos de la Estación Experimental del Zaidín. (C.S.I.C.)



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

**Toro, N., Molina-Sánchez, M.D. y Fernández-López, M.** (2002) Identification and characterization of bacterial class E group II introns. *Gene*, **299**; 245-250.

**Molina-Sánchez, M.D., Martínez-Abarca, F. y Toro, N.** (2006) Excision of *Sinorhizobium meliloti* group II intron as circles *in vivo*. *Journal of Biological Chemistry*, **281**; 28737-28744.

**Molina-Sánchez, M.D., Martínez-Abarca, F. y Toro, N.** (2008) Characterization of the C-terminal maturase region and C-tail extension of the *Sinorhizobium meliloti* RmInt1 group II intron-encoded protein. En preparación.

**Molina-Sánchez, M.D. y Toro, N.** (2008) Role of the Branch-site and the Coordination Loop in the *in vivo* excision of the bacterial group II intron, RmInt1. En preparación.





# *INDICE*



<b>INTRODUCCIÓN</b>	3
<b>1. MECANISMOS DE MADURACIÓN DE INTRONES DEL GRUPO II</b>	6
1.1. ESTRUCTURA DE LA RIBOZIMA	7
<b>1.1.1. Características estructurales de cada dominio</b>	9
1.1.1.1. Dominio I	9
1.1.1.2. Dominio II	11
1.1.1.3. Dominio III	12
1.1.1.4. Dominio IV	13
1.1.1.5. Dominio V	13
1.1.1.6. Dominio VI	15
1.1.1.7. Dominio VII	15
1.1.1.8. Regiones interdominio	16
<b>1.1.2. Plegamiento general del intrón: conformación del centro catalítico</b>	17
<b>1.1.3. Clasificación estructural de intrones del grupo II</b>	20
1.2. MECANISMOS DE ESCISIÓN EN INTRONES DEL GRUPO II	22
1.2.1. Mecanismo de branching o de formación de intrón <i>lariat</i>	24
1.2.2. Mecanismo de hidrólisis o de formación de intrón lineal	26
1.2.3. Mecanismo de escisión de intrón circular	27
<b>1.2.4. Otras reacciones asociadas a la maduración de intrones del grupo II</b>	28
1.3. ESCISIÓN <i>IN VIVO</i> DE INTRONES DEL GRUPO II ASISTIDA POR PROTEÍNAS	29
1.3.1. Proteínas codificadas por el intrón, IEPs	29
1.3.2. Otros factores de <i>splicing</i> codificados en el genoma hospedador	31
<b>2. MECANISMOS DE MOVILIDAD EN INTRONES DEL GRUPO II</b>	33
2.1. PROTEÍNAS CODIFICADAS POR INTRONES DEL GRUPO II	33
2.1.1. Dominio reverso transcriptasa (RT)	35
2.1.2. Dominios de unión a ADN (D) y ADN endonucleasa (En)	36
2.1.3. Modelización tridimensional de una IEP	38
2.1.4. Clasificación de los intrones en función de su IEP	40
2.2. MECANISMOS DE <i>RETROHOMING</i>	41
2.2.1. Reconocimiento del ADN diana	42
2.2.2. Homing en intrones mitocondriales de levadura	44
2.2.3. Retrohoming en intrones bacterianos	46

2.2.3.1. <i>Retroboming</i> mediado por TPRT	48
2.2.3.2. Movilidad independiente de actividad ADN endonucleasa	49
2.3. MECANISMO DE RETROTRANSPOSICIÓN	51
<b>3. RmInt1, EL INTRÓN DEL GRUPO II DE <i>Sinorhizobium meliloti</i></b>	52
<b>4. ELPASADO Y EL FUTURO DE INTRONES DEL GRUPO II</b>	54
4.1. EVOLUCIÓN DE ESTOS RETROELEMENTOS	55
4.2. APLICACIONES EN BIOTECNOLOGÍA	57
<b>OBJETIVOS</b>	65
<b>MATERIAL Y MÉTODOS</b>	66
<b>1. ESPECIES Y CEPAS BACTERIANAS</b>	68
<b>2. PLÁSMIDOS Y VECTORES DE CLONAJE</b>	68
<b>3. OLIGONUCLEÓTIDOS</b>	72
<b>4. CULTIVOS CELULARES</b>	72
4.1. MEDIOS DE CULTIVO Y ANTIBIÓTICOS	72
4.2. CONDICIONES DE CULTIVO	74
4.3. CONSERVACIÓN DE CULTIVOS BACTERIANOS	74
<b>5. AISLAMIENTO DE ÁCIDOS NUCLEICOS</b>	74
5.1. EXTRACCIÓN DE ADN total	74
5.1.1. AISLAMIENTO DE ADN total USANDO PRONASA	74
5.1.2. EXTRACCIÓN DE ADN total MEDIANTE KIT COMERCIAL	75
5.1.3. LISADOS DE COLONIA	76
5.2. EXTRACCIÓN ADN PLASMÍDICO	76
5.2.1. AISLAMIENTO DE PLÁSMIDOS POR LISIS ALCALINA	76
5.2.2. EXTRACCIÓN DE ADN PLASMÍDICO POR PRECIPITACIÓN CON SALES DE MAGNESIO	77
5.2.3. AISLAMIENTO DE ADN PLASMÍDICO MEDIANTE KIT COMERCIAL	78
5.2.4. EXTRACCIÓN DE PLÁSMIDOS A PARTIR DE GRANDES VOLÚMENES DE CULTIVO (MIDIPREPARACIONES)	78
5.3. EXTRACCIÓN DE ARN total	78
5.3.1. AISLAMIENTO DE ARN total DE <i>S. meliloti</i>	79
5.3.1.1. Modificación por la adición de proteinasa K	79
5.3.1.2. Modificación por el uso de tiocianato de guanidina	79
5.3.2. EXTRACCIÓN DE ARN total MEDIANTE KIT COMERCIAL	80

5.4. EXTRACCIÓN DE PARTÍCULAS RIBONUCLEOPROTEICAS (RNPs)	80
5.4.1. PRECIPITACIÓN DEL ARN EN LAS DISTINTAS SUBFRACCIONES DE LAS RNPs	81
5.5. CUANTIFICACIÓN DE LA CONCENTRACIÓN DE ARN Y ADN	82
<b>6. MANIPULACIÓN Y AMPLIFICACIÓN DE ADN</b>	82
6.1. DIGESTIÓN CON ENZIMAS DE RESTRICCIÓN	82
6.2. REACCIÓN DE DEFOSFORILACIÓN	82
6.3. PCR	83
6.3.1. PCR DE COLONIA	83
<b>7. ELECTROFORESIS</b>	84
7.1. ELECTROFORESIS EN GELES DE AGAROSA NO DESNATURALIZANTES	84
7.2. ELECTROFORESIS EN GELES DE AGAROSA DESNATURALIZANTES	84
7.3. ELECTROFORESIS EN GELES DE POLIACRILAMIDA NO DESNATURALIZANTES	85
7.4. ELECTROFORESIS EN GELES DE POLIACRILAMIDA DESNATURALIZANTES	85
7.5. MARCADORES DE PESO MOLECULAR	86
<b>8. CLONAJE DE FRAGMENTOS</b>	86
8.1. AISLAMIENTO DE FRAGMENTOS DE ADN DE GEL DE AGAROSA	86
8.2. LIGACIÓN DE FRAGMENTOS DE RESTRICCIÓN	87
8.2.1. ADENILACIÓN DE AMPLIFICADOS DE PCR	87
8.3. PREPARACIÓN DE CÉLULAS COMPETENTES	87
8.3.1. CÉLULAS COMPETENTES POR MÉTODOS QUÍMICOS: RbCl	87
8.3.2. CÉLULAS ELECTROCOMPETENTES	88
8.4. TRANSFORMACIÓN DE CÉLULAS COMPETENTES	88
8.4.1. TRANSFORMACIÓN POR CHOQUE TÉRMICO	88
8.4.2. ELECTROTRANSFORMACIÓN	89
<b>9. SECUENCIACIÓN Y ANÁLISIS DE SECUENCIAS</b>	89
<b>10. TÉCNICAS DE HIBRIDACIÓN DE ÁCIDOS NUCLEICOS</b>	90
10.1. HIBRIDACIÓN ADN-ADN (SOUTHERN BLOT)	90

10.1.1. TRANSFERENCIA ALCALINA POR VACÍO	90
10.1.2. MARCAJE DE LA SONDA	90
10.1.3. HIBRIDACIÓN, LAVADOS Y REVELADO	91
10.2. HIBRIDACIÓN ARN-ADN ( <i>NORTHERN BLOT</i> )	92
10.2.1. TRANSFERENCIA DEL ARN	92
10.2.1.1. Transferencia por capilaridad	92
10.2.1.2. Transferencia semi-seca o electro-transferencia	92
10.2.2. MARCAJE DE LA SONDA	93
10.2.3. HIBRIDACIÓN, LAVADOS Y REVELADO	93
10.3. HIBRIDACIÓN EN COLONIA	93
<b>11. ENSAYOS DE MOVILIDAD</b>	94
11.1. CONJUGACIONES EN <i>S. meliloti</i>	94
11.2. DETECCIÓN POR HIBRIDACIÓN	95
11.3. DETECCIÓN POR SELECCIÓN DE COLONIAS TcR	96
<b>12. TÉCNICAS CON ARN</b>	96
12. 1. SÍNTESIS DE ARNs TRANSCRITOS IN VITRO	96
12.2. ESTUDIOS DE SPLICING IN VITRO	97
12.3. DIGESTIÓN DEL ARN	97
12.3.1. DIGESTIÓN CON UNA DNazyme	97
12.3.2. DIGESTIÓN CON ARNasa H	98
12.4. EXTENSIÓN DE CEBADOR	98
12.4.1. EXTENSIÓN DE CEBADOR ENVENENADA CON ddNTPs (PPE)	99
12.5. RT-PCR	100
12.5.1. cRT-PCR	101
12.6. ENSAYOS DE PROTECCIÓN DEL ARN FRENTE A DIGESTIÓN CON ARNasas (RPA)	102
12.7. ENSAYO DE LIGACIÓN DE OLIGONUCLEÓTIDOS (OLA)	103
12.8. AMPLIFICACIÓN DE LOS EXTREMOS DE UN ARN (RACE)	104
12.8.1. RACE 5'	104
12.8.2. RACE 3'	104
12.8.3. MODIFICACIONES EN RACE	105
<b>13. ACTIVIDADES BIOQUÍMICAS DE LAS RNPs</b>	106
13.1. ACTIVIDAD RT EXÓGENA	106
13.2. CORTE DE SUSTRATOS ADN	107

13.2.1. MARCAJE DE SUSTRATOS	107
13.2.1.1. Marcaje de sustratos de ADN de cadena simple	107
13.2.1.1.1. Marcaje en 5' con polinucleótido kinasa y [ $\gamma$ -32P]ATP	107
13.2.1.1.2. Marcaje en 3' con terminal transferasa y [ $\alpha$ -32P]ddATP	108
13.2.1.2. Marcaje de sustratos de ADN de cadena doble	108
13.2.1.2.1. Marcaje en 5' con polinucleótido kinasa y [ $\gamma$ -32P]ATP	108
13.2.1.2.2. Marcaje interno mediante PCR y [ $\alpha$ -32P]dCTP	108
13.2.2. REACCIÓN DE CORTE/SPLICING REVERSO	109
13.2.3. REACCIÓN DE UNIÓN DE LAS RNPs AL SUSTRATO ADN	109
13.2.4. ESCALERAS DE SECUENCIACIÓN: MÉTODO DE SANGER	110
<b>14. CONSTRUCCIÓN DE MUTANTES</b>	<b>110</b>
14.1. MUTAGÉNESIS MEDIANTE KIT COMERCIAL	111
14.2. INTRODUCCIÓN DE MUTACIONES POR PCR SOLAPANTE	111
<b>RESULTADOS Y DISCUSIÓN</b>	<b>123</b>
<b>I. NUEVOS RETROELEMENTOS CARACTERIZADOS EN <i>S. meliloti</i> GR4:</b>	
<b>INTRONES DE LA CLASE E</b>	<b>125</b>
I.1. BÚSQUEDA DE NUEVOS RETROELEMENTOS EN EL PLÁSMIDO CRÍPTICO pRmeGR4b	129
I.1.1. ANÁLISIS DE SECUENCIAS	132
<b>I.1.1.1. La secuencia completa del intrón</b>	<b>134</b>
<b>I.1.1.2. La secuencia del retrón</b>	<b>135</b>
I.2. IDENTIFICACIÓN DE UNA NUEVA CLASE DE INTRONES DEL GRUPO II	137
I.2.1. AGRUPAMIENTO FILOGENÉTICO DE ALGUNAS IEPs BACTERIANAS	139
I.2.2. ESTRUCTURA CONSENSO PARA LA RIBOZIMA DEL NUEVO GRUPO DE INTRONES	141
I.3. DISCUSIÓN	143
<b>ANEXO CAPÍTULO I</b>	<b>147</b>
<b>II. FORMAS DE ESCISIÓN GENERADAS EN LA MADURACIÓN DE RmInt1</b>	<b>151</b>
II.1. RmInt1 FORMA CÍRCULOS IN VIVO	152
II.1.1. DETECCIÓN DE MOLÉCULAS CIRCULARES	153
<b>II.1.1.1. Escisión in vivo del intrón RmInt1 determinada por extensión a partir de cebador</b>	<b>153</b>



II.1.1.2. Detección de dos productos de escisión diferentes por RT-PCR	154
II.1.1.3. RmInt1 se escinde in vivo dando lugar a intrón <i>lariat</i> y circular	156
II.1.1.4. ¿Contradicciones en los resultados?	157
II.1.1.5. Identificación de moléculas de ARN generadas por cRT-PCR	159
II.1.1.6. Detección directa de ARN circular	160
II.1.1.6.1. Ensayo de ligación de oligonucleótidos (OLA)	161
II.1.1.6.2. Ensayos de protección frente a digestión con ARNasas (RPA)	164
II.1.1.6.3. Extensión de cebador modificada por la adición de un ddNTP	166
II.1.1.6.4. Digestión mediante el uso de ARNasa H o de <i>DNAzymes</i> seguida de extensión de cebador	169
II.1.1.6.5. <i>Northern blot</i>	171
II.1.1.6.5.1. ¿Son realmente moléculas lineales?	173
II.1.2. LA ACTIVIDAD MADURASA DE LA IEP DE RmInt1 ES NECESARIA EN LA ESCISIÓN TANTO LARIAT COMO CIRCULAR.	175
II.1.3. ESTRUCTURAS DE LA RIBOZIMA QUE INTERVIENEN EN LA ESCISIÓN DE RmInt1 IN VIVO	178
II.1.3.1. Determinación de la eficiencia de escisión in vivo de los mutantes de RmInt1 mediante extensión de cebador	181
II.1.3.2. Naturaleza de los productos generados en la escisión in vivo de los mutantes de RmInt1	182
II.1.3.3. Movilidad de los mutantes en la ribozima	184
II.1.3.4. Estudio estructural de los dominios I y VI de la ribozima de los intrones del grupo IIB	186
II.2. LA LIGACIÓN DE EXONES ES UN PROCESO INEFICIENTE EN RmInt1	188
II.2.1. DETECCIÓN DE EXONES LIGADOS POR PPE	188
II.2.2. DETECCIÓN DE EXONES LIGADOS POR RT-PCR	190
II.3. DISCUSIÓN	193
II.3.1. LAS FORMAS GENERADAS POR RmInt1 EN SU ESCISIÓN	193

<b>II.3.1.1. El mecanismo de escisión de moléculas circulares en RmInt1</b>	195
<b>II.3.2. IMPLICACIÓN DE LA IEP EN LA ESCISIÓN DE RmInt1</b>	195
<b>II.3.3. EL PAPEL DE LA A DESAPAREADA DEL dVI Y LA REGIÓN DE COORDINACIÓN EN LA ESCISIÓN IN VIVO</b>	196
<b>II.3.3.1. Papel de la A en la escisión de RmInt1</b>	196
<b>II.3.3.2. Relevancia funcional de la región de coordinación en la conformación del sitio catalítico de la ribozima de RmInt1</b>	198
<b>II.3.3.3. Conservación filogenética de estas estructuras en intrones</b>	
<b>IIB</b>	200
<b>ANEXO CAPÍTULO II</b>	203
<b>III. CARACTERIZACIÓN DE LA REGIÓN C-terminal DE LA PROTEÍNA CODIFICADA POR RmInt1</b>	209
III. 1. ESTUDIO FILOGENÉTICO DE LA REGIÓN C-terminal DE LA IEP EN INTRONES IIB3	211
III. 2. ANÁLISIS DE LAS MUTACIONES EN EL C-terminal DE LA IEP DE RmInt1	214
III. 2.1. EFECTO DE LAS DIFERENTES DELECCIONES	215
III. 2.2. ANÁLISIS DE MUTACIONES EN RESIDUOS CONSERVADOS DE LA HÉLICE $\alpha_4$	218
III. 2.3. RESULTADO DE LA MUTACIÓN DE RESIDUOS CONSERVADOS EN LA CADENA $\beta_2$	223
III. 3. DISCUSIÓN	224
<b>ANEXO CAPÍTULO III</b>	229
<b>CONCLUSIONES</b>	241
<b>ABREVIATURAS Y ANGLICISMOS</b>	245
<b>AGRADECIMIENTOS</b>	



# *INTRODUCCIÓN*



## INTRODUCCIÓN

Desde la definición esencial dada por Mendel en 1865<sup>1</sup>, el concepto molecular de gen ha sufrido multitud de matizaciones. El trabajo inicial realizado en procariotas, principalmente en la bacteria *Escherichia coli*, llevó a los investigadores hace más de medio siglo a promulgar lo que se ha dado a llamar el Dogma Central de la Biología Molecular: “la información genética, codificada en el ADN, se transcribe en forma de moléculas intermediarias de ARN, que a su vez se traducen en proteínas, macromoléculas que, amén de sus funciones estructurales y enzimáticas, constituyen los principales agentes funcionales y reguladores de la actividad génica” (Crick, 1970). Otro premio Nobel, Jacques Monod, universalizaba este dogma central con la afirmación: “Lo que vale para *E. coli*, vale también para el elefante” (Fiedman, 2004). Monod llevaba parte de razón, pero los investigadores de la época no tardarían en darse cuenta que una aseveración tan simplista resultará insuficiente para describir la compleja biología molecular de los eucariotas (y en nuestros días, también la de procariotas).

Una de las primeras señales de debilidad de esta doctrina sobrevino en 1977 de la mano de varios investigadores que, en laboratorios y trabajos independientes, demostraron que los genes eucariotas no estaban formados por bloques contiguos de secuencias codificadoras de proteínas (Berget *et al.*, 1977; Chow *et al.*, 1977; Jeffreys y Flavell, 1977). Por el contrario, se trata de secuencias de ADN que codifican fragmentos de proteínas, lo que se ha denominado **exones**, interrumpidos por secuencias, a menudo extensas, a las que se llamó **intrones**. Estos intrones deben ser eliminados del ARNm antes de que sea traducido a proteína, proceso que se conoce como maduración o *splicing*. Este descubrimiento explicaba una de las principales cuestiones que preocupaba a los investigadores a principios de los años 70: “la paradoja o el enigma del valor C<sup>2</sup>” (Thomas, 1971; Gregory, 2001). Durante mucho tiempo se ha considerado que el ARN intrónico era inservible, pero si fuese así, no se explica que el 98,5% del genoma humano esté compuesto de secuencias no codificantes. Hoy día sabemos que los intrones sobrevivieron en los organismos complejos porque desempeñan funciones tan importantes como, por ejemplo, facilitar la reordenación de segmentos de ADN para crear nuevas proteínas, maduración alternativa para dar lugar a diferentes proteínas a partir de un mismo ARNm o la regulación de la expresión génica. También se ha

---

<sup>1</sup> Mendel describe gen como: “The gene was recognized as a particulate factor that passes unchanged from parent to progeny” (Lewin, 2007).

<sup>2</sup> El valor C se define como la cantidad total de ADN en el genoma haploide característico de cada especie. La paradoja o el enigma del valor C se refiere a la relación entre el tamaño del genoma y la complejidad de los organismos; suponen un ejemplo claro las células de algunas salamandras, que contienen hasta 40 veces más ADN que las células humanas. Hay amebas que cuentan con 1000 veces más cantidad de ADN que algunos mamíferos. El nemátodo *Caenorhabditis elegans*, que consta sólo de unas 1000 células, porta unos 19.000 genes codificadores de proteínas, casi el 50% más que los insectos (13.500) y casi tantos como los humanos (alrededor de 25.000).

asumido que la presión competitiva del entorno microbiano ha eliminado los intrones del genoma procariota (cuanto menos ADN que replicar, menos tiempo tardarán en colonizar nuevos nichos; Ochman *et al.*, 2000). Es cierto que no son tan abundantes como en eucariotas superiores, pero también existen, y tienen peculiaridades que los distinguen de los intrones clásicos definidos en eucariotas. Uno de los tipos de intrones que se encuentran en procariotas son los intrones autocatalíticos del grupo II.

Hasta nuestros días, se han descrito cinco tipos principales de intrones presentes en los organismos vivos en función de diferencias estructurales y de su mecanismo de maduración. Además de los intrones nucleares eucariotas (también conocidos como intrones espliceosómicos) que hemos mencionado anteriormente (Burge *et al.*, 1999; Reed, 2000), encontramos un grupo de intrones específicos de arqueobacterias (Lykke-Andersen *et al.*, 1997), un tipo particular de intrones nucleares que se localizan en los ARNt (Abelson *et al.*, 1998) y dos grupos de intrones cuya particularidad es que son capaces de auto-madurar en condiciones de laboratorio: intrones del grupo I y del grupo II (Michel *et al.*, 1982; Davies *et al.*, 1982; Schmelzer *et al.*, 1982)<sup>3</sup>. A su vez, dentro de los intrones del grupo II se ha hecho una distinción al referirse a intrones del grupo III, que no son más que intrones del grupo II simplificados estructural y funcionalmente (Christopher y Hallick, 1989; Copertino y Hallick, 1993).

Los intrones del grupo II, objeto de estudio en esta Tesis Doctoral, fueron inicialmente descritos a partir de únicamente cuatro secuencias procedentes de dos genes mitocondriales de la levadura *Saccharomyces cerevisiae* (Michel *et al.*, 1982). Uno de los principales motivos por los que se convirtieron en interesantes objetos de estudio fue el singular parecido entre su mecanismo de maduración y el que presentaban los intrones nucleares, hecho por el que muchos autores han considerado a los intrones del grupo II como los “padres inquietos” de los intrones espliceosómicos (Padgett *et al.*, 1984; Ruskin *et al.*, 1984; Sharp, 1985; 1991; Cech, 1986; Cavalier-Smith, 1991; Moore *et al.*, 1993; Eickbush, 1994; Koonin, 2006). Por este motivo, entre otros que veremos más adelante, los investigadores han buscado y caracterizado extensamente este tipo de intrones en los genomas de todos los organismos vivos (Michel y Ferat, 1995; Bonen y Vogel, 2001). Se han usado principalmente dos tipos de estrategias en la localización de intrones del grupo II. Una aproximación experimental mediante PCR usando oligonucleótidos degenerados en regiones conservadas en este tipo de elementos (Ferat y Michel, 1993; Ferat *et al.*, 1994); pero también, gracias al creciente número de genomas secuenciados, se han realizado extensas búsquedas de identidades entre secuencias de las bases de datos y dominios conservados en la secuencia/estructura de estos intrones (Knoop y Brennicke, 1994; Knoop *et al.*, 1994). Así, el número de secuencias clasificadas como intrones del grupo II ha ido aumentando desde 1982 hasta alcanzar en nuestros días la cifra aproximada de 700 miembros (Fedorova y Zingler, 2007).

---

<sup>3</sup> La distinción entre intrones del grupo I y II se lleva a cabo en 1982 por Michel y colaboradores cuando, al analizar las secuencias completas de diez intrones mitocondriales de levadura y dos incompletas (una de levadura y otra de *Aspegillus nidulans*), encuentran claras diferencias estructurales (Michel *et al.*, 1989).

Como hemos dicho, se encontraron originariamente en los genomas de los orgánulos de eucariotas inferiores y plantas y posteriormente se han descrito en bacterias y arqueas (Lambowitz y Zimmerly, 2004; Lehmann y Schmidt, 2004; Toro *et al.*, 2007). No suelen ser muy abundantes en mitocondrias de hongos; son muy numerosos en el genoma tanto de mitocondrias (Oda *et al.*, 1992; Knoop y Brennicke, 1993) como de cloroplastos (Ohyama *et al.*, 1986; Shinozaki *et al.*, 1986) de plantas superiores y están presentes en el genoma de orgánulos de algas (Kück *et al.*, 1987; Kück, 1989; Kono *et al.*, 1991; Manhart y von der Haar, 1991). En este sentido, es de destacar la especial abundancia de intrones del grupo II (y también del grupo III) en los cloroplastos de *Engelma gracilis* donde han llegado a identificarse hasta 155 intrones, lo que supone el 40% del genoma del orgánulo (Hallick *et al.*, 1993; Copertino y Hallick, 1993; Doetsch *et al.*, 1998; 2001). Así mismo, la magnitud de esta abundancia puede comprobarse al comparar este número de intrones con el que encontramos en el genoma del cloroplasto de *Arabidopsis* (26 intrones; Sato *et al.*, 1999) o en los genomas mitocondriales de ésta (22 intrones; Marienfeld *et al.*, 1999) y de *Marchantia polymorpha* (25 intrones; Oda *et al.*, 1992). Curiosamente, los intrones identificados en orgánulos se encuentran en la mayoría de los casos interrumpiendo genes que codifican proteínas (Open Reading Frames, ORFs) y, en algunos casos, los genes de ARNt y ARNr. Se observan ciertas tendencias en los genomas: en plantas con flor, todos los intrones mitocondriales del grupo II se encuentran interrumpiendo ORFs, mientras que un tercio de los presentes en el cloroplasto se han insertado en ADNt (los dos tercios restantes interrumpen ORFs); en cambio, los dos intrones presentes en la mitocondria del alga roja *Porphyra* se localizan en el ADN que codifica los ARNr (Bonen y Vogel, 2001). Llama la atención que aún no se han encontrado intrones del grupo II en mitocondrias animales o interrumpiendo genes nucleares eucariotas, aunque se ha descrito que el genoma nuclear de *Arabidopsis thaliana* (concretamente el cromosoma 2) contiene la inserción de una secuencia de ADN mitocondrial (270 kb) que a su vez incluye un intrón del grupo II (Lin *et al.*, 1999).

A principios de los años 90, Ferat y colaboradores identificaron por primera vez intrones del grupo II en proteobacterias (*Azotobacter*, *Escherichia*) y cianobacterias (*Calothrix*, *Anabaena*, *Nostoc*) (Ferat y Michel, 1993; Ferat *et al.*, 1994), bacterias que se supone que han dado origen a las mitocondrias y los cloroplastos, respectivamente. Más tarde, se identificaría uno de los intrones del grupo II más estudiados y caracterizados, L1.ltrB en *Lactococcus lactis* (Mills *et al.*, 1996; Shearman *et al.*, 1996). Aproximadamente, el 25% de los genomas bacterianos secuenciados contienen intrones del grupo II y en algunos casos se han descrito hasta 12 copias de estos elementos en el genoma (Lambowitz y Zimmerly, 2004), lo que nos proporciona una idea de la amplia distribución y presencia de este tipo de elementos entre los procariotas (Martínez-Abarca y Toro, 2000a; Dai y Zimmerly, 2002). Al contrario de lo que ocurre con los intrones de orgánulos, los intrones bacterianos suelen localizarse entre genes (que no interrumpiendo ORFs) o bien, con mucha frecuencia, dentro de elementos móviles como secuencias de inserción o plásmidos (Ferat *et al.*, 1994; Martínez-Abarca *et al.*, 1998; Mills *et al.*, 1996; Shearman *et al.*



*al.*, 1996; Klein y Dunny, 2002). Estas últimas localizaciones podrían propiciar la dispersión de estos elementos en el mundo procarionta (Martínez-Abarca y Toro, 2000a; Lambowitz y Zimmerly, 2004).

Mucho más reciente es la descripción de este tipo de elementos en arqueobacterias (Galagan *et al.*, 2002; Deppenmeier *et al.*, 2002). Se han encontrado únicamente en dos especies de arqueas metanogénicas estrechamente relacionadas, *Methanosarcina acetivorans* y *Methanosarcina mazei*, y se piensa que su presencia podría estar relacionada con transferencia horizontal desde eubacterias (Rest y Mindell, 2003; Dai y Zimmerly, 2003; Toro, 2003).

Otra de las peculiaridades de este tipo de intrones es que pueden codificar proteínas pertenecientes a la familia de las reverso-transcriptasas (Michel y Lang, 1985; Xiong y Eickbush, 1990). Genéricamente, estas proteínas reciben el nombre de IEPs (Intron-Encoded Proteins), aunque alguna como la del intrón Ll.ltrB de *L. lactis* tiene nombre propio: LtrA. Un dato curioso es que, a diferencia de los intrones del grupo II en orgánulos<sup>4</sup>, casi todos los descritos en bacterias presentan este tipo de ORFs (Lambowitz y Zimmerly, 2004). Se sabe que estas proteínas asisten específicamente la maduración de su intrón y participan activamente en otros procesos en los que se ven involucrados estos intrones (Lambowitz, 1989; Lambowitz y Belfort, 1993; Lambowitz *et al.*, 1999).

A las muchas particularidades nombradas hasta el momento, debemos sumar una más: estos intrones presentan capacidad de movimiento. Ya se intuía, por su presencia y distribución, que podríamos estar hablando de elementos genéticos móviles (Lambowitz, 1989) pero no fue hasta 1990 que se demostró su capacidad para invadir una nueva secuencia específica de ADN (Meunier *et al.*, 1990; Skelly *et al.*, 1991). No todos los intrones del grupo II son móviles, de hecho sólo se ha podido demostrar esta capacidad en algunos de ellos: aI1 y aI2 (también conocidos como *coxI-I1* y *coxI-I2*) presentes en *S. cerevisiae* (Meunier *et al.*, 1990; Lazowska *et al.*, 1994); Ll.ltrB de *L. lactis* (Mills *et al.*, 1997) y RmInt1 en *Sinorhizobium meliloti* (Martínez-Abarca *et al.*, 2000). Recientemente, se han identificado un grupo de intrones que se comportan de un modo similar, en cuanto a maduración y movilidad, a Ll.ltrB en *E. coli*: Na.1 y Na.2 de *Novosphingobium aromaticivorans*, y Ec.5 de *E. coli* (Karberg, 2005).

## 1. MECANISMOS DE MADURACIÓN DE INTRONES DEL GRUPO II

Los intrones del grupo II, al igual que el resto de intrones conocidos, deben escindir del ARNm para que la traducción de éste a proteínas tenga éxito, o bien para que el ARN sea funcional en el caso de los ARNt y ARNr. Este proceso se conoce como maduración o *splicing*. Mientras que los intrones espliceosómicos precisan de una compleja maquinaria para su maduración, los intrones del grupo II se caracterizan por catalizar su propia escisión. Esta afirmación es cierta en condiciones no fisiológicas, pero *in vivo* todo el proceso se encuentra asistido por la IEP o bien, por factores de *splicing* presentes en el

---

<sup>4</sup> Sólo un tercio de los intrones del grupo II de orgánulos presentan IEP (Toor *et al.*, 2001)

organismo hospedador (Lambowitz y Belfort, 1993). En cualquier caso, este proceso consiste en dos reacciones consecutivas de transesterificación en las que un nucleófilo (por lo general un radical hidroxilo activado) ataca los sitios de *splicing* o maduración. Para entender mecanísticamente el proceso de maduración de estos intrones es preciso conocer más en detalle su secuencia, su estructura y organización.

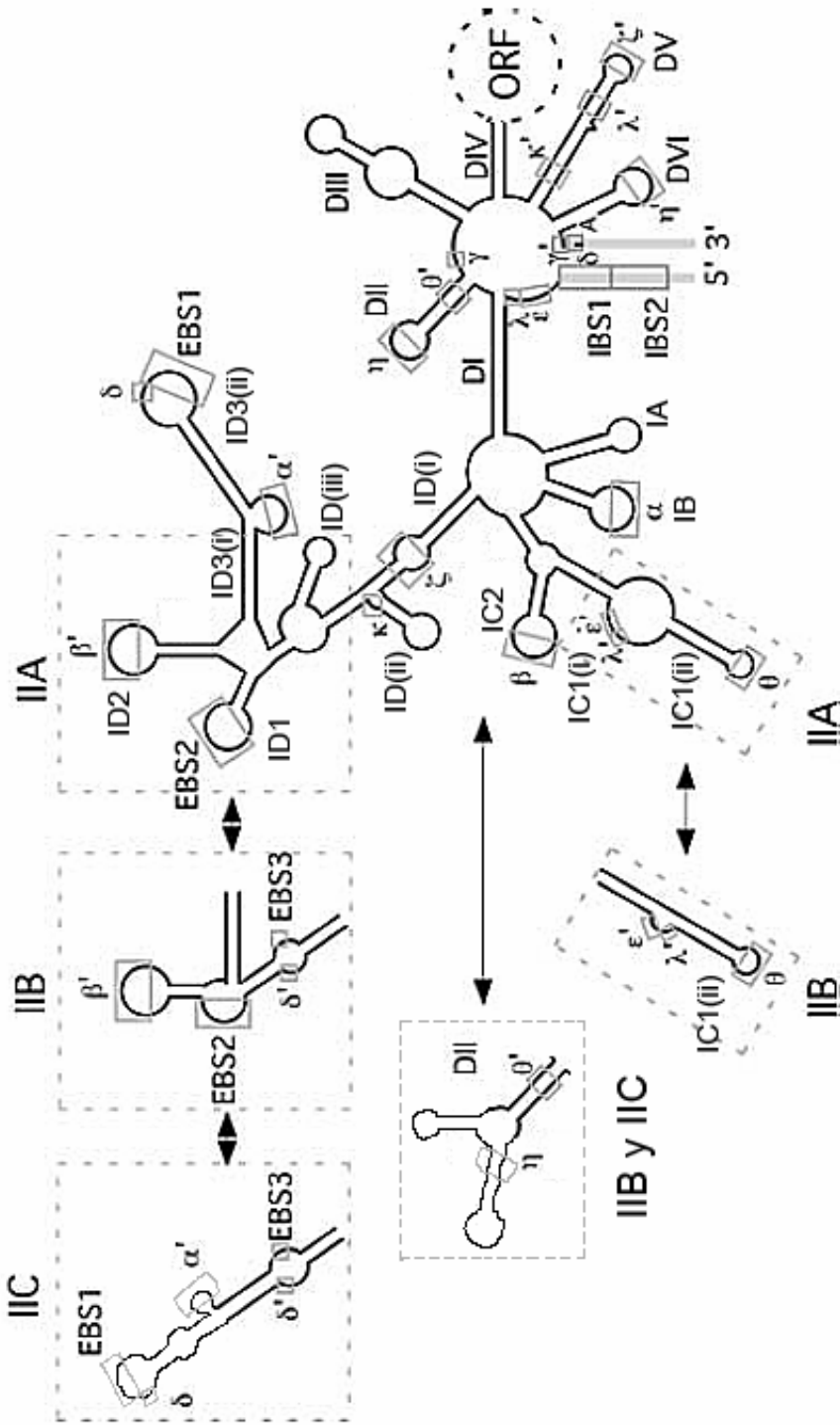
### 1.1. ESTRUCTURA DE LA RIBOZIMA

Acabamos de introducir el término ribozima para referirnos al componente ARN de los intrones del grupo II, y diferenciarlo así del componente proteico al que venimos denominando IEP. Una ribozima es una molécula de ARN con capacidad catalítica. Su nombre deriva de “ARN enzima” aunque no son enzimas verdaderas ya que, en el proceso que catalizan, suelen modificarse a sí mismas<sup>5</sup>. En la naturaleza se han identificado ocho tipos de ribozimas diferentes y se han clasificado en dos grandes grupos en función de su tamaño y su mecanismo de reacción (Tanner, 1999; Lehmann y Schmidt, 2004; Gordon *et al.*, 2007). Por un lado encontramos las ribozimas pequeñas o endonucleolíticas, donde se engloban los motivos catalíticos *hammerhead* y *hairpin*, el ARN VS, la ribozima HDV (Hepatitis Delta Virus) y glmS. Este tipo de ARNs catalíticos tienen entre 35 y 155 nucleótidos y llevan a cabo reacciones en las que se generan productos con 5'-OH y 2'-3'-fosfato cíclico. En el otro extremo están las grandes ribozimas, cuyo tamaño oscila entre los pocos cientos y los 3000 nucleótidos, y entre las que se encuentran la RNasa P<sup>6</sup> y los intrones del grupo I y II. En este caso se generan más variedad de productos: 5'-P y 3'-OH (RNasa P), 5'-G/3'-OH intrón y exones ligados 5'-3' (intrones del grupo I) y 2'-5'/3'-OH intrón *lariat* y exones ligados 5'-3' (intrones del grupo II). En los últimos años, se ha demostrado que el componente catalítico de muchos complejos ribonucleoproteicos relevantes en el funcionamiento celular es el ARN (Cech *et al.*, 2006); tal es el caso de los ribosomas (ARNr; Moore y Steitz, 2006), el espliceosoma (snARN; Will y Lührmann, 2006; Valadkhan, 2007), la maquinaria de procesamiento del ARNr (snoARN; Tycowski *et al.*, 2006) o la de edición del ARN (ARNguía; Simpson, 2006). Es importante resaltar que el descubrimiento de las ribozimas supuso para Thomas R. Cech y Sidney Altman el premio Nobel de química en 1989.

Los escasos datos filogenéticos disponibles en 1982 permitieron a Michel y colaboradores establecer una estructura consenso para los intrones del grupo II (Michel *et al.*, 1982). En esta estructura se diferencian claramente seis dominios de diferente longitud en los que se intercalan segmentos en doble hélice y bucles, denominados dominio I al VI (dI-dVI). En una representación esquemática bidimensional se disponen en forma radial a partir de una estructura central que comprende las regiones interdominio. El elevado número de interacciones entre dominios sugiere una estructura 3D compleja (Qin y Pyle, 1998).

<sup>5</sup> Con la excepción de la RNasa P, que procesa el extremo 5' de los precursores de los ARNt.

<sup>6</sup> Realmente se trata de una RiboNucleoPartícula (RNP) formada por un ARN de 350-400 nt y una proteína básica de 14 kDa.



**Figura 1.** Estructura secundaria consenso para intrones del grupo II. Consiste en seis dominios con forma de doble hélice (DI-DVI) que se disponen de forma radial alrededor de una estructura central. Los distintos subdominios se indican con la denominación del dominio seguida de un número, y, si procede, entre paréntesis se indican nuevas subdivisiones (p.e., ID(i)). La proteína (ORF) se representa con una línea discontinua en el DIV de la ribozima. En letras griegas se señalan las principales interacciones terciarias descritas. Las diferencias más representativas entre las tres clases de intrones, IIA, IIB y IIC, se muestran en cuadros. Figura adaptada de Lambowitz y Zimmerly, 2004.

### 1.1.1. Características estructurales de cada dominio

A pesar de tener una estructura secundaria muy conservada, no se puede decir lo mismo de su secuencia primaria (Michel *et al.*, 1989). Los residuos que sabemos son importantes en el establecimiento de la estructura adecuada para la catálisis o para la propia reacción, se encuentran diseminados a lo largo de toda la secuencia del intrón. Por otro lado, la división de la estructura en dominios y su nomenclatura no sigue ningún criterio funcional, sino que trata de simplificar la compleja organización de estos elementos (Qin y Pyle, 1998). Gracias a estudios en *trans*<sup>7</sup> con los diferentes dominios, hoy podemos asociar funciones concretas a determinados dominios.

#### 1.1.1.1. Dominio I

Este dominio se considera el de mayor tamaño de los seis, si nos olvidamos del dIV de aquellos intrones que contienen IEP, y, a su vez, el de mayor complejidad estructural. Esta complejidad ha llevado a los autores a dividirlo en varios subdominios: A, B, C (C1 y C2) y D (D1, D2 y D3, a su vez también subdivididos). Paralelamente, también es el dominio que concentra el mayor número de interacciones implicadas en los procesos desarrollados por el intrón, por lo que, desde muy pronto, se ha considerado indispensable para la catálisis (Michel *et al.*, 1989; Koch *et al.*, 1992; Michels y Pyle, 1995). En este dominio se han identificado las secuencias implicadas en el reconocimiento de los exones, por lo que juega un papel fundamental en la selección de los sitios de maduración así como en el reconocimiento de la secuencia diana necesario en el proceso de movilidad (tanto en sustratos ARN como ADN). Por otro lado, se dice que es el “andamio” sobre el que se sostiene toda la estructura de estos intrones y que conforma el sitio catalítico al interaccionar con los otros dominios, básicamente con el dV; es más, mantiene contactos terciarios con casi todos los dominios restantes (Pyle y Lambowitz, 2006).

Michel y colaboradores verificaron la existencia de secuencias que co-variaban filogenéticamente con el exón 1 en dos regiones separadas dentro el subdominio ID (Michel *et al.*, 1989); se las denominó secuencias de reconocimiento de los exones 1 y 2 (EBS1 y 2). Estos motivos interaccionan mediante apareamiento de bases con regiones complementarias presentes en el extremo 3' del exón 1 (IBS1 y 2, Intron Binding Sites) (Jacquier y Rosbash, 1986; Jacquier y Michel, 1987). Inicialmente se estableció que el sitio de maduración 5' (5'-ss, 5'-splice site) estaba determinado por encontrarse enfrentado al primer nucleótido apareado del EBS1 (Jacquier y Jacquesson-Breuleux, 1991) pero posteriores investigaciones apuntan a que el sitio de escisión viene marcado por la transición estructural entre cadena simple y doble (Su *et al.*, 2001). Aunque se ha establecido que la interacción EBS2-IBS2 es “prescindible” para el *splicing* (Jacquier y Jacquesson-Breuleux, 1991; Toor *et al.*, 2006; Barrientos-Durán, 2008), no es el caso del

<sup>7</sup> Se trata de estudios *in vitro* en los que los dominios se transcriben como moléculas independientes y, cuando se combinan con otras regiones del intrón y se analizan las reacciones que son capaces de catalizar, podemos deducir las funciones en las que están directamente involucrados.

EBS1; parece ser que la desestabilización espacial o la delección de la región donde se encuentra el EBS1 provoca una fuerte disminución de la eficiencia de escisión del intrón induciendo incluso la elección de sitios de *splicing* erróneos en 5' (Hetzer *et al.*, 1997a; Lehmann y Schmidt, 2003). Así, la selección del 5'-ss podría ser el resultado de un efecto cooperativo de varias interacciones.

Una de estas interacciones se piensa que podría ser la  $\epsilon$ - $\epsilon'$ ; este contacto terciario tiene lugar entre los nucleótidos 3 y 4 del intrón (usualmente GY,  $\epsilon$ ) y dos nucleótidos situados en una región desapareada del dominio IC1 (RC,  $\epsilon'$ ). Se piensa que esta interacción no sólo está implicada en el reconocimiento del 5'-ss sino que es fundamental en el posicionamiento del primer nucleótido del intrón en el sitio catalítico (Jacquier y Michel, 1990; Boudvillain y Pyle, 1998). En este sentido, la naturaleza de este primer nucleótido (suele ser una G) es elemental, pues su sustitución por cualquiera de los otros tres nucleótidos provoca una fuerte disminución de la eficiencia de maduración (Chanfreau y Jacquier, 1993; Peebles *et al.*, 1993; Holländer y Kück, 1999a).

Además de albergar regiones críticas en la identificación del exón 1, en el dI también se localizan algunos de los elementos que contribuyen al reconocimiento del exón 2. Las interacciones implicadas en este reconocimiento varían entre los diferentes tipos de intrones del grupo II. En intrones del grupo IIA, el nucleótido que inmediatamente precede al EBS1, denominado  $\delta$ , se encuentra apareado con el primer residuo del exón 2, conocido como  $\delta'$  (Jacquier y Jacquesson-Breuleux, 1991; Costa *et al.*, 2000). Sin embargo, en los intrones IIB la posición  $\delta'$  se encuentra situada en una región desapareada presente en el dI y conocida como región de coordinación. Así pues, la interacción responsable del reconocimiento del exón 2 en intrones IIA ( $\delta$ - $\delta'$ ) se convierte en un contacto intradominio en los intrones IIB. Curiosamente en la región de coordinación encontramos la posición EBS3 en intrones IIB, que es justamente la que interacciona con el nucleótido +1<sup>8</sup> del exón 2 (denominado en este caso IBS3; Costa *et al.*, 2000). De este modo, la región de coordinación se encarga, al menos en intrones del grupo IIB, de que las secuencias de reconocimiento de los dos exones se encuentren próximas espacialmente. Es más, la región de coordinación recibe este nombre porque se piensa que coordina el ensamblaje del sitio catalítico de estas ribozimas (Hamill y Pyle, 2006; este trabajo).

Si los EBSs son importantes para que tenga lugar el *splicing*, no lo son menos para que se produzca la reacción inversa; el *splicing* reverso es el mecanismo básico para que estos intrones puedan invadir nuevas localizaciones en el genoma de su hospedador (Eskes *et al.*, 1997; Guo *et al.*, 1997). La fidelidad a la hora de reconocer el sustrato ha sido estudiada en detalle tanto desde el punto de vista mecanístico como energético, observándose un alto grado de especificidad en la unión a su diana y la inducción de un

---

<sup>8</sup> Si consideramos que el intrón está insertado en la posición 0, las secuencias de los exones se nombran -X si pertenecen al exón 1 y +X si se encuentran en el exón 2.

cambio conformacional en la estructura de la ribozima derivada del acoplamiento intrón/sustrato (Michels y Pyle, 1995; Xiang *et al.*, 1998; Qin y Pyle, 1999; Su *et al.*, 2001). El reagrupamiento conformacional del dI es independiente del de los otros dominios y su adecuada disposición supone además un paso clave para el plegamiento del resto de la estructura del intrón. Parece ser que una vez plegado presenta una serie de “bolsillos” en los que encajan perfectamente el resto de los dominios y estructuras que al unirse podrían generar ligeros cambios en la conformación local (Qin y Pyle, 1997; Su *et al.*, 2005). En la formación y estabilización de esta estructura participan tanto interacciones intradominio como contactos con el resto de los dominios. Las interacciones intradominio más universales son:  $\alpha$ - $\alpha'$  y  $\beta$ - $\beta'$ . La primera de ellas está muy conservada<sup>9</sup> y tiene lugar entre el bucle terminal del subdominio IB y una región desapareada interna en el subdominio ID3 (Jacquier y Michel, 1897); se ha visto que podría facilitar la unión de los exones (Harris-Kerr *et al.*, 1993; Qin y Pyle, 1998) o participar activamente en el plegamiento del intrón (Waldsich y Pyle, 2007). En cambio, la interacción  $\beta$ - $\beta'$  entre dos regiones desapareadas de los subdominios IC2 y ID2 no está presente en todos los intrones del grupo II (Michel y Ferat, 1995); parece ser que podría ser una interacción dispensable para el funcionamiento de estos elementos (Ferodova y Zingler, 2007) pero cuando está presente juega un papel importante en la compactación de la ribozima (Waldsich y Pyle, 2007). Una interacción que únicamente se ha descrito en intrones IIB es  $\delta$ - $\delta'$ ; como ya hemos explicado, podría permitir el posicionamiento del exón 1 en el sitio catalítico al acercar el EBS1 a los sitios donde se anclan los dominios V y VI (Costa *et al.*, 2000). Finalmente, existen una serie de contactos que el dI establece con distintas regiones del intrón que son claves para su actividad, principalmente aquellos con el dV (Ferodova y Zingler, 2007):  $\zeta$ - $\zeta'$ ,  $\kappa$ - $\kappa'$  y  $\lambda$ - $\lambda'$ . Los dos primeros son críticos para el ensamblaje del sitio catalítico porque permiten la unión del dV al dI (Costa y Michel, 1995; Costa *et al.*, 1998; Boudvillain y Pyle, 1998; Konforti *et al.*, 1998a), mientras que la interacción  $\lambda$ - $\lambda'$  permitiría el acercamiento del dV y el 5'-ss (gracias a su vez a la proximidad con la interacción  $\epsilon$ - $\epsilon'$ ), jugando un papel directo en la catálisis (Boudvillain *et al.*, 2000).

Por último, es importante resaltar que en el dI de Ll.ltrB (intrón IIA de *L. lactis*) se han identificado algunos contactos potenciales con LtrA (su IEP), concretamente en las subestructuras IC2 y ID4 (Matsuura *et al.*, 2001). Por otro lado, recientes estudios con el factor de maduración CRS1 y el intrón atpF en maíz han revelado que el sitio de unión de esta proteína se emplaza en una inserción situada justo en el tallo que da inicio al dI (Ostersetzer *et al.*, 2005), localización que había sido postulada previamente como sitio potencial de unión de proteínas (Adamidi *et al.*, 2003).

### 1.1.1.2. Dominio II

Es la región menos conservada tanto desde el punto de vista de la secuencia como de la estructura, produciéndose claras diferencias entre los distintos tipos de intrones (Michel *et*

<sup>9</sup> De hecho fue definida a partir de análisis filogenéticos.

*al.*, 1989). Este dominio establece contactos terciarios del tipo tetraloop/receptor<sup>10</sup> con el dI ( $\theta$ - $\theta'$ ) y con el dVI ( $\eta$ - $\eta'$ ). La interacción  $\theta$ - $\theta'$  implica la región basal del dII y la zona terminal de la región IC1, variando el tipo de asociación tetraloop/receptor entre intrones IIA (GAAA/11nt) y IIB (GNRA/G-C). Se mantiene durante todo el proceso de maduración lo que indica que podría estar implicada en la estabilización de la estructura nativa del intrón (Chanfreau y Jacquier, 1996; Costa *et al.*, 1997b). Esta interacción también se considera importante para acercar tanto el dIII como la región entre dII/dIII ( $J_{II/III}$ ) al centro catalítico (Fedorova *et al.*, 2003).

La interacción con el dVI tiene mayores implicaciones en la catálisis que la anterior aunque en ningún caso participa directamente. El contacto tetraloop/receptor  $\eta$ - $\eta'$  se establece entre la región central del dII y el bucle terminal del dVI (Chanfreau y Jacquier, 1996). Al igual que la interacción  $\theta$ - $\theta'$ , se manifiesta de modo diferencial entre los dos grupos mayoritarios de intrones del grupo II: en intrones IIA el tetraloop se localiza en el dII y el receptor en el dVI, mientras que en IIB se sitúan a la inversa (Costa *et al.*, 1997b). La eliminación del dII (Kwakman *et al.*, 1989; Bachl y Schmelzer, 1990) o bien la disrupción de esta interacción conlleva la drástica reducción de la eficiencia de la segunda reacción de *splicing*, por lo que se piensa que este contacto podría facilitar un cambio conformacional entre los dos pasos de la reacción, posibilitando la ligación de los exones (Chanfreau y Jacquier, 1996). Este reagrupamiento del sitio catalítico debe ser muy leve (probablemente sólo implique al dVI), ya que todos los elementos que participan en ambas reacciones se mantienen cerca durante todo el proceso conformando un único sitio activo (Chanfreau y Jacquier, 1994; 1996; de Lencastre *et al.*, 2005).

### 1.1.1.3. Dominio III

Es uno de los dominios más conservados. Consiste en una estructura corta formando una doble hélice que contiene dos regiones desapareadas, una interna y el bucle terminal. No es un elemento requerido estrictamente en la catálisis (Koch *et al.*, 1992), sin embargo se ha descrito que mejora tanto la eficiencia de escisión (Podar *et al.*, 1995a) como la de inserción del intrón (Griffin *et al.*, 1995; Xiang *et al.*, 1998). Se pliega de manera independiente al resto de los dominios y se une al intrón con una alta afinidad (Podar *et al.*, 1995a), fundamentalmente a los dominios I y V (Jestin *et al.*, 1997; Fedorova *et al.*, 2003; Fedorova y Pyle, 2005). Se ha demostrado que queda imbuido en la estructura y que probablemente su función esté en relación con la compactación de ésta, impidiendo que moléculas de agua puedan entrar a forma parte del centro activo (Swisher *et al.*, 2001; Fedorova *et al.*, 2003; Su *et al.*, 2005).

---

<sup>10</sup> Son contactos que suelen estar implicados en la estabilización de la estructura terciaria mediante puentes de hidrógeno. Hay 2 tipos principales descritos en intrones del grupo II: (i) GNRA/receptor G-C, en el que las posiciones 3 y 4 del tetraloop interactúan con el surco menor del ARN receptor; y (ii) GAAA/receptor motivo 11 nt (CCUAAG...UAUGG) (Costa y Michel, 1995; 1997).

#### 1.1.1.4. Dominio IV

Es la región más variable de los intrones del grupo II y no se le ha asignado ningún papel catalítico. Sin embargo es importante porque codifica la IEP en aquellos intrones que presentan. Parece encontrarse fuera de la estructura formada por el resto de los dominios de la ribozima y presenta el sitio de anclaje principal para la proteína codificada por el intrón (Matsuura *et al.*, 2001; Singh *et al.*, 2002; Watanabe y Lambowitz, 2004) así como para otras proteínas (Ostersetzer *et al.*, 2005). La unión de la IEP va a facilitar los procesos de *splicing* (Wank *et al.*, 1999) y movilidad de estos elementos (Huang *et al.*, 2003). Profundizaremos en todos estos aspectos en posteriores apartados.

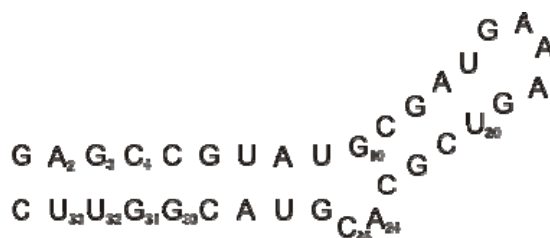
#### 1.1.1.5. Dominio V

Es, sin lugar a dudas, el dominio más conservado de toda la estructura tanto a nivel de secuencia primaria como de estructura secundaria. Consiste en una horquilla corta (normalmente 34 nt) terminada en un bucle GNRA y que aproximadamente a dos tercios de la base presenta una región desapareada asimétrica (Michel *et al.*, 1989). Este dominio junto con el dI son los únicos dos considerados indispensables para la actividad del intrón, razón por la cual es uno de los más estudiados (Koch *et al.*, 1992).

Uno de los motivos más conservados del dV es lo que se ha denominado la triada catalítica situada en la base de este dominio: A2, G3 y C4. Cambios en cualquiera de los residuos bloquean completamente el *splicing*, mientras que si se modifican los nucleótidos con los que aparean (G31 U32 U33) el efecto es leve (Chanfreau y Jacquier, 1994; Boulanger *et al.*, 1995; Peebles *et al.*, 1995).

De los 3, la G es esencial para la maduración del intrón tanto *in vitro* como *in vivo* (Boulanger *et al.*, 1995; Peebles *et al.*, 1995) y podría jugar un papel dual: hacia el surco menor, estabiliza la unión del dV al dI, y hacia el surco mayor, el O6 y el N7 de esta G participarían activamente en la química de reacción (Konforti *et al.*, 1998b). Ver figura 2.

Otro de los elementos considerados esenciales para el adecuado funcionamiento del dV es la región desapareada interna. Inicialmente se había descrito que esta región estaba comprendida por los residuos C25 y G26, pero estudios posteriores han demostrado que, al menos, la componen los nucleótidos A24 y C25 (Costa *et al.*, 1998; Konforti *et al.*, 1998b; Zhang y Doudna, 2002). El análisis de la región mediante mutagénesis en el intrón aI5 $\gamma$  llevó a la conclusión de que la naturaleza de los nucleótidos no era tan importante como su número (Schmidt *et al.*, 1996), pues la distorsión generada en la estructura varía en función del número de nucleótidos desapareados (Lehmann y Schmidt, 2003). Estudios de modelización del centro activo de la ribozima del intrón aI5 $\gamma$  basados en ensayos de entrecruzamiento con UV parecen indicar que cada uno de los nucleótidos



**Figura 2.** Secuencia primaria propuesta para el intrón aI5 $\gamma$ .

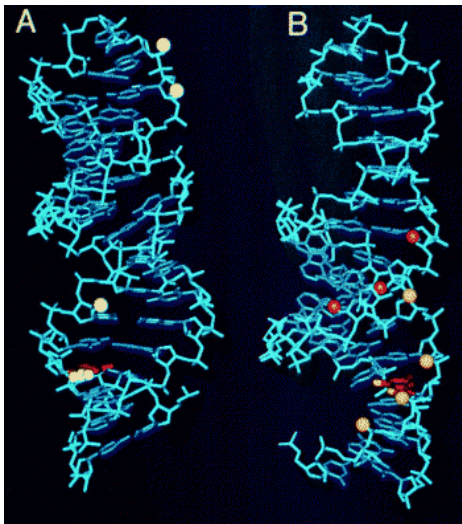


desapareados interacciona con una región diferente en el sitio catalítico. Ambos nucleótidos se encuentran fuera de la hélice formada por el dV y mientras que la A24 queda próxima al quinto nucleótido del intrón, el otro residuo desapareado, C25, se dispone hacia la región interdominio J<sub>II/III</sub> (de Lencastre *et al.*, 2005; de Lencastre y Pyle, 2008). Sin embargo, trabajos recientes de RMN sugieren que este dinucleótido desapareado pudiera abarcar más bien una región con cuatro nucleótidos: U9, A24, C25 y G26 (Sigel *et al.*, 2004; Eldho y Dayie, 2007). Todos podrían establecer algún tipo de interacción con los nucleótidos adyacentes que los situarían dentro de la hélice excepto la G26, que adopta una configuración diferente que la expulsa fuera de la estructura y hacia el surco mayor. La caracterización mediante RMN de los cambios en la estructura del dV durante el proceso de unión al dI, ha permitido describir la existencia de un cambio conformacional en la región interna desapareada del dV necesario para la consecución de la estructura catalítica activa; esta teoría ha recibido la denominación de modelo de activación estructural (Gumbs *et al.*, 2005).

Se sabe que el dV establece interacciones no W-C con otras partes del intrón (Franzen *et al.*, 1993; Pyle y Green, 1994; Peebles *et al.*, 1995). En su interacción con el dI se han identificado tres contactos, dos del tipo tetraloop/receptor ( $\zeta$ - $\zeta'$  y  $\kappa$ - $\kappa'$ ) y un tercero que implica dos pares de bases G-C consecutivos junto a la región interna desapareada del dV con dos nucleótidos desapareados adyacentes al motivo  $\epsilon'$  ( $\lambda$ - $\lambda'$ ). Mientras que los primeros contribuirían sólo estructuralmente, la interacción  $\lambda$ - $\lambda'$  tendría importantes implicaciones en el proceso catalítico (ver dominio I; Boudvillain *et al.*, 2000). Recientemente, se ha identificado la interacción  $\mu$ - $\mu'$  entre el bucle terminal del dIII y la región comprendida entre la triada catalítica y la región desapareada interna del dV, que probablemente esté implicada en la fijación del dIII al sitio catalítico (Fedorova y Pyle, 2005).

En el dV también se han identificado dos regiones que podrían coordinar iones divalentes que se sabe que son indispensables para que se lleve a cabo la actividad así como el plegamiento de estas grandes ribozimas (Steitz y Steitz, 1993; Sontheimer *et al.*, 1999; Sigel, 2005). Estos dos sitios son la región interna desapareada (Sigel *et al.*, 2000) y la A de la triada catalítica AGC (Gordon y Piccirilli, 2001; Sigel *et al.*, 2004).

La enorme cantidad de datos disponibles han permitido desarrollar modelos tridimensionales del dV (Abramovitz *et al.*, 1996; Konforti *et al.*, 1998b; Zhang y Doudna, 2002). En estas estructuras se pueden distinguir dos regiones claramente diferenciadas desde el punto de vista funcional: la *región catalítica* donde se concentran los grupos funcionales importantes para la catálisis y que comprende el surco mayor de la hélice; y la *región de unión* donde se encuentran los grupos implicados en la interacción con otras partes del intrón. Ver figura 3.



**Figura 3.** Modelo tridimensional del DV. En esta representación se destacan los sustituyentes más importantes que definen la superficie de unión y la superficie catalítica: G3 está representada en rojo, mientras las esferas con distintas tonalidades representan grupos 2'-OH. (A) Muestra el surco menor que está implicado en la unión de este dominio al resto del intrón (B) Se trata de la misma estructura girada 180° que nos muestra el surco mayor donde se localizan el N7 y el O6 (puntos amarillos) de G3 cerca de una serie de sustituyentes (círculos amarillos) importantes en la química de la reacción (Figura copiada de Konforti *et al.*, 1998b).

#### 1.1.1.6. Dominio VI

Este dominio es importante porque en él encontramos una adenosina desapareada (A) a 7-8 nt del final del intrón muy conservada (Michel *et al.*, 1989). Esta A participa en la primera reacción de *splicing* proporcionando el 2'-OH que actúa como nucleófilo y ataca al primer nucleótido del intrón (5'-ss). Por ser un residuo puntual, conservado e importante en la catálisis, se han desarrollado multitud de estudios relacionados con su mecanismo de acción, el mecanismo por el cual se reconoce, su implicación en la selección del 3'-ss, etc. (Revisado en Lehmann y Schmidt, 2003 y en capítulos posteriores de este trabajo). Brevemente decir que se postula que es reconocida mediante interacciones de Van der Waals y de exclusión (Liu *et al.*, 1997) y que hay controversia para situarla en la estructura: algunos autores consideran que se encuentra fuera de la hélice (Chu *et al.*, 1998; Zhang y Doudna, 2002; Schlatterer *et al.*, 2006), mientras otros piensan que podría encontrarse dentro aunque ligeramente expuesta (Erat *et al.*, 2007).

Durante mucho tiempo se han tratado de determinar las regiones en el resto del intrón con las que este dominio interacciona. Como dijimos en el apartado 1.1.1.2. (dominio II) se sabe que interacciona con el dII (Chanfreau y Jacquier, 1996), pero este contacto se considera insuficiente para posicionar la A en el sitio catalítico. De modo que se consideró que su posición relativa al dV, contribuye a ese posicionamiento (Dib-Hajj *et al.*, 1993; Chin y Pyle, 1995; Boudvillain y Pyle, 1998; Podar y Perlman, 1999; Chu *et al.*, 2001). Recientemente se ha identificado la región de coordinación en el dI y se postula que podría ser el receptor de la A del dVI (Hamill y Pyle, 2006; este trabajo).

#### 1.1.1.7. Dominio VII

La existencia de este séptimo dominio es algo excepcional que sólo se ha descrito en el intrón *Bc.I4* de *Bacillus cereus* (Fedorova y Zingler, 2007). Se trata de una inserción de 56 nt en el extremo 3' del dVI, justo donde debía encontrarse el 3'-ss (3 nt después del dVI; Tourasse *et al.*, 2005). Forma una estructura en doble hélice coronada por un bucle

desapareado. Se habla de séptimo dominio porque las interacciones  $\gamma$ - $\gamma'$  y EBS3-IBS3 se encuentran tras esta inserción, aunque el nucleófilo preciso en la primera reacción de transesterificación no se altera; esto, se traduce en un intrón escindido en forma de lazo con una cola 3' más prolongada y estructurada (Stabell *et al.*, 2007). Este intrón debe haberse adaptado perfectamente a esta estructura extra porque su presencia no afecta a su maduración (Fedorova y Zingler, 2007).

#### 1.1.1.8. Regiones interdominio

Los nucleótidos que forman parte de esa especie de “rueda central” en la estructura secundaria de los intrones del grupo II son de los más conservados y, en algunos casos, desempeñan un papel esencial en la catálisis. Como hemos mencionado con anterioridad (ver dI y dV), los primeros nucleótidos de estas ribozimas participan activamente en el establecimiento de la estructura terciaria necesaria durante la catálisis:  $\lambda$ - $\lambda'$ , interacción tripartita entre el quinto nucleótido del intrón, la región desapareada interna del subdominio IC1 y la base del segundo tallo del dV; y  $\epsilon$ - $\epsilon'$ , entre los residuos 3 y 4 del intrón y el dinucleótido que sigue al residuo  $\lambda$  que se encuentra en la región desapareada del subdominio IC1. Además el primer nucleótido del intrón (G1) se sabe esencial en los dos pasos del *splicing* (Chanfreau y Jacquier, 1993; Peebles *et al.*, 1993). Recientes trabajos de entrecruzamiento con UV han permitido situar el 5'-ss en las proximidades de la triada catalítica<sup>11</sup>, conformación indispensable para que tenga lugar la primera reacción del proceso de maduración (de Lencastre y Pyle, 2008). Además, se ha descrito que, durante la segunda reacción de transesterificación, este nucleótido forma un apareamiento no canónico (no W-C) con el penúltimo residuo del intrón (Chanfreau y Jacquier, 1993).

Entre los dominios II y III se encuentra una región rica en purinas, (A/G)GA muy conservada en todos los intrones del grupo II denominada J<sub>II/III</sub> (Michel *et al.*, 1989). Se ha comprobado que forma parte del centro activo de estas ribozimas (Podar *et al.*, 1998a; de Lencastre *et al.*, 2005) y que sólo funciona en cooperatividad con la región proximal del dominio II (Fedorova *et al.*, 2003). Inicialmente se pensó que sólo intervenía en el segundo paso de la reacción de maduración ( $\gamma$ - $\gamma'$ ; Jacquier y Michel, 1990; Jacquier y Jacquesson-Breuleux, 1991; Mikheeva *et al.*, 2000) pero posteriores datos apuntan a que es crítico en todo el proceso (Podar *et al.*, 1998a; Konforti *et al.*, 1998a; [ $\psi$ - $\psi'$ ; Fedorova *et al.*, 2003]). De hecho, parece ser que el N7 del residuo central (G) tiene un papel funcional dentro del sitio activo en el primer paso del *splicing* (de Lencastre y Pyle, 2008). Asimismo, Sigel y colaboradores, regidos por los patrones de hidrólisis de ARN inducida por metales en el intrón aI5 $\gamma$ , concluyeron que esta región supone un sitio de unión específico para iones metálicos divalentes (Sigel *et al.*, 2000), aunque no queda claro si el Mg<sup>2+</sup> participa directamente en la catálisis o sólo facilita la conformación del centro activo (de Lencastre y Pyle, 2008).

---

<sup>11</sup> Se ha observado que el primer nucleótido del intrón aI5 $\gamma$  (G1) se entrecruza con un residuo situado en la región interdominio J<sub>IV/V</sub> (U814).

### 1.1.2. Plegamiento general del intrón: conformación del centro catalítico

Después de enumerar cada uno de los componentes estructurales y los contactos terciarios que los interrelacionan estamos preparados para dilucidar el complejo proceso que tiene lugar hasta la conformación del centro activo de estas grandes ribozimas (excelente revisión Pyle *et al.*, 2007). La gran mayoría de los estudios que han permitido determinar tanto la configuración del centro catalítico como el proceso de plegamiento de estos intrones se han llevado a cabo con el intrón de *S. cerevisiae* aI5y a partir de moléculas que contienen los dominios I, III y V.

En una primera aproximación a lo que sería el proceso de plegamiento de estos intrones, se hicieron estudios en condiciones no fisiológicas con alta concentración de  $Mg^{2+}$ , en ocasiones también de cationes monovalentes como  $K^+$  y elevadas temperaturas<sup>12</sup> (el  $Mg^{2+}$  se encarga de estabilizar los complejos intermediarios). Los intrones del grupo II forman estructuras muy compactas en presencia de  $Mg^{2+}$  de manera que el 40% de la molécula se encuentra aislada del solvente que la rodea (Swisher et al, 2001). El plegamiento de estos intrones pasa por dos etapas diferentes: una primera etapa bastante lenta en la que la molécula no plegada alcanza un estado intermedio donde sólo el dI está colapsado; y una segunda mucho más rápida, en la cual los otros dominios del intrón, que se habrán plegado independientemente, interaccionan con el dI alcanzando la conformación nativa (Su *et al.*, 2003; 2005). Una explicación a la lentitud de la primera etapa, es la existencia en el dI de una subestructura que limite y controle el plegamiento; estudios de mapeo por interferencia con análogos de nucleótidos han permitido situar esta estructura hacia la mitad del dI, concretamente en la región de las interacciones  $\kappa/\zeta$  (Waldsich y Pyle, 2007). A la compactación podrían ayudar regiones adyacentes, como la región de coordinación, aunque la configuración final de este intermediario compactado podría venir estabilizada por las interacciones  $\alpha-\alpha'$  y  $\beta-\beta'$ . La secuencia de acontecimientos podría relatarse siguiendo las siguientes pautas: una vez que se une el  $Mg^{2+}$ , la región  $\kappa/\zeta$  adopta una conformación que induce la curvatura del eje principal del dI. Esta estructura probablemente se extienda y refuerce por el plegamiento específico de las regiones adyacentes; así, las dos mitades del dI se acercan permitiendo que tengan lugar contactos como  $\alpha-\alpha'$  y  $\beta-\beta'$ . Este tipo de interacciones se encargarán de estabilizar definitivamente este andamio funcional que va a suponer el dI y en el que el resto de los dominios del intrón, así como los exones, se van a colapsar (Qin y Pyle, 1997; Su *et al.*, 2005). Así pues, los contactos  $\kappa/\zeta$  presentan una doble actividad en los intrones del grupo II: por un lado, en etapas iniciales en el plegamiento del intrón, actúan de elementos de control del plegamiento, y en las etapas finales, participan como receptor del sitio activo interaccionando con lo que hemos llamado región de unión del dV. Este empaquetamiento posibilita que la región catalítica del dV establezca el resto de

---

<sup>12</sup> Por lo general, las condiciones son 100 mM de  $Mg^{2+}$  y 42°C.

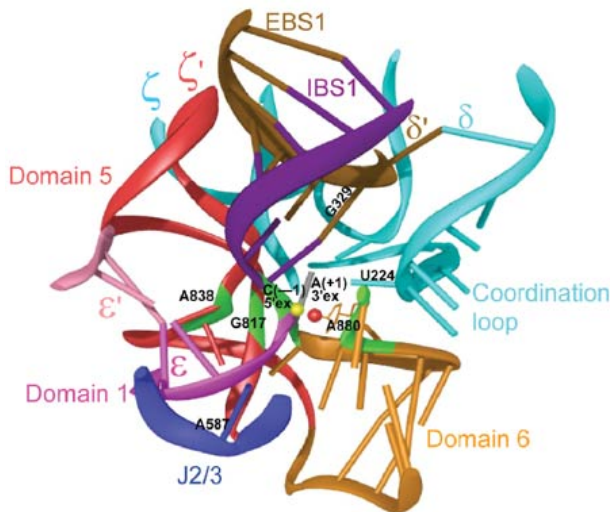
interacciones que configurarán el centro activo de la ribozima ( $\epsilon\text{-}\epsilon'$  y  $\lambda\text{-}\lambda'$ ) (Waldsich y Pyle, 2007). Este mecanismo supone una estrategia para el estricto control de la función de estas ribozimas, ya que no son reactivas hasta que no se pliegan completa y adecuadamente (Pyle *et al.*, 2007).

Aunque los experimentos realizados en condiciones no fisiológicas han permitido interpretar un modelo para el plegamiento de estas ribozimas, es evidente que el interior celular no reúne estas características. Así pues, se realizaron una serie de estudios simulando condiciones cercanas al medio intracelular (30°C y 1 mM  $\text{Mg}^{2+}$ ; Fedorova *et al.*, 2007), de los que se desprende que, aunque el intrón es capaz de plegarse casi completamente (siguiendo una cinética muy lenta), esta conformación es inestable; esto lleva a pensar que *in vivo* deben intervenir otros factores (como proteínas) que agilicen el proceso y estabilicen la ribozima.

En a15y, se han realizado estudios de unión entre moléculas que contienen los dominios I, III y V y un oligonucleótido o sustrato que representa las secuencias IBSs (Qin y Pyle, 1997). Demostraron que la unión del sustrato dependía de la formación de la interacción  $\alpha\text{-}\alpha'$ . Estudios posteriores de mapeo con DMS en el intrón P1LSU/2 de *Pylaiella littoralis* permitieron determinar que la unión del sustrato no sólo depende de la adecuada conformación del dI, sino que la afinidad de unión se incrementa  $\sim 100$  veces si además está adecuadamente situado el dV completo (Costa y Michel, 1999). Esta observación podría tener un significado biológico: el hecho de que se una un posible sustrato antes de la maduración podría dar lugar a que el resto de la estructura del intrón tenga dificultades para formarse, lo que conduciría a un retraso en la producción del ARNm, que directamente se traduce en problemas en la síntesis de proteínas. Esta cadena de acontecimientos supondría una desventaja para la dispersión y el mantenimiento de estos elementos en el genoma de su hospedador (Costa y Michel, 1999).

Una amplia variedad de análisis bioquímicos y minuciosos estudios filogenéticos junto con aproximaciones computacionales han permitido a los investigadores obtener modelos detallados de la arquitectura global de los intrones del grupo II, en especial del centro catalítico (Boudvillain y Pyle, 1998; Costa *et al.*, 2000; Swisher *et al.*, 2001; Noah y Lambowitz, 2003; de Lencastre *et al.*, 2005; Gumbs *et al.*, 2006; de Lencastre *et al.*, 2008). Las regiones directamente implicadas en la catálisis en intrones del grupo II son, además de los dos nucleótidos desapareados y la triada AGC del dV, la región interdominio  $J_{II/III}$ , la interacción  $\epsilon\text{-}\epsilon'$  y la región de coordinación en el dI. Por supuesto, elementos que no pueden faltar en este centro activo son los exones, apareados con los EBSs del dI, y la A desapareada del dVI, indispensable para la primera reacción de maduración el intrón (escisión en forma de lazo, *lariat*). Se ha especulado que entre las dos reacciones de transesterificación podrían ocurrir grandes cambios conformacionales similares a los que se han descrito en el *splicing* de intrones del grupo I, aunque trabajos recientes apuntan hacia un único sitio activo en el que tienen lugar pequeñas reorganizaciones (de Lencastre *et al.*, 2005; Lambowitz y Pyle, 2006). En el caso de aquellos intrones que precisan la

asistencia de su IEP en el proceso de maduración, la unión de esta proteína<sup>13</sup> parece estabilizar determinadas interacciones terciarias críticas en el proceso pero permite que el resto del intrón quede conformacionalmente flexible (Matsuura *et al.*, 2001; Noah y Lambowitz, 2003). Ver figura 4.

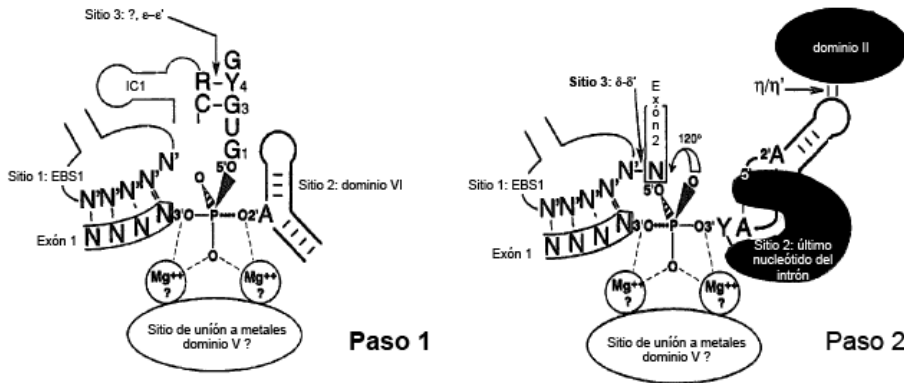


**Figura 4.** Modelo que representa la región del sitio activo del intrón del grupo II a<sub>i</sub>5<sub>y</sub>. Resaltadas en verde se encuentran las regiones que se consideran indispensables: la región desapareada interna del DV, la adenosina desapareada del DVI, los primeros nucleótidos de los exones 1 y 2, y la G de la triada catalítica del DV. Las esferas roja (2'-OH) y amarilla (P) representan los grupos reactivos que participan en la primera reacción de *splicing*, y la línea gris se corresponde con el primer nucleótido del exón 2. En morado se representa el IBS1 apareado con el EBS1 en marrón. Los primeros nts del intrón se distinguen en magenta, incluyendo la posición  $\epsilon$  que aparece con la  $\epsilon'$  en rosa. En azul celeste se distingue la región de coordinación y en azul oscuro se

muestra la región interdominio J<sub>II/III</sub>. El dominio V se representa en rojo y la región proximal del dominio VI en naranja. Figura copiada del de Lencaste *et al.*, 2005.

En paralelo, se ha investigado el papel de los iones  $Mg^{2+}$  en el proceso de plegamiento del intrón así como en su capacidad catalítica (Sontheimer *et al.*, 1999; Sigel *et al.*, 2000; Gordon *et al.*, 2000a y b; Sigel, 2005; Gordon *et al.*, 2007). Parece ser que en el centro activo de estas ribozimas se encuentran dos iones divalentes que se encargarían de activar el nucleófilo en cada una de las reacciones de transesterificación, así como de estabilizar los grupos reactivos en los estados de transición. Aún no se ha determinado el mecanismo preciso por el cual tiene lugar la ligación de exones, pero comienza a cobrar fuerza la teoría de “los dos iones metálicos” propuesta por Steitz y Steitz en 1993 (Steitz y Steitz, 1993; Jacquier, 1996; Sontheimer *et al.*, 1999). Estos iones  $Mg^{2+}$  coordinan el 3'-oxianión de los nucleótidos reactivos situados en el centro catalítico dividido en 3 partes, sin contar con el principal punto de anclaje, el dV: sitios 1, 2 y 3. El sitio 1 estaría situado en el dI cerca de las regiones EBSs y permanecería constante a lo largo de los dos pasos de la reacción. Los otros dos puntos de coordinación van a variar dependiendo de los grupos reactivos necesarios en cada reacción. En el sitio 2 encontraremos la A desapareada del dVI durante la primera reacción de *splicing* y el último nucleótido del intrón en la segunda. En el sitio 3 se situaría el primer nucleótido del intrón para que tenga lugar la primera transesterificación, probablemente influenciado por la interacción  $\epsilon$ - $\epsilon'$ , que se cambiará por el exón 3' durante la segunda reacción. Ver figura 5.

<sup>13</sup> Se ha descrito que LtrA interactúa con su intrón, L1.LtrB, fundamentalmente a través del dIV, pero también se han descrito contactos con los dominios I, II y VI (Matsuura *et al.*, 2001).

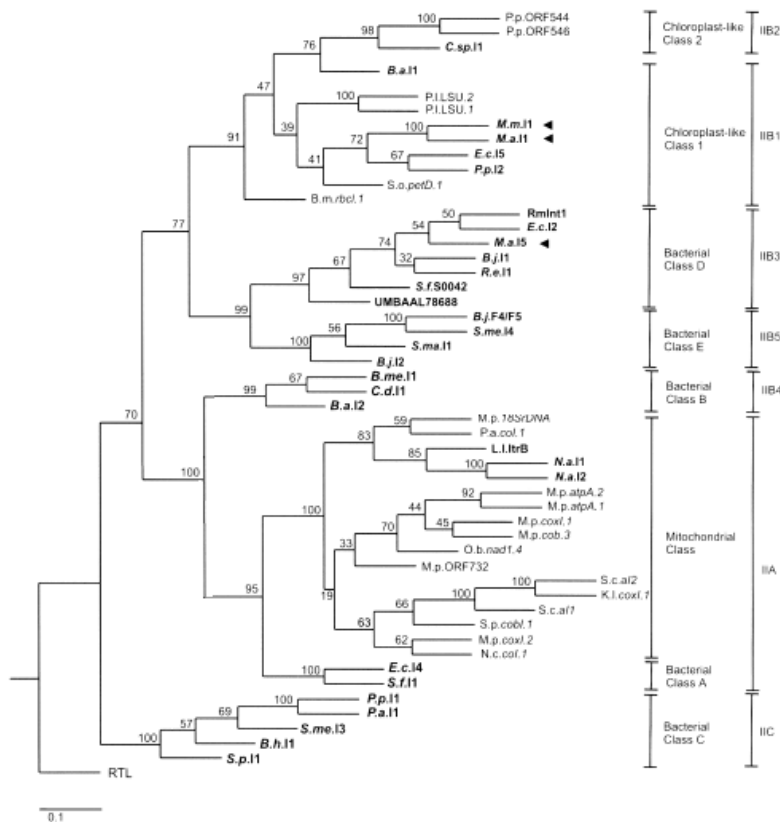


**Figura 5.** Mecanismo de *splicing* de intrones del grupo II. Se representa la constitución del sitio activo durante las dos reacciones de transesterificación que tiene lugar durante el *splicing*. Este esquema está basado en la propuesta de Steitz y Steitz, 1993. Los tres nucleótidos activos en cada paso se identifican bajo los epígrafes sitio 1, 2 y 3. También se muestran los iones  $Mg^{2+}$  coordinados por el dV de la ribozima que participan en la química de reacción. N y N' son nucleótidos que covarian, por tanto, que establecen apareamiento de bases entre ellos. Y representa a pirimidinas (U o C), mientras R se corresponde con purinas (A o G). La flecha marcada con  $120^\circ$  simboliza el cambio de configuración que tiene lugar durante la primera reacción (de Sp a Rp). Figura adaptada de Jacquier, 1996.

### 1.1.3. Clasificación estructural de intrones del grupo II

Desde prácticamente los inicios de las investigaciones con intrones del grupo II, se observaron diferencias estructurales y de secuencia que han llevado a los investigadores a distinguir tres clases de intrones: IIA, IIB y IIC (Toro, 2003; Ferat *et al.*, 2003; Toor *et al.*, 2006). Inicialmente, sólo se hizo la distinción entre subgrupos IIA y IIB (Michel *et al.*, 1982; Schmelzer *et al.*, 1983), subdividiéndolos a su vez en IIA1, IIA2, IIB1 y IIB2, atendiendo básicamente a detalles estructurales (Michel *et al.*, 1989). Entre esas diferencias se pueden destacar: (i) la región desapareada interna donde se encuentran las interacciones  $e'$  y  $\lambda$  es de  $\sim 14$  nt en el caso de los intrones IIA y de 4-5 en los IIB; (ii) en los intrones IIA encontramos una adenosina desapareada 7 nt antes del final del intrón (excepcionalmente, 8 nt), mientras que en los IIB suele ser 8 nt; (iii) la región donde se encuentran una de las señales de reconocimiento de los exones (EBS1, Exon Binding Site) suele ser asimétrica en el caso de los IIA, sin embargo, los intrones IIB conservan cierta simetría; (iv) la región donde se localiza la otra secuencia para el reconocimiento del exón 5' (EBS2) es diferente en los dos casos y está en relación con la longitud de la región distal del dominio ID; (v) el tamaño de la región desapareada del dominio III varía entre los dos subgrupos, así como la longitud de las regiones interdominio comprendidas entre los dominios I al IV; (vi) por último, los intrones IIA terminan con una secuencia consenso YAY mientras que los IIB lo hacen con RAY. La estructura A2 se diferencia de la A1 principalmente porque presenta una inserción entre los dominios I y II, mientras que la B2 difiere de la B1 en una inserción al inicio del intrón, justo en la región opuesta a la inserción que se encuentra en los intrones IIA2 (Michel *et al.*, 1989; Toor *et al.*, 2001).

Con posterioridad los investigadores se percataron de que estas diferencias no eran sólo estructurales sino que se traducen en diferente comportamiento catalítico. En reacciones de auto-maduración llevadas a cabo *in vitro*, se ha determinado que los requerimientos de  $Mg^{2+}$  así como de cationes monovalentes son mayores en el caso de los IIA que, por lo general, suelen ser menos reactivos<sup>14</sup> (Lehmann y Schmidt, 2003); por otro lado, la incidencia de reapertura de los exones<sup>15</sup> una vez ligados es mayor (casi exclusivo) en IIB que en IIA (Schmidt *et al.*, 1990; Hebbar *et al.*, 1992). Además, estos dos subgrupos tienen en común el mecanismo por el cual determinan el sitio de maduración 5', pero difieren en la selección del sitio de maduración 3' (ver interacciones EBS3-IBS3 y  $\delta$ - $\delta'$  en apartado 1.1.1.1. Dominio I). Ver figura 6.



**Figura 6.** Relaciones filogenéticas en intrones del grupo II. Aunque la filogenia representada está basada en el alineamiento del dominio RT de las IEP, se puede observar una correspondencia con las distintas clases estructurales: IIA, IIB y IIC. Figura adaptada de Toro, 2003 (Fig. 2, ver detalles en el pie de la figura original).

<sup>14</sup> Esta afirmación deriva de comparaciones entre intrones IIB que no codifican ORF (~ 1 kb) con intrones IIA que sí lo hacen (~ 2,5 kb). Para conseguir equiparar las eficiencias de auto-maduración, lo que se hizo fue eliminar la mayor parte de esa secuencia, trabajándose con intrones de tamaño similar.

<sup>15</sup> La reapertura de los exones (SER, Spliced Exons Reopening) es una reacción que tiene lugar *in vitro* (aún no se ha determinado *in vivo*), en la cual el propio intrón escindido cataliza el ataque hidrolítico en el punto de ligación de los exones dando lugar a la liberación de los exones (Jarrell *et al.*, 1988b; Daniels *et al.*, 1996).

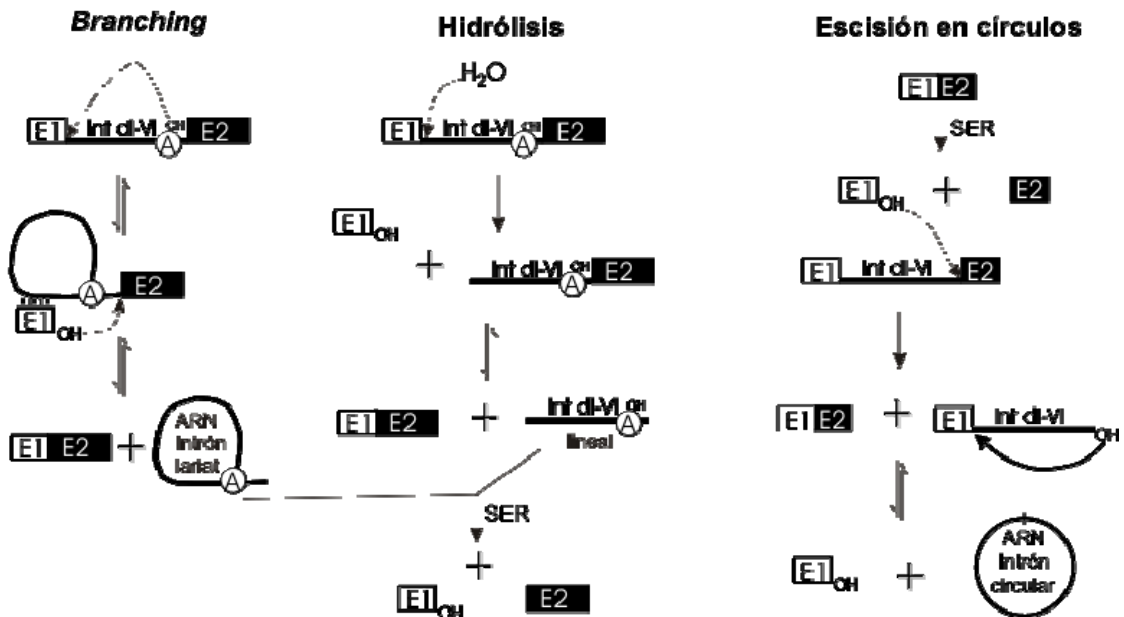


Más recientemente se han sumado a esta clasificación los intrones IIC (Grandlund *et al.*, 2001). Estos intrones se caracterizan por no presentar EBS2 y tener una secuencia EBS1 de menor extensión que los IIA y IIB (3-4 nt en vez de 6 nt). Además suelen encontrarse insertados aguas abajo de terminadores de transcripción independientes de  $\rho$ , estructuras en horquilla que se piensa podrían conceder parte de la especificidad perdida con el EBS2 para el reconocimiento de una nueva localización donde insertarse (Martínez-Abarca y Toro, 2000a; Grandlund *et al.*, 2001; Toor *et al.*, 2006; Michel *et al.*, 2007). Además se han descrito otra serie de diferencias estructurales: (i) el dominio V es más corto que en IIA y IIB, y en lugar de la triada catalítica característica, AGC, presentan los residuos CGC; y, (ii) la región donde se describe la interacción  $\epsilon'$  tiene 11 nt y conserva el motivo AGGAA.

## 1.2. MECANISMOS DE ESCISIÓN EN INTRONES DEL GRUPO II

Los intrones del grupo II se han considerado un excelente sistema modelo para el estudio de la reacción de maduración (Jacquier, 1990) y, en general, para el estudio de las ribozimas porque llevan a cabo una amplia variedad de reacciones químicas (Pyle y Lambowitz, 2006). Aunque los estudios *in vivo* han aportado información relevante de la biología y el mecanismo de reacción de estos intrones, los resultados más esclarecedores proceden del análisis de su actividad catalítica *in vitro*. A pesar del gran número de intrones conocidos, sólo unos pocos muestran capacidad de automadurar a partir del pre-ARNm *in vitro* (Lehmann y Schmidt, 2003; Toro *et al.*, 2007). En estos experimentos, las ribozimas deben someterse a condiciones no fisiológicas, con alta concentración de sales y elevadas temperaturas, que le van a permitir desarrollar una actividad catalítica eficiente. Por esta razón, muchos autores coinciden en que, incluso los intrones que son capaces de madurar eficientemente *in vitro*, dependen de proteínas para llevar a cabo el *splicing* de una manera eficaz *in vivo* (Lambowitz y Perlman, 1990). El potencial catalítico se localiza claramente en el ARN, mientras que las proteínas podrían encargarse de ayudar al plegamiento o estabilizar la conformación activa de la ribozima (Lehmann y Schmidt, 2003). En ningún caso, los datos disponibles sugieren diferencias químicas o mecánicas entre la reactividad *in vitro* e *in vivo* una vez que el intrón se ha plegado adecuadamente (Fedorova y Zingler, 2007).

El mecanismo de *splicing* descrito originalmente, más complejo y, a su vez, mejor estudiado se conoce como mecanismo de *branching* (Peebles *et al.*, 1986; van der Veen *et al.*, 1986). Cuando el intrón madura siguiendo esta ruta, se libera en forma de ARN lazo o *lariat*; esta configuración le va a proporcionar mayor estabilidad haciéndolo menos susceptible a degradación (Daniels *et al.*, 1996; Conklin *et al.*, 2005) y fijando la estructura tridimensional de la ribozima (Jacquier y Jacquesson-Breuleux, 1991). Esta forma de maduración es el motivo principal por el considera que los intrones del grupo II podrían ser los precursores de los intrones nucleares eucarióticos; a este hecho, se han ido sumando con el tiempo otra serie de observaciones que afianzan esta hipótesis como veremos en el apartado 4.1. de esta memoria. Ver figura 7.



**Figura 7.** Mecanismos de escisión descritos en intrones del grupo II. Los exones están representados por recuadros blancos (exón 1, E1) y negros (exón 2, E2) y el ARN del intrón se muestra como una línea. La flecha en los dos sentidos refleja la reversibilidad de algunas reacciones. Con flechas discontinuas hacemos referencia a los ataques nucleofílicos que tienen lugar en las diferentes rutas.

Como hemos mencionado anteriormente, la maduración de los intrones es vital para la viabilidad celular. Por eso, los intrones del grupo II han buscado rutas alternativas para llevar a cabo este proceso y asegurar así su supervivencia en los genomas que invaden (Daniels *et al.*, 1996; Chu *et al.*, 1998; Pyle y Lambowitz, 2006). Una de estas vías alternativas de maduración conlleva la liberación del ARN del intrón en forma lineal y se conoce como mecanismo de hidrólisis (van der Veen *et al.*, 1987a; Jarrell *et al.*, 1988b). Esta forma de intrón escindido limita el potencial infectivo de estos elementos: por un lado, es más fácilmente accesible a la digestión con ARNasas, pero lo más importante es que, al contrario que las transesterificaciones que tienen lugar durante la ruta de *branching*, la hidrólisis es una reacción irreversible, lo que va a impedir que el intrón lineal sea capaz de invadir una nueva diana (Daniels *et al.*, 1996; Chu *et al.*, 1998; Podar *et al.*, 1998b; D'Souza y Zhong, 2002).

Finalmente, existe una tercera vía por la cual los intrones del grupo II pueden escindirse de sus ARNm precursores, pero está mucho menos estudiada. Se trata de la ruta de escisión en forma de círculos (Murray *et al.*, 2001). Son pocos los intrones del grupo II en los que se ha demostrado que maduran siguiendo esta vía (Murray *et al.*, 2001; Vogel y Börner, 2002; Li-Pook-Than y Bonen, 2006; Molina-Sánchez *et al.*, 2006), por lo que existen muchas incógnitas en torno a este mecanismo de escisión: en qué consiste, las características de los productos generados, qué regiones del intrón intervienen,... y, uno de los principales puntos clave, cuál es su significado biológico.

### 1.2.1. Mecanismo de *branching* o de formación de intrón *lariat*

Los intrones del grupo II se escinden de los exones que los flanquean gracias a un proceso en dos pasos catalizado por el ARN del propio intrón. Consiste en dos reacciones sucesivas de transferencia de grupos fosfato en cada uno de los sitios de maduración, induciendo la liberación del intrón y la unión de los exones (Peebles *et al.*, 1986; van der Veen *et al.*, 1986; Schmelzer y Schweyen, 1986). El nucleófilo que actúa durante la primera transesterificación es el 2'-OH de la A desapareada del dVI de la ribozima, residuo que también se conoce como punto/sitio de ramificación o de *branch* (van der Veen *et al.*, 1986; 1987a). Este hidroxilo ataca el grupo fosfato presente en el 5'-ss, quedando la A unida covalentemente al primer nucleótido del intrón mediante un enlace 2'-5'. En este punto, el intrón todavía se encuentra unido al exón 2 y el exón 1 permanece apareado con el ARN del intrón (mediante las interacciones EBS-IBS) con un extremo 3'-OH libre (Jacquier y Michel, 1987; Jacquier y Jacquesson-Breuleux, 1991). Este hidroxilo libre del exón 1 será el nucleófilo de la segunda reacción, atacando el fosfato en el 3'-ss. Así, ambos exones quedan unidos covalentemente mediante un enlace 3'-OH/5'-P y el intrón se libera en forma de lazo o *lariat*. Esta particular forma de escisión es precisamente la que caracteriza la principal ruta de maduración de los intrones del grupo II y la que los relaciona con los intrones nucleares eucariotas (Padgett *et al.*, 1984; Ruskin *et al.*, 1984).

Inicialmente, datos de microscopía electrónica hacían pensar que el intrón se liberaba en forma de círculos verdaderos donde el principio y el final del intrón quedaban unidos mediante un enlace covalente (Arnberg *et al.*, 1980); incluso, se llegaron a identificar formas intermediarias del proceso de maduración en las que el intrón aún permanece unido un exón 2. Aunque la relación entre estos “círculos” y la reacción de maduración de estos intrones era en esos momentos una incógnita, parecía evidente que el enlace fosfodiéster que cerraba estas moléculas no era una unión convencional 3'-5' (van der Veen *et al.*, 1986). El trabajo de Peebles y colaboradores determinaba la naturaleza de este enlace, describiéndola como una unión 2'-5' que probablemente tuviese lugar entre el primer nucleótido del intrón y algún residuo cercano al extremo 3' del intrón (Peebles *et al.*, 1986). En un estudio prácticamente paralelo, van der Veen y colaboradores señalan a la A desapareada (y fuertemente conservada entre los distintos intrones del grupo II) como la responsable del desencadenamiento de las reacciones de maduración (van der Veen *et al.*, 1986; 1987a). La particularidad de este nucleótido en relación con la primera reacción de transesterificación en el proceso de *splicing*, lo ha convertido en uno de los residuos más y mejor estudiados en estos intrones.

Como en muchos otros aspectos relacionados con estos retroelementos, los estudios *in vitro* han permitido vislumbrar el posible mecanismo químico de las reacciones de *splicing*. Se trata de reacciones de sustitución nucleofílica del tipo S<sub>N</sub>2 (Padgett *et al.*, 1994; Podar *et al.*, 1995b; 1998c; Jacquier, 1996). Los dos oxígenos que no participan directamente en el enlace fosfodiéster no son estereoquímicamente equivalentes y, para distinguirlos, se

denominan pro-R (Rp) y pro-S (Sp). La sustitución de uno u otro de estos oxígenos por azufre ha permitido definir el curso estereoquímico de la reacción, llegando a la conclusión de que ambos pasos de la reacción de *splicing* muestran la misma estereoespecificidad (Padgett *et al.*, 1994). Así, la sustitución del oxígeno pro-R bloquea las dos reacciones de transesterificación. Sin embargo, la sustitución del oxígeno pro-S tiene como consecuencia una reducción sustancial de la actividad, comportándose de modo diferencial dependiendo del paso del *splicing* de que se trate, esto es, el efecto es más acusado en la primera que en la segunda reacción de transesterificación (Podar *et al.*, 1995b; 1998c). De manera que se ha propuesto que los dos oxígenos, pro-R y pro-S, del fosfato en el 5' ss se encuentran interaccionando con el sitio activo de la ribozima, aunque estas interacciones en ningún caso serían equivalentes<sup>16</sup>; en el caso del 3' ss, sólo el oxígeno pro-R interaccionaría con la ribozima. Además se observa una inversión de la configuración del fosfato en las dos reacciones, esto es, el oxígeno pro-S cambia su configuración a pro-R. El papel de estos oxígenos se desconoce aunque lo más probable es que intervengan en la coordinación de iones  $Mg^{2+}$  (Padgett *et al.*, 1994; Podar *et al.*, 1995b; 1998c). Estas diferencias en el mecanismo indujeron a pensar que podríamos estar hablando de dos centros activos diferentes, pero trabajos posteriores apuntan a la existencia de un único centro activo en el que tendrían lugar algunas reorganizaciones entre los dos pasos del *splicing* (ver apartado 1.1.2. Configuración del sitio catalítico; Chanfreau y Jacquier, 1994; Steitz y Steitz, 1993; Jacquier, 1996; Sontheimer *et al.*, 1999; de Lencastre *et al.*, 2005).

Estudios cinéticos *in vitro* han permitido determinar que la primera reacción de transesterificación limita la velocidad de maduración de estos intrones (Jacquier y Jacquesson-Breuleux, 1991). Algunos autores apuntan al pH como un factor clave en el proceso, de manera que suponen que un cambio conformacional relacionado directamente con la concentración de  $H^+$  en el medio es el responsable de la limitación en la primera reacción de transesterificación (Podar *et al.*, 1998c). La segunda reacción es extraordinariamente rápida (Chin y Pyle, 1995). En cualquier caso, ambas reacciones de transesterificación son reversibles (Augustin *et al.*, 1990; Mörl y Schmelzer, 1990a) hasta el punto que la velocidad de la primera reacción es tan rápida como la de su reversa (Chin y Pyle, 1995) mientras que la segunda reacción es mucho más rápida que su reversa (Jacquier y Jacquesson-Breuleux, 1991; Aizawa *et al.*, 2003). Estas observaciones podrían dar sentido a que la segunda reacción sea mucho más rápida que la primera: el que la reacción se lleve a término previene que las moléculas que han llevado a cabo con éxito la primera reacción de transesterificación no reviertan el proceso (Chin y Pyle, 1995). Otro dato importante es que estas transesterificaciones no requieren de energía externa para llevarse a cabo (Peebles *et al.*, 1986; Jacquier, 1996), sólo es necesario la activación de los correspondientes nucleófilos (van der Veen *et al.*, 1987a).

---

<sup>16</sup> Se ha postulado que el oxígeno pro-S podría participar en la interacción con iones divalentes previo a la catálisis, bien acelerando la consecución de la conformación activa de la ribozima o bien estabilizando esta conformación. Mientras, el oxígeno Rp participa directamente en la reacción química (Podar *et al.*, 1995b).

### 1.2.2. Mecanismo de hidrólisis o de formación de intrón lineal

Diversos estudios de maduración *in vitro* mostraban la existencia de dos bandas de diferente movilidad electroforética derivadas de intrón escindido. En un principio se pensó que se trataba de formas degradadas de intrón escindido *lariat* pero enseguida se demostró que se trataba de formas lineales de intrón (van der Veen *et al.*, 1987a; Schmidt *et al.*, 1987; Jarrell *et al.*, 1988; Daniels *et al.*, 1996). Así, se propuso la existencia de una ruta alternativa de *splicing* que generaba como productos finales intrón lineal y exones ligados (Jarrell *et al.*, 1988; Daniels *et al.*, 1996). Esta vía de maduración consiste en la rotura por hidrólisis del 5'ss debido a la presencia en el sitio activo de la ribozima de moléculas de agua o de iones hidroxilo. El *splicing* se completa con una segunda reacción de transesterificación idéntica a la que tiene lugar por la ruta de *branching*. No hay razones para pensar que ambas rutas tengan lugar en distinto sitio activo (Daniels *et al.*, 1996), sobre todo si tenemos en cuenta que la hidrólisis conserva la misma estereoespecificidad que la primera reacción de la ruta de *branching* (Podar *et al.*, 1995b).

La existencia de esta ruta alternativa podría tener implicaciones interesantes. La más importante deriva de la naturaleza irreversible de la reacción de hidrólisis: podría funcionar regulando la movilidad de estos intrones. Los eventos de movilidad asociados a estos retroelementos dependen de la reversión de la reacción de maduración de las formas *lariat* sobre nuevos sustratos de ADN (Michel y Ferat, 1995; Bonen y Vogel, 2001; Belfort *et al.*, 2002); por tanto, si en el mecanismo hidrolítico la primera reacción es irreversible, los intrones que sólo puedan escindirse en forma lineal serán incapaces de transponerse (Daniels *et al.*, 1996; Podar *et al.*, 1998b; Lehmann y Schmidt, 2003). Otro posible papel de estas formas escindidas lineales podría estar relacionado con un aumento en la velocidad de *splicing* derivada de una eficiente traducción de la madurasa que algunos de estos intrones codifican. Parece ser que las formas lineales presentan menos dificultades para ser traducidas que las formas circulares o en lazo, por lo que los intrones resultantes de la vía hidrolítica suponen un intermediario potencial para la producción en serie de proteína; estas madurasas van a participar en la maduración más rápida y eficiente del intrón a partir de los pre-ARNm (Daniels *et al.*, 1996; Lehmann y Schmidt, 2003).

Desde el punto de vista cinético, los resultados extraídos de los ensayos de *splicing in vitro*, muestran la existencia de dos conformaciones del ARN precursor (Chin y Pyle, 1995; Daniels *et al.*, 1996). Una población de moléculas, A, que reacciona de forma rápida siguiendo las dos vías descritas, pero con una marcada preferencia por la formación de *lariat*. La población B tarda más en reaccionar, usa también las dos vías pero, en esta ocasión, prefiere la hidrólisis. Generalmente, estas reacciones *in vitro* se desarrollan en presencia de alta concentración de iones en el medio (divalentes, pero fundamentalmente monovalentes<sup>17</sup>) porque se sabe que una elevada cantidad de cationes posibilita el

---

<sup>17</sup> En general, los iones Mg<sup>2+</sup> son indispensables tanto para el plegamiento como para la catálisis y suelen usarse a concentración 100mM. Sin embargo, se habla de condiciones de reacción con alta concentración de sal (HS; 0,5-1,5M) o con baja concentración de sal (LS; 100mM) dependiendo de la cantidad de iones

adecuado plegamiento del ARN. Así la eficiencia de reacción de la población A depende de estas condiciones de alta concentración de sales, sin embargo, la población B reacciona eficientemente aún en condiciones de concentración de sales similares a las fisiológicas (20mM MgCl<sub>2</sub> y 140mM KCl) (Daniels *et al.*, 1996). La reactividad de los dos confórmeros también se ha relacionado con el pH del medio de reacción (Chin y Pyle, 1995). Finalmente decir que se han descrito intrones del grupo II que sólo son capaces de madurar por la vía hidrolítica *in vitro* (GBSi1 en *Streptococcus*, Granlund *et al.*, 2001).

Las dos rutas pueden darse *in vitro* aunque el *branching* parece más eficiente que la hidrólisis<sup>18</sup> (Jarrell *et al.*, 1988; Daniels *et al.*, 1996), por lo que se especuló que la formación de intrón lineal pudiera ser un artefacto de las condiciones a las que se sometía el precursor *in vitro*. Poco tiempo después, quedó demostrado que ambas rutas coexisten también *in vivo* (Podar *et al.*, 1998b; Vogel y Börner, 2002). Además, se han descrito intrones que carecen de la A del dVI en el gen del ARNt<sup>Val</sup> en el cloroplasto de varios organismos (*Nicotiana tobacum*, *Marchantia polymorpha*, *Zea mays* y *Hordeum vulgare*; Michel *et al.*, 1989). Concretamente, el intrón presente en el gen *trnV* (ARNt<sup>Val</sup>) de cloroplastos de cebada se ha demostrado que se escinde únicamente vía hidrólisis *in vivo* (Vogel y Börner, 2002).

### 1.2.3. Mecanismo de escisión de intrón circular

Además de los dos anteriores mecanismos muy bien estudiados, se ha descrito una tercera vía, probablemente minoritaria, por la que los intrones del grupo II pueden madurar: escisión como círculos (en intrones mitocondriales de levadura, Murray *et al.*, 2001; en intrones de mitocondrias de trigo, Li-Pook-Than y Bonen, 2006; en intrones bacterianos, Molina-Sánchez *et al.*, 2006). Estos círculos estarían formados por moléculas en las que el primer nucleótido del intrón se une covalentemente mediante un enlace fosfodiéster 2'-5' al último nucleótido del intrón. Esta forma de escisión tendría lugar tanto *in vitro* (Murray *et al.*, 2001; Costa *et al.*, 2006a) como *in vivo* (Murray *et al.*, 2001; Li-Pook-Than y Bonen, 2006; Molina-Sánchez *et al.*, 2006). El mecanismo de reacción más sencillo sería pensar que, debido a la proximidad en el centro activo de la ribozima de los dos sitios de maduración, el 2'OH del último nucleótido del intrón podría atacar el fosfato del primer nucleótido del intrón. El problema es explicar el que no se hayan encontrado intermediarios intrón circular unido a exón 2 (Murray *et al.*, 2001). Así, el mecanismo de reacción propuesto precisa que el exón 2 sea liberado primero, de modo que el 2'OH del último residuo del intrón ataca el 5'ss originando el intrón escindido circular (Murray *et al.*, 2001). Aún no está muy claro la naturaleza del nucleófilo que ataca

---

monovalentes presentes en el medio. Por otro lado, es importante resaltar que existe una dependencia de la ruta de maduración seleccionada *in vitro* en función del ión monovalente en reacción; en general, KCl produce una mayor cantidad de productos escindidos en forma lineal; (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> desplaza la preferencia por la ruta de *branching*, y otras sales como (NH<sub>4</sub>)Cl, generan la misma cantidad de ambas formas (Lehmann y Schmidt, 2003).

<sup>18</sup> Dependiendo de las condiciones de reacción pero también de la subclase de intrón del grupo II.

el 3' ss y si, finalmente, se generan o no exones ligados. Murray y colaboradores proponen que sería exón 1 libre presente en el medio el que atacaría la unión intrón 3'/exón 2 en el precursor (Murray *et al.*, 2001). Por otro lado, se ha postulado que secuencias externas puedan participar en este ataque nucleofílico, pero sin generar ligación de exones, o también que podrían ser el resultado de la circularización de moléculas escindidas por la vía hidrolítica gracias a ARN ligasas celulares (Li-Pook-Than y Bonen, 2006). En cualquier caso, es una reacción catalizada por la ribozima y cuyo significado biológico sigue siendo un enigma (Murray *et al.*, 2001). De lo que no se tiene ninguna duda es que supone un mecanismo para estabilizar el ARN del intrón (Pyle y Lambowitz, 2006).

Además de las moléculas de ARN circular, también se han detectado ADN circular que contiene el intrón pero se desconoce su procedencia y función (Murray *et al.*, 2001; Molina-Sánchez *et al.*, 2006). Se ha especulado que podría ser el resultado de la reverso transcripción de las moléculas de ARN circular llevado a cabo por la IEP. En *Podospora anserina*, la presencia de ADN circular derivado de intrones del grupo II se ha asociado a senescencia celular (Osiewacz y Esser, 1984; Schmidt *et al.*, 1994; Begel *et al.*, 1999), pero no se ha demostrado que sea éste el caso en ninguno de los intrones que maduran vía círculos<sup>19</sup>.

#### 1.2.4. Otras reacciones asociadas a la maduración de intrones del grupo II

Los intrones del grupo II son ribozimas muy versátiles, ya que pueden usar durante la catálisis varios tipos de nucleófilos (2'-OH, 3'-OH, hidroxilos libres o agua) así como de sustratos (son las únicas ribozimas capaces de usar con similar eficiencia sustratos ARN o ADN) (Lehmann y Schmidt, 2003; Pyle y Lambowitz, 2006). Además de las que ya hemos nombrado, se han descrito otra serie de reacciones, asociadas o no con los procesos más estudiados de estos retroelementos: maduración y movilidad. En estas reacciones, los intrones funcionan como verdaderas ribozimas: actúan en *trans* sobre un sustrato sin sufrir ningún tipo de modificación y pueden repetir múltiples ciclos (Lehmann y Schmidt, 2003). Entre las reacciones no asociadas a *splicing* y movimiento podemos citar: ARN recombinasa (Mörl y Schmelzer, 1990b), Terminal Transferasa (Mueller *et al.*, 1991), 3'-5' ARN polimerasa (Mueller *et al.*, 1993a) y ADN polimerasa (Hetzer *et al.*, 1997b).

La reacción más importante por su significación biológica, descrita sólo *in vitro*, asociada con el proceso de maduración de intrones del grupo II responde al acrónimo SER (Spliced Exons Reopening) (Peebles *et al.*, 1987; Jarrell *et al.*, 1988; Podar *et al.*, 1995b; Daniels *et al.*, 1996). Una vez producido el *splicing*, el intrón (tanto en su forma lineal como *lariat*) reconoce los exones ligados y los hidroliza exactamente en el punto de

---

<sup>19</sup> El modelo propuesto para la formación de círculos en *Podospora anserina* es muy diferente. Implica, en primer lugar, que deben existir dos copias de intrón en tándem en el ADN mitocondrial del hongo. La escisión de una de las copias por recombinación homóloga generaría la molécula de ADN circular (Mueller *et al.*, 1993b; Sellem *et al.*, 1993; Sainsard-Chanet *et al.*, 1994).

ligación. Inicialmente se pensó que podría estar acoplada con la vía de hidrólisis puesto que su eficiencia parece mayor en presencia de KCl (Jarrell *et al.*, 1988), pero datos posteriores apuntan a que no existe tal acoplamiento (Daniels *et al.*, 1996). Atendiendo al mecanismo de reacción (hidrólisis), no se puede decir que esta reapertura de exones sea la reacción reversa de la segunda reacción de maduración (transesterificación), pero llama la atención que la estereoespecificidad de estas dos reacciones es inversa, como correspondería a una reacción reversible; es decir, la reacción SER, al contrario que la segunda reacción de transesterificación en *branching* o hidrólisis, se bloquea si sustituimos el oxígeno Sp por azufre pero sigue adelante con una sustitución en el oxígeno pro-R (Daniels *et al.*, 1996). Esta reacción parece ser específica de intrones del grupo IIB (Lehmann y Schmidt, 2003; Fedorova y Zingler, 2007) y, como hemos dicho, no se tiene constancia de que tenga lugar *in vivo*; de cualquier modo, su relevancia es fundamental, puesto que hace disminuir la proporción de exones ligados casi invalidando uno de los principios de la reacción de *splicing*: la ligación de exones (Pyle y Lambowitz, 2006). Algunos autores han postulado que el exón 1 libre generado en esta reacción podría estar implicado en la escisión del intrón como ARN circular (Murray *et al.*, 2001).

Dada la adaptabilidad y diversidad de intrones del grupo II, no se descarta que se puedan seguir describiendo nuevas actividades asociadas a estos elementos genéticos móviles (Pyle y Lambowitz, 2006).

### 1.3. ESCISIÓN *IN VIVO* DE INTRONES DEL GRUPO II ASISTIDA POR PROTEÍNAS

Como mencionamos anteriormente, aunque algunos intrones son capaces de madurar *in vitro* en condiciones no fisiológicas, todos ellos precisan proteínas para madurar eficientemente *in vivo* (Lambowitz y Belfort, 1993; Lambowitz *et al.*, 1999; Lambowitz y Zimmerly, 2004; Pyle y Lambowitz, 2006). Los mecanismos de maduración del ARN intrónico son los descritos en apartados anteriores, sólo que *in vivo* necesitan proteínas que ayudarían al correcto plegamiento y la estabilización de la estructura catalítica activa de la ribozima (Lehmann y Schmidt, 2003; Pyle y Lambowitz, 2006). La proteína que más comúnmente participa en la reacción de maduración de intrones del grupo II es aquella que está codificada por el propio intrón, IEP (Lambowitz y Zimmerly, 2004). Sin embargo, se considera que otra serie de proteínas codificadas en el genoma nuclear del hospedador podrían participar en el proceso de *splicing* de estos intrones en levaduras, algas y plantas superiores (Lambowitz *et al.*, 1999; Lehmann y Schmidt, 2003).

#### 1.3.1. Proteínas codificadas por el intrón, IEPs

Estos particulares factores de *splicing* suelen presentar cuatro dominios claramente diferenciados. Tres de ellos, los dominios reverso transcriptasa (RT), de unión a ADN (D) y ADN endonucleasa (En), están directamente implicados en el proceso de movilidad de estos intrones mientras un cuarto dominio se ha ligado con el proceso de maduración,



el dominio madurasa o dominio X (Mohr *et al.*, 1993). Por la temática que nos ocupa, en este apartado nos centraremos en este último dominio y en un capítulo posterior hablaremos extensamente de los tres restantes (ver apartado 2.1.).

La presencia del dominio madurasa es universal entre aquellos intrones que codifican proteínas, de manera que existen IEPs que carecen de dominio En, D o que tienen un dominio RT muy degenerado, pero no se ha encontrado ninguna en la que el dominio X esté ausente (Mohr *et al.*, 1993). El primer indicio de la existencia de este dominio en intrones del grupo II provino de la caracterización genética de mutaciones en la proteína que tenían como consecuencia defectos en *splicing* en intrones de levadura (aI1 y aI2), y que eran compensados al proporcionar en *trans* la proteína salvaje (Carignani *et al.*, 1983; Moran *et al.*, 1994). Estudios bioquímicos posteriores con la proteína del intrón Ll.ltrB reafirman estos resultados: LtrA purificada promueve (en ausencia de ATP) la maduración del intrón Ll.ltrB *in vitro* en condiciones cercanas a las fisiológicas (5mM Mg<sup>2+</sup>) en las que el intrón no es capaz de auto-escindir-se (Matsuura *et al.*, 1997; Saldanha *et al.*, 1999). Derivado de su implicación en el proceso de movilidad, no sorprende que la delección de los dominios D y En no afecte en absoluto la eficiencia de la reacción de maduración *in vivo* (Zimmerly *et al.*, 1995a; 1995b; Matsuura *et al.*, 1997). Bien es cierto que el dominio RT parece estar ligado en cierta medida con el *splicing*, ya que mutaciones en esta región reducen la actividad sin llegar en ningún caso a inhibir completamente el *splicing* (Moran *et al.*, 1994; 1995; Zimmerly *et al.*, 1995b).

Al igual que observamos en los estudios de plegamiento de la ribozima en condiciones de laboratorio, la interacción ARN↔IEP se produce de manera secuencial. La mayoría de los resultados a este respecto han sido obtenidos para Ll.ltrB/LtrA, gracias al desarrollo de un sistema de expresión heteróloga en *E. coli* que facilita enormemente el análisis bioquímico (Matsuura *et al.*, 1997). La IEP, gracias a la región N-terminal del dominio RT y al dominio X, se une inicialmente a una estructura presente al comienzo del dominio IV de la ribozima, designada como dominio IVa (Wank *et al.*, 1999; Singh *et al.*, 2002; Watanabe y Lambowitz, 2004; Cui *et al.*, 2004). Esta región contiene la secuencia Shine-Dalgarno y el sitio de iniciación de la traducción de LtrA, por lo que, al impedir que el ribosoma acceda a su secuencia de reconocimiento, la proteína es capaz de auto-regular su producción (Singh *et al.*, 2002). La unión de la proteína tiene lugar en forma de dímero ( $K_d$  a 30°C ≤ 0,12 pM) y parece inducir algunos cambios conformacionales en el ARN del intrón así como en la propia madurasa; es más, se piensa que en solución la proteína puede ser un monómero y que dimeriza cuando se une al ARN (Saldanha *et al.*, 1999; Rambo y Doudna, 2004). Una vez que se ha unido al dVIa, LtrA establece contactos débiles, pero indispensables, con regiones conservadas del centro catalítico (dominios I, II y VI) que permiten un *splicing* residual *in vivo* e *in vitro* cuando se deleciona completamente el dominio IV (Wank *et al.*, 1999; Matsuura *et al.*, 2001; Singh *et al.*, 2002; Huang *et al.*, 2003; Rambo y Doudna, 2004; Cui *et al.*, 2004).

Esta extensa región de unión a ARN reconoce caracteres estructurales únicos de su propio precursor no maduro y proporciona a ese ARN las condiciones óptimas en su

plegamiento activo, por lo que se dice que son factores de maduración altamente específicos (Lambowitz *et al.*, 1999; Matsuura *et al.*, 1997; Noah y Lambowitz, 2003). Sin embargo, se ha sugerido que, en algunos organismos, estas madurasas podrían participar en el *splicing* de varios intrones. Es el caso de una proteína codificada por un intrón del grupo III en el cloroplasto de *Euglena*: se considera que está implicada en la maduración de otros intrones del mismo tipo presentes en el organismo (Copertino *et al.*, 1994; Doestch *et al.*, 1998). Otro factor que respalda la observación anterior es que la inhibición de la síntesis de proteínas en el cloroplasto de algunas gramíneas bloquea el *splicing* de varios intrones del grupo II, incluso de aquellos que no codifican su propia proteína (Hess *et al.*, 1994; Hübschmann *et al.*, 1996; Jenkins *et al.*, 1997; Vogel *et al.*, 1997). Una publicación reciente pone de manifiesto que una IEP de un intrón procedente de *Trichodesmium erythraeum* funciona como madurasa para varios intrones de diferentes genes de la misma especie expresados en *E. coli* (Meng *et al.*, 2005). Asimismo, se ha descrito una IEP degenerada (conservan sólo el dominio X y parte del dominio RT), MatK<sup>20</sup>, que facilita el *splicing* de intrones presentes en cloroplastos de plantas superiores (Vogel *et al.*, 1999). Además se ha demostrado que LtrA puede unirse, en determinadas condiciones (100mM NaCl), a otros intrones (aI1, aI2 y aI5 $\gamma$  de levadura, IntB de *E. coli*, Cal.X1 de *Calothrix*, o Nc.LSU de *Neurospora crassa*) y moléculas de ARN, aunque la maduración de estos intrones resulta ineficiente (Saldanha *et al.*, 1999).

### 1.3.2. Otros factores de *splicing* codificados en el genoma hospedador

La maquinaria implicada en la maduración de estos intrones identificada hasta la fecha varía en características y especificidad, siendo además dependiente del organismo donde se encuentren (Bonen y Vogel, 2001). Este tipo de proteínas se ha descrito fundamentalmente en levaduras, algas y plantas superiores (Lambowitz *et al.*, 1999; Lehmann y Schmidt, 2003). Algunas de estas proteínas necesitan la ayuda de otras proteínas, como madurasas, para completar su función. Como característica común, destacar que suelen tener una doble funcionalidad, por lo que se puede deducir que han sido reclutadas recientemente en la evolución por estos intrones para funcionar en *splicing* (Lambowitz y Perlman, 1990; Lambowitz *et al.*, 1999). La mayoría de los factores identificados participan en el *splicing* de manera indirecta pero algunos se sabe que interaccionan directamente con el ARN del intrón (Lehmann y Schmidt, 2003).

Los primeros candidatos a tener en cuenta en el proceso de maduración de intrones del grupo II serían proteínas con actividad madurasa. nMat-1a, -1b, -2a y -2b se encuentran codificadas por genes nucleares de plantas superiores (*A. thaliana* y *Oryza sativa*) pero tienen secuencias señal que las redirigen a la mitocondria. Presentan características similares a MatK, esto es, dominio madurasa conservado pero dominios RT y En

<sup>20</sup> Como curiosidad que después puede ayudar a entender los entresijos evolutivos en que se ven envueltos estos elementos genéticos móviles, matK es retenido en el genoma de los plastidios de plantas no fotosintéticas (Wolfe *et al.*, 1992; Ems *et al.*, 1995).

degenerados o ausentes. Se piensa que podrían participar en el *splicing* de uno o varios intrones del grupo II carentes de IEP presentes en la mitocondria (Mohr y Lambowitz, 2003)

Dos proteínas presentes en levaduras han centrado la atención de los investigadores: MSS116 y MRS2. La proteína de unión a ARN MSS116 (proteína celular de la familia “DEAD box”, motivo característico en ARN helicasas) parece funcionar como una ARN chaperona durante el plegamiento del intrón (Huang *et al.*, 2005), además de tener otros papeles que afectan a la función mitocondrial (Séraphin *et al.*, 1989). Los defectos en el procesamiento de ARN originados en la mitocondria en los mutantes MSS116 pueden complementarse con otra proteína del mismo tipo presente en *N. crassa*, CYT-19 (Mohr *et al.*, 2002; Huang *et al.*, 2005). En ambos casos, se trata de proteínas que precisan en su funcionamiento de energía en forma de ATP. El segundo factor de *splicing*, MRS2, es homólogo a transportadores de  $Mg^{2+}$  unidos a membrana y su mutación parece afectar específicamente el *splicing* de intrones del grupo II (no afecta la maduración de intrones del grupo I). Se le ha asignado un papel esencial en la homeostasis de  $Mg^{2+}$  en la mitocondria, ión que se sabe indispensable para el correcto plegamiento y catálisis de este tipo de intrones (Gregan *et al.*, 2001; Lehmann y Schmidt, 2003). Curiosamente, en *Arabidopsis*, una proteína homóloga a MRS2 puede complementar mutantes de levadura defectuosos en *splicing* (Schock *et al.*, 2000). Se han descrito otras proteínas en levadura cuya mutación tiene un efecto indirecto en la maduración de los intrones presentes en la mitocondria: MSS51, MRF1, MRH4 y varios genes NAM y MRS (Lehmann y Schmidt, 2003).

El análisis de mutantes defectuosos en la reacción de maduración en maíz ha permitido identificar varios factores relacionados con el *splicing* de intrones del grupo II. La proteína nuclear CRS1 está implicada en la traducción en el cloroplasto así como en la maduración de un único intrón del subgrupo IIA en el gen *atpF* (Jenkins *et al.*, 1997) y, en este último proceso, podría funcionar junto con MatK y/u otras ARN chaperonas (Till *et al.*, 2001). Parece ser que se une a los dominios I y IV promoviendo el plegamiento del centro catalítico de modo similar a las madurasas codificadas por los intrones del grupo II (Ostersetzer *et al.*, 2005). Nueve intrones del subgrupo IIB presentes en cloroplasto de maíz dependen de otro factor nuclear para su maduración, CRS2 (Jenkins *et al.*, 1997; Vogel *et al.*, 1999; Jenkins y Barkan, 2001). Esta proteína muestra identidad con peptidil-ARNt hidrolasas bacterianas y forma complejos con otras dos proteínas que determinan la especificidad por un intrón u otro, CAF1 y CAF2 (Ostheimer *et al.*, 2003). De estudios estructurales se deduce que no funcionan como tales hidrolasas (aunque no son capaces de complementar esta actividad en *trans* en *E. coli*) y que su papel en *splicing* puede estar mediado por la interacción directa con el ARN (Ostheimer *et al.*, 2005).

La maduración de algunos intrones está afectada por la mutación de determinados genes; es el caso de un intrón del gen *psaA* del cloroplasto de *Chlamydomonas reinhardtii*, cuyo *splicing* está afectado por mutación en 14 genes nucleares y cuyos productos parecen formar parte de grandes complejos (Goldschmidt-Clermont *et al.*, 1990; Perron *et al.*,

2004). Se especula que tanto en orgánulos como en procariotas podrían actuar similares factores de *splicing* que pueden haber contribuido a la dispersión de estos intrones (Holländer y Kück, 1999b). Podemos estar encontrándonos con el “eslabón perdido” en la evolución de intrones del grupo II hacia intrones espliceosómicos (Pyle y Lambowitz, 2006).

## 2. MECANISMOS DE MOVILIDAD EN INTRONES DEL GRUPO II

Hasta el momento hemos hablado de las propiedades de los intrones del grupo II como ARN catalíticos pero sabemos que algunos son, además, elementos genéticos móviles específicos de sitio (bien es cierto que sólo unos pocos pueden actuar como retroelementos). Los intrones pueden insertarse en nuevos sitios del genoma que son idénticos a aquellos a partir de los cuales se escindieron, en este caso el proceso se conoce como *homing* (Skelly *et al.*, 1991; Lazowska *et al.*, 1994; Moran *et al.*, 1995); pero también pueden moverse a sitios ectópicos, diferentes a su ADN diana habitual, proceso que denominamos transposición (Mueller *et al.*, 1993b; Sellem *et al.*, 1993). Como parece lógico, el *homing* se produce con una frecuencia cercana al 100%, esto es, casi siempre que la diana específica del intrón no esté invadida, el intrón va a encargarse de interrumpirla, mientras que la transposición es un proceso mucho más infrecuente (Cousineau *et al.*, 2000; Muñoz *et al.*, 2001). Uno de los aspectos más reveladores en la caracterización de estos procesos deriva de un sencillo diseño experimental: la interrupción de un intrón del grupo II con un intrón del grupo I (ambos intrones se escinden tras la transcripción). De este modo, la maduración del intrón del grupo II está condicionada al *splicing* del intrón del grupo I. Sin embargo, el proceso de *homing* tiene lugar y el análisis del producto de invasión mostró que el intrón del grupo I había desaparecido (Cousineau *et al.*, 1998; Martínez-Abarca *et al.*, 2004). Por tanto, con objeto de dejar patente que el proceso de movilidad está mediado por un intermediario de ARN, se añade el prefijo *retro* al nombre de ambos procesos, así la denominación pasa a ser *retrohoming* y retrotransposición (Curcio y Belfort, 1996; Ichianagi *et al.*, 2002).

Esta capacidad de insertarse en nuevas localizaciones del genoma de su hospedador reside en los complejos conocidos como RiboNucleoProteín-partículas (RNPs). Una vez que se ha producido la transcripción del intrón, tiene lugar la traducción de la IEP; ésta interacciona con el ARN transcrito precursor y facilita su maduración. Una vez han ocurrido todos estos procesos, la proteína permanece asociada al ARN escindido formando los complejos ribonucleoproteicos que denominamos RNPs.

### 2.1. PROTEÍNAS CODIFICADAS POR INTRONES DEL GRUPO II

Como hemos mencionado anteriormente, los intrones del grupo II tienen la peculiaridad de codificar una proteína que ayuda al ARN a catalizar los procesos de maduración y movilidad si bien, no todos los intrones de este tipo codifican proteínas, ni todas las proteínas codificadas por estos elementos son idénticas. Cuando está presente, la

secuencia que codifica estas IEPs (de una longitud que ronda las 1,2-2 kb, dependiendo de los dominios que presente) se localiza siempre en del dominio IV de la ribozima, de manera que la mayoría de la secuencia codificante queda fuera del centro catalítico de la ribozima (Belfort *et al.* 2002; Lehmann y Schmidt, 2003). Curiosamente, en mitocondrias de levaduras la IEP se traduce siguiendo la misma pauta de lectura que los exones flanqueantes, de manera que se produce una proteína de fusión que posteriormente es procesada por proteólisis<sup>21</sup> (Saldanha *et al.*, 1993; Michel y Ferat, 1995; Lehmann y Schmidt, 2003). Sin embargo, en bacterias (también en algunos intrones mitocondriales y de cloroplastos) la organización es diferente: la IEP se traduce de forma independiente de los exones gracias a que en el mismo dominio IV encontramos tanto el Shine-Dalgarno como el codón de iniciación de la traducción (Lambowitz *et al.*, 1999; Singh *et al.*, 2002).

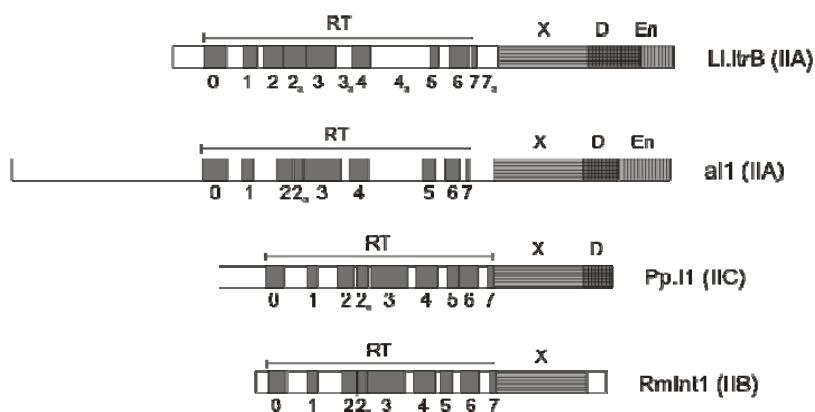
Los intrones móviles de los que se tienen más datos son los intrones de mitocondrias de levaduras, aI1 y aI2, pero fundamentalmente, del intrón de *L. lactis*, Ll.ltrB. Todos ellos presentan una IEP con 4 dominios diferencialmente conservados: el dominio reverso transcriptasa (RT), el dominio madurasa (X), el de unión a ADN (D) y el dominio ADN endonucleasa (En) (Matsuura *et al.*, 1997; Lambowitz y Zimmerly, 2004). La región N-terminal de la IEP (en torno a 250-300 aa) contiene un dominio RT característico de retrovirus y otros retroelementos conocidos (Michel y Lang, 1985; Kennell *et al.*, 1993). El dominio X, de 100 aa de longitud aproximada, se localiza aguas abajo del dominio RT; como hemos explicado con anterioridad, está implicado en ARN *splicing* (Mohr *et al.*, 1993; Cui *et al.*, 2004). El siguiente dominio en dirección 3' (o C-terminal) es el dominio D; su extensión ronda los 55 aa y su secuencia primaria apenas si está conservada. Finalmente, en la región C-terminal de muchas IEP se ha descrito la presencia del dominio En; su longitud suele estar en torno a los 50-60 aa y presenta motivos conservados de unión a ADN y otros con similitud a endonucleasas bacterianas (San Filippo y Lambowitz, 2002). Sin embargo, más de la mitad de las IEPs bacterianas anotadas hasta la fecha y un cuarto de las presentes en orgánulos carecen de dominio En, y muchas también han perdido el dominio D (Martínez-Abarca y Toro, 2000a; Belfort *et al.*, 2002; Lambowitz y Zimmerly, 2004; Toro *et al.*, 2007). Ver figura 8.

Fusiones traduccionales de LtrA con la proteína fluorescente GFP y microscopio de inmunofluorescencia, han permitido situar la IEP en los polos celulares de organismos como *L. lactis* o *E. coli* (Zhao y Lambowitz, 2005). En células de gran tamaño, a punto de dividirse, también hay focos hacia mitad de la célula. La localización se mantiene independientemente del tipo de fusión (N- o C-terminal), de la co-expresión del ARN de Ll.ltrB y de la velocidad de crecimiento y los niveles de expresión (Zhao y Lambowitz, 2005). En este trabajo relacionan la localización con posibles interacciones con la maquinaria replicativa (como la Pol III) que faciliten el acceso de las RNPs al ADN durante la replicación; también con componentes de membrana en la región de los polos

---

<sup>21</sup> De hecho, estos intrones presentan un dominio P aguas arriba del dominio RT que se asemeja a los dominios proteasa que se encuentran en retrovirus (Moran *et al.*, 1995). Esta actividad podría participar en la proteólisis de la proteína de fusión para dar lugar a una IEP activa.

relacionados o no con la localización polar del origen de replicación; otra posibilidad hipotética podría ser que las RNPs se asociasen con la maquinaria implicada en división celular permitiendo la segregación de estos intrones entre la descendencia. En un estudio posterior queda patente que la localización polar de LtrA es consecuencia de su exclusión del nucleoide<sup>22</sup>, que depende del momento del ciclo celular y que se mantiene en diferentes mutantes de *E. coli* relacionados con la formación del nucleoide, el reparto cromosómico y la replicación (Beauregard *et al.*, 2006). El mantener las RNPs en regiones donde el ADN está ausente podría ser un mecanismo desarrollado por la bacteria para protegerse de las invasiones sobre el ADN genómico que llevan a cabo los intrones del grupo II.



**Figura 8.** Representación esquemática de las proteínas codificadas por algunos intrones del grupo II. En este esquema podemos observar parte de las variadas IEPs que encontramos asociadas a estos retroelementos y pueden ofrecernos una idea de su evolución. Las regiones conservadas del dominio RT se muestran como rectángulos sombreados identificados en la parte inferior (0-7). El resto de dominios se representan igualmente como rectángulos pero con distintas matrices de relleno: el dominio X relleno con rayas horizontales, el dominio de unión a ADN (D) con una matriz cuadriculada y el dominio endonucleasa (En) con líneas verticales.

### 2.1.1. Dominio reverso transcriptasa (RT)

El dominio RT de intrones del grupo II contiene una serie de bloques de aminoácidos conservados que definen siete regiones (1-7) características de reverso transcriptasas de retrovirus (Michel y Lang, 1985). Generalmente, en la región N-terminal presentan otra región conservada (región Z o 0) que las relaciona con retrotransposones (rTn) no-LTR (Xiong y Eickbush, 1990; Zimmerly *et al.*, 2001). En el bloque 5 se encuentra el motivo catalítico para la actividad RT, YXDD (siendo X generalmente una A), como lo demuestra el que el nivel de actividad RT de un mutante del intrón aI2, en el que este motivo ha sido modificado a YAHH, está cercano al valor umbral del ensayo (Zimmerly *et al.*, 1999). Estudios desarrollados en la RT del virus de inmunodeficiencia humana

<sup>22</sup> El nucleoide en bacterias ocupa la mayoría del espacio celular, limitando el espacio citosólico a las regiones polares de la célula. Como curiosidad, apuntar que la mayoría de las proteínas de unión a ADN (con distintas funciones biológicas) suelen encontrarse asociadas al nucleoide (Beauregard *et al.*, 2006).

(HIV-1) ponen de manifiesto que los dos aspárticos conservados (Asp o D) del motivo YXDD, junto con un tercer Asp, podrían participar en la coordinación de  $Mg^{2+}$  (Lowe *et al.*, 1991). Los otros dos aminoácidos (Tyr y X, que en retrovirus suele ser Met) parece que participan en la unión del dNTP que va a incorporarse al ADNc que se está sintetizando (Huang *et al.*, 1998).

Como su nombre indica, son ADN polimerasas dependientes de ARN que precisan de la existencia de un cebador de ADN (Zimmerly *et al.*, 1995b). La primeras evidencias bioquímicas de una actividad reverso transcriptasa asociada a las RNPs de intrones del grupo II se obtuvieron del estudio de los intrones mitocondriales aI1 y aI2 de *S. cerevisiae* (Kennell *et al.*, 1993; Zimmerly *et al.*, 1999) y COX1I1 de *P. anserina* (Faßbender *et al.*, 1994); posteriormente, se les ha unido *cobI1* de *Schizosaccharomyces pombe* (Schäfer *et al.*, 2003). También se describe en intrones bacterianos como Ll.ltrB de *L. lactis* (Matsuura *et al.*, 1997), Xln6 de *Pseudomonas alcaligenes* (Yeo *et al.*, 1997), RmInt1 de *S. meliloti* (Martínez-Abarca *et al.*, 1999) o EA1 en la bacteria termófila *Bacillus caldolyticus* (Ng *et al.*, 2007).

Reiterar que esta actividad está asociada a las RNPs aunque algunas IEPs como LtrA muestran cierto nivel de actividad RT cuando la proteína purificada se encuentra en presencia de un sustrato externo poli(rA)/oligo(dT)<sub>18</sub>; sin embargo, esta actividad decae con el tiempo de incubación, lo que sugiere que la proteína en ausencia del ARN se desestabiliza probablemente por desnaturalización (Saldanha *et al.*, 1999). Por otro lado, se ha descrito en aI2 que tanto mutantes en la IEP (P714I) como en el ARN del intrón ( $\Delta$ dV) alteran la funcionalidad de las RNPs formadas, de manera que se estimula la síntesis de un ADNc inespecífico como consecuencia de problemas para el reconocimiento del cebador correcto (Morozova *et al.*, 2002); en ningún caso va a tener lugar el *homing* (Moran *et al.*, 1995).

Finalmente, recordar que los dominios RT y X actúan juntos a la hora de unirse al ARN del intrón no sólo para dar lugar a su escisión, sino también para permitir el posicionamiento de la IEP en un momento tan crucial como es la síntesis del ADNc usando el ARN del intrón como molde. La unión al dominio IVa, que mencionábamos en el apartado 1.3.1., permite que el dominio RT de la IEP tenga acceso al cebador para la síntesis de este ADNc como veremos más adelante (Matsuura *et al.*, 1997; Wank *et al.*, 1999; Cui *et al.*, 2004; Cui, 2005).

### 2.1.2. Dominios de unión a ADN (D) y ADN endonucleasa (En)

La definición de estos dominios se ha conseguido gracias al análisis de proteínas truncadas en su extremo C-terminal (Guo *et al.*, 1997; Singh y Lambowitz, 2001). Intrones deletados en el dominio En eran incapaces de producir corte en el ADN, aunque todavía mantenían la capacidad de unión a su ADN diana. Si la delección se extiende al dominio D, siguen reconociendo su diana (esta propiedad reside en su mayor parte en el ARN del intrón), pero son incapaces de unirse cuando la diana se encuentra como doble cadena; se ha observado una actividad residual cuando el ADN diana se encuentra como

cadena sencilla, lo que confiere al dominio D cierto papel en el desenrollamiento de la doble cadena.

Por otro lado, esta región C-terminal no se precisa para la actividad madurasa pero si parece afectar a la actividad RT (Zimmerly *et al.*, 1995b; 1999; San Filippo y Lambowitz, 2002). Cuando se modifican residuos clave que inhiben el corte en el ADN diana, también disminuyen los niveles de actividad RT de las RNPs. Esta disminución de actividad no parece deberse a graves alteraciones estructurales en la proteína, sino más bien a problemas para el posicionamiento respecto al cebador o a alteraciones locales del dominio RT. Del mismo modo, este dominio no parece ser necesario en la dimerización, pero podría interactuar en algún momento de la reacción con el dominio RT, activándolo. La interacción podría producirse una vez que ha tenido lugar el corte del ADN diana para permitir el reposicionamiento del sitio activo RT respecto de su cebador para iniciar la reverso transcripción (Blocker *et al.*, 2005).

El dominio de unión a ADN no está muy conservado en cuanto a secuencia, pero presenta dos regiones importantes desde el punto de vista funcional: una zona muy rica en aminoácidos básicos y una región cuya estructura secundaria predicha es una  $\alpha$ -hélice. Estas dos regiones se conservan (no a nivel de secuencia primaria) en otras IEPs y análisis de evolución unigénica<sup>23</sup> así como estudios bioquímicos en mutantes ponen de relieve su importancia. Llama la atención que en la naturaleza la variabilidad es mayor que en los estudios de evolución unigénica. Esta observación es fácilmente explicable si tenemos en cuenta que los dominios D de las distintas IEPs han debido adaptarse al reconocimiento de diferentes ADNs diana, sin embargo cuando una IEP se enfrenta a su diana específica la variabilidad debe ser menor (San Filippo y Lambowitz, 2002).

El dominio ADN endonucleasa contiene dos motivos característicos de este tipo de proteínas estructurados de manera similar a como se encuentran en la endonucleasa VII del fago T4 y a las colicinas E7 y E9 (Gorbalenya, 1994; Shub *et al.*, 1994). En la naturaleza existen cuatro tipos de ADN endonucleasas, en su mayoría asociadas con otros elementos genéticos móviles, los intrones del grupo I: LAGLIDADG, GIY-YIG, His-Cys y H-N-H (Belfort y Roberts, 1997; Jurica y Stoddard, 1999; Chevalier y Stoddard, 2001). Sin embargo, las ADN endonucleasas que se encuentran asociadas con intrones del grupo II son, en su mayoría, las H-N-H endonucleasas, descritas en aI1, aI2 y Ll.ltrB (Zimmerly *et al.*, 1995b; Yang *et al.*, 1996; Mills *et al.*, 1996). Se trata de proteínas con una secuencia consenso que ronda los 30 aa, con dos pares de residuos conservados de histidina (H) flanqueando una asparragina (N). Del grupo de cuatro, las H-N-H endonucleasas son las menos estudiadas y caracterizadas; se sabe que presentan una organización estructural que comprende dos cadenas  $\beta$  seguidas de dos hélices  $\alpha$ , que se

---

<sup>23</sup> Se trata de construir librerías de intrones con mutaciones introducidas al azar por PCR. Se seleccionan aquellas variantes de la proteína que todavía sean capaces de desarrollar una determinada actividad (en este caso se seleccionaron mutantes que retienen la capacidad de movimiento) y se analizan las mutaciones que contiene. Aquellas posiciones en las que no se detecten cambios, hipomutables, son posiciones de relevancia funcional (Cadwell y Joyce, 1992).



organizan alrededor del enlace fosfodiéster de la molécula que cortan y que coordinan un ión metálico sirviéndose de tres histidinas. Así, el sitio catalítico de la actividad ADN endonucleasa definido para Ll.ltrB a partir del análisis de evolución unigénica y del comportamiento de mutantes en posiciones clave es el motivo H-N-H que contiene un único pero esencial ión  $Mg^{2+}$  (San Filippo y Lambowitz, 2002). Se ha descrito la existencia de intrones del grupo II que están asociados con ADN endonucleasas del tipo LAGLIDADG en mitocondrias de diferentes hongos; curiosamente, estas IEPs carecen de otros dominios típicos, como RT o X (Toor y Zimmerly, 2002).

Flanqueando este motivo encontramos dos parejas de residuos cisteína (Cys) conservados y característicos de proteínas de unión a ADN, que se conocen como motivo “dedos de Zinc”<sup>24</sup>. En el caso de Ll.ltrB, este motivo parece encargarse de la estabilización de la estructura de la región y, al menos, LtrA purificada no parece contener ningún ión  $Zn^{2+}$ ; se especula que ha podido perderlo una vez que la proteína se ha plegado correctamente (San Filippo y Lambowitz, 2002).

### 2.1.3. Modelización tridimensional de una IEP

La caracterización bioquímica de proteínas se completa con la obtención del cristal proteico que permita la visualización de las interacciones que tienen lugar en los distintos procesos en que la proteína se ve involucrada. Al contrario que algunas endonucleasas de intrones del grupo I, aún no se ha conseguido cristalizar ninguna de las proteínas multifuncionales codificadas por intrones del grupo II. Por esto algunos investigadores tratan de modelizar la organización tridimensional que presentarían las IEPs, basándose en proteínas ya cristalizadas con las que presentan similitudes. Así, un trabajo reciente muestra la predicción de estructura cuaternaria de LtrA usando como proteínas modelo la RT de HIV-1, la endonucleasa VII del bacteriófago T4 y la colicina E7 (Blocker *et al.*, 2005). Ver figura 9.

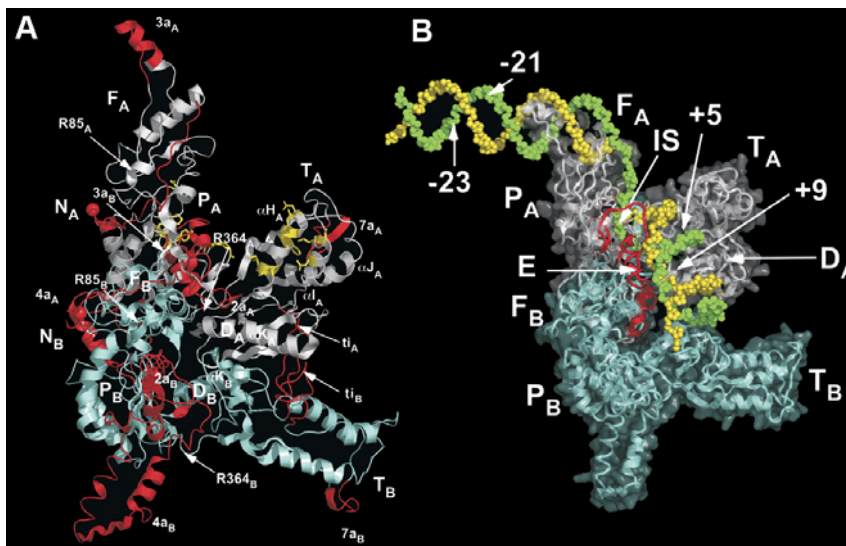
Hace catorce años que se conoce la estructura de RT de HIV-1; se trata de un heterodímero formado por dos subunidades, p66 y p51. La organización tridimensional de p66 se ha asemejado con una mano donde los dedos, la palma y el pulgar forman una hendidura o surco que contiene el sitio activo de la RT y a través del cual va a interactuar con sus sustratos; a continuación encontramos el dominio de conexión, que consiste en una hélice  $\alpha$ , y el dominio ARNasa H, ausente en p51 (Kohlstaedt *et al.* 1992). Desde el punto de vista catalítico p51 es dispensable, pero se requiere para el correcto plegamiento de p66 y parece contribuir en cierta medida a la unión del complejo molde/cebador (Ding *et al.* 1998; Huang *et al.* 1998). Esta RT muestra una conformación flexible entre dos estados: abierto, cuando se une el molde/cebador, y cerrada, al sumarse el dNTP. En ambas conformaciones hay una reorganización de la estructura del complejo

---

<sup>24</sup> Dedos de Zinc hace alusión a dos observaciones: por un lado, la representación de su estructura secundaria recuerda los dedos de la mano; es precisamente esta conformación la que permite a estas proteínas unirse al ADN. Por otro, los residuos Cys suelen encontrarse coordinando iones  $Zn^{2+}$ .

ARN molde/ADN cebador en las proximidades del sitio activo RT, sustrato que además sufre una curvatura de  $\sim 40^\circ$  (Ding *et al.* 1998; Huang *et al.* 1998).

LtrA se presenta como un homodímero pero cada monómero muestra variaciones en la estructura terciaria similares a las presentes en p66 y p51. Al igual que la RT de HIV-1, los dominios RT 1 al 7 representan una serie de hélices  $\alpha$  que se alternan con cadenas  $\beta$  que comprenden los dedos y la palma. El extremo N-terminal de LtrA no se ha conseguido modelizar puesto que carece de similitud de secuencia con la misma región de la proteína de HIV-1. El dominio X de intrones del grupo II está compuesto por tres hélices  $\alpha$  y se corresponde con el pulgar de la RT viral, aunque se extiende algo hacia el dominio de conexión. Finalmente, el dominio de unión a ADN comprende la parte central del dominio de conexión de p66 coincidiendo con una hélice  $\alpha$  presente en esta proteína. La identidad entre LtrA y p66 se pierde hacia las dos terceras partes del dominio de conexión de esta última, justo donde comienza el dominio ADN endonucleasa definido en intrones del grupo II (Blocker *et al.*, 2005). A partir de aquí el modelo se estructura en función de la endonucleasa VII del fago T4 con el motivo H-N-H posicionado para producir el corte y la colicina E7 con su sustrato ADN unido. Esta modelización predice que la interacción con el ADN diana va a precisar de la curvatura de este ADN para que la proteína sea capaz de abarcar todas sus posiciones de reconocimiento en ambos exones, hecho que posteriormente se comprueba mediante microscopía de fuerza atómica (Noah *et al.*, 2006).



**Figura 9.** Modelo tridimensional del dímero formado por LtrA obtenido por comparación con la RT de HIV-1 y la endonucleasa VII del fago T4 realizado por Blocker y cols. Cada monómero se representa con un color diferente: la subunidad A en blanco y la B en celeste. Se señalan elementos

importantes en la estructura como la palma (P), dedos (F) y pulgar (T), así como el dominio de unión a ADN (D), identificados con el subíndice de cada monómero. (A) En rojo se destacan las principales diferencias respecto de la RT de HIV-1 y en amarillo las coincidencias más fuertes. El extremo N-terminal de ambas subunidades está señalado con una esfera roja y denotado  $N_A$  y  $N_B$ . (B) LtrA unida a su ADN diana. El dominio En está representado en rojo (E) y cada una de las hebras del ADN en amarillo (inferior) y verde (superior). Figura adaptada de Blocker *et al.*, 2005.

#### 2.1.4. Clasificación de los intrones en función de su IEP

Al igual que se ha establecido una clasificación filogenética atendiendo a la estructura de la ribozima, se ha trabajado intensamente en la clasificación de los intrones del grupo II de acuerdo a la organización de la proteína que codifican (Martínez-Abarca y Toro, 2000a; Zimmerly *et al.*, 2001; Toor *et al.*, 2001; Toro, 2003). Las IEPs representan un criterio de clasificación válido puesto que los intrones del grupo II que codifican proteínas se encuentran tanto en orgánulos eucariotas como en bacterias y arqueas. Casi todos los intrones procariotas descritos hasta nuestros días presentan IEP; sólo algunos intrones identificados en la bacteria termófila *Thermosynechococcus elongatus* y en la arquea *Methanosarcina acetivorans* carecen de proteína en su secuencia (Dai y Zimmerly, 2003). En eucariotas inferiores, aproximadamente la mitad de los intrones del grupo II contienen IEP, pero en plantas superiores el porcentaje de intrones con IEP es considerablemente menor. Este patrón de distribución podría marcar la tendencia de los intrones en su evolución: el origen de los intrones del grupo II que codifican su propia proteína probablemente sea bacteriano y su dispersión hacia organismos con mayor complejidad estructural ha ido asociada a la pérdida de la IEP (Toor *et al.*, 2001). Otra explicación posible es que estos intrones en orgánulos funcionan como verdaderos intrones, mientras que en bacterias se comportan como retroelementos ávidos por dispersarse (Dai y Zimmerly, 2002). La presencia de intrones con dominios RT degenerados en *Englena*, *Chlamydomonas* y una amplia variedad de plantas terrestres (Copertino y Hallick, 1993; Rochaix, 1996; Toor *et al.*, 2001), o de intrones en cuyo dominio IV quedan remanentes de lo que pudo ser una IEP (se han encontrado los codones de inicio y final de traducción además de algunos aminoácidos en *M.a.I6* de *M. acetivorans*; Dai y Zimmerly, 2003), apoyan las teorías anteriores.

Los dominios RT 0-7 se encuentran bastante bien conservados entre intrones bacterianos e intrones de mitocondrias, sin embargo el dominio X apenas se conserva entre estos dos subgrupos. Esta divergencia puede ser el resultado de una fuerte presión evolutiva sobre el dominio X en su adaptación para reconocer su intrón de forma específica y asistirlo en sus funciones (Zimmerly *et al.*, 2001). Resaltar que la mayoría de los intrones degenerados se localizan en intrones de cloroplastos. En ninguno de los estudios realizados para la clasificación de las IEP se ha considerado el dominio En porque se considera que se ha adquirido en múltiples ocasiones por distintos intrones durante la evolución (Zimmerly *et al.*, 2001).

En los primeros árboles filogenéticos se distinguían dos grandes grupos de IEPs, junto con los cuales se entremezclaban algunos intrones bacterianos: los intrones de cloroplastos y los mitocondriales (Zimmerly *et al.*, 2001). Otros análisis hablaban de 3 grupos principales (Martínez-Abarca y Toro, 2000a; Toro, 2003). Finalmente, las IEP se han clasificado en ocho linajes: mitocondrial, cloroplástico de tipo 1 y de tipo 2, y cinco linajes bacterianos denotados de la A-E (Zimmerly *et al.*, 2001; Toro *et al.*, 2002; Toro, 2003). Ver figura 6. En este sentido, lo más destacable es que cada linaje proteínico se

corresponde con una subclase de estructura de ARN, sugiriendo una coevolución de ambos componentes (Martínez-Abarca y Toro, 2000a; Toor *et al.*, 2001). Tanto la clase mitocondrial como las dos cloroplásticas incluyen algún intrón presente en bacterias (por ejemplo, Ll.ltrB pertenece al linaje mitocondrial) lo que podría apuntar a que los intrones de organelas tiene un posible ancestro bacteriano derivado de su origen como endosimbiontes en la célula eucariota.

Cabe destacar que una determinada clase de intrón en función de su IEP no es específica de una especie bacteriana particular. Como ejemplo cercano tenemos los intrones descritos la bacteria *Sinorhizobium meliloti* cepa 1021: uno de ellos pertenece a la clase bacteriana D, el otro a la C y el tercero aún no se ha determinado (Toro, 2003). Otra conclusión que puede extraerse del análisis de los árboles filogenéticos es que los fenómenos de transferencia horizontal son muy habituales en procariotas (englobando bacterias y arqueas) (Zimmerly *et al.*, 2001; Dai y Zimmerly, 2002; Belhocine *et al.*, 2004). Sobre aspectos evolutivos hablaremos más adelante (Apartado 4.1. de esta memoria).

## 2.2. MECANISMOS DE RETROHOMING

Los mecanismos de movilidad usados por los intrones del grupo II comenzaron a descifrarse en la década de los 90 con la caracterización de la reacción de corte e inserción en el ADN diana por parte de los intrones mitocondriales de levadura, aI1 y aI2. Como ya avanzamos, la capacidad de movimiento está mediada por complejos ribonucleoproteicos, RNPs, y participan activamente tanto el componente ARN como la IEP. Sin embargo, la complejidad de la regulación de la recombinación en levaduras no facilita la comprensión del proceso. Se precisa de un sistema más sencillo, mejor estudiado y más controlado. En bacterias, estos mecanismos de regulación están perfectamente descritos y además los genes que los rigen pueden modificarse en función de las necesidades del estudio (Eickbush, 1999). La posibilidad de trasladar los estudios de movilidad a un sistema bacteriano viene de la mano del intrón Ll.ltrB. Este intrón no sólo es móvil en el hospedador donde se describió, *L. lactis* (Mills *et al.*, 1997), sino que también es capaz de moverse en *E. coli* (Matsuura *et al.*, 1997). El desarrollo de sistemas de expresión y de movilidad en *E. coli* ha permitido la caracterización de todos los procesos que intervienen en el *retrohoming* de Ll.ltrB (Cousineau *et al.*, 1998).

La principal diferencia entre el *retrohoming* de intrones móviles presentes en mitocondrias de levaduras y los intrones bacterianos es el fenómeno conocido como co-conversión de los exones flanqueantes (Belfort *et al.*, 2002). En levaduras se han descrito dos mecanismos de movilidad en los que, o bien sólo el exón 1 o ambos exones, pueden acompañar al intrón en su movimiento; por el contrario, en bacterias no existe este fenómeno, el intrón se inserta limpiamente en su nueva diana.

Son muy pocos los intrones del grupo II que se haya demostrado que son móviles a pesar de que muchos de ellos son capaces de madurar (*in vivo* y/o *in vitro*) y formar RNPs. Pero lo más llamativo a este respecto es que cada uno de estos intrones presenta

particularidades en el mecanismo de *retrobombing* que lo diferencian de los demás. Resulta evidente que los intrones se han adaptado a las características de su organismo hospedador y que esta capacidad “camaleónica” les ha permitido dispersarse en sistemas muy dispares: arqueas, levaduras e incluso plantas superiores.

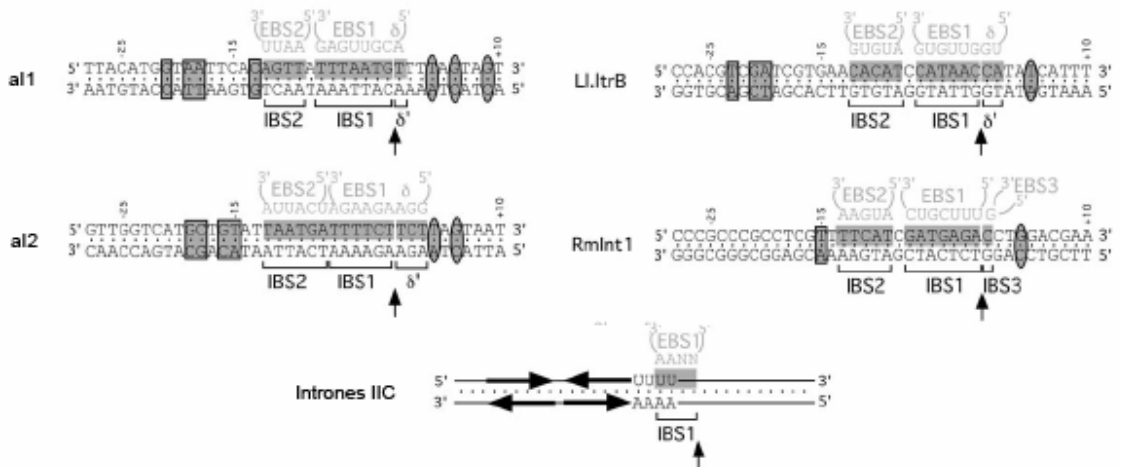
### 2.2.1. Reconocimiento del ADN diana

El reconocimiento del ADN diana por parte de las RNPs implica la participación activa de ambos componentes, el ARN y la IEP. Se ha estudiado en todos los intrones móviles bien caracterizados hasta el momento y la conclusión más clara es que la especificidad del reconocimiento debe venir dada por la ribozima, puesto que la IEP reconoce relativamente pocas posiciones (Yang *et al.*, 1996). De modo similar al reconocimiento de los exones que tiene lugar durante el *splicing*, el ARN del intrón, a través de las secuencias EBS presentes en el dI (EBS1 y 2, y  $\delta$  en intrones del grupo IIA; EBS1, 2 y 3 en IIB), interacciona por apareamiento de bases con las posiciones correspondientes de los exones en el ADN diana (IBSs y/o  $\delta'$ ). El tipo de interacción (apareamiento de bases) se pone de manifiesto al comprobar que la capacidad de movimiento sobre dianas mutadas puede restituirse usando mutantes en el dI que compensen esos cambios en el ADN diana (Guo *et al.*, 2000; Mohr *et al.*, 2000).

Pero el ADN se suele encontrar en los organismos vivos en forma de doble cadena (ADNdc), de manera que el reconocimiento por apareamiento de bases precisa de un paso previo o, al menos, paralelo: el desenrollamiento de la doble hélice. Esta parece ser la principal implicación de la IEP en su interacción inicial con el ADN diana, establece contactos con posiciones distales de ambos exones provocando el desenrollamiento local de las dos hebras. Este proceso ha sido descrito en detalle para el intrón de *L. lactis*, Ll.ltrB: el dominio D de LtrA reconoce vía surco mayor una serie de nucleótidos clave en la región distal del exón 1, T-23, G-21 y A-20, y, probablemente, otra serie de posiciones adyacentes (-24 a -13) vía surco menor (Singh y Lambowitz, 2001). A esta conclusión se llega tras comprobar que las RNPs pueden mediar la inserción del ARN en sustratos mutados en esas posiciones únicamente cuando se encuentran como ADN de cadena simple (ADNsc) *in vitro*, nunca como cadena doble<sup>25</sup> (Mohr *et al.*, 2000; Zhong y Lambowitz, 2003). Por otro lado, es preciso el reconocimiento del exón 2 para que el proceso de movilidad de estos intrones se complete. En Ll.ltrB, la interacción más importante se establece con la T+5, que se piensa que aún se encuentra dentro de la región desenrollada (Mohr *et al.* 2000; Singh y Lambowitz, 2001). Resultados similares se extraen del análisis de la diana del intron aI2 (Guo *et al.*, 1997). Ver figura 10.

---

<sup>25</sup> Se usaron RNPs reconstituidas *in vitro* a partir de sus componentes purificados que se enfrentan en condiciones adecuadas de reacción a sustratos de ADNsc (oligonucleotidos de longitud variable) o a sustratos de ADNdc (por lo general, productos de PCR). Los productos generados se revelan en geles de poliacrilamida desnaturizantes (veremos más adelante en Materiales y Métodos).



**Figura 10.** Reconocimiento del ADN diana en algunos ejemplos representativos de intrones móviles, así como la propuesta lanzada para intrones de la clase IIC. Con una flecha se indica el punto de inserción del intrón. Las posiciones reconocidas por la IEP en el ADN diana se representan enmarcadas con  si se encuentran en el exón 1 o con  si se localizan en el exón 2. Las posiciones correspondientes con reconocimiento por apareamiento de bases están sombreadas en gris (). a11, a12 y Ll.ltrB pertenecen a los intrones del grupo IIA; RmInt1 se encuentra entre los intrones IIB. Figura adaptada de Lambowitz *et al.*, 2004.

En general, se considera que las RNPs se unen de forma inespecífica al ADN celular, proceso facilitado por mecanismos de difusión, y luego se encargarían de recorrer el cromosoma hasta identificar la secuencia de su sitio diana, sistema similar al que utilizan otras proteínas de unión a ADN específicas de sitio (Aizawa *et al.*, 2003). Reconocen la región distal del exón 1 provocando, junto con la interacción del ARN, la separación de las dos cadenas. A continuación, tiene lugar el reconocimiento de la región del exón 2; si tenemos en cuenta que la proteína se encuentra anclada al exón 1, el establecimiento de contactos con el exón 2 (recordar la modelización de la interacción LtrA con su ADN diana) va a provocar que la diana se curve en esta región (Noah y Lambowitz, 2006). De esta manera, el ADN se posiciona para que la actividad ADN endonucleasa de la IEP produzca el corte en una de las cadenas (Lambowitz y Zimmerly, 2004; Pyle y Lambowitz, 2006; Toro *et al.*, 2007).

En este sentido, cabe destacar que la longitud de las dianas reconocidas por intrones cuya IEP presenta dominio En es considerablemente mayor (30-35 bp) si se compara con la secuencia de reconocimiento de intrones cuyas RNPs carecen de esta actividad (25 bp) (Mohr *et al.*, 2000; Jiménez-Zurdo *et al.*, 2003). Un caso particular lo constituyen los intrones pertenecientes a la clase IIC; una de sus características definitorias es que carecen de interacción EBS2-IBS2 pero se localizan insertados aguas abajo de estructuras en horquilla, generalmente terminadores de transcripción independientes de  $\rho$ . Así, en estos intrones se ha especulado que las RNPs podrían reconocer específicamente estas estructuras además de sus IBSs para el *homing* (Grandlund *et al.*, 2001; Toor *et al.*, 2006; Michel *et al.*, 2007).

### 2.2.2. *Homing* en intrones mitocondriales de levadura

Las primeras evidencias de la movilidad de intrones del grupo II derivan del análisis de la progenie de cruces entre cepas haploides de levadura con diferentes combinaciones de intrones en su genoma mitocondrial: durante el cruce, las mitocondrias se fusionan, y al segregarse, se observa que los intrones presentes en una de las mitocondrias han invadido las dianas libres de la otra. De este modo, se obtiene una población casi homogénea en la que aproximadamente el 90% de las dianas se encuentran invadidas (Skelly *et al.*, 1991; Lambowitz y Zimmerly, 2004). Estudios posteriores han permitido desentrañar los procesos que median la movilidad de estos intrones (Lazowska *et al.*, 1994; Moran *et al.*, 1995; Zimmerly *et al.*, 1995a; 1995b; Yang *et al.*, 1996; Eskes *et al.*, 1997; 2000).

El primer estudio detallado de la movilidad de los intrones aI1 y aI2 pone de manifiesto que ambos son capaces de moverse independientemente del otro, pero, en todos los casos, se observa la cotransferencia asimétrica de parte de los exones junto con el intrón a la nueva localización (Lazowska *et al.*, 1994), concretamente >50 bp en el exón 1 y <25 bp dentro del exón 2. La observación previa de que las RNPs comienzan la síntesis del ADNc muy cerca del extremo 5' del exón 2 cuando usan como molde un ARN precursor no maduro *in vitro* (Kennell *et al.*, 1993), podría explicar este patrón de co-conversión: el ADNc comienza en el exón 2, continúa a través del intrón hasta bastante avanzado el exón 1 y luego se inserta por recombinación homóloga en la nueva diana. Pero entonces, ¿qué papel jugarían el intrón escindido y la actividad ADN endonucleasa de las RNPs?

Un trabajo posterior apunta la existencia de dos posibles rutas para el *homing* de estos intrones en levaduras *in vivo* (Moran *et al.*, 1995). Observaron que mutaciones en el centro activo de la actividad RT reducen la movilidad sólo al 40% en aI2, pero que mutaciones que afectan la actividad ADN endonucleasa inhiben por completo el *homing*. Estos eventos de movilidad en ausencia de actividad RT deben producirse a nivel de ADN, probablemente tengan lugar por recombinación mediada por corte producido por la actividad endonucleasa de la IEP. En este caso, el *homing* va acompañado de co-conversión en ambos exones. El proceso dependiente de RT se asemejaría al descrito para el elemento R2 de *Bombyx mori* (Luan *et al.*, 1993): el corte en una de las cadenas (en el exón 2, pero muy próximo al sitio de inserción) permitiría a la RT copiar el intrón usando el 3'OH libre como cebador y el ARN precursor no maduro como molde. Así se explicaría que se produzca co-conversión en el exón 1 pero no en el exón 2.

Sin embargo, los detalles del proceso se extrajeron con estudios bioquímicos *in vitro* usando preparaciones enriquecidas en RNPs (Zimmerly *et al.*, 1995a; 1995b; Yang *et al.*, 1996). De acuerdo con lo observado por Moran y colaboradores, una de las rutas de movilidad se asemeja al mecanismo TPRT (Target-primed Reverse Transcription) descrito para retrotransposones no-LTR (R2Bm) (Zimmerly *et al.*, 1995a). Con estos trabajos demuestran que el *homing* está mediado por corte en las dos cadenas: por un lado, el ARN del intrón se inserta en una de las cadenas (en adelante, la cadena de arriba, superior o sentido por analogía con los esquemas) gracias al proceso que se ha definido

como *splicing* reverso; mientras, la actividad ADN endonucleasa de la IEP produce un corte en la otra hebra ligeramente desplazado aguas abajo del sitio de inserción del intrón (Zimmerly *et al.*, 1995b).

Cuando hablamos de *splicing*, decíamos que las dos transesterificaciones que tienen lugar durante la maduración de los intrones del grupo II por la ruta de *branching* son reversibles (Chin y Pyle, 1995). De hecho, varios autores han sido capaces de demostrar *in vitro* la capacidad del ARN *lariat* de insertarse en otra molécula de ARN que represente los exones unidos (o que al menos contenga secuencias complementarias a los EBSs del intrón; Augustin *et al.*, 1990; Mörl y Schmelzer, 1990a); además, esta inserción tiene lugar justo en el punto de unión de los exones. Esta reacción no está limitada a sustratos ARN, sino que se ha visto que estos intrones catalizan con una eficiencia relativa su inserción sobre ADN *in vitro* (Mörl *et al.*, 1992; Griffin *et al.*, 1995; Yang *et al.*, 1996). En cualquier caso, parece ser que la reacción directa (*splicing*) está favorecida sobre la inversa (Daniels *et al.*, 1996), pero en determinadas condiciones (por ejemplo, que existan otras reacciones acopladas) se puede impulsar la reacción inversa (Aizawa *et al.*, 2003). Tras la reacción de reversión de la maduración, el ARN del intrón queda en forma lineal entre los dos exones, entonces decimos que el *splicing* reverso es completo. Pero también puede ser parcial: el intrón rompe la unión entre los dos exones pero permanece unido en forma de lazo al exón 2. En levaduras se observa un comportamiento diferencial *in vitro* respecto a inserción completa o parcial: mientras que aI2 se inserta mayoritariamente por *splicing* reverso parcial, aI1 se suele insertar completo en el ~50% de los casos (Zimmerly *et al.*, 1995b; Yang *et al.*, 1996; Eskes *et al.*, 1997; 2000).

En cuanto al corte generado por la ADN endonucleasa de la IEP de estos intrones de levadura tiene lugar en la posición +10 dentro del exón 2 pero en la cadena opuesta (cadena de abajo, inferior o antisentido) a la que se ha invadido por el ARN del intrón (Yang *et al.*, 1996; Zimmerly *et al.*, 1995a; 1995b). Yang y colaboradores demostraron con RNPs de aI1 *in vitro* que la proporción de *splicing* reverso completo se encuentra entre 38-63% y en > 95% de estas moléculas también se había producido el corte de la cadena de abajo (Yang *et al.*, 1996). El extremo 3'OH libre que se ha generado sirve de cebador a la reverso transcriptasa para la síntesis del ADNc, de hecho, se ha propuesto que la reverso-transcripción a través del 3'ss bloquea la escisión y desplaza el equilibrio hacia una inserción completa del intrón (Aizawa *et al.*, 2003).

Finalmente, la confirmación definitiva de los mecanismos de movilidad que operan en levaduras deriva del análisis genético de un gran número de eventos de *homing in vivo*, atendiendo a los patrones de co-conversión (Eskes *et al.*, 1997; 2000). Así pues se describen dos rutas principales para la movilidad de estos intrones:

\* Una ruta minoritaria de *homing* (~40% de los eventos analizados) independiente de actividad RT. La integración del intrón tiene lugar por recombinación homóloga a nivel de ADN: la doble rotura generada por las RNPs en el ADN diana se repara mediante el sistema DSBRR (Double Strand Break Repair). En estos casos, el intrón se integra acompañado de parte de ambos exones.



\* La ruta mayoritaria de *retrohoming* (~60%) tiene como resultado la cotransformación sólo del exón 1 e implica la cooperación de varios mecanismos: por un lado, TPRT para la síntesis del ADNc a partir del intrón y, por otro, recombinación homóloga (DSBR) para la integración definitiva del ADNc del intrón en la nueva diana. Existe una enorme controversia al tratar de explicar la resolución definitiva del proceso, ya que en levaduras se dan dos situaciones: intrón insertado parcialmente y la inserción completa. En las dianas en que la inserción del ARN ha sido completa, todos los autores coinciden en que el intrón insertado es el molde y que en algún momento, probablemente cuando la RT haya concluido, se produce la reparación de esta región mediante recombinación (un sistema muy activo en levaduras). Cuando la inserción ha tenido lugar por *splicing* reverso parcial la RT iniciaría la síntesis usando intrón insertado al menos hasta el enlace modificado 2'-5' presente en el *lariat*<sup>26</sup> (por eso no existe co-conversión del exón 2), el problema es explicar qué sucede después. Hay autores que consideran que podría producirse un cambio de molde al ARN precursor no maduro (Zimmerly *et al.*, 1995b), pero ninguno descarta la posibilidad de que la RT atravesase el enlace y continúe con la síntesis a partir del intrón insertado, resolviendo finalmente mediante recombinación aunque la síntesis del ADNc no se haya completado (Eskes *et al.*, 1997; 2000). Además se establece un gradiente de co-conversión para aI2 que se fija entre las posiciones -387 (exón 1) y +23 (exón 2); así, el ~70% recombina entre la posición -387 y el sitio de inserción del intrón, mientras el 95% de las recombinaciones en eventos independientes de RT se producen antes de la posición +23 (Eskes *et al.*, 2000). Ver figura 11.

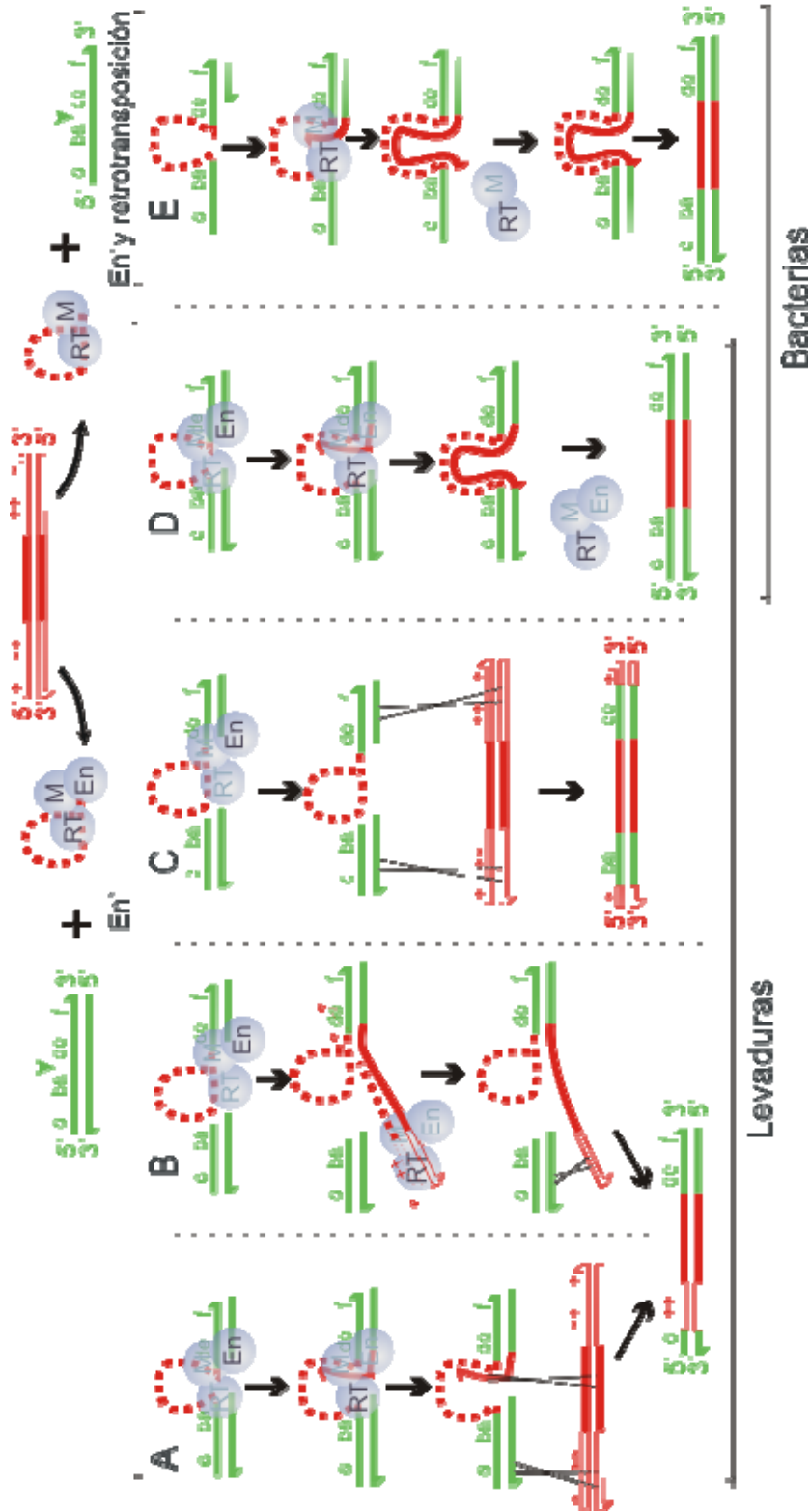
Se han propuesto muchos factores que determinan la elección de la ruta de movilidad en levaduras: la proporción de integración parcial/completa, la velocidad de iniciación de la síntesis del ADNc, la disociación de la RT, la degradación del ARN, el corte en la cadena de abajo, los sistemas de recombinación/reparación de la mitocondria, e incluso, la naturaleza del ADN diana y del propio intrón (Eskes *et al.*, 2000).

### 2.2.3. *Retrohoming* en intrones bacterianos

Aunque el trabajo desarrollado en levaduras ha permitido esclarecer muchos aspectos de la movilidad de los intrones del grupo II, su estudio ha estado supeditado a procesos muy particulares que tienen lugar en las mitocondrias. Los sistemas procariotas presentan menos complejidad, pero los intrones bacterianos también presentan sus particularidades. Una de las principales nos obliga a distinguir un mecanismo de *retrohoming* diferente a los anteriormente descritos: hay intrones cuya IEP carece de dominio ADN endonucleasa. La otra (ausencia de co-conversión) advierte de la existencia de una nueva ruta para la movilidad de los intrones bacterianos.

---

<sup>26</sup> Este tipo de enlaces suponen una fuerte parada de muchas RT (Lorsch *et al.*, 1995); algunas de ellas se desenganchan, pero otras son capaces de atravesarlo.



**Figura 11.** Mecanismos de movilidad propuestos para intrones del grupo II. Los tres mecanismos de la izquierda se han descrito únicamente en levaduras; en todos los casos dependen de recombinación homóloga. Se observa co-transcripción sólo del exón 5' (A y B) o de ambos exones (C). A (mayoritariamente asignado a aI1) y/o B (aI2 presenta un mayor porcentaje de invasión del ARN por *splicing* reverso parcial) se corresponden con la ruta mayoritaria (60%) que se describe en levaduras, mientras C representa la ruta minoritaria (40%) independiente de RT. El diagrama D se corresponde con la ruta mayoritaria descrita para *L.L.itrB* y, en determinadas circunstancias, también se manifiesta en levaduras. Se trata de un mecanismo iniciado por TPRT que se resuelve gracias a la maquinaria de reparación celular, independiente de recombinación. En E se representa el mecanismo de *retrohoming* para intrones que carecen de dominio ADN endonucleasa y la hipótesis más aceptada para la retrotransposición. Se produce invasión sobre ADNsc y probablemente está ligado a replicación.

Por otro lado, el desarrollo de sistemas de doble plásmido para el estudio de la movilidad de estos intrones ha facilitado enormemente el trabajo (Cousineau *et al.*, 1998; Martínez-Abarca *et al.*, 2000; Guo *et al.*, 2000). Se trata de un plásmido donador de intrón que contiene la secuencia del intrón completo flanqueado por sus exones correspondientes para permitir un eficiente *splicing*, y de un plásmido receptor que presenta la diana del intrón, es decir, los exones ligados. El diseño del sistema es muy diverso y ha ido evolucionando hacia plásmidos cada vez más complejos que faciliten en mayor medida la detección de los productos de invasión. Las líneas fundamentales de trabajo han ido encaminadas a la construcción de plásmidos donadores más competentes que permitan una mayor frecuencia y eficiencia de *homing*<sup>27</sup> y al diseño de plásmidos aceptores invadidos más fácilmente seleccionables. En cuanto a la mejora de los donadores, se ha sustituido el intrón completo por construcciones en las que el intrón se divide en sus dos componentes: la ribozima se deletorea de la mayor parte del dominio IV (es lo que se ha denominado  $\Delta$ ORF), mientras que la IEP se expresa en *cis* (aguas arriba o aguas abajo de la secuencia del  $\Delta$ ORF) (Guo *et al.*, 2000; Plante y Cousineau, 2006; Nisa-Martínez *et al.*, 2007). Esta modificación permite la introducción en el dIV de nueva información como marcadores de resistencia (resistencia a kanamicina, Km<sup>R</sup>, Cousineau *et al.*, 1998) o secuencias promotoras (promotor de la ARN polimerasa del fago T7, pT7, Guo *et al.*, 2000) que ayuden a la selección del producto de *homing*. Además la reducción de las secuencias de los exones al mínimo necesaria para la maduración del intrón, aumenta la eficiencia del proceso (Guo *et al.*, 2000; Nisa-Martínez *et al.*, 2007). Todas estas modificaciones en el donador parecen ir encaminadas a una mayor estabilización del sistema, particularmente podrían hacer al ARN menos susceptible de degradación (Guo *et al.*, 2000).

### 2.2.3.1. Retrohoming mediado por TPRT

Este mecanismo está extensamente estudiado para el intrón de *L. lactis*, Ll.ltrB. Es similar al que se ha descrito en levaduras y, por consiguiente, al de retrotransposones no-LTR (Curcio y Belfort, 1996). La principal diferencia respecto al primero es que no presenta transferencia de exones (Cousineau *et al.*, 1998); en relación al segundo, hay que puntualizar que la movilidad en Ll.ltrB está mediada por rotura en las dos cadenas en una localización muy específica, mientras que la mayoría de las endonucleasas codificadas por los rTn no-LTR cortan de manera inespecífica permitiéndoles la inserción en cualquier lugar del genoma (Eickbush, 1999).

Ll.ltrB es un intrón móvil del grupo IIA descrito inicialmente en *L. lactis* (Mills *et al.*, 1996; 1997; Shearman *et al.*, 1996), pero que también es capaz de madurar y moverse en *E. coli* (Matsuura *et al.*, 1997; Karberg, 2005). La IEP que codifica, LtrA, presenta

---

<sup>27</sup> La frecuencia de invasión se considera el porcentaje de la población total de células que ha sufrido, al menos, un evento de *homing*. Se habla en términos de eficiencia para referirnos al nivel de invasión que presentan las dianas disponibles dentro de una misma célula.

actividades RT, madurasa y ADN endonucleasa al organizarse en las RNPs (Saldanha *et al.*, 1999). La reacción de movilidad se inicia, una vez reconocida la diana, con la inserción completa del intrón en el ADN justo en el punto de unión de los exones. A continuación, tienen lugar una serie de reorganizaciones conformacionales (Aizawa *et al.*, 2003) que desembocan en el corte de la cadena antisentido en la posición +9 por parte del dominio En de LtrA (Matsuura *et al.*, 1997). Entonces, la RT posicionada para iniciar la reversotranscripción se sirve del 3'OH generado como cebador para la síntesis del ADNc y usa como molde el ARN del intrón insertado en la hebra sentido. La síntesis del ADNc dentro del exón 1 puede estar facilitada bien por la acción de una helicasa que desplace la cadena de abajo, o bien por la acción degradadora de nucleasas celulares (Belfort *et al.*, 2002; Cousineau *et al.*, 1998). En virtud de que las RTs no son buenas ADN polimerasas dependientes de ADN (Muñoz-Adelantado *et al.*, 2003; Zhong y Lambowitz, 2003), lo más probable es que las nucleasas celulares se encarguen de eliminar el extremo 3' protuberante de la cadena de abajo hasta el sitio de inserción del intrón (recordar que la ADN endonucleasa corta en la posición +10 del exón 2, por lo que encontramos estos 10 nt desapareados unidos al exón 1), de manera que la RT sólo polimeriza hasta el inicio del intrón. El ARN insertado en la cadena superior es degradado probablemente por una ARNasa H celular y tiene lugar la síntesis del ADN complementario a la hebra de abajo (el ADNc recién sintetizado por la RT) a partir del 3'OH libre en el exón 1. Este proceso correría a cargo de una ADN polimerasa del hospedador. Finalmente, sólo queda la reparación de las muescas en los extremos de las polimerizaciones que, de la misma manera que los dos procesos anteriores, también será llevada a cabo por la maquinaria celular. A diferencia de levaduras, los últimos pasos de integración del intrón en su diana ocurren por procesos de reparación del ADN (Smith *et al.*, 2005) y no por recombinación: el *retroboming* de Ll.ltrB es independiente de RecA (Cousineau *et al.*, 1998). Es conveniente mencionar que este mismo mecanismo de movilidad sin co-conversión de los exones ha sido también descrito en levaduras, pero representa niveles ínfimos en condiciones normales. Lo más llamativo es que la proporción de estos eventos aumenta hasta el 43% cuando el ADN diana se encuentra mutado en determinadas posiciones: -8 en el exón 1 y +2 y +5 en el exón 2. Se sabe además que estas mutaciones potencian el *splicing* reverso en aI2 y que, a su vez, incrementan la proporción del proceso completo respecto al parcial, por lo que se deduce que cuando tiene lugar la inserción completa del intrón, la vía descrita para el intrón de *Lactococcus* es la más factible también en mitocondrias de levaduras (Eskes *et al.*, 2000).

### 2.2.3.2. Movilidad independiente de actividad ADN endonucleasa

Ninguno de los mecanismos descritos anteriormente son susceptibles de ser llevados a cabo por intrones cuya IEP carezca de dominio ADN endonucleasa como consecuencia de su manifiesta incapacidad para generar el cebador que permita iniciar el *retroboming* por TPRT. A pesar de ello, son muchos los intrones bacterianos en los que este dominio está

ausente, pero esto no impide que algunos se muevan eficientemente a nuevas localizaciones.

Para tratar de caracterizar el mecanismo por el estos intrones son capaces de moverse, Lambowitz y colaboradores construyeron un mutante de L1.ltrB que, sin tener afectadas otras actividades, era incapaz de producir corte en la cadena de abajo (San Filippo y Lambowitz, 2002; Zhong y Lambowitz, 2003). En este mutante, YRT, se reduce considerablemente el *retroboming* y a ello hay que sumar la particularidad de que pasa a depender de la replicación celular. Además muestra una clara tendencia a insertarse en la cadena que sirve como molde para la síntesis de la hebra conductora durante la replicación, lo que sugiere que esta cadena naciente podría ser utilizada por la RT como cebador para iniciar la síntesis del ADNc. También existe movilidad sobre dianas que se encuentran en la otra hebra, pero es residual. En cualquier caso, parece ser que la ADN polimerasa debería desengancharse al encontrarse con la RNP, dejando la cadena naciente en posición para permitir la reverso-transcripción a la IEP (el *retroboming* es absolutamente dependiente de RT), para después volver a la horquilla de replicación y completar el proceso (Zhong y Lambowitz, 2003).

Sin embargo, un estudio de la localización respecto del origen de replicación de los intrones del grupo II que carecen de dominio En de forma natural, muestra que la mayoría de estos intrones se encuentran insertados en la hebra que sirve de molde para la síntesis de la cadena retardada durante la replicación (justo la orientación opuesta a la predicha con el mutante YRT de L1.ltrB; Ichiyana *et al.*, 2002). La explicación para esta observación, es que los intrones En<sup>-</sup> son capaces de insertarse sólo cuando el ADN se encuentra como cadena sencilla, una vez que la horquilla de replicación ha atravesado el sitio de inserción, y que usan como cebador la cadena retardada naciente (no se ha determinado si usarían los cortos fragmentos de ARN sintetizados por la primasa o los propios fragmentos de Okazaki; Martínez-Abarca *et al.*, 2004). Se ha especulado que podrían existir algún tipo de interacciones con la maquinaria de replicación del hospedador que faciliten el *retroboming* a estos intrones, pero aún no se ha encontrado más relación que ser uno de los momentos en el que parte del genoma se encuentra como ADNsc. Se han sugerido otros mecanismos que permitan iniciar la síntesis del ADNc como, por ejemplo, muescas al azar en la cadena de abajo (intrón *cob-I1* de *Schizosaccharomyces pombe*, Schäfer *et al.*, 2003) o iniciación de novo (RT codificada por el plásmido Mauriceville en mitocondria de *Neurospora*, Wank y Lambowitz, 1993).

El caso más estudiado de este tipo de intrones es RmInt1, localizado en una secuencia de inserción en el genoma de *Sinorhizobium meliloti* (Martínez-Abarca *et al.*, 1998). Su frecuencia de *retroboming* es muy alta, similar al del intrón de *Lactococcus* o a los de levadura (Martínez-Abarca *et al.*, 2000). Sin embargo, aunque el mecanismo de reconocimiento de la diana también es parecido, la IEP podría ser menos determinante (Jiménez-Zurdo *et al.*, 2003). Aunque se ha demostrado que es capaz de insertarse *in vitro* en ADN de cadena simple y doble (con mucha menor eficiencia, y tanto menos cuanto más largo es el sustrato), se ha comprobado que no es capaz de cortar la cadena de abajo (Muñoz-

Adelantado *et al.*, 2003). Este intrón también muestra una clara tendencia a insertarse en la hebra molde para la síntesis de la hebra retardada, aunque en menor frecuencia también se inserta en la otra cadena (Martínez-Abarca *et al.*, 2004; Nisa-Martínez *et al.*, 2007).

Otro caso no menos curioso es el de los intrones de la clase IIC, que también carecen de dominio ADN endonucleasa aunque conservan parcialmente el dominio de unión a ADN. El reconocimiento tan particular de su diana ha despertado el interés de los investigadores que han llegado a proponer un posible mecanismo de movilidad relacionado no sólo con la replicación sino con la naturaleza funcional de los terminadores bajo los que se localizan (Granlund *et al.*, 2001; Michel *et al.*, 2007).

### 2.3. MECANISMO DE RETROTRANSPOSICIÓN

A pesar de las múltiples rutas de *homing* propuestas en los diferentes organismos, distintos estudios han llevado a la conclusión que, en la mayoría los casos, existe una única vía para la transposición ectópica de los intrones del grupo II (Dickson *et al.*, 2001; Ichiyangi *et al.*, 2002; 2003). La identificación de los primeros eventos de retrotransposición en intrones mitocondriales se realizó por PCR y una de las observaciones más importantes fue que, a diferencia del *homing*, no existía co-conversión de los exones flanqueantes (Mueller *et al.*, 1993b; Sellem *et al.*, 1993). Así, el mecanismo inicialmente propuesto para este fenómeno dependiente de RT fue la integración del ARN del intrón por *splicing* reverso sobre un ARN que contenga los exones ligados, seguido de la síntesis del ADNc por parte de la RT IEP y la resolución por recombinación homóloga. Los primeros estudios en Ll.ltrB parecían refrendar esta hipótesis (Cousineau *et al.*, 2000). Sin embargo, trabajos posteriores proporcionan una base sólida para afirmar que, también en la retrotransposición, la inserción del ARN tiene lugar a nivel de ADN (Yang *et al.*, 1998; Cousineau *et al.*, 2001 erratum; Dickson *et al.*, 2001; Ichiyangi *et al.*, 2002). Además, al menos en intrones bacterianos, se ha demostrado que es un proceso independiente de recombinación homóloga (Martínez-Abarca y Toro, 2000b; Ichiyangi *et al.*, 2002). Pronto se dieron cuenta de que, en la mayoría de estos eventos, la única secuencia de reconocimiento que se observa es la interacción EBS1-IBS1; la conservación del EBS2 y de las posiciones distales en los exones 1 y 2 es mínima en la mayoría de los casos (Yang *et al.*, 1998; Cousineau *et al.*, 2000). En este punto debemos recordar que la interacción de la proteína con las regiones distales de ambos exones permite la separación de la doble hélice y el corte en la cadena de abajo por parte del dominio En (Singh y Lambowitz, 2001), por lo que el dato anterior nos indicaría que para la transposición ectópica se precisa el ADNsc. Una observación adicional es que este proceso se produce en ausencia de actividad ADN endonucleasa (Cousineau *et al.*, 2000). Si a todos los resultados anteriores le sumamos que existe una clara tendencia a la invasión de sitios no específicos situados en la cadena que sirve como molde para la síntesis de la hebra retardada durante la replicación (Ichiyangi *et al.*, 2002; 2003), la conclusión parece obvia: el mecanismo principal por el cual tiene lugar la retrotransposición en todos los intrones hasta ahora

caracterizados se asemeja al *retrohoming* de los intrones que carecen de dominio En y se le ha llamado mecanismo de *retrohoming* ilegítimo (ver apartado anterior 2.2.3.2.) (Dickson *et al.*, 2001; Ichiyanagi *et al.*, 2002; 2003; Belfort *et al.*, 2002).

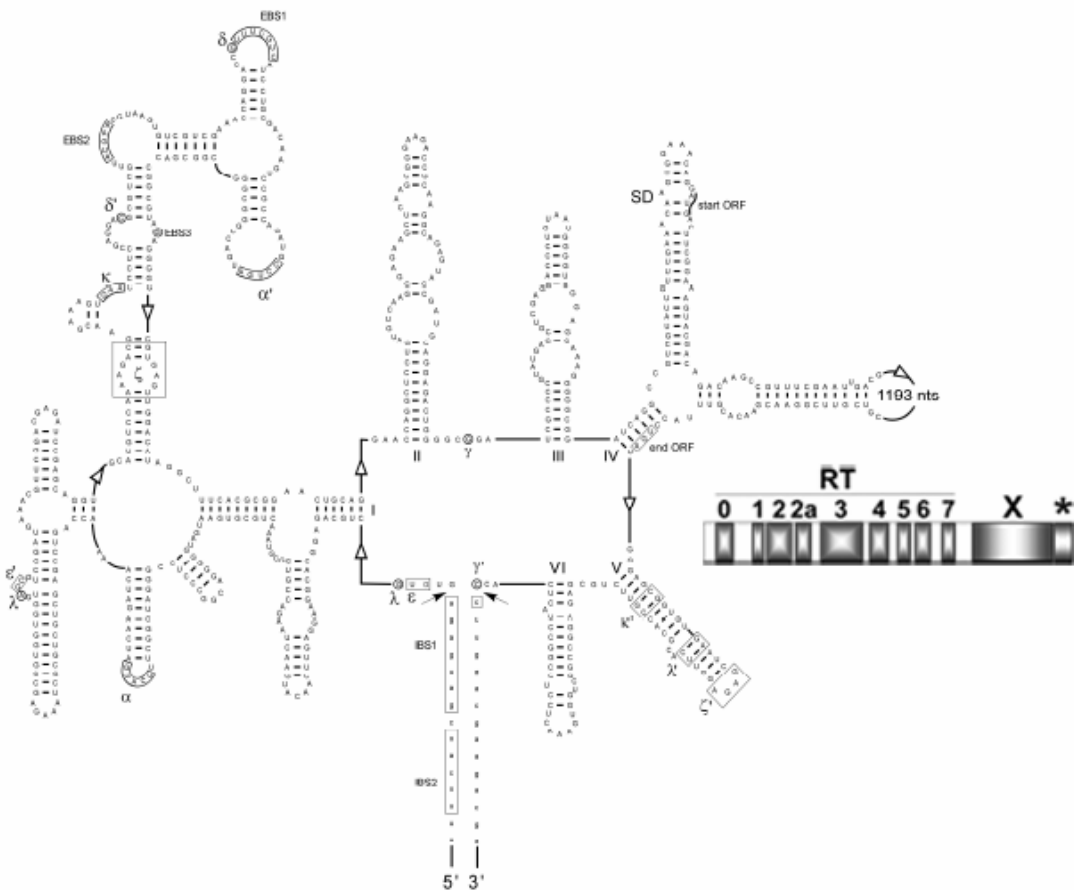
Finalmente, estudios muy recientes apuntan a que el mecanismo de retrotransposición depende del organismo hospedador ya que se ha observado un comportamiento diferencial de Ll.ltrB en *L. lactis* y en *E. coli* (Coros *et al.*, 2005). Mientras en *Lactococcus* sigue la vía descrita, en *E. coli* parece producirse la inserción del intrón también sobre ADNdc. La explicación de esta observación es que a diferencia de lo descrito anteriormente, la conservación tanto de los EBSs como de las posiciones más relevantes para el reconocimiento por parte de la IEP es mucho mayor que la descrita en *Lactococcus*. Se describe una preferencia de ruta en *E. coli* ligada a las condiciones de crecimiento de las células. Cuando se replican lentamente, prefieren integrarse en ADN de doble cadena mostrando un curioso patrón, cerca del origen y el final de la replicación. Este patrón se ha justificado en virtud de la localización polar de la IEP en este tipo de bacterias y de la posición de estas regiones durante la replicación (Zhao y Lambowitz, 2005; Beauregard *et al.*, 2006). Si crecen rápidamente, el mecanismo de invasión sobre ADNsc es más frecuente y las inserciones se localizan dispersas a lo largo de todo el genoma; en esta ocasión, replicación y división celular ocurren simultáneamente y el cromosoma está mucho menos compacto (Coros *et al.*, 2005).

### 3. RmInt1, EL INTRÓN DEL GRUPO II DE *Sinorhizobium meliloti*

RmInt1 es el primer intrón descrito en el orden Rhizobiales. Se identificó al analizar la secuencia de una región del plásmido críptico pRmeGR4b presente en la cepa GR4 de la bacteria *Sinorhizobium meliloti*, un microorganismo capaz de fijar de nitrógeno en simbiosis con alfalfa (Martínez-Abarca *et al.*, 1998). Siguiendo el comportamiento de los intrones bacterianos (presumiblemente para facilitar su dispersión), se suele encontrar asociado a la secuencia de inserción ISRm2011-2, aunque también se puede encontrar invadiendo otras ISs relacionadas (ISRm10-1 e ISRm10-2, Martínez-Abarca y Toro, 2000b; ISRm10-3; Muñoz-Adelantado, 2003) e incluso se ha localizado otro sitio de transposición ectópica en el gen *oxi1* (Martínez-Abarca y Toro, 2000b). Es interesante destacar que RmInt1 se ha identificado en el 90% de los aislados de suelo de *S. meliloti* y generalmente suele encontrarse en alto número de copias por genoma (de 0 a 11 copias) (Martínez-Abarca *et al.*, 1998). Además se encuentra ampliamente distribuido entre otras especies de la misma familia de acuerdo a patrones de herencia vertical y transferencia horizontal, aunque se observa una tendencia a la inactivación por pérdida del extremo 3' del intrón (Fernández-López *et al.*, 2005).

La estructura secundaria de la ribozima se ajusta a la descrita para intrones del grupo IIB3 (Martínez-Abarca *et al.*, 1998). Se ha puesto de manifiesto su capacidad para madurar *in vitro* en ausencia de la proteína dando lugar a moléculas en forma *lariat* pero también en forma lineal, dependiendo de las condiciones de reacción (Costa *et al.*, 2006a). *In vivo*, se

ha determinado la presencia de formas escindidas por extensión a partir de cebador, pero esta técnica no permite discriminar el tipo de producto de *splicing* ni tampoco detectar la presencia de exones ligados (Muñoz-Adelantado *et al.*, 2003). Por otro lado, la proteína codificada por RmInt1 presenta los 7 dominios característicos de este tipo de RT junto con el dominio X (madurasa), pero carece de los dos dominios presentes en el C-terminal de intrones IIA: dominio de unión a ADN y dominio ADN endonucleasa (Martínez-Abarca *et al.*, 2000). En su lugar presenta una corta extensión de 20 aminoácidos cuya función se desconoce (San Filippo y Lambowitz, 2002), aunque se ha propuesto que podría participar en el reconocimiento de las regiones distales del ADN diana (Jiménez-Zurdo *et al.*, 2003). En función de la distribución de dominios RT, RmInt1 se enmarca dentro de los intrones bacterianos de la clase D (Toro *et al.*, 2002).



**Figura 12.** Estructura secundaria propuesta para RmInt1. Junto a ésta se muestran los dominios que conforman la IEP de este intrón. La región C-terminal se señala con un \* porque se desconoce su función.

Con todo, RmInt1 es uno de los pocos intrones que conservan la capacidad de movimiento en el genoma de su hospedador (Martínez-Abarca *et al.*, 2000). Las primeras pruebas de este fenómeno se obtuvieron en sistemas de salto desde el genoma a un plásmido que contenían su secuencia diana (Martínez-Abarca *et al.*, 2000; Jiménez-Zurdo



*et al.*, 2003). El siguiente paso fue el diseño de un sistema sencillo de *homing* de plásmido (donador) a plásmido (receptor) (Martínez-Abarca *et al.*, 2000), sobre el que se sigue trabajando con el fin de mejorar sus prestaciones (Nisa-Martínez *et al.*, 2007; Barrientos-Durán, 2008). En todos los casos, se comprobó que se trataba de un mecanismo independiente de RecA (Martínez-Abarca *et al.*, 2000; Martínez-Abarca y Toro, 2000b) y mediado por un intermediario de ARN (Martínez-Abarca *et al.*, 2004).

En nuestro laboratorio se han desarrollado numerosos estudios para profundizar en los mecanismos moleculares implicados en la maduración y el movimiento del intrón, tanto *in vivo* (Jiménez-Zurdo *et al.*, 2003) como *in vitro* (Muñoz-Adelantado *et al.*, 2003). Los estudios *in vitro* muestran que las RNPs de este intrón tienen actividad reverso transcriptasa usando un sustrato exógeno poli(rA)/Oligo(dT), aunque la mera adición de cebadores es suficiente para que se produzca polimerización (Muñoz-Adelantado *et al.*, 2003). Se ha comprobado que son capaces de insertarse en sustratos de ADN de cadena doble y de cadena sencilla mediante una reacción de *splicing* reverso, pero, como cabría esperar, no producen corte de la cadena de abajo (Muñoz-Adelantado *et al.*, 2003). En cuanto a los requerimientos para el *retrohoming in vivo*, se ha determinado que la diana mínima necesaria para unos niveles adecuados de movilidad se extiende desde la posición -20 en el exón 1 hasta la posición +5 en el exón 2. La mutación de los IBSs en esas dianas ha permitido comprobar que el IBS1 es menos restrictivo para el reconocimiento durante el *homing* que el IBS2 o que el IBS3 (Jiménez-Zurdo *et al.*, 2003). Por el contrario, los estudios desarrollados en los últimos años en relación con los requerimientos de secuencia para el *splicing in vivo* muestran justamente la especificidad inversa: la eficiencia de maduración de RmInt1 no está tan condicionada por la identidad del IBS2 o el IBS3 como por la secuencia IBS1 (Barrientos-Durán, 2008).

Estudios basados en la pauta de invasión de plásmidos con replicación unidireccional tipo  $\theta$  (Martínez-Abarca *et al.*, 2004) así como en la dispersión del intrón en el genoma de la bacteria tras haber sido transformada con un plásmido donador (Nisa-Martínez *et al.*, 2007), muestran una clara tendencia para la invasión de la hebra que sirve como molde en la síntesis de la cadena retardada durante la replicación. También es capaz de insertarse en la otra cadena pero con mucha menor eficiencia. Por tanto, RmInt1 es el principal representante de los intrones cuya ruta de *retrohoming* es independiente de actividad En (Pyle y Lambowitz, 2006; Toro *et al.*, 2007).

#### **4. EL PASADO Y EL FUTURO DE INTRONES DEL GRUPO II**

Una vez que sabemos cómo funcionan resultará más sencillo comprender su evolución: desde su origen hasta su posible fin como elementos móviles (pero no como secuencias que proporcionan plasticidad a los genomas). No menos interesante resulta la elaboración de herramientas que nos permitan el estudio y modificación de los genomas, sirviéndonos de esas cualidades que les han permitido dispersarse en los distintos reinos de la vida.

#### 4.1. EVOLUCIÓN DE ESTOS RETROELEMENTOS

En general, las cuestiones evolutivas suelen ser muy controvertidas y este caso no podía ser una excepción. Nuestro propósito se ha centrado en buscar una versión coherente que explique todos los estadios en la evolución de los intrones, desde su origen hasta sus posibles “descendientes”.

Uno de los aspectos en que muchos autores parecen coincidir es que, de acuerdo al patrón de distribución filogenética, los intrones del grupo II han evolucionado en bacterias y luego se han dispersado por los genomas eucariotas y las arqueas. Esto es lo que se conoce como la “hipótesis del retroelemento ancestral”: todos los intrones derivan de otros presentes en bacterias que tienen asociadas proteínas (concretamente del tipo RT); posteriormente, tiene lugar la divergencia en los diferentes linajes y/o la pérdida o degeneración de estas IEPs (Toor *et al.*, 2001; Hausner *et al.*, 2006; Robart y Zimmerly, 2005). Se cree que la transmisión de los intrones a eucariotas se produjo en concomitancia con la endosimbiosis que dio origen a las mitocondrias y los cloroplastos (Cavalier-Smith, 1991; Palmer y Logsdon, 1991), sin embargo su aparición en arqueas se ha atribuido a recientes eventos de transferencia horizontal (Rest y Mindel, 2003; Toro, 2003; Dai y Zimmerly, 2003). Fenómenos de herencia lateral también han sido extensamente descritos entre las propias eubacterias (Martínez-Abarca y Toro, 2000a; Zimmerly *et al.*, 2001; Belhocine *et al.*, 2004; Fernández-López *et al.*, 2005) y recientemente en un anélido del género *Nephtys* (Vallès *et al.*, 2008).

El problema surge cuando se intenta encontrar el origen de los intrones en bacterias. Algunos autores los consideran como auténticas reliquias de un mundo ARN (Lambowitz y Belfort, 1993); se trata de ARNs catalíticos preparados para llevar a cabo la reacción de maduración pero que han debido adecuarse para funcionar como elementos móviles (reclutamiento de proteínas que completen estas funciones) (Kennell *et al.*, 1993; Wank *et al.*, 1999). Otra posibilidad es que se trate de un retroelemento que, tratando de minimizar su efecto deletéreo para el hospedador, se insertó dentro de una secuencia capaz de madurar, un intrón del grupo II (Curcio y Belfort, 1996). Una de las alternativas recientemente planteadas más interesantes (quizá relacionada con la primera posibilidad apuntada) es que estas ribozimas se encontraran inicialmente en microorganismos termófilos o halófilos atendiendo a las necesidades que muestran estos intrones para una eficiente auto-maduración *in vitro*; la adquisición de la IEP les abre el camino a la invasión de nuevos genomas bacterianos (Lambowitz y Zimmerly, 2004).

Cualquiera que fuese el origen de estos intrones, la proteína parece haberse adquirido con un claro propósito, convertirlos en elementos móviles (Kennell *et al.*, 1993). En paralelo, la interacción tan específica que mantienen con el ARN del intrón, convierte estas IEPs en piezas clave necesarias para su maduración. La adquisición de estas proteínas parece haber ocurrido una única vez puesto que en todos los intrones del grupo II se encuentran codificadas en el mismo dominio, el dIV, interfiriendo en el menor grado posible con la maduración (Curcio y Belfort, 1996). Se cree que las proteínas más ancestrales son

aquellas que únicamente contienen los dominios RT y X, permitiendo que los intrones que las portan se muevan sobre ADNsc según el mecanismo descrito para intrones que carecen de actividad ADN endonucleasa. La adquisición de las actividades asociadas a los dominios D y E favorecen la inserción de estos intrones sobre ADNdc, de manera que dejan de depender de la replicación de su hospedador para generar el cebador que les permita transponerse (Zimmerly *et al.*, 2001; Lambowitz y Zimmerly, 2004).

Hasta ahora, podemos deducir que el comportamiento de estos intrones ha ido encaminado a incrementar su supervivencia en la población (Dai y Zimmerly, 2002). Pero, ¿qué pasaría si uno de estos elementos se inserta dentro de un gen funcional? La respuesta no ha lugar a dudas: ese intrón está obligado a madurar, debe conservar esta capacidad o de lo contrario su hospedador no puede sobrevivir. Así pues, la estrategia desarrollada por los diferentes hospedadores fue sustituir la actividad de la IEP por factores celulares que permitieran a estos organismos un mayor control sobre el *splicing* de los intrones. Esto es justamente lo se deduce del análisis de algunos de los intrones presentes en orgánulos, donde la IEP ha sido sustituida por una proteína con función madurasa (por ejemplo, nMat en plantas superiores). Además, parece ser que se producen distintos patrones de degeneración estructural del ARN dependiendo de las proteínas que el hospedador ofrezca a los intrones para su maduración. De manera que cuanto mayor sea la variación de la estructura respecto al consenso para intrones del grupo II, mayor será el número de proteínas requeridas para estabilizar el proceso, proteínas que además deberán coordinarse y formar complejos (Lambowitz y Zimmerly, 2004). Esta afirmación está basada en el estudio de los intrones fragmentados presentes en cloroplastos de algas (*Chlamydomonas*; Goldschmidt-Clermont *et al.*, 1991; Perron *et al.*, 2004) y plantas terrestres (Bonen, 1993; Bonen, 2008) y en mitocondrias de angiospermas (petunia y trigo, Bonen, 1993; *Oenothera*, Knoop *et al.*, 1997; Qiu y Palmer, 2004). Se trata de intrones que se encuentran divididos en dos o tres partes localizadas en diferentes regiones del genoma del orgánulo, se ensamblan asistidos por proteínas y son capaces de funcionar. Se ha postulado que esta fragmentación podría explicar una de las principales ventajas que hoy se asocian a la presencia de intrones en los genomas eucariotas, el barajamiento de exones (Hiller *et al.*, 2000; Belfort *et al.*, 2002). En cuanto a degeneración de la estructura general del intrón (que no fragmentación), el ejemplo más espectacular viene representado por los intrones del cloroplasto de *E. gracilis*, donde en ocasiones sólo se han logrado identificar los dominios V y VI de la ribozima (Copertino y Hallick, 1993).

La pérdida de la IEP está inevitablemente asociada a la inmovilidad de estos intrones. La fragmentación y la pérdida de movilidad es lo que presumiblemente ha ocurrido en los intrones espliceosómicos (Cech, 1986; Sharp, 1991). La hipótesis de que los intrones del grupo II han dado lugar a los intrones espliceosómicos nucleares se encuentra sustentada fundamentalmente por el paralelismo funcional/catalítico existente entre ellos (Sharp, 1985; Cech, 1986). Esta teoría cobra cada vez más fuerza basándose en las similitudes estructurales que se han descrito entre los diferentes motivos del centro activo de los intrones del grupo II y las subestructuras formadas por los distintos ARN pequeños

nucleares (ARNsn) al interactuar en el espliceosoma (Jacquier, 1990; Madhani y Guthrie, 1992; Sashital *et al.*, 2004). De hecho, se han descrito evidencias de complementación funcional entre distintas estructuras del espliceosoma e intrones del grupo II (Hetzer *et al.*, 1997a; Shukla y Padgett, 2002). Muchos autores consideran que los intrones se encontraron con una importante barrera para la maduración y la movilidad en organismos eucariotas: la membrana nuclear. Esta membrana separa la transcripción de la traducción desde el punto de vista espacial pero también temporal, de manera que la IEP no podría unirse al ARN del intrón justo tras la transcripción. Esto podría haber conllevado el desarrollo de un aparato de *splicing* común para todos los genes formado por proteínas del hospedador que funcionan eficientemente en *trans* (Cavalier-Smith, 1991; Palmer y Logsdon, 1991). Otros investigadores se atreven a afirmar que fueron precisamente los intrones los que obligaron al primitivo proto-eucariota a adquirir muchas de las características que hoy usamos para definir la célula eucariota con el fin de defenderse de la indiscriminada invasión por parte de estos elementos genéticos móviles (Koonin, 2006).

Por otro lado, aquellos intrones que no se hayan insertado dentro de genes no se encuentran bajo una presión selectiva tan fuerte para mantener el *splicing*, por lo que se considera que podrían haber evolucionado para dar lugar a los retrotransposones no-LTR (Zimmerly *et al.*, 1995b; Eickbush, 1994; 2000). Si las asunciones expuestas aquí son correctas, la impronta que los intrones del grupo II han dejado, por ejemplo en el genoma humano, es digna de consideración: la mitad de nuestro material genético está formado por la suma de estos dos tipos de secuencias, intrones y RTn no-LTR (LINEs, SINEs) (Pyle y Lambowitz, 2006).

#### 4.2. APLICACIONES EN BIOTECNOLOGÍA

Como en el caso de muchas otras ribozimas (Mills, 2001; Lewin y Hauswirth, 2001; Burt, 2003; Puerta-Fernández *et al.*, 2003; Long *et al.*, 2003), los intrones del grupo II se están desarrollando con el fin de ser usados como herramientas biotecnológicas eficientes, de aplicabilidad tanto en procariontes como en eucariotas (Cui y Davies, 2007). El uso de estos retroelementos proporciona una serie de ventajas sobre otras tecnologías ya desarrolladas para la manipulación del ADN (Cui y Davies, 2007). Estas ventajas están en gran medida relacionadas con las características particulares de estos intrones (Belfort *et al.*, 2002; Lambowitz y Zimmerly, 2004; Robart y Zimmerly, 2005; Pyle y Lambowitz, 2006; Plante y Cousineau, 2006; Toro *et al.*, 2007). En primer lugar, son elementos móviles capaces de integrarse específicamente en su ADN diana con alta eficiencia mediante un proceso independiente de recombinación homóloga. A este hecho podemos sumar el que la inserción es muy estable, de manera que se mantiene en sucesivas generaciones. Por otro lado, el reconocimiento de la diana se conoce relativamente bien;

se produce principalmente por apareamiento de bases<sup>28</sup>, por lo que la especificidad de la diana puede ser modificada al cambiar las secuencias EBSs dentro del ARN del intrón. Además, en las versiones modificadas más recientes (aquellas en que la IEP es expresada en *trans* fuera del ARN del intrón, versiones  $\Delta$ ORF), el intrón puede portar información dentro de determinadas regiones del dominio IV de la ribozima. Por último, aunque es muy probable que la finalización del proceso dependa de funciones celulares, estos requerimientos son mínimos y probablemente impliquen a genes del hospedador esenciales en otros procesos.

Los intrones del grupo II se están usando con éxito para la modificación de genes en un amplio número de especies bacterianas, tanto Gram-positivas como Gram-negativas, y se cree que es extensible a muchas más. Además se ha erigido como un sistema especialmente útil en aquellas bacterias en las que resulta complicado la manipulación genética (Chen *et al.*, 2005). La utilidad potencial de estos intrones puede abarcar desde la interrupción al azar de genes (Guo *et al.*, 2000; Zhong *et al.*, 2003; Perutka *et al.*, 2004; Yao *et al.*, 2005) hasta la invasión dirigida de una secuencia particular (Guo *et al.*, 2000; Karberg *et al.*, 2001; Frazier *et al.*, 2003; Zhong *et al.*, 2003; Chen *et al.*, 2005; Yao *et al.*, 2006) pasando por la introducción de información genética de utilidad (por ejemplo, un gen funcional) (Zhong *et al.*, 2003; Jones *et al.*, 2005; Rawsthorne *et al.*, 2006; Plante y Cousineau, 2006) o la inducción de la reparación de un gen mediante recombinación estimulada por rotura de la doble cadena (DSBR) (Karberg *et al.*, 2001; Zhong *et al.*, 2003). Hasta el momento, el único intrón que se ha manipulado con éxito y que incluso se ha comercializado bajo el nombre de TargeTron® es el intrón de *L. lactis*, Ll.ItrB ([http://www.sigmaaldrich.com/Area\\_of\\_Interest/Life\\_Science/Functional\\_Genomics\\_and\\_RNAi/TargeTron.html](http://www.sigmaaldrich.com/Area_of_Interest/Life_Science/Functional_Genomics_and_RNAi/TargeTron.html)).

La primera demostración de que la modificación dirigida de los EBSs de Ll.ItrB conduce a la inserción del intrón en el gen que nos interese (en este caso fue el gen de la timidilato sintasa, *thyA*) se realizó con un sistema de plásmido a plásmido utilizando como marcador de selección el gen de resistencia a kanamicina (Km<sup>R</sup>) portado por el intrón en su dIV (Cousineau *et al.*, 1998; Mohr *et al.*, 2000). El proceso fue bastante ineficiente por lo que se propusieron modificar el sistema para mejorar el rendimiento del proceso. En muchos casos, se buscó mejorar el límite de detección de los eventos de *homing* (selección de colonias resistentes a tetraciclina, Tc<sup>R</sup>, Guo *et al.*, 2000), en otros se trató de rentabilizar el tiempo para llevar a cabo la modificación de los intrones (PCR en lugar de clonaje, Karberg *et al.*, 2001; Zhong *et al.*, 2003) y en muchos casos se prescindió de un sistema de selección (mapeo por PCR con cebadores específicos, Frazier *et al.*, 2003; Yao *et al.*, 2005). Uno de los pasos más importantes fue la creación de una librería de intrones en *E. coli* cuyos EBSs, la posición  $\delta$  y los IBSs presentes en el exón 1 (con el fin de no afectar el *splicing*) se encontraban modificados aleatoriamente (Karberg *et al.*, 2001).

---

<sup>28</sup> La proteína reconoce determinadas posiciones para facilitar el desenrollamiento del ADN y para producir el corte en la cadena de abajo, pero el escaso número de interacciones que establece permite encontrar potenciales dianas en casi todos los genes (Perutka *et al.*, 2004).

Originalmente se trataba de un sistema de doble plásmido basado en la activación de la resistencia a tetraciclina ( $Tc^R$ ) cuando el intrón se inserta en la diana (Guo *et al.*, 2000). El plásmido receptor contiene una diana (el gen que deseamos interrumpir) y, a continuación, se encuentra la secuencia de un gen de resistencia a tetraciclina carente de promotor<sup>29</sup>. Por su parte, el plásmido donador expresa el intrón bajo el promotor inducible T7lac; este intrón (en versión  $\Delta$ ORF con EBSs,  $\delta$  e IBSs modificados al azar y la IEP expresada aguas abajo del exón 2) porta en su dIV el promotor de la ARN polimerasa del fago T7 de manera que cuando se produce el salto del intrón sobre la diana se activa la resistencia a tetraciclina. De este modo, la eficiencia de *homíng* se puede evaluar directamente a partir de la razón entre el número de colonias  $Tc^R Amp^R$  (plásmido receptor invadido) y el número de colonias  $Amp^R$  (plásmido receptor sin invadir). A partir de las formas invadidas, por secuenciación, se pueden seleccionar las combinaciones de intrones modificados más interesantes, complementar, si fuese necesario, las EBSs con las IBSs del exón 1 para optimizar el *splicing* y, finalmente, utilizarlos para interrumpir la secuencia deseada en el genoma de la célula. Este sistema es muy eficiente, consiguiéndose frecuencias de invasión entre el 1 y el 80%, sin embargo conlleva demasiado tiempo encontrar el intrón óptimo. Por esto, se ha diseñado y validado un programa informático que permite encontrar los posibles sitios diana presentes en el gen de interés y que incluso proporciona los cebadores para la construcción del intrón modificado que allí se inserte (Perutka *et al.*, 2004).

Por otro lado, la introducción de un sistema de selección ligado al salto del intrón y su inserción cromosómica aumenta la frecuencia de detectar la interrupción de un gen al 100%. Con este fin, Zhong y colaboradores insertaron un gen de resistencia a trimetoprima,  $Tp^R$ , en el dIV del intrón en orientación opuesta a la transcripción del  $\Delta$ ORF; este gen, a su vez, se encuentra interrumpido por un intrón del grupo I (*tdI*) dispuesto en la misma dirección de la transcripción (al conjunto se le ha llamado marcador de selección activado por retrotransposición, RAM). Así, la  $Tp^R$  no se activa hasta que el intrón se ha insertado en el cromosoma tras la maduración del *tdI* (Zhong *et al.*, 2003). Posteriormente, el RAM fue flanqueado por sitios FRT para la recombinasa F<sub>1</sub>p, de manera que podrían hacerse múltiples rondas de inserciones usando el mismo gen marcador para la selección. Esto es, tras una primera inserción se expresa en *trans* la recombinasa que va a eliminar el gen de resistencia; ahora tenemos células sensibles al antibiótico que pueden ser sometidas a nuevas inserciones y seleccionar los eventos de salto mediante  $Tp^R$ , repitiendo el proceso cuantas veces consideremos oportuno (Zhong *et al.*, 2003; Cui y Davies, 2007). La mejora más reciente en el desarrollo de los “TargeTrons” ha sido el traslado del sistema a un vector de amplio rango de hospedador

---

<sup>29</sup> Para disminuir el ruido de fondo, rodeando a estas secuencias se encuentran distintos terminadores de transcripción: la diana se encuentra flanqueada aguas arriba por el doble terminador de la ARN polimerasa de *E. coli* y la ARN polimerasa del fago T7 (T1) y aguas abajo por tres copias del terminador de la ARN polimerasa de *E. coli* (T2); mientras, el gen de resistencia está seguido del terminador  $\Phi$  (que termina la transcripción a partir de la ARN polimerasa del fago T7 pero no de la de *E. coli*) (Karberg *et al.*, 2001).

que puede ser potencialmente usado en todas las bacterias que sean capaces de replicarse (Yao y Lambowitz, 2007).

Hasta el momento se ha logrado la inserción cromosómica de intrones diseñados “a la carta” en *E. coli* (*thyA*, *lacZ*, *trpE*, *dadA* y *proA*, Karberg *et al.*, 2001; *lacZ* y *argF*, Zhong *et al.*, 2003; hasta 27 genes que codifican proteínas relacionadas con ADN helicasas y proteínas con motivo DExH/D, Perutka *et al.*, 2004; múltiples genes a lo largo del cromosoma; Yao *et al.*, 2005; *lacZ*, Yao *et al.*, 2007; *cynX*, *cynS* y *cynT*, Cui y Davies, 2007), *Shigella flexneri* y *Salmonella typhimurium* (*thyA*, Karberg *et al.*, 2001), *L. lactis* (*mleS* y *tetM*, Frazier *et al.*, 2003; *tra904*, *tra981* y *tra983*, Rawsthorne *et al.*, 2006), *Staphylococcus aureus* (*hsa* y *seb*, Yao *et al.*, 2006), *Clostridium perfringens* (*plc*, Chen *et al.*, 2005), *Pseudomonas aeruginosa* (*pqsA* y *pqsH*, Yao y Lambowitz, 2007) y *Agrobacterium tumefaciens* (*aopB* y *chvI*, Yao y Lambowitz, 2007). En algunos casos se han usado sistemas basados en RAM, pero, en muchos otros, las inserciones genómicas han conseguido detectarse sin utilizar ningún gen marcador, valiéndose de mapeo por PCR y en algunos casos del propio fenotipo mutante.

Las inserciones pueden clasificarse en dos tipos dependiendo de su reversibilidad. Hablamos de inserciones definitivas o irreversibles cuando el intrón interrumpe la cadena antisentido del gen, esto es, en ningún caso se va transcribir y no podrá haber *splicing*. Sin embargo, cuando la inserción de la versión  $\Delta$ ORF del intrón tiene lugar en la cadena sentido podemos adicionar elementos en *trans* que posibiliten la maduración del intrón, hablándose entonces de inserción condicional o reversible. El factor a añadir en *trans* más evidente es la IEP expresada a partir de vectores inducibles (Frazier *et al.*, 2001; Yao *et al.*, 2006), pero recientes estudios apuntan a que el aumento de la concentración extracelular de  $Mg^{2+}$  libre deriva, en algunos casos, en la maduración del intrón (Cui, 2005). En cualquier caso, es conveniente destacar que estos sistemas de mutagénesis son altamente específicos y que, por regla general, cuando tratan de hacerse inserciones al azar, se produce una inserción por genoma aunque en raras ocasiones se han contabilizado dos (Karberg *et al.*, 2001; Zhong *et al.*, 2003; Perutka *et al.*, 2004; Yao *et al.*, 2005).

Se ha observado un cierto sesgo en la inserción de intrones modificados en el genoma de *E. coli*. La mayoría de ellos se inserta cerca del origen de replicación sugiriendo una relación con la replicación del ADN (Zhong *et al.*, 2003). Algunos autores consideran que, independientemente de la existencia de este sesgo, el sistema es lo suficientemente eficiente para generar inserciones a lo largo del todo el genoma (Yao *et al.*, 2005).

Como ya dijimos, las aplicaciones pueden ser variadas pero debemos tener presente que estos sistemas también presentan limitaciones. Se ha descrito que pueden portar con éxito información en el dIV (Zhong *et al.*, 2003; Jones *et al.*, 2005; Rawsthorne *et al.*, 2006) pero también se ha demostrado que cuando estas secuencias son >1kb la eficiencia de movilidad se ve drásticamente reducida, reflejando probablemente una disminución de la estabilidad de estos intrones (Plante y Cousineau, 2006). Bien diseñados, se presentan como una estrategia válida para construir cepas superproductoras; si se introduce un gen de interés en el dIV del intrón y éste se modifica para insertarse en secuencias que se

encuentren muy repetidas en el genoma hospedador, tras sucesivas rondas podemos conseguir una cepa que exprese varias copias de nuestro gen (Rawsthorne *et al.*, 2006). Se está trabajando intensamente en la adaptación de estos sistemas para su uso en células eucariotas (Guo *et al.*, 2000; Cui, 2005). La mayoría de los eucariotas carecen de sistemas eficientes de recombinación que permitan la modificación de genes por las técnicas convencionales, de modo que la capacidad de los intrones del grupo II de funcionar en fondos genéticos RecA<sup>-</sup> los convierte en una herramienta potencial para la manipulación de genomas eucariotas (Cui y Davies, 2007). Como en el caso de procariotas, esta actividad depende de la formación de RNPs funcionales. Ya se han usado con éxito diferentes estrategias para introducir RNPs modificadas en cultivos celulares eucariotas: bien como complejos purificados reconstituidos *in vitro* que luego se implantan en el núcleo (Guo *et al.*, 2000; Cui, 2005) o bien en forma de construcciones que expresen ambos componentes para formar las RNPs (Cui, 2005). Por ahora, son sistemas bastante ineficientes (Cui y Davies, 2007), pero en el momento que se consiga que los intrones del grupo II con sus ventajas asociadas (especificidad, eficiencia y flexibilidad) funcionen en eucariotas habremos conseguido una herramienta poderosa para estudios de genómica funcional, terapia génica e ingeniería genética.





## *OBJETIVOS*



## OBJETIVOS

Desde su descubrimiento en 1998, nuestro grupo de investigación está trabajando en la caracterización del intrón del grupo II de *Sinorhizobium meliloti*, RmInt1. Los diferentes trabajos se han abordado desde muy diferentes perspectivas: molecular (*in vitro*), funcional (*in vivo*) y evolutiva. En los últimos años, además, se está trabajando en el diseño de herramientas de manipulación genética basadas en las propiedades de este intrón. Para llegar a comprender a la perfección todos los procesos en los que RmInt1 se ve involucrado y ser capaz de modificarlos, es preciso caracterizar las actividades que este intrón lleva a cabo. Así pues, el objetivo general de este trabajo es profundizar en la biología básica de RmInt1 particularizando en los siguientes objetivos:

1. Sabemos que el genoma de *S. meliloti* contiene una amplia variedad de secuencias de inserción y que otros elementos transponibles, los intrones del grupo II, pueden encontrarse en ocasiones en alto número de copias. Con el fin de profundizar en nuestro conocimiento de la biología evolutiva de los intrones, pero también para determinar la influencia de este tipo de elementos en el genoma bacteriano, nos propusimos la búsqueda y caracterización de nuevos retroelementos presentes en el genoma de *S. meliloti*.

2. La mayoría de los intrones del grupo II es capaz de madurar principalmente siguiendo dos rutas posibles (al menos *in vitro*). En el caso particular de RmInt1 se ha demostrado que es capaz de madurar tanto *in vivo* como *in vitro*. Los datos procedentes de la caracterización molecular de la reacción de *splicing in vitro* muestran que este intrón es capaz de madurar generando tres posibles formas: ARN *lariat*, ARN lineal y ARN circular. Sin embargo, desconocemos cuáles son los mecanismos de maduración que RmInt1 utiliza *in vivo*. La comprensión del proceso de *splicing* de este intrón podría ayudarnos a mejorar el rendimiento del proceso y, por consiguiente, la eficiencia de este elemento para trasponerse a otras localizaciones en el genoma. Por tanto, la siguiente meta perseguida en esta Tesis Doctoral fue la caracterización de los productos generados durante la maduración de RmInt1 *in vivo*.

3. Cuando nos referimos a la proteína codificada por RmInt1, hablábamos de la identificación clara de los dominios RT y madurasa, no obstante parece carecer de dominio de unión a ADN y dominio En. Estas carencias se traducen en particularidades mecánicas en la reacción de *retrobombing*. Si compramos la IEP de RmInt1 con otras proteínas, podemos distinguir una corta región de 20 aminoácidos sin identidad aparente y cuya funcionalidad, por ende, se desconoce. Así, nuestro primer objetivo se centra en la caracterización funcional del extremo C-terminal de la proteína codificada por RmInt1.



# *MATERIAL Y MÉTODOS*



## 1. ESPECIES Y CEPAS BACTERIANAS

Todas las especies y cepas utilizadas en los ensayos de laboratorio se incluyen en la siguiente tabla.

CEPA BACTERIANA	CARACTERÍSTICAS	REFERENCIA
<i>Sinorhizobium meliloti</i> GR4	Aislado de suelo de la EEZ; Nod+, Fix+	Casadesús y Olivares, 1979
<i>S. meliloti</i> GRM6	Derivado de GR4 curada del plásmido pRmeGR4b	Mercado-Blanco <i>et al.</i> , 1993
<i>S. meliloti</i> RMO17	Aislado de suelo de Salamanca; Nod+, Fix+	Villadas <i>et al.</i> , 1995
<i>Escherichia coli</i> DH5 $\alpha$	F <sup>-</sup> , $\phi$ 80d <i>lacZ</i> $\Delta$ M15, $\Delta$ ( <i>lacZYA-argF</i> )U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rK <sup>-</sup> , mK <sup>+</sup> ), <i>pboA</i> , <i>supE44</i> , $\lambda$ -, <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Bethesda Research Lab.
<i>E. coli</i> HB101	<i>supE44</i> , $\Delta$ ( <i>mcrC-mrr</i> ), <i>recA13</i> , <i>ara-14</i> , <i>proA2</i> , <i>lacY1</i> , <i>galK2</i> , <i>rpsL20</i> , <i>xyf-5</i> , <i>mtl-1</i> , <i>leuB6</i> , <i>thi-1</i>	Promega
<i>E. coli</i> ES1301 mutS	<i>lacZ53</i> , <i>mutS201::Tn5</i> , <i>thyA36</i> , <i>rba-5</i> , <i>metB1</i> , <i>deoC</i> , IN( <i>rrnD-rrnE</i> )	Promega
<i>E. coli</i> HMS174 (DE3)	F <sup>-</sup> <i>recA1</i> <i>hsdR</i> (r <sub>K12</sub> <sup>-</sup> m <sub>K12</sub> <sup>+</sup> ) (DE3) (Rif <sup>R</sup> )	Novagen

## 2. PLÁSMIDOS Y VECTORES DE CLONAJE

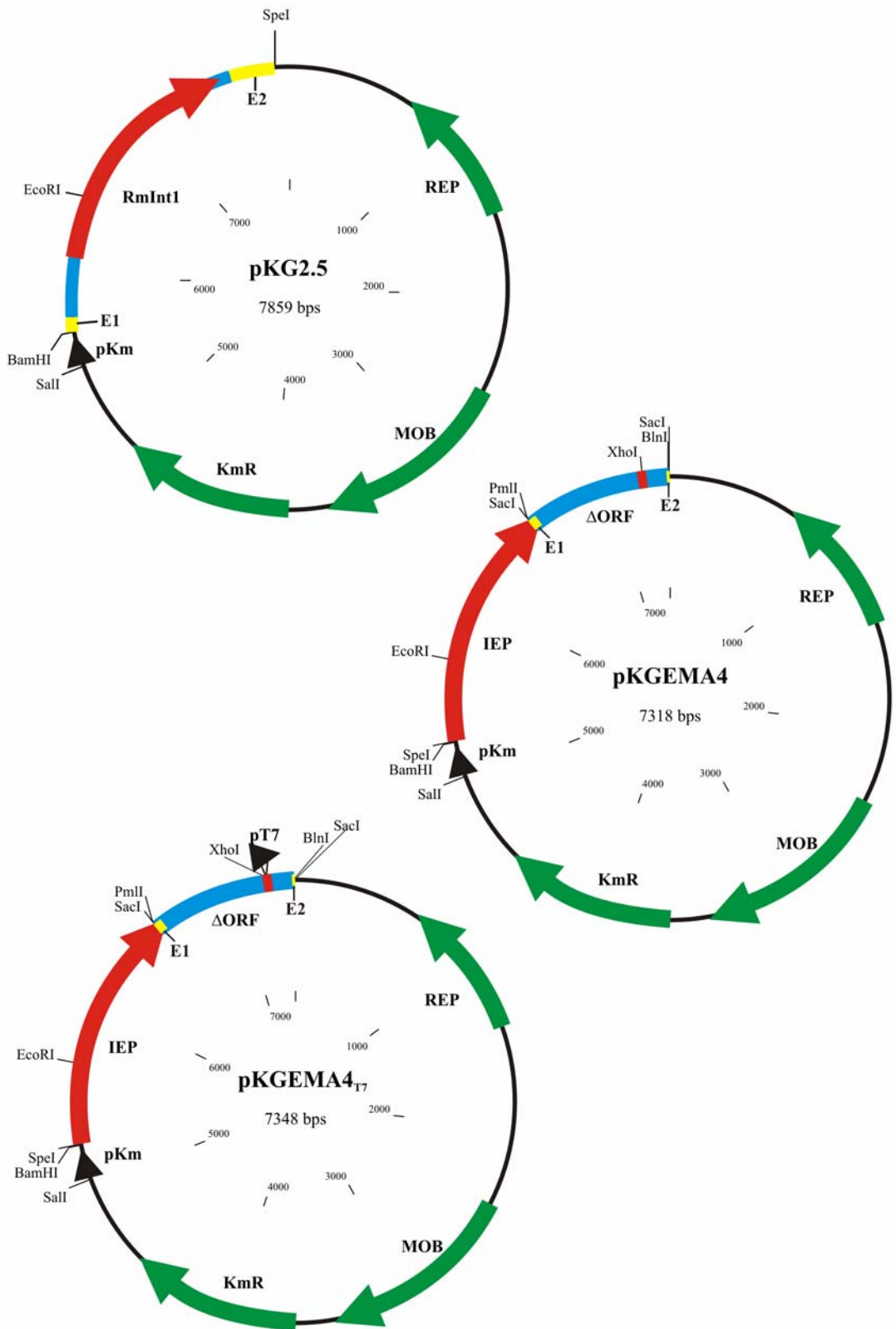
Se utilizaron una serie de vectores para el clonaje de fragmentos dependiendo de la estrategia diseñada en cada momento. Por lo general, se utilizaron vectores de alto número de copia y que permiten la selección de colonias blancas como pGEM-T (Promega) o pKSBluescript (Stratagene).

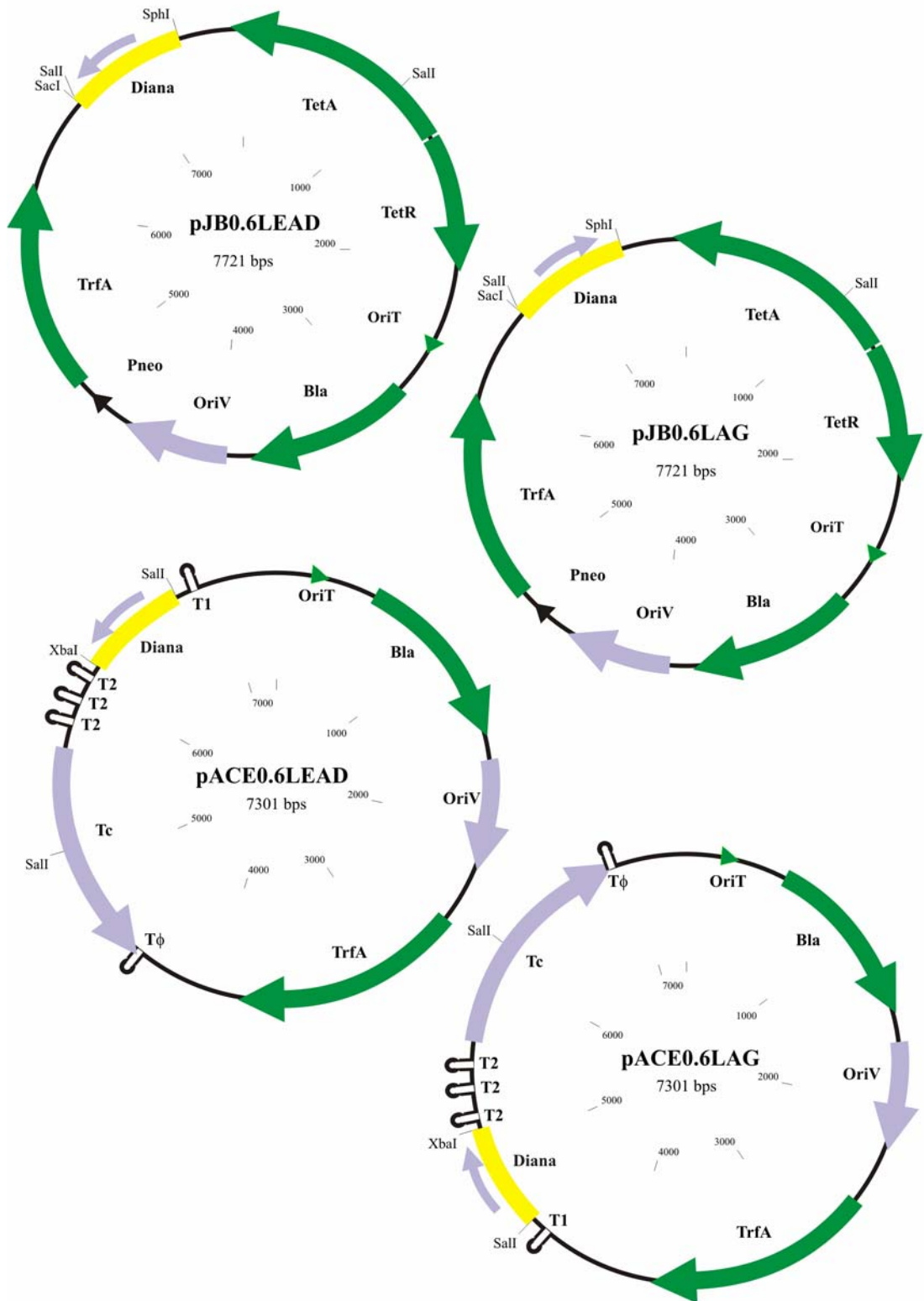
Por otro lado, los plásmidos en los que se han construido los mutantes presentados en este trabajo, además de algunos otros utilizados, fueron construidos otros miembros del grupo de investigación, por lo que, a título informativo se adjuntan los mapas a escala donde se señalan las características más relevantes a las que se hace mención a lo largo de este trabajo.





	NOMBRE	USO	REFERENCIA
VECTOR	pBlueScript KS/SK	Clonaje del fragmento genómico que contiene la secuencia 135F. Proporciona Ap <sup>R</sup>	Stratagene
	pGEM-T/pGEM-T easy	Múltiples construcciones intermedias usadas y para enviar a secuenciar. Ap <sup>R</sup>	Promega
	pRK290	Como control de digestión de los pRmNTs. Proporciona Tc <sup>R</sup> .	Ditta <i>et al.</i> , 1985
PLÁSMIDOS	pRmNTs	Conjunto de plásmidos que conforman la librería del plásmido críptico pRmcGR4b. Tc <sup>R</sup> .	Toro, 1985
	pKG2.5	Construcción donadora de intrón en la que RmInt1 se encuentra flanqueado por exones -174/+466. Resistencia a Km.	Martínez-Abarca <i>et al.</i> , 2000
	pKGEMA4	Construcción donadora de intrón en la que la ribozima de RmInt1 ( $\Delta$ ORF) se encuentra flanqueado por exones -20/+5. Km <sup>R</sup>	Nisa-Martínez <i>et al.</i> , 2007
	pKGEMA4 <sub>17</sub>	Construcción derivada de la anterior que contiene el promotor de la ARN polimerasa del fago T7 dentro del $\Delta$ ORF. Km <sup>R</sup>	Barrientos-Durán, 2008
	pJB0.6 LAG/LEAD	Plásmido receptor derivado de pJB3Tc19 que contiene 640 bp de la ISR $m2011-2$ clonada en una u otra orientación respecto del <i>oriV</i> . Se ha usado en ensayos de <i>homing</i> por hibridación. Ap <sup>R</sup>	Martínez-Abarca <i>et al.</i> , 2000 ; 2004
	pACE0.6 LAG/LEAD	Plásmido receptor derivado de pJBRBS (Ap <sup>R</sup> ) que contiene 640 bp de la ISR $m2011-2$ clonada en una u otra orientación respecto del <i>oriV</i> y de un gen de Tc <sup>R</sup> . Contiene distintos terminadores de transcripción flanqueando ambos componentes. Se usa para evaluar la capacidad de movilidad por selección de Tc <sup>R</sup>	Fernández-López, sin publicar





### 3. OLIGONUCLEÓTIDOS

Para la síntesis de oligonucleótidos se ha utilizado el servicio que proporciona el Instituto de Parasitología y Biomedicina “López-Neyra” (C.S.I.C., Granada). Según indicaciones del jefe de servicio, los oligonucleótidos empleados en este trabajo se han sintetizado con *DNA Synthesizer*® 3400 (Applied Biosystem).

En los anexos tras cada capítulo de resultados se presentan tablas que contienen los oligonucleótidos utilizados en la construcción de los mutantes descritos en el correspondiente capítulo. Aquellos oligonucleótidos específicos de cada técnica se detallarán junto con la técnica en la que se utilizaron.

### 4. CULTIVOS CELULARES

#### 4.1. MEDIOS DE CULTIVO Y ANTIBIÓTICOS

El medio completo utilizado para el crecimiento rutinario de *Sinorhizobium meliloti* fue TY (Beringer, 1974) preparado en agua desionizada:

CaCl <sub>2</sub> x 2H <sub>2</sub> O	0,9 g/l
Triptona (DIFCO)	5 g/l
Extracto de levadura (DIFCO)	3 g/l

Para la preparación del medio sólido se añadió agar (PANREAC) al 1,6%. La esterilización se realizó en autoclave durante 20 minutos a 120°C.

Trabajos anteriores demostraron que la modificación de la composición del medio rico TY mejoraba el rendimiento de obtención de RNPs en *S. meliloti* reduciendo la mucosidad de la cepa. Así, para este protocolo en concreto se utilizó el medio denominado TyA preparado con agua bidestilada:

NaCl	1,25 g/l
Peptona Triptona (DIFCO)	8 g/l
Extracto de levadura (DIFCO)	5 g/l

Para la preparación del medio sólido se añadió agar (PANREAC) al 1,6%. La esterilización se realizó en autoclave durante 20 minutos a 120°C.

En protocolos en los que resulta necesario reducir el crecimiento de *E. coli* sin alterar en demasía el crecimiento de *S. meliloti* se utilizó la siguiente modificación del medio mínimo (MM) descrito por Robertsen *et al.* (1981) preparado en agua desionizada:



K <sub>2</sub> HPO <sub>4</sub>	0,3 g/l
KH <sub>2</sub> PO <sub>4</sub>	0,3 g/l
MgSO <sub>4</sub> ·7H <sub>2</sub> O	0,15 g/l
CaCl <sub>2</sub>	0,05 g/l
FeCl <sub>3</sub>	0,006 g/l
NaCl	0,05 g/l
Glutamato sódico	1,1 g/l
Manitol	10 g/l
Biotina	0,0002 g/l
Pantotenato cálcico	0,0001 g/l

Para solidificar el medio se añadió agar purificado (OXOID) al 1,6%. El pH se ajustó de 6,8 a 7,0 y se esterilizó durante 20 minutos a 120°C.

Como medio de cultivo rutinario de las cepas de *Escherichia coli* se utilizó el medio de Luria-Bertani (LB) (Sambrook *et al.*, 1989) preparado con agua desionizada:

NaCl	5 g/l
Triptona (DIFCO)	10 g/l
Extracto de levadura	5 g/l

Para solidificar el medio se añadió agar (PANREAC) al 1,6%. El medio se esterilizó igualmente a 120°C durante 20 minutos.

Como medio de cultivo restrictivo para coliformes se utilizó el Endo Agar™ LES (DIFCO). Su uso se restringe a la comprobación de que las colonias de *S. meliloti* seleccionadas de MM tras las conjugaciones están limpias de *E. coli* y a comprobar que los cultivos para la extracción de RNPs no se han contaminado con esta bacteria. Se prepara disolviendo 51 g por litro de agua desionizada, calentando hasta disolución completa y luego autoclavado a 120°C durante 20 minutos.

Las concentraciones finales de los antibióticos precisados en los diferentes protocolos fueron:

ANTIBIÓTICO	EFEECTO	<i>E.coli</i>	<i>S. meliloti</i>
Tetraciclina (Tc, Sigma)	Bactericida	10 mg/l	10 mg/l
Kanamicina (Km, Roche)	Bactericida	50 mg/l	180 mg/l
Ampicilina (Ap, Sigma)	Bacterioestático	200 mg/l	200 mg/l

La adición de antibióticos a los medios de cultivo se hizo a partir de soluciones 100 veces concentradas preparadas en agua desionizada (Km, Ap) o en agua-etanol al 50% (Tc). Las soluciones preparadas en agua se esterilizaron por filtración utilizando unidades Minisart® NML (Sartorius) de 0,2 µm de tamaño de poro.

## 4.2. CONDICIONES DE CULTIVO

Las distintas cepas de *S. meliloti* se crecieron a 28°C, siendo su tiempo de generación en medio líquido de aproximadamente 2,5 h a 190 r.p.m. en agitador orbital.

*E. coli* se incubó a 37°C y su tiempo de generación en medio líquido ronda los 30 min a 190 r.p.m. en agitador orbital.

## 4.3. CONSERVACIÓN DE CULTIVOS BACTERIANOS

Para la conservación prolongada de los cultivos bacterianos se utilizó la congelación. Este método se basa en la paralización del metabolismo celular al disminuir la disponibilidad de agua. Para que la viabilidad celular se mantenga durante el periodo de conservación se precisa de lo que se denominan crioprotectores que evitan el daño que los cristales de agua podrían ocasionar en las paredes celulares; en nuestro caso, el crioprotector elegido fue glicerol 20% (v/v). Este glicerol estéril dispuesto en criotubos se mezcló con cultivos crecidos hasta fase logarítmica tardía. Los viales se congelaron rápidamente y se almacenaron a una temperatura de -80°C.

## 5. AISLAMIENTO DE ÁCIDOS NUCLEICOS

En este apartado vamos a describir cómo se obtuvieron la mayoría de las materias primas utilizadas en los diferentes ensayos posteriores. Hemos decidido incluir también el protocolo de obtención de complejos ARN-proteína.

### 5.1. EXTRACCIÓN DE ADN total

#### 5.1.1. AISLAMIENTO DE ADN total USANDO PRONASA

En este protocolo se produce una lisis celular progresiva tratando de mantener en el mayor grado posible la integridad del material genético, por eso es muy importante tener cuidado con la agitación. Se utilizó para la extracción del ADN genómico de *S. meliloti* GR4.

Se recogen 1,5 ml de cultivo en fase logarítmica de crecimiento por centrifugación durante 2 min a 13000 r.p.m.. Al tratarse de cultivos de rizobios productores de exopolisacáridos es conveniente lavar el cultivo recogido con 0,5 ml de N-lauroil sarcosina (sarcosil) al 0,1% en TE pH 8 (10 mM Tris, 1 mM EDTA; ajustar el pH con NaOH). El sedimento se resuspende en 1 ml de NaCl 1 M y se mantiene durante 1 h a 4°C, realizando movimientos de inversión cada 15'. Volvemos a centrifugar las bacterias



durante 2 min y retiramos el sobrenadante. Resuspendemos hasta homogeneidad en 0,25 ml de sacarosa al 20% en TE esterilizado conservado en frío. En este momento se añaden 0,25 ml de una solución de lisozima (Roche) 5 mg/ml con ARNasa 1 mg/ml disueltas en TE (preparación fresca) y agitamos por inversión del tubo. Incubamos durante 30 min a 37°C para permitir la acción de la enzima sobre la pared celular (peptidoglucano). Para completar la degradación de la pared (proteínas), se adicionan 0,1 ml de una mezcla compuesta de sarcosil al 5% disuelta en TE y pronasa (Roche) 10 mg/ml (conservada alícuotas congeladas). Se incuba una hora más a 37°C. Se agregan 70 µl de acetato sódico 3 M pH 5,2 (precipitación de proteínas) junto con 0,2 ml de fenol<sup>30</sup> y 0,2 ml de una mezcla fenol:cloroformo:alcohol isoamílico (25:24:1, ΦCIA)<sup>31</sup>. Se mezcla vigorosamente durante 2 min y se centrifuga durante 5 min a máximas revoluciones. Se recoge la fase acuosa superior con cuidado de no alterar la interfase y se mezcla con 0,3 ml de cloroformo:alcohol isoamílico (CIA). Agitar vigorosamente y volver a centrifugar 5 min. A la fase superior recogida en un tubo nuevo se le añaden 0,7 ml de isopropanol preenfriado a -20°C, mezclamos por inversión y se deja precipitar a -70°C durante 15 min. Transcurrido este tiempo, o más si lo consideramos oportuno, se centrifuga durante 15 min y retiramos el sobrenadante. El exceso de sales en la preparación se elimina lavando con 0,5 ml de etanol al 70% frío. Secamos y resuspendemos en 0,1 ml de TE diluido diez veces (para evitar los posibles efectos inhibitorios que el EDTA puede tener sobre determinadas enzimas, por ejemplo, la Taq polimerasa). Permitir la correcta resuspensión del ADN si es preciso dejando las muestras varias horas a 37°C.

### 5.1.2. EXTRACCIÓN DE ADN total MEDIANTE KIT COMERCIAL

Para este propósito se utilizó el kit comercial *AquaPure genomic DNA isolation* (Bio-Rad). Recogimos 1ml de cultivo de *Rhizobium* y lo lavamos con 0,2 ml de Sarcosil al 0,1% en TE. La lisis se lleva a cabo por la adición de 0,6 ml de “*Lysis Solution*” incubando 10 min a 80°C. Se deja enfriar a temperatura ambiente y añadimos 0,003 ml de “*Solución RNase*”. Se mezcla invirtiendo 25 veces e incubamos una hora a 37°C. Pasado este tiempo se adicionan 0,2 ml de “*Protein Precipitation*” y agitamos vigorosamente durante 20 seg. Incubar en hielo 10 min. La muestra se centrifuga a máximas revoluciones durante 5 min y recogemos el sobrenadante en un tubo nuevo. El ADN se precipita añadiendo 0,6 ml de isopropanol frío invirtiendo hasta la mezcla total durante 10 min a -20°C.

---

<sup>30</sup> El fenol pH neutro se equilibra mezclando una solución de fenol 100% disponible comercialmente (Amresco) con 0,1% 8-hidroxiquinoleína en Tris-HCl 0,1 M pH 8 (ambos componentes se venden juntos). Se deja reposar toda la noche a 4°C y se conserva a esa misma temperatura. Se toma de la fase inferior. Este componente es muy volátil y peligroso, por lo que es conveniente trabajar en campana de extracción de gases siempre que se manejen este tipo de productos.

<sup>31</sup> Para su preparación se usa el fenol descrito anteriormente más las cantidades correspondientes de cloroformo e isoamil alcohol. Se debe preparar con antelación porque necesita reposo para separar las fases. También se toma de la fase inferior.

Centrifugamos 10 min y lavamos el sedimento con 0,2 ml de etanol 70%. Se resuspende con 0,04 ml de agua bidestilada.

### 5.1.3. LISADOS DE COLONIA

La siguiente técnica consiste en la obtención rápida de ADN con suficiente calidad y cantidad para protocolos como PCR pero insuficiente para otras técnicas (Villadas *et al.*, 1995). Consiste en resuspender en 0,1 ml de agua desionizada<sup>32</sup> estéril una colonia picada de placa en condiciones de esterilidad. Se sedimentan las células por centrifugación durante 2 min a 13000 r.p.m. Se elimina el sobrenadante y se lava 2 veces más con 0,1 ml de agua. Tras el último lavado, las células se resuspenden en 0,1 ml de agua y se incuban en agua hirviendo durante 10 min. Centrifugamos 2 min y el sobrenadante resultante (concretamente 2-4  $\mu$ l) directamente se usa de molde para el PCR.

## 5.2. EXTRACCIÓN ADN PLASMÍDICO

Dependiendo de las necesidades de cada ensayo, se utilizó uno u otro protocolo. En líneas generales, el aislamiento mediante kit comercial se restringió al envío de muestras para secuenciación y en protocolos que precisasen de un ADN libre de ARNasas (por ejemplo, transcripción *in vitro*). El procedimiento de lisis seguida de precipitación con sales de magnesio funciona muy bien en células de *E. coli*; sin embargo, el procedimiento rutinario para la extracción plasmídica en *S. meliloti* fue la lisis alcalina. Determinadas aplicaciones precisan del uso de mucha cantidad de plásmido; para éstas recurrimos a las midipreparaciones de plásmidos mediante kit comercial.

### 5.2.1. AISLAMIENTO DE PLÁSMIDOS POR LISIS ALCALINA

El aislamiento de ADN plasmídico de bajo peso molecular se realizó siguiendo el protocolo modificado por Sambrook y colaboradores (Birnboim y Doly, 1979; Ish-Horowicz y Burke, 1981; Sambrook *et al.*, 1989). Se recogen 1,5 ml de cultivo bacteriano crecido hasta fase exponencial y se sedimentan a 13000 r.p.m. durante 2 min. Lavamos las células con 500  $\mu$ l de sarcosil 0,1% en TE y volvemos a centrifugar, retirando el sobrenadante. Resuspendemos las células con 100  $\mu$ l de lisozima 4 mg/ml disuelta en solución I (glucosa 50 mM, Tris-HCl 25 mM pH 8, EDTA 10 mM; conservada a 4°C). Se incuban 5 min a temperatura ambiente, tras los cuales se añaden 200  $\mu$ l de solución II (NaOH 0,2 M, 1% SDS). Se homogeniza la preparación agitando el tubo por inversión varias veces y se mantiene en hielo durante 5 min. Entonces se adicionan 150  $\mu$ l de solución III (acetato potásico 5 M pH 4,8) que ayudará a la neutralización de la solución y la precipitación de proteínas. Se mezcla invirtiendo el tubo y se incuban en hielo 5 min para minimizar el posible daño con ADNasas celulares. Centrifugamos a 13.000 r.p.m.

---

<sup>32</sup> Debemos tener presente que si estamos lisando células de rizobios este agua debe contener sarcosil 0,1% (p/v).





durante 5 para sedimentar los restos celulares y recogemos el sobrenadante en un tubo nuevo. Añadimos un volumen (~400  $\mu$ l) de  $\Phi$ CIA y mezclamos vigorosamente. Nuestra muestra va a separarse en dos fases tras una centrifugación de 5 min a 13.000 r.p.m. Recogemos la fase superior acuosa y la mezclamos con un volumen de CIA (elimina posibles restos de fenol en las preparaciones). Volvemos a mezclar y centrifugamos durante 5 min. Pasamos la fase acuosa a un tubo nuevo y precipitamos el ADN con 2,5 volúmenes de etanol 100% frío. Podemos optar por precipitar 30 min a  $-20^{\circ}\text{C}$  o 15 min a  $-70^{\circ}\text{C}$ , pero lo más habitual para evitar la precipitación de ADN genómico son 5 min a temperatura ambiente. Centrifugamos el ADN precipitado y lavamos el sedimento con 200  $\mu$ l de etanol al 70%. Dejamos secar y resuspendemos en 25  $\mu$ l de agua desionizada en la que habremos diluido una solución 100 veces concentrada de ARNasa A<sup>33</sup> (10  $\mu$ g/ml de concentración final).

### 5.2.2. EXTRACCIÓN DE ADN PLASMÍDICO POR PRECIPITACIÓN CON SALES DE MAGNESIO

Esta técnica está basada en el método de Studier (1991), permite el aislamiento rápido de plásmidos de *E. coli*. Recogemos 1,5 ml de cultivo bacteriano por centrifugación y, tras retirar el sobrenadante, se resuspenden las células en 100  $\mu$ l de agua desionizada. La lisis tiene lugar por la adición de 100  $\mu$ l de una solución compuesta por NaOH 0,1 M, EDTA 10 mM y 2% SDS. Se homogeneiza por agitación y se hierven durante 2 min. El ADN lineal, normalmente cromosómico, se precipita por la adición de 50  $\mu$ l de  $\text{MgCl}_2$  1 M. Tras homogeneizar completamente (aparición de precipitado blanco), se incuba en hielo de 2-5 min. Centrifugamos 1 min para sedimentar el precipitado en el fondo del tubo y añadimos 50  $\mu$ l de acetato potásico 5 M para eliminar proteínas. El tubo invertido se mezcla con vórtex evitando que se levante el precipitado blanco del fondo y se incuba 5 min en hielo. Centrifugamos durante 5 min y recogemos el sobrenadante a un tubo nuevo donde precipitaremos el ADN plasmídico con 0,6 ml de etanol 100% guardado a  $-20^{\circ}\text{C}$ . Se deja 5 min a temperatura ambiente y se centrifuga 5 min. Lavamos el sedimento con 200  $\mu$ l de etanol al 70% y dejamos secar a vacío. Resuspender en el volumen adecuado dependiendo de la concentración deseada o necesaria para las diferentes aplicaciones; lo más habitual para un plásmido de alto número de copias es resuspender en 25  $\mu$ l agua desionizada adicionada de ARNasa 10  $\mu$ g/ml.

---

<sup>33</sup> La solución ARNasa 10 mg/ml se prepara diluyendo, en las cantidades indicadas, la enzima (liofilizada, Sigma) en una solución con 10 mM Tris-HCl pH 7.5 y 15 mM NaCl. Hervir durante 15' y dejar enfriar a RT. Se conserva a  $-20^{\circ}\text{C}$ . Debemos distribuirla en alícuotas porque la congelación-descongelación hace que pierda actividad (Sambrook *et al.*, 1989).

### 5.2.3. AISLAMIENTO DE ADN PLASMÍDICO MEDIANTE KIT COMERCIAL

Durante este periodo hemos usado distintos kits comerciales para el aislamiento de ADN plasmídico en todos los casos se han seguido las indicaciones de uso del proveedor: *Wizard® Plus SV Minipreps DNA Purification Systems* (Promega), *QIAprep Spin Miniprep Kit* (Qiagen) y *PerfectPrep® plasmid mini* (Eppendorf). En todos los casos consistió en la lisis celular seguida de la unión del ADN en disolución a una resina/membrana y elución con agua. La ventaja fundamental es la limpieza del ADN y la ausencia de ARNasa en los pasos finales.

### 5.2.4. EXTRACCIÓN DE PLÁSMIDOS A PARTIR DE GRANDES VOLÚMENES DE CULTIVO (MIDIPREPARACIONES)

Con este fin también se usó un kit comercial, *QIAGEN Plasmid Midi Kit* (Qiagen). Es un protocolo rápido y sencillo de aislamiento de ADN plasmídico ultrapuro superenrollado con alto rendimiento. Está basado en protocolos de purificación plasmídica por el procedimiento de lisis alcalina modificada, seguido de la unión del plásmido a una resina de intercambio aniónico en condiciones de baja sal y pH apropiado. Todo tipo de posibles contaminantes son eliminados mediante el lavado con soluciones media sal y posteriormente el ADN es eluido en tampones con alta sal. Tras lo cual precisa ser precipitado. Partimos de 100 ml de cultivo que recogemos por centrifugación a 7000 r.p.m. (rotor GSA en centrífuga *Sorvall*) durante 15 min a 4°C. Se siguieron las instrucciones del proveedor hasta eluir nuestro ADN. La precipitación se llevó a cabo en tubos de 1,5 ml en alícuotas de 1 ml a las que se añadieron 700 µl de isopropanol frío. Centrifugamos a máximas revoluciones durante 30 min y lavamos el exceso de sales con 200 µl de etanol 70% frío. Dejamos secar y el ADN contenido en cada tubo se resuspende en 20 µl de agua bidestilada (De 100 ml de cultivo de partida se obtienen ~100 µl de plásmido ultrapuro).

## 5.3. EXTRACCIÓN DE ARN total

Es la primera ocasión en que nos vamos a referir a este tipo de material, pero las indicaciones dadas deben respetarse en adelante. El ARN es bastante más lábil que el ADN y las ARNasas muy estables. Por esta razón debemos ser especialmente cautelosos en el manejo de material y soluciones que vayan a destinarse a este tipo de protocolos (usar siempre guantes y reservar material para uso exclusivo). Otra de las precauciones que debemos tener está en referencia al almacenamiento de muestras; mientras el ADN se conserva perfectamente a -20°C, el ARN es preferible mantenerlo a -80°C. Durante el manejo es conveniente conservarlas el mayor tiempo posible a 4°C (en hielo).



### 5.3.1. AISLAMIENTO DE ARN total DE *S. meliloti*

Este protocolo presenta algunas modificaciones sobre el descrito por Cabanes *et al.* (2000). Se lleva a cabo la inoculación de 0.5 ml. del cultivo bacteriano saturado en un tubo que contiene 9.5 ml. de medio con el correspondiente antibiótico. Se crece hasta el final de su fase exponencial, esto es  $\sim 0,6$  U de  $A_{600}$ . Recogemos los 10 ml de cultivo en dos tubos de 1,5 ml gracias a varias centrifugaciones de 1-2 min a 14000 r.p.m. Resuspendemos las células sedimentadas con 300  $\mu$ l de solución compuesta por 1.4% SDS y 4mM EDTA. Se incuba durante 10 min a 65°C, agitando de manera esporádica (cada 3 min). Pasado este tiempo, los tubos se ponen en hielo y se adicionan 150  $\mu$ l de NaCl 5M a 4°C, se mezclan vigorosamente y se mantienen en hielo durante 10 min. Centrifugamos 15 min a 4°C y 13000 r.p.m. Se pone el sobrenadante en un tubo nuevo y se precipita con 1 ml de etanol 100% frío (en este punto podemos dejarlo varios días). Se centrifuga el ácido nucleico precipitado durante 30 min a la máxima velocidad de la centrífuga y 4°C. Eliminamos muy bien todos los restos de etanol y dejamos secar ligeramente a temperatura ambiente. Resuspendemos los precipitados correspondientes a los 10 ml de cultivo inicial en 85  $\mu$ l de agua bidestilada mezclando los duplos para el tratamiento con ADNasa I. La digestión del ADN en nuestras preparaciones tiene lugar por la incubación durante una hora a 37°C de estos 85  $\mu$ l con 50 U de ADNasa I libre de ARNasas (Roche) en presencia de 60 U de inhibidor de ARNasas (GE healthcare) en el tampón adecuado (50 mM Tris-HCl pH 7.5, 1 mM  $MgCl_2$ ). Podemos agitar de vez en cuando. Transcurrido este tiempo, se eliminan las proteínas mediante dos fenolizaciones sucesivas: en la primera se añade 1V (100  $\mu$ l) de  $\Phi$ CIA<sup>34</sup>, se toma la fase superior y se mezcla con 1V de CIA. En el primer caso se centrifuga durante 10 min y en el segundo es suficiente con 5 min a 4°C y máxima velocidad. La fase acuosa superior se precipita con 600  $\mu$ l de etanol 100%/75 mM acetato sódico 3 M pH 5,2 frío. Se mantiene al menos una hora a -80°C. Sedimentar por centrifugación a 4°C durante 30 min y lavar con 0,5 ml de etanol 70% frío. Secar la muestra sin temperatura y resuspender en 20  $\mu$ l de agua bidestilada.

#### 5.3.1.1. Modificación por la adición de proteinasa K

El protocolo es básicamente como el que se ha explicado pero se introduce la modificación de adicionar proteinasa K a una concentración final de 42 mg/ml en la solución de lisis (1.4% SDS y 4mM EDTA). El resto se mantuvo igual.

#### 5.3.1.2. Modificación por el uso de tiocianato de guanidina

En este caso la modificación introdujo en la fenolización tras el tratamiento con ADNasa I; en lugar de 1V de  $\Phi$ CIA, se utilizó una solución compuesta por 25  $\mu$ l de tiocianato de guanidina 5 M/ 2,5  $\mu$ l de acetato sódico 3 M pH 4,8/ 93  $\mu$ l de fenol ácido /

---

<sup>34</sup> Al trabajar con ARN es conveniente que utilicemos fenol a pH ácido (amresco) porque este tipo de fenol arrastra con más facilidad al ADN, dejando en la fase acuosa el ARN (Sambrook *et al.*, 1989).

30 µl de una mezcla 49:1 de cloroformo:alcohol isoamílico. Esta mezcla, preparada fresca, se incuba durante 15 min en hielo con nuestro ARN, se centrifuga 10 min a 4°C y máximas revoluciones. A continuación, la fase superior se mezcla con un volumen de CIA y el ARN se precipita como se ha indicado previamente. El tiocianato de guanidina se va a encargar de desnaturalizar las proteínas (Chomczynski y Sacchi, 1987; Kingston *et al.*, 1996).

### 5.3.2. EXTRACCIÓN DE ARN total MEDIANTE KIT COMERCIAL

En ocasiones puntuales ha sido necesaria la extracción de ARN total mediante el uso de kit comerciales, exclusivamente para el aislamiento de ARN de cultivos de *E. coli*. En concreto, se han usado *RNeasy Protect Bacteria Mini Kit* (Qiagen) y *Aurum total RNA Mini kit* (Bio-Rad) en cada ocasión siguiendo las indicaciones del proveedor.

### 5.4. EXTRACCIÓN DE PARTÍCULAS RIBONUCLEOPROTEICAS (RNPs)

En este caso no se trata de un aislamiento sino de un enriquecimiento en complejos ARN-proteína presentes en la célula. Este protocolo está basado en el descrito por Matsuura *et al.* (1997) modificado para *S. meliloti* (Muñoz-Adelantado *et al.*, 2003). Todas las soluciones se filtran en el momento de ser utilizadas mediante filtros *Minisart® NML* (Sartorius) de 0,45 µm. Partimos de 200 ml de bacteria crecida en TyA hasta alcanzar las 0.8 U de A<sub>600</sub>. Recogemos las células mediante una centrifugación de 12 min a 5000 r.p.m. y 4°C (*Sorvall*; Rotor GSA) y hacemos un primer lavado con 20 ml de sarcosil 0.1% TE transfiriendo las células resuspendidas a un tubo cónico de 50 ml. Lavamos dos veces más con 20 ml de una disolución de NaCl 150 mM, centrifugando en cada ocasión 5 min a 5000 r.p.m. 4°C (*Eppendorf 5810/R*; rotor F-34-6-38); en las dos ocasiones las células deben encontrarse bien resuspendidas. Tras retirar el sobrenadante del lavado final, las bacterias deben resuspenderse en 5 ml. de Buffer A frío [50mM Tris pH 7.5, 1mM EDTA, 10% glicerol, 1mM DTT<sup>35</sup> (se pone en el momento de usar el tampón)] suplementada con 500 µl de lisozima 20 mg/ml<sup>36</sup> y se congelan a -80°C (en este punto podemos dejar el protocolo). Una vez congeladas, las células se someten a 3 ciclos de congelación-descongelación (si es necesario dar alguno más) en hielo seco y un baño a 28°C, alternativamente, hasta obtener una solución mucosa (lisado). Adicionamos 20 ml. de HKCTD frío [500 mM KCl, 50 mM CaCl<sub>2</sub>, 25 mM Tris pH 7.5, 5 mM DTT (añadir en el momento de utilizar la solución)] y homogeneizamos. Mantenemos en hielo hasta proceder con la sonicación. Los restos celulares se someten a 5-8 pulsos de 5 seg y

---

<sup>35</sup> Se añade a partir de una solución preparada DTT 1M que se mantiene congelada en alícuotas que evitaremos congelar-descongelar de forma reiterada. Se prepara disolviendo 3,09 gr de DTT sólido en 20 ml de acetato sódico 0,01 M pH 5,2 y se esteriliza por filtración.

<sup>36</sup> Se prepara disolviendo la correspondiente cantidad en buffer A, esterilizamos por filtración y se guarda congelada en alícuotas de 500 µl a la concentración final.



potencia 5 con una punta fina (*Sonifier 250*, Branson) para eliminar terminar el proceso de lisis y eliminar en cierta medida la mucosidad ocasionada por el ADN en solución. Una vez aclarado el lisado, se centrifuga durante 15 min a 10000 r.p.m. El sobrenadante se pasa a un tubo nuevo y se conserva en hielo. Entretanto habremos preparado tres tubos de ultracentrifuga de policarbonato con tapón de rosca con 5 ml de un colchón de sacarosa 1,85 M preparado en HKCTD<sup>37</sup> por cada muestra de la que queramos extraer RNPs. Se rellena el tubo con el lisado procurando no distorsionar la interfase (5,5 ml aprox.) y se centrifuga a 50.000 r.p.m. durante  $\approx$  20 h. a 4°C (Ultracentrífuga *Beckman L-70*; rotor 50Ti). Transcurrido este tiempo, se retira el sobrenadante por aspiración y el sedimento se lava 3 veces con 1 ml de agua bidestilada filtrada. Resuspenderemos el sedimento en función de su tamaño durante al menos 2 h a 4°C; un volumen estándar podrían ser 100  $\mu$ l de tampón de dilución de RNPs [10 mM Tris pH 7.5, 1 mM DTT]. La eliminación de material insoluble requiere de una última centrifugación a 4°C en un tubo de 1,5 ml durante 5 min a 13000 r.p.m. Se reparte en alícuotas de 10  $\mu$ l en tubos preenfriados y se guardan a -80°C donde se conservan estables varios años.

#### 5.4.1. PRECIPITACIÓN DEL ARN EN LAS DISTINTAS SUBFRACCIONES DE LAS RNPs

Para los experimentos mostrados en la figura II.14 se precisó la precipitación del ARN presente en las distintas subfracciones de las RNPs. Tras la ultracentrifugación se obtienen tres fracciones: una fase superior más laxa (S1), 5 ml de sobrenadante más denso (S2) y el sedimento (P). Cada una de estas fases fue recogida y tratada por diferentes métodos que optimizaron la precipitación del material de interés. La fracción S1 se mezcló con 1,5 volúmenes de una solución compuesta de tiocianato de guanidina 0,8 M/acetato sódico 50 mM pH 4,8/fenol (pH 4,5):cloroformo:alcohol isoamílico (150:49:1). La mezcla se mantuvo en hielo durante 15 min y luego se centrifugó a 13000 r.p.m. durante 15 min. La fase superior se sometió a una segunda extracción con 1 V de CIA seguido de precipitación con 1 V de isopropanol. El sedimento resultante se resuspendió en agua y fue tratado como si se tratase de una muestra de ARN (digestión con ADNasa I, fenolizaciones y precipitación).

El sobrenadante más denso (S2) se precipitó con 3 V de etanol 100%/LiCl 0,8 M. El sedimento resultante de la centrifugación fue resuspendido en agua y precipitado nuevamente con 3V de etanol 100% suplementado con acetato sódico 0,25 M pH 5,2.

El sedimento (P) se trató según el procedimiento habitual.

---

<sup>37</sup> Esta solución se prepara a partir de los componentes sólidos (excepto el DTT) debido a que la sacarosa es muy higroscópica. Se disuelve por agitación (proceso lento) y se esteriliza por filtración. Se suele guardar alícuoteado (en volúmenes múltiplo de 15 ml) a -80°C.

## 5.5. CUANTIFICACIÓN DE LA CONCENTRACIÓN DE ARN Y ADN

El cálculo de la concentración de ácidos nucleicos se estimó siguiendo el método espectrofotométrico descrito en Sambrook *et al.* (1989). Se midió la absorbancia de las muestras en solución acuosa a longitudes de onda 260 y 280 nm utilizando cubetas de cuarzo. La concentración se determinó en función de los valores estándar: 1 U a  $D.O_{260nm}$  se corresponde con 50  $\mu\text{g/ml}$  de ADN de cadena doble, 40  $\mu\text{g/ml}$  de ADN de cadena sencilla y ARN, y 33  $\mu\text{g/ml}$  de oligonucleótidos. La relación  $D.O_{260nm} / D.O_{280nm}$  se usó como estimador de la pureza de la preparación, considerándose valores por debajo de 1,8 como indicadores de contaminación por proteínas o fenol.

En los últimos años, se ha usado un espectrofotómetro que integra todos estos datos con sólo poner 1  $\mu\text{l}$  de nuestra muestra, *NanoDrop*® *ND-1000*.

## 6. MANIPULACIÓN Y AMPLIFICACIÓN DE ADN

### 6.1. DIGESTIÓN CON ENZIMAS DE RESTRICCIÓN

La digestión completa del ADN se realizó siguiendo las indicaciones de temperatura y tampón óptimos recomendadas por los proveedores (Roche y New England Biolabs). Lo más común es utilizar 5 U de enzima/ $\mu\text{g}$  ADN a digerir aunque, en el caso de muestras complejas como ADN total, esta cantidad debe elevarse al doble. A esto debemos sumarle una prolongación del tiempo de digestión de manera que en ocasiones la digestión se ha realizado durante la noche, aunque lo habitual es una hora. En el caso de querer realizar digestiones con 2 enzimas de restricción simultáneamente hay que tener precaución con el tampón elegido. Debemos tratar de buscar un tampón compatible para ambas (instrucciones del proveedor) y, en caso de no ser posible, realizar primero la digestión con la enzima que precise el tampón con menor fuerza iónica. Si aún así no fuese factible, el ADN se limpiaría de sales y enzima tras la primera digestión, y se procedería con la segunda, así sucesivamente. Hay dos alternativas para eliminar las sales: la primera es el uso de kits comerciales como *QLAquick PCR Purification Kit* (Qiagen), pero también podemos optar por fenolizar y precipitar nuestro ADN.

### 6.2. REACCIÓN DE DEFOSFORILACIÓN

En ocasiones es necesaria la eliminación de los grupos fosfato presentes en los extremos 5' de vectores linearizados para evitar que vuelvan a ligarse. Esto nos evita el análisis de muchas colonias que sean vector re-ligado en vez de contener el inserto que nos interesa. La defosforilación de 5  $\mu\text{g}$  de plásmido digerido precisa 1 U de fosfatasa alcalina (CIP, *Calf Intestine Phosphatase*, Roche) en tampón correspondiente durante 1 h a 37°C. La reacción separa añadiendo 55  $\mu\text{l}$  de una solución compuesta por tampón STE 1x [1 mM Tris-HCl pH8, 100 mM NaCl y 1 mM EDTA] y SDS 1% e incubando 15 min



a 68°C. Para obtener mejor rendimiento en la ligación es conveniente fenolizar y precipitar.

### 6.3. PCR

La reacción en cadena de la polimerasa (Polymerase Chain Reaction, PCR) consiste en la amplificación en sucesivos ciclos de un material de partida gracias al uso de una pareja de cebadores y una polimerasa termoestable (*Taq* polimerasa). Puede tener distintas finalidades y en nuestro caso la hemos utilizado para amplificar insertos que después necesitamos clonar añadiendo determinados enzimas en los extremos hasta para la construcción de mutantes, pasando por comprobaciones rutinaria de la presencia de un determinado inserto. Consiste en 3 pasos fundamentales repartidos en 3 ciclos: desnaturalización, anillamiento y extensión. La desnaturalización suele realizarse a 94°C 30 seg aunque el tiempo depende del tipo de molde; al comienzo del programa de PCR suele incluirse un primer ciclo de desnaturalización de 2 min para disponer de todo el ADN como cadena simple. La temperatura de anillamiento depende de la composición de los cebadores pero suele rondar los 55°C y el tiempo estándar es 30 seg. La extensión no es más que la temperatura óptima para la actividad de la enzima; el tiempo de este ciclo suele estar en función del tamaño del amplificado y del tipo de enzima usado (por ejemplo, *Taq* polimerasa 1Kb/30 seg). Además el programa de PCR suele terminarse con un ciclo prolongado de extensión de 5-10 min para completar los fragmentos. Depende de la finalidad perseguida se utilizaron distintas polimerasas: *TripleMaster*® (Eppendorf), *High Fidelity*® (Roche) o *Pfu* polimerasa (Promega), si era necesaria una baja tasa de errores en el producto final, mientras que para comprobaciones de rutina se usó una *Taq* polimerasa purificada en el laboratorio.

Las reacciones se llevaron a cabo en un volumen final de 25 µl conteniendo 10 a 100 ng del ADN molde, 25 pmol de cada uno de los cebadores específicos, 0,1 mM de deoxinucleósidos trifosfato (1 M si se trata de ADN total o de lisados celulares), el tampón especificado para cada aplicación y 2 unidades de ADN polimerasa. La amplificación se realizó en un termociclador *Eppendorf Mastercycler*® y el programa empleado varió dependiendo del objetivo.

La limpieza de este tipo de productos (eliminación de cebadores y sales) se realizó mediante columnas *MicroSpin*™ S-300HR (GE Healthcare) o con el uso del *QLAquick PCR Purification Kit* (Qiagen).

#### 6.3.1. PCR DE COLONIA

Es una estrategia muy rápida para la comprobación de la presencia de determinados insertos a partir de las colonias en placa. Es más sucia que utilizar lisados de colonias pero de efectividad similar. Se trata de repartir los 25 µl de la mezcla de reactivos de PCR y, en lugar de poner el ADN o el lisado, picar directamente una colonia de la placa con palillo y agitar dentro de la mezcla.

## 7. ELECTROFORESIS

Consiste en la separación de macromoléculas en función de su tamaño y carga. Los ácidos nucleicos, con carga negativa, migran hacia el polo positivo.

### 7.1. ELECTROFORESIS EN GELES DE AGAROSA NO DESNATURALIZANTES

El análisis de productos de PCR, de ADN total o plasmídico o de fragmentos de restricción se llevó a cabo mediante electroforesis horizontal en geles de agarosa (agarosa *SeaKem® LE*, Cambrex/Iberlabo) al 1% en TAE 1x: [TrisHCl 40 mM, EDTA Na<sub>2</sub> 2 mM, ácido acético glacial 0,1142%(v/v)]. El voltaje habitual de trabajo fue de 120 V aunque determinados procedimientos requirieron velocidades inferiores. Como tampón de carga se utilizó una solución 6x (1µl por cada 5 de muestra) de 0,25% (p/v) de azul de bromofenol y glicerol 45% (v/v).

El revelado de los geles de agarosa se realizó por inmersión de los mismos en una solución de bromuro de etidio 1 mg/ml en agua durante 15 a 20 min. Este intercalante (cancerígeno) es visible en un transiluminador que emite luz UV de 365 nm de longitud de onda. Las fotografías de los geles se realizaron con *Gelprinter* y *Gelstation* de TDI para grabar y procesar las imágenes.

### 7.2. ELECTROFORESIS EN GELES DE AGAROSA DESNATURALIZANTES

Esta técnica se ha utilizado tanto para visualizar la integridad del ARN total extraído como para una de las modalidades de *Northern blot* que se detallan más adelante. Consiste en la adición durante la preparación del gel de agarosa de un agente desnaturalizante como es el formaldehído. El porcentaje del gel varía entre las aplicaciones, 1,4% para la primera y 1% para el *Northern blot*. Se prepara disolviendo (calentando) la agarosa (Roche) en agua; cuando se disipe el calor se añade MOPS<sup>38</sup> [4X: MOPS 80 mM, acetato sódico 20 mM y EDTA 4 mM; ajustar a pH 7 con NaOH] y formaldehído hasta concentración final 1X y 1,875% (v/v), respectivamente. El tampón de electroforesis es MOPS 1X. Cuando se trata de ver simplemente un ARN, para no tener que sumergirlo en la solución de bromuro de etidio, éste se adiciona directamente en el tampón de carga 6X [glicerol 45% (v/v), azul de bromofenol 0,09% (p/v) y bromuro de etidio 0,25 mg/ml].

En cambio, protocolos como el *Northern blot* no necesita de la adición bromuro de etidio en el tampón de carga, pero precisan que éste sea desnaturalizante. Así, los ARNs se incuban 10 min a 65°C, tras lo cual se añade 1V de tampón A [Formamida desionizada 50%, Formaldehído 2,5%, MOPS 1X] y se incuba otros 15 min a 55°C. Para finalizar se adiciona 0,1 V de tampón B [Glicerol 50%, EDTA 1 mM, 0,25% azul de bromofeno y 0,25% xilencianol] y se carga la muestra. El volumen final de muestra fueron ~10 µl

---

<sup>38</sup> El MOPS es sensible al calor y a la luz, perdiendo parte de su actividad.





conteniendo 15  $\mu\text{g}$  de ARN (5  $\mu\text{g}$  de RNPs). La electroforesis se llevó a cabo a 60 V hasta que el azul de bromofenol alcanzó el último tercio del gel.

### 7.3. ELECTROFORESIS EN GELES DE POLIACRILAMIDA NO DESNATURALIZANTES

En ocasiones la agarosa no resulta demasiado resolutive por lo que se recurre a electroforesis vertical en geles de poliacrilamida. En este caso el tampón de electroforesis suele ser TBE 1X [Tris-HCl 0,089 M pH 8, ácido bórico 0,089 M y EDTA 0,002 M]. Hemos trabajado con disoluciones 40% (29:1) de acrilamida:bis-acrilamida (Bio-Rad). El porcentaje y el grosor del gel varía según las necesidades y con éstos también cambian las proporciones relativas del agente polimerizante (TEMED, Sigma) y el catalizador (persulfato amónico, APS; Sigma).

El revelado de estos geles suele hacerse con tinción con nitrato de plata (Heuer *et al.*, 1997). Se fija con una solución compuesta por etanol 10% (v/v) y ácido acético glacial 0,5% (v/v) (2 veces durante 3 min); a continuación se incuba en agitación durante 30 min con una solución de nitrato de plata 0,2% en agua. Se lava 4 veces con agua para eliminar el exceso de plata y se revela con una mezcla compuesta por tetrahidrobórato sódico 10% (p/v)/formaldehído 0,4% (v/v)/NaOH 1,5% (p/v) controlando la aparición de las bandas. Se lava con agua destilada y el revelado se para por la incubación con carbonato sódico 0,75%. Previo a ser secado es conveniente mantener el gel toda la noche en una solución de etanol 25% y glicerol 10%.

En el caso de que las muestras que se corran estén marcadas radiactivamente, el revelado se hace por exposición en película fotográfica (veremos más adelante).

### 7.4. ELECTROFORESIS EN GELES DE POLIACRILAMIDA DESNATURALIZANTES

Para conseguir aún más capacidad resolutive recurrimos a electroforesis vertical en geles de poliacrilamida desnaturalizantes por la adición de urea 8 M (Roche). Como en el caso anterior el tampón de electroforesis también fue TBE 1X. Este tipo de electroforesis se han llevado a cabo tanto para la separación de productos marcados radiactivamente como para una de las modalidades de *Northern blot*. Se preparan secuencialmente, esto es, primero se procede a la disolución completa de la urea aplicando calor pero impidiendo que hierva; después se mezclan el resto de componentes en las proporciones adecuadas para finalmente añadir el agente iniciador y el catalizador. Como ejemplo práctico: 25,2 g de urea, 6 ml TBE 10x, 9 ml de Acri/Bis 40% completando con agua hasta un volumen final de 60 ml; 20  $\mu\text{l}$  de TEMED y 600  $\mu\text{l}$  de APS 10%.

A la hora de cargar este tipo de geles hay que tener dos precauciones: que el gel haya precorrido durante 1h a 1500 V ( $\sim 45\text{W}$ ) y haya adquirido temperatura (en torno a los  $50^\circ\text{C}$ ) y que las muestras antes de cargarlas estén desnaturalizadas. Para la

desnaturalización se requiere, además de estar resuspendidas en un tampón desnaturalizante [Tampón 2x: 0,3% Azul de bromofenol, 0,3% Xilencianol, 10 mM EDTA pH7,5 y 97,5% formamida desionizada], el calentamiento de las muestras durante 5 min a 95°C y luego se pasan a hielo. El gel se corre a 1700 V (~45W); la migración de los colorantes es un indicativo del tiempo de electroforesis. Para el revelado de los geles (exceptuamos el *Northern*) es preciso secar los geles a 80°C y vacío (*Gel Dryer*® 583, Bio-Rad). A continuación se exponen a una pantalla de exposición especialmente indicada para este uso (*Imagin Plate 2040*, Fujifilm) y se revelan tras 12-48 h (*Personal Molecular Imager*® FX sustentado por el programa *Quantity One*®, ambos de Bio-Rad).

## 7.5. MARCADORES DE PESO MOLECULAR

Los marcadores de peso molecular fueron en todos los casos comerciales:

- Marcador II: ADN del fago  $\lambda$  digerido con la enzima HindIII. Está compuesto por 8 fragmentos que abarcan desde los cientos de nucleótidos hasta más de 20Kbp (Roche; Universidad Autónoma de Madrid). También se dispone marcado con digoxigenina que marcan el peso molecular en las hibridaciones ADN-ADN.
- Marcador III: ADN del fago  $\lambda$  digerido con las enzimas HindIII y EcoRI. En este caso el patrón de digestión comprende 13 fragmentos que abarcan desde los cientos de nucleótidos hasta más de 20Kbp (Roche). También marcado con digoxigenina.
- Marcador pGEM: ADN del plásmido pGEM-T digerido con las enzimas HinfI y EcoRI generando 15 fragmentos que se encuentran entre las pocos nt hasta las ~2500 bp (Promega).
- Marcador  $\Phi$ 29: ADN del fago con el mismo nombre digerido con HindIII. El patrón de digestión se compone de 14 fragmentos entre las 72 y la 4370 bp (UAM).

## 8. CLONAJE DE FRAGMENTOS

En este apartado se incluyen distintos protocolos utilizados habitualmente para el clonaje de insertos de interés en vectores.

### 8.1. AISLAMIENTO DE FRAGMENTOS DE ADN DE GEL DE AGAROSA

Para el aislamiento de fragmentos de restricción separados por electroforesis horizontal en gel de agarosa se utilizaron kits comerciales de acuerdo con las directrices proporcionadas por los proveedores: *PerfectPrep Gel Cleanup*® (Eppendorf) y/o *QIAquick Gel Extraction Kit* (Qiagen). El paso previo para el uso de kit precisa cortar la banda de



interés, para lo que nos servimos del transiluminador UV una vez se ha teñido el gel con bromuro de etidio.

## 8.2. LIGACIÓN DE FRAGMENTOS DE RESTRICCIÓN

La unión covalente de dos moléculas de ADN lineal se lleva a cabo gracias a la T4 ADN ligasa (Roche). La proporción molar más habitual vector:inserto suele ser 1:3 aunque puede variar dependiendo del tamaño de los fragmentos a clonar. Las reacciones se llevaron a cabo en un volumen final de 10-15  $\mu$ l a 14-16°C durante toda la noche.

### 8.2.1. ADENILACIÓN DE AMPLIFICADOS DE PCR

En ocasiones el producto que se precisa ligar no es un fragmento de restricción sino un producto de PCR. En el mercado se encuentran disponibles vectores que permiten la clonación de amplificadores con T protuberantes (extremos que deja la *Taq* polimerasa), pero algunas polimerasas con actividad exonucleasa 3'-5' dejan extremos romos. Por eso es un procedimiento habitual la adición de adeninas en los extremos de estos fragmentos; este procedimiento está catalizado por la *Taq* polimerasa (5 U) en presencia de 0,2 mM de dATP (70°C durante 30 min). Este producto puede ser utilizado directamente en la reacción de ligación.

## 8.3. PREPARACIÓN DE CÉLULAS COMPETENTES

Una vez que tenemos el ADN circularizado necesitamos que se replique en la bacteria. Con este fin se utilizan un tipo células que está preparado para captar ADN exógeno, las células competentes. Su preparación puede estar mediada por métodos químicos o físicos que alteran la pared celular de la bacteria. En nuestro caso sólo se han preparado células competentes de distintas cepas de *E. coli* y observamos una mayor eficiencia de transformación en células electrocompetentes.

### 8.3.1. CÉLULAS COMPETENTES POR MÉTODOS QUÍMICOS: RbCl

En general, este tipo de protocolos deben llevarse a cabo en condiciones de esterilidad; es aconsejable que el proceso sea rápido y en frío para que el metabolismo celular se enlentezca en el mayor grado posible y consigamos el mayor número de células viables para la transformación.

Se cultivan las células de *E. coli* en 100 ml de medio LB hasta que se alcance una  $A_{600}$  de 0,4 (fase logarítmica). Entonces se detiene el crecimiento incubando el cultivo en hielo durante 15 min. Transcurrido este tiempo, las bacterias se sedimentan por centrifugación (6000 r.p.m., 10 minutos, 4 °C). Con cuidado, se resuspende el sedimento con 32 ml de solución RF1 estéril y preenfriada a 4 °C [RF1 (por 100 ml): 1,2 g de RbCl; 0,99 g de  $MnCl_2 \times 4 H_2O$ ; 0,294 g de acetato potásico; 0,15 g de  $CaCl_2 \times H_2O$ ; 11,9 ml de glicerol; agua destilada hasta 100 ml. Ajustar el pH a 5,8 con una solución de ácido acético 0,2 M.

Esterilizar por filtración y conservar a 4°C]. Se incubaron en hielo durante 15 minutos y después se centrifugaron 6.000 r.p.m. durante 10 minutos. Se retira el sobrenadante y las células se resuspenden en 4 ml de solución RF2 preenfriada a 4 °C [RF2 (por 50 ml): 0,1046 g de ácido morfolino propanosulfónico (MOPS); 0,06 g de RbCl; 0,55 g de CaCl<sub>2</sub> x H<sub>2</sub>O; 5,95 ml de glicerol; agua destilada hasta 50 ml. Ajustar el pH a 6,8 con una solución diluída de NaOH. Esterilizar por filtración y conservar a 4°C]. Se reparten en alícuotas de 100 µl en tubos preenfriados a 4°C y se congelan inmediatamente con nitrógeno líquido. Se conservan por un periodo limitado a -80°C.

### 8.3.2. CÉLULAS ELECTROCOMPETENTES

Para este protocolo se precisa la limpieza completa de las sales del medio de cultivo de manera que no interfieran con la electroporación, el verdadero causante de la desorganización de la pared celular. A su vez, se añade un agente crioprotector que va a permitir la viabilidad de las células conservadas a -80°C, glicerol. Consiste en sucesivos lavados con volúmenes cada vez más reducidos de glicerol 10%; partimos de 500 ml de cultivo a una A<sub>600</sub> de 0,5 cuyo metabolismo se retrasa por incubación en hielo durante 20 min. Se recogen a 6.000 r.p.m. durante 15 minutos y 4°C y se resuspenden en 100 ml de glicerol 10% frío, volviéndose a centrifugar. En el siguiente paso se resuspenden en 20 ml de glicerol para acabar, tras este segundo lavado, con las bacterias resuspendidas en 2 ml de glicerol 10%. Se reparten en alícuotas de 50 µl que inmediatamente se congelan en N<sub>2</sub> líquido. Se conservan a -70°C durante un periodo limitado.

## 8.4. TRANSFORMACIÓN DE CÉLULAS COMPETENTES

Se pueden usar diferentes maneras de inducir la desestabilización de las membranas celulares. En células cuya pared se encuentra dañada se suele usar un choque térmico para abrir poros en la membrana de la pared; en cambio, las células electrocompetentes deben mantener intacta su pared celular por lo que nos ayudamos de una descarga eléctrica para cambiar el potencial de membrana y permitir la captación del ADN.

### 8.4.1. TRANSFORMACIÓN POR CHOQUE TÉRMICO

Se sacaron las células competentes a hielo hasta su descongelación total (~20 min). Se añadió en condiciones de esterilidad el ADN, procedente de ligación o no, y se dejó incubando con las células durante otros 20-25 min. Las células son sometidas a un choque térmico de 42°C durante 90 seg, para seguidamente devolverlas a hielo durante otros 5 min en los que se permitirá la captación del ADN. Ahora debemos permitir la recuperación de las células por la adición de 900 µl de LB fresco durante 1 h a 37°C (en agitación a bajas revoluciones). Transcurrido este tiempo, se siembra parte o todo el cultivo en medio LB sólido, dependiendo de la eficiencia de transformación que esperemos.



## 8.4.2. ELECTROTRANSFORMACIÓN

En este protocolo también debimos descongelar las células en hielo tras lo cual añadimos el ADN con el que deseamos transformar. Transferimos la suspensión celular a una cubeta de electroporación que se somete a un pulso eléctrico según las recomendaciones hechas por el fabricante. En nuestro caso usamos un electroporador *Eppendorf 2510* que recomienda un pulso de 1800 V durante 3-5 milisegundos. Como en el caso anterior, se recompone la estabilidad celular por la adición de 950  $\mu$ l de LB fresco. A continuación se procede como en el apartado anterior.

## 9. SECUENCIACIÓN Y ANÁLISIS DE SECUENCIAS

Para la secuenciación de las construcciones presentadas en este trabajo hemos recurrido al servicio de secuenciación de ADN/Genómica del Instituto de Parasitología y Biomedicina “López-Neyra”. Este servicio amplifica las muestras con un termociclador *PE-9600* (Perkin-Elmer) y las secuencia mediante el secuenciador *ABI 373 XL Stretch* (Perkin-Elmer) usando el kit *ABI PRISM Big Dye™ Terminator Cycle Sequencing Ready Reaction* (Perkin-Elmer).

La visualización de los picos de secuencia se realizó mediante el programa *Chromas Lite 2.01*. El trabajo rutinario en las secuencias (búsqueda de dianas de restricción, diseño de cebadores, comparativa de secuencias, búsqueda de ORFs, etc.) se llevó a cabo con el programa *Clone Manager Professional Suite®*.

Para la comparativa e identificación de nuestras secuencias, así como para la búsqueda de nuevas secuencias en las bases de datos, se utilizaron distintos algoritmos BLAST disponibles en la página web del NCBI (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) (Pearson y Lipman, 1988; Altschul *et al.*, 1990). Para el alineamiento múltiple de secuencia se han usado dos versiones del programa Clustal disponibles en red (<http://www.bioweb.pasteur.fr> y <http://clustalw.genome.ad.jp/>) (Thompson *et al.*, 1994; Higgins *et al.*, 1996; Thompson *et al.*, 1997). El análisis filogenético de las secuencias en función de las identidades en el alineamiento se llevó a cabo mediante el programa PHYLIPS versión 3.573c (*Seqboot*, *Protdist*, *Neighbour*, *Protpars* y *Consense*) (Felsenstein, 1985; Felsenstein, 1995).

La predicción de la estructura secundaria de la proteína se realizó con el servidor JPred (<http://www.compbio.dundee.ac.uk/~www-jpred/>). La secuencia logo basada en alineamientos Clustal W se obtuvo en la página (<http://weblogo.berkeley.edu>).

## 10. TÉCNICAS DE HIBRIDACIÓN DE ÁCIDOS NUCLEICOS

### 10.1. HIBRIDACIÓN ADN-ADN (*SOUTHERN BLOT*)

#### 10.1.1. TRANSFERENCIA ALCALINA POR VACÍO

Una vez que hemos separado el ADN total o plasmídico en geles de agarosa tenemos que transferir y fijar este material a una membrana de *biodine*® cargada positivamente. Para ello nos servimos de un sistema de transferencia alcalina (NaOH 1 M) por vacío (sistema *VacuGene*™ XL de Pharmacia). El proceso se ve facilitado por la rotura del ADN provocada por la exposición prolongada (15-20 min) a los rayos UV en el transiluminador. La membrana debe ser de un tamaño superior al de la ventana de transferencia y ésta, a su vez, más pequeña que el gel. Para evitar el depósito de sales, la membrana se embebió en agua desionizada para luego equilibrarla con 20xSSC [1xSSC: NaCl 150 mM, citrato sódico 15 mM, pH 7,0] durante 5 min. Se prepara el sistema de transferencia según las instrucciones de los proveedores con cuidado de que no se formen burbujas de aire entre la membrana y el gel. Se conecta el vacío y, una vez comprobada la estanqueidad del sistema, se cubre con NaOH 1M. Debemos vigilar que el sistema se mantenga constante en torno a 55 mbar durante 1-2 h. El siguiente paso fue lavar la membrana con una solución de 2xSSC durante 10 minutos en suave agitación y dejar secar a temperatura ambiente. La fijación del ADN a la membrana se llevó a cabo por calor a vacío envuelta en papel Whatman 3MM (120°C 35 min, 70 cm Hg). Se puede conservar a temperatura ambiente.

#### 10.1.2. MARCAJE DE LA SONDA

En este tipo de experimentos el marcaje se lleva a cabo con digoxigenina acoplada a un dNTP que en nuestro caso fue dUTP. La sonda se sintetizó por PCR y su tamaño osciló entre los 200-600 bp. Un procedimiento que aumenta en gran medida la especificidad de la sonda es utilizar como molde un ADN purificado de gel amplificado con la misma pareja de cebadores con la que queremos obtener nuestra sonda marcada. Con este molde se preparó una reacción de PCR que contenía los componentes habituales de un PCR pero además se adicionó digoxigenina-11-dUTP (Roche) en una concentración final de 0,1 mM. Se completan 30 ciclos del PCR y el amplicón se limpia con columnas *MicroSpin*™ S-300HR (GE healthcare). Se ponen en paralelo controles sin DIG para comprobar el retraso en la migración en gel como consecuencia de la incorporación de la marca.

Esta sonda se diluye a 60 ng del ADN /ml de solución de prehibridación. Entre 5-10 ml son suficientes para cubrir completamente la membrana. Estas soluciones de hibridación se conservan a -20°C y pueden ser reutilizadas en sucesivas ocasiones.



### 10.1.3. HIBRIDACIÓN, LAVADOS Y REVELADO

Este proceso se llevó a cabo en un horno de hibridación *Hybridiser® HB-1D* (Techne). Se incubó el filtro en una solución de prehibridación (20 ml/100 cm<sup>2</sup>) compuesta por 5xSSC, sarcosil 0,1%, SDS 0,02%, formamida 50% y solución de bloqueo (Roche) al 2% durante 2 horas a 42°C. Entretanto, habremos desnaturalizado la sonda mediante ebullición durante 10 min y posterior paso a hielo (5-10 min). Se desecha la solución de prehibridación y se añade la solución de hibridación (ver apartado anterior) con la que se incuba a 42°C toda la noche.

Se retiró la sonda y se procedió con los lavados y el revelado de la hibridación. Se realizaron dos lavados de 5 min a temperatura ambiente con 50 ml de 2xSSC/SDS al 0,1% (p/v), seguido de 2 lavados durante 15 min a 68°C con 0,1xSSC/SDS 0,1% (p/v). Estos lavados han servido para eliminar la sonda pegada inespecíficamente. Se disminuye la temperatura del rulo y la del horno de hibridación que trabajará en adelante a temperatura ambiente. Se hace un enjuague con 50 ml de tampón de lavado [0,3% de Tween-20 en tampón 1 (ácido málico 0,1 M, 0,15 M NaCl, pH 7,5)] para eliminar los restos de detergente antes de proceder con el bloqueo de la membrana.

La detección de la sonda se basa en el uso de anticuerpos anti-digoxigenina conjugados con la enzima fosfatasa alcalina. Esta enzima va a encargarse de defosforilar sustratos quimioluminiscentes como es el CSPD que da lugar a un intermediario que emite luz. Para que estos anticuerpos no se fijen a lo largo de toda la membrana es preciso evitar el reconocimiento inespecífico mediante el bloqueo de la superficie del filtro. Con este fin se añaden 30 ml de una solución de bloqueo [Blocking Reagent (Roche) 1% (p/v) en tampón 1] con la que se incuba la membrana durante 30 min. Pasado este tiempo se retira esta solución y se añaden 20 ml de solución de bloqueo en los que se habrá diluido 1:10000 el anticuerpo Anti-Digoxigenin-AP (Roche) con el que se incubará otros 30 min. El exceso de anticuerpo no unido a la sonda se retira mediante dos lavados sucesivos con tampón de lavado. A continuación la membrana se preparó para la adición del CSPD por incubación con 50 ml de tampón 2 [Tris-HCl 0,1 M, NaCl 0,1 M, MgCl<sub>2</sub> 50 mM, pH 9,5] durante 5 min. Tras retirar esta solución se añaden 5 ml de una dilución 1:100 del CSPD (Roche) en tampón 3 y se incuba nuevamente 5 min. Se deja secar perfectamente la membrana, se envuelve en plástico transparente y se incuba a 37°C durante 15 min para desencadenar la reacción de la enzima.

La luz emitida por el sustrato es captada por una película Kodak X-Omat y se realizan dos tipos de exposiciones: una de corta duración que capta los momentos de mayor emisión del sustrato (4-6 h) y otra que permite la detección de señales débiles (24 h). El revelado de este tipo de películas se realiza del modo tradicional con una solución de revelado y una solución de fijación diluidas según las indicaciones del fabricante.

Sobre estas membranas podrían realizarse hasta cuatro hibridaciones pero como el sistema de detección es el mismo, se precisó de eliminar la sonda de una hibridación a la siguiente mediante dos lavados a 37°C de 15 min en NaOH 0,2 M, SDS 0,1% (p/v). La limpieza de restos de la solución anterior se realiza con un lavado de 5 min con 2xSSC. En este punto podría secarse y guardar el filtro, o comenzar añadiendo la solución de prehibridación.

## 10.2. HIBRIDACIÓN ARN-ADN (*NORTHERN BLOT*)

En este trabajo se han llevado a cabo dos protocolos distintos para la separación de los ARNs en gel, electroforesis en gel de agarosa y en geles de poliacrilamida, en ambos casos, desnaturizantes. Por esto, la forma de proceder con la transferencia fue diferente.

### 10.2.1. TRANSFERENCIA DEL ARN

#### 10.2.1.1. Transferencia por capilaridad

Esta técnica se usó para transferir a membrana los ARNs separados en geles de agarosa. Se siguieron las indicaciones descritas por distintos autores (Sambrook *et al.*, 1989; Brown *et al.*, 2004). Tanto el gel como la membrana se equilibraron durante 10 min en 10xSSC. El sistema de transferencia se estableció en el siguiente orden: en la parte inferior se sitúa un papel Whatman 3MM en contacto por los laterales con el tampón de transferencia (10xSSC); sobre éste 3 papeles Whatman 3MM del mismo tamaño del gel, el gel, la membrana y sobre ella, procurando que no se formen burbujas, otros 3 papeles Whatman 3MM. Sobre este sistema se colocan una cantidad considerable de toallas de papel y algo de peso. Se deja durante 12-18 h a temperatura ambiente.

Dada por finalizada la transferencia, se deja secar la membrana y se fija a 80°C durante 1 h a vacío (70 cm Hg). Podemos comprobar cómo ha transcurrido la transferencia tiñendo con azul de metileno 0,03% que nos va a revelar, al menos, la presencia de los ARNs ribosómicos.

#### 10.2.1.2. Transferencia semi-seca o electro-transferencia

Este procedimiento se utilizó para fijar en la membrana los ARNs separados en geles de poliacrilamida 6% desnaturizantes (8M urea). Se llevó a cabo con la unidad *Trans-Blot SD Semi-Dry Transfer Cell* (BioRad) siguiendo las indicaciones del fabricante. El tampón de transferencia fue TBE 1x. El montaje del sistema consistió en la colocación de 4 papeles Whatman 3MM bien humedecidos en TBE1x; a continuación se puso la membrana también muy húmeda (esto ayuda a que se formen menos burbujas), el gel y finalmente otros 4 papeles Whatman 3MM muy húmedos. La transferencia se llevó a cabo a 400 mA durante 50 min. Se fijó una hora a 80°C a vacío (70 cm Hg).





### 10.2.2. MARCAJE DE LA SONDA

Para la hibridación se utilizaron oligonucleótidos de 20-25 nt marcados radiactivamente en el extremo 5' mediante polinucleótido kinasa (PNK) y [ $\gamma$ - $^{32}$ P]ATP. Se marcaron 75 pmol de oligonucleótido con 10 U de T4 PNKinasa y 20  $\mu$ Ci de [ $\gamma$ - $^{32}$ P]ATP (3000Ci/mmol, 10  $\mu$ Ci/ $\mu$ l; GE healthcare). Esta reacción se lleva a cabo en el medio de reacción adecuado a 37°C durante 90 min. Posteriormente, se elimina el exceso de nucleótido marcado no incorporado mediante el uso de columnas de sephadex *MicroSpin™ G25* (GE healthcare). Las distintas sondas utilizadas fueron:

LOCALIZACIÓN	NOMBRE	SECUENCIA	REFERENCIA
Exón 1	Int2	5'-CCGTCCATAGTAGGCAATCC-3'	Martínez-Abarca <i>et al.</i> , 1998
Intrón	P	5'-TGAAAGCCGATCCCGGAG-3	Muñoz-Adelantado <i>et al.</i> , 2003
Exón 2 proximal	4IS+4	5'-GCGAATTGGAGCTCCGAGGC-3'	Este trabajo
Exón 2 distal	5'RTSp	5'-CGCGCGTAATACGACTCACT-3'	Este trabajo

### 10.2.3. HIBRIDACIÓN, LAVADOS Y REVELADO

Se incubaron las membranas en 20 ml de solución de prehibridación [Tampón fosfato 0,5 M pH 7,2, SDS 7%, EDTA 10mM] precalentada a 42°C en un horno de hibridación con agitación rotatoria durante, al menos, 30 min. Transcurrido este tiempo se añadió la sonda marcada y desnaturalizada (95°C 5 min) directamente a la solución de prehibridación (10<sup>6</sup> CPM/ml) y se incubó durante 4-24 h a 42°C. Se retiró la sonda y se hizo un primer lavado en 2xSSC/0,1%SDS a temperatura ambiente para eliminar el exceso de sonda. A continuación se hicieron 3 lavados sucesivos de 15 min a 42°C con 100 ml de las siguientes soluciones (precalentadas): 2xSSC/0,1%SDS, 1xSSC/0,1%SDS y 0,1xSSC/0,1%SDS. Tras el último lavado, dejamos secar la membrana, se envolvió en plástico transparente y se expuso a pantalla (*Imagin Plate 2040*, Fujifilm).

El revelado se realizó con el escáner Personal Molecular Imager® FX sustentado por el programa *Quantity One®* (Bio-Rad).

### 10.3. HIBRIDACIÓN EN COLONIA

Esta técnica se basa en la obtención de lisados celulares fijados a membranas de nylon que posteriormente van a hibridarse con sondas marcadas con DIG en nuestro caso. Se pico por triplicado un número representativo de clones de la genoteca y se dejan crecer

12 h a 37°C. Trascurrido este tiempo, colocamos sobre el medio sólido una membrana de nylon con un diámetro ligeramente inferior al de la placa e incubamos a 37°C durante 2h (debemos reservar una de las placas del triplo como referencia). Se van retirando los filtros de manera que tengamos tiempo de ir pasándolos consecutivamente por las siguientes soluciones embebidas en papel Whatman 3MM: (1) Mantener 2-5 min en solución de lisis [SDS 10%]; (2) Incubar 5-15 min en solución de desnaturalización [NaOH 0,5M y NaCl 1,5M]; y finalmente, 5-15 min en solución de neutralización [Tris-HCl 0,5M pH 8, NaCl 1,5M]. Una vez terminado este proceso, las membranas se lavan en 2xSSC durante 5 min, se dejan secar sobre papel Whatman 3MM y fijamos el ADN a vacío (70 cm Hg) y temperatura (120°C) durante 35 min.

## 11. ENSAYOS DE MOVILIDAD

Independientemente de cuál fue el sistema de detección de los eventos de *retrohoming* producidos por RmInt1, la capacidad de transposición de este intrón fue evaluada en sistemas de plásmido a plásmido en la cepa *S. meliloti* RMO17. Esta cepa tiene la particularidad de carecer de copias de intrón en su genoma, aunque no de la secuencia de inserción hospedadora natural para RmInt1, *ISRm2011-2* (hasta 13 copias). El ensayo consiste en la introducción de dos plásmidos en una misma bacteria: un plásmido donador de intrón (pKG2,5, pKGEMA4 o pKGEMA4<sub>T7</sub>) y un plásmido receptor que contiene 640 bp de la *ISRm2011-2*, entre las que se encuentra el sitio de inserción de RmInt1 (detección por hibridación: pJB0,6LEAD o pJB0,6LAG; detección por resistencia a tetraciclina: pACE0,6LEAD o pACE0,6LAG). Si el intrón invade su diana se va a observar la aparición de un plásmido recombinante derivado del plásmido receptor que contiene el intrón interrumpiendo su secuencia diana. Esto se correspondería, dependiendo del método de detección, con la aparición de una banda nueva, de mayor tamaño, en la hibridación o con la presencia en placa de colonias Tc<sup>R</sup>.

### 11.1. CONJUGACIONES EN *S. meliloti*

La incorporación de los plásmidos donador y receptor en *S. meliloti* se realiza por conjugación triparental (Ditta *et al.*, 1980). En este tipo de conjugación, *S. meliloti* RMO17 va a recibir el plásmido que le aporta *E. coli* DH5 $\alpha$  gracias a los genes de movilización presentes en el plásmido RK2013 (*E. coli* HB101;pRK2013). Son necesarias dos rondas de conjugación para la introducción de ambos plásmidos por lo que una forma de simplificar el proceso fue utilizar como cepa aceptora *S. meliloti* RMO17 con el plásmido receptor ya introducido, así sólo precisamos de una conjugación.

Estas conjugaciones se llevaron a cabo durante 12-16 h en medio rico TY (se mezclaron cantidades similares de cada tipo bacteriano) incubadas a 28°C. Tras lo cual se hizo una extensión en placas de MM con los correspondientes antibióticos para el aislamiento de colonias de *S. meliloti* libres de *E. coli* (ver resistencias en apartado 2; debemos seleccionar aquellas que contengan los dos plásmidos, por tanto, Km<sup>R</sup> Ap<sup>R</sup>). Se



incubaban a 28°C hasta la aparición de colonias (2-3 días). Volvemos a realizar una nueva selección en MM pero ya de colonias aisladas, comprobando en paralelo la posible contaminación con coliformes mediante el uso de Endo-Agar sólido.

## 11.2. DETECCIÓN POR HIBRIDACIÓN

El conjunto plásmidos aislados por lisis alcalina de *S. meliloti* RMO17 se digiere con la endonucleasa de restricción *SaI* y los fragmentos resultantes se separan mediante gel de agarosa 1% en TAE. Esta enzima lineariza el plásmido donador pero tiene dos dianas en el plásmido receptor que flanquean el sitio de inserción del intrón (ver esquema en figura III.8 y apartado 2 de este material y métodos). En el caso de que se haya producido invasión del intrón el patrón de digestión *SaI* nos va a generar 4 bandas: una correspondiente con el donador (7.854 bp para pKG2,5 o 7318 bp para pKGEMA4); otra que deriva del plásmido receptor (5.360 bp); una tercera que se corresponde con el fragmento que contiene la diana no invadida (2.380 bp); y, finalmente, la diana invadida por el intrón (4.264 bp o 3120 bp, dependiendo de si se invade por el intrón completo o por la versión  $\Delta$ ORF).

Para la detección de los eventos de *homing* se utilizaron dos sondas diferentes, una específica del intrón que nos va a detectar el plásmido donador y la banda correspondiente a la invasión del plásmido receptor, y la otra en el exón 2, que nos permitirá detectar tanto la banda derivada de la invasión del intrón como el fragmento correspondiente a la fracción de plásmido receptor no invadido. Con la sonda en el exón 2 podremos evaluar de la eficiencia de invasión según la fórmula:

$$\text{Eficiencia de Homing} = \frac{\text{P. receptor invadido}}{\text{P. receptor invadido} + \text{P. receptor no invadido}} \times 100$$

OLIGO	SECUENCIA	APLICACIÓN	REF.
ε1	5'-GTGAGCGTCGGATGAAAC-3'	Amplificación de la sonda específica de la ribozima de RmInt1 (posiciones 147 a 594 de la secuencia del intrón)	Muñoz-Adelantado, 2003.
18R.0	5'-ACGTTTCTCAATTCGAAACG-3'		
2011-b1	5'-TGGACGAAGACGAACATGG- 3	Amplificación de la sonda específica del exón 2 (ISRm2011-2) (posiciones +3 a +466)	Martínez-Abarca <i>et al.</i> , 1998
2011-b2	5'-TTGAAGTAGGCTGCGCATT-3'		

El análisis y tratamiento de imágenes, así como la cuantificación de las bandas se realizó mediante el programa *Quantity One*® 4.6.2. (BioRad).

### 11.3. DETECCIÓN POR SELECCIÓN DE COLONIAS Tc<sup>R</sup>

Este sistema para poner de manifiesto eventos de movilidad está basado en el diseño experimental propuesto por Guo *et al.* (2000) y modificado por Karberg *et al.* (2001). Consiste en un ensayo con doble plásmido similar al descrito anteriormente con particularidades tanto en el plásmido donador como en el receptor. El plásmido donador (versión ΔORF rodeado por exones -20/+5 y precedido por la IEP, pKGEMA4) presenta el promotor de la ARN polimerasa del fago T7 sustituyendo la mayor parte de la secuencia de la IEP (Barrientos-Durán, 2008); entretanto, el plásmido receptor presenta la diana de inserción del intrón (las mismas 640 bp de la ISRm2011-2) aguas arriba de un gen de resistencia a tetraciclina que carece de promotor, flanqueados en ambos casos por diferentes terminadores de transcripción (Fernández-López, no publicado) (ver esquema en el apartado 2 de este material y métodos). Cuando se produce el salto de RmInt1 sobre su diana, el promotor que porta activa la transcripción del gen de Tc<sup>R</sup> (ver esquema en figura II.18).

La T7 ARN polimerasa no es funcional en *S. meliloti* por lo que el revelado del ensayo de *homing* debe realizarse en células de *E. coli* que presenten la enzima (HMS174 (DE3))<sup>39</sup>. Se sembraron distintas diluciones de la electrotransformación en placas LB/Ap (resistencia que confiere el plásmido receptor *per se*) pero también LB/Ap/Tc (resistencia que adquiere la bacteria si contiene copias del plásmido receptor invadidas). Así la eficiencia de movilidad se puede calcular según la fórmula:

$$\text{Eficiencia de Homing} = \frac{\text{N}^{\circ} \text{ colonias Ap}^{\text{R}} \text{ Tc}^{\text{R}}}{\text{N}^{\circ} \text{ colonias Ap}^{\text{R}}} \times 100$$

## 12. TÉCNICAS CON ARN

En este apartado se engloban todas las técnicas que tienen ARN como material de partida. Además de los procedimientos experimentales, también se han realizado estudios computacionales de predicción de estructuras secundarias en nuestras secuencias de ARN gracias al servidor web *mfold* versión 3.2 (<http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi>) (Zuker, 2003; Mathews *et al.*, 1999).

### 12. 1. SÍNTESIS DE ARNs TRANSCRITOS *IN VITRO*

Para la síntesis *in vitro* de moléculas de ARN se utilizó el kit comercial *Riboprobe*® *Combination System SP6/T7 RNA Polymerase* (Promega) siguiendo las instrucciones del

---

<sup>39</sup> No se hace directamente en *E. coli* porque disponemos de evidencias que apuntan a que la movilidad de nuestro intrón en este fondo genético se encuentra muy disminuida respecto de *S. meliloti*.



proveedor. Todas las secuencias que precisamos transcribir se encontraban bajo el promotor de la T7 ARN polimerasa y la integridad de los productos sintetizados *in vitro* se comprobó mediante geles de agarosa desnaturalizantes.

En el caso de necesitar estos ARNs marcados radiactivamente se utilizó [ $\alpha$ - $^{32}$ P]UTP 3000 Ci/mmol; 10  $\mu$ Ci/ $\mu$ l.

## 12.2. ESTUDIOS DE *SPLICING IN VITRO*

Los productos generados a partir de reacciones de *splicing in vitro* se utilizaron como marcadores para la identificación de las bandas obtenidas en el *Northern blot*. Se usó un transcrito precursor de 946 nt sintetizado con la T7 ARN polimerasa según las condiciones descritas por Costa *et al.* (2006a). El plásmido molde (pLM1) es un derivado de pUC119 que contiene la secuencia del intrón  $\Delta$ ORF precedida por los últimos 15 nt del exón 1 clonado aguas abajo de la secuencia promotora de la T7 ARN polimerasa. La linearización del plásmido se llevó a cabo con *Bam*HI y después de la transcripción *in vitro* sólo observamos una banda en gel de agarosa desnaturalizante. La reacción se llevó a cabo a 45°C en un medio tamponado por 40 mM Tris-HCl pH 7,5 en presencia de 50-100 mM de MgCl<sub>2</sub> y 1 M de KCl o (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Costa *et al.*, 2006a).

## 12.3. DIGESTIÓN DEL ARN

En uno de los diseños experimentales realizados en este trabajo para la detección de moléculas circulares se precisa la digestión de nuestras muestras de ARN. Para ello se eligieron dos posibles alternativas que se explican a continuación.

### 12.3.1. DIGESTIÓN CON UNA *DNazyme*

Las *DNazymes* son oligonucleótidos ADN que presentan capacidad catalítica y permiten la manipulación del ARN una vez sintetizado el transcrito. Se han descrito varios tipos entre los que destacamos las *DNazymes* tipo 10-23 (Santoro y Joyce, 1997); se caracterizan por presentar un pequeño centro catalítico (15 nt) cuyas regiones flanqueantes se encuentran apareadas en la secuencia que se desea cortar (estabilidad de la doble cadena de 10-12 kcal/mol; Santoro y Joyce, 1998). Rompen el enlace entre una purina y una pirimidina en presencia de Mg<sup>2+</sup> dando lugar a un extremo 5' hidroxilo y un 2',3'-fosfato cíclico (Pyle *et al.*, 2000). En nuestro caso la *DNazyme* (5'-CTCGATTCACAGGCTAGCTACAACGACGCTCCCATC-3') se diseñó para producir el corte entre las posiciones 1815 y 1816 del ARN del intrón (brazo proximal del dominio V. Aunque hubo mucho trabajo de puesta a punto con transcritos *in vitro* marcados radiactivamente, las condiciones de digestión elegidas fueron 20  $\mu$ M *DNazyme* en presencia de MgCl<sub>2</sub> 100 mM, Tris-HCl 40 mM pH8, NaCl 150 mM durante 2 h a 37°C. Se llevó a cabo un paso previo de anillamiento de los dos componentes (90°C 3 min, 4°C 5 min) tras lo cual la adición de Mg<sup>2+</sup> desencadena la reacción. El uso en pasos

posteriores precisó la precipitación de las muestras con 3 V etanol 100%, 0,05V acrilamida lineal y 0,05V acetato amónico 3M.

### 12.3.2. DIGESTIÓN CON ARNasa H

El sustrato de la ARNasa H es el ARN en una molécula híbrida ARN-ADN (Agrawal *et al.*, 1990; Nakamura *et al.*, 1991). Por esta razón se diseñó un oligonucleótido ADN que hibrida con el ARN del intrón en la región entre el dIV y el dV (RNaseH: 5'-CACCGCTCCCATCAGGTAAAC-3'). La comprobación de la capacidad de corte con ARNs transcritos *in vitro* no fue la esperada por lo que, ante los resultados con la DNazyme, se abandonó esta aproximación. En cualquier caso, las condiciones de reacción con los transcritos *in vitro* consistieron en la incubación de nuestro oligonucleótido (1-20 pmol) anillado al ARN (50-300 nM) con 2 U de ARNasa H en su correspondiente tampón [Tris-HCl 40 mM pH 8, 3 mM MgCl<sub>2</sub>, DTT 1 mM, KCl 60 mM] durante 2 h a 30°C. La eliminación de la ARNasa H precisó de la digestión con proteinasa K y posterior precipitación.

### 12.4. EXTENSIÓN DE CEBADOR

Esta técnica se emplea de forma rutinaria para detectar sitios de inicios de transcripción o sitios de procesamiento de formas precursoras (Boorstein y Craig, 1989). Consiste en mezclar un ARN problema con un exceso de oligonucleótido complementario a una región cercana al que se supone su extremo 5'; este cebador marcado radiactivamente será utilizado como cebador por una RT que va a producir un ADNc cuya longitud debe corresponder con la localización del extremo de la molécula. A esto debemos sumar que la mayoría de las RTs descritas no son capaces de superar enlaces modificados 2'-5' (Lorsch *et al.*, 1995); precisamente este tipo de enlaces son los que unen el extremo 5' de los intrones maduros a la A desapareada del dominio VI de la ribozima. Así, esta técnica se plantea como un sistema adecuado para la detección y cuantificación del *splicing* en intrones del grupo II.

Nuestro diseño experimental se plantea en la figura II.2. El cebador utilizado fue el oligonucleótido P (5'-TGAAAGCCGATCCCGGAG-3') que hibrida entre las posiciones 97 a 80 del ARN de RmInt1 en dirección hacia la región 5'. Desde el punto de vista experimental se siguieron las directrices marcadas por Marqués *et al.* (1993). Marcamos 10 pmol del cebador P con 8 U de T4 PNKinasa (New England Biolabs) en presencia de 10 µCi de [γ-<sup>32</sup>P]ATP. Aproximadamente 0,2 pmol del oligonucleótido P marcado (300.000 CPM) se hibridaron con 15 µg de ARN total o 5 µg de RNPs en tampón PIPES 10 mM pH 7,5/ NaCl 400 mM; tras 5 min a 85°C, se hace caer de forma rápida la temperatura a 60°C para después dejar enfriar lentamente hasta 44°C temperatura a la cual se añade la mezcla de extensión que contiene tampón AMV RT 1X, dNTPs 1mM, actinomicina D (SIGMA) 60 µg/ml, 2 U de inhibidor de ARNasa (ROCHE) y 7 U de AMV-RT (Roche); se incuba durante a 42°C. La reacción se para



añadiendo 150  $\mu\text{l}$  de etanol 100% preenfriado + 5  $\mu\text{l}$  de acetato sódico 3 M pH 5,2. Los productos se analizaron en gel desnaturalizante (8 M urea) de poliacrilamida al 6%.

El análisis y tratamiento de imágenes, así como la cuantificación de las bandas se realizó mediante el programa *Quantity One*® 4.6.2. (BioRad).

#### 12.4.1. EXTENSIÓN DE CEBADOR ENVENENADA CON ddNTPs (PPE)

Esta es una variación de la técnica anterior que consiste en parar la reverso transcripción por introducción en la reacción de un ddNTPs (Cui *et al.*, 2004). Este sistema se pensó para discriminar entre formas de intrón *lariat* y circular, pero también se usó para distinguir entre precursores no maduros y exones ligados, en virtud de las diferencias de secuencia entre ambos tipos de moléculas. Aunque el proceso es similar al descrito anteriormente el protocolo varía, por ejemplo el volumen de reacción es menor y las enzimas que se utilizaron son también diferentes. Se pueden adoptar dos criterios para el marcaje del ADNc: marcar el oligonucleótido en 5' con PNKinasa y  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  o marcar la elongación del producto con un  $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$ . Se utilizaron ambos tipos de aproximaciones. En protocolos en los que se utilizó oligo marcado, 15  $\mu\text{g}$  de ARN total o 5  $\mu\text{g}$  de RNPs se anillaron con  $\sim 10$  pmol del cebador junto con el tampón de reacción en un volumen final de 5  $\mu\text{l}$  (65°C 5 min, 4°C 5 min); luego se añade el resto de componentes de reacción: 15 U de inhibidor de ARNasas, DTT 10 mM, ddNTP 1,25-0,5 mM, 3 dNTPs 1, 25 mM y 200 U de *SuperScript*®II o de *M-MLV* RT<sup>40</sup> (Invitrogen) y el conjunto se incubaba 90 min a 42°C (SsII) o a 37°C (M-MLV). También se comprobó el comportamiento de la enzima en ausencia de un dNTP (en lugar de añadir un ddNTP); en principio, no afecta a la reacción sólo que se genera un producto un nucleótido más corto.

El protocolo varía ligeramente si lo que deseamos es marcar la incorporación de un  $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$  (se realizó con  $[\alpha\text{-}^{32}\text{P}]\text{dATP}$  y  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$ ). Las proporciones del resto de componentes se mantienen pero, además de la adición de ddNTP y 2 dNTP, hay que añadir 10  $\mu\text{Ci}$  del  $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$  (3.000 Ci/mmol).

Se ensayaron muchas condiciones posibles, combinaciones de cebadores y ddNTP (por eso es difícil, definir un sólo protocolo) y algunas de las mejores se muestran en las figuras II.8 y II.21.

Los cebadores utilizados en estos protocolos fueron:

---

<sup>40</sup> Cuando tratamos de distinguir moléculas circulares y *lariat* precisamos de una enzima cuya capacidad para sobrepasar los enlaces 2'-5' se haya demostrado, *SuperScript*®II RT. Enzimas como la *AMV* RT son muy ineficientes a la hora de superar este tipo de enlaces y cuando se encuentran ante ellos se desenganchan. En cambio, la utilización de *M-MLV* RT para la detección de precursor frente a exones ligados responde a un criterio de disponibilidad.

CEBADOR	SECUENCIA	POSICIÓN <sup>a,b</sup>	APLICACIÓN
LCP1	5'-TCATGTTGAACTCCTTCCGTGCCTCTGCAG - 3'	Intrón 25 a 1 nt	Diferenciación de formas circulares y <i>lariat</i> . Hibridan en el extremo 5' del intrón hacia el inicio
LCP2	5'-AGTTCATGTTGAACTCCTTCCGTGCCTCTG - 3'	Intrón 35 a 6 nt	
RmInt15'	5'-CTCCTTCCGTGCCTCTGCAGCACAC -3'	Intrón 38 a 9 nt	
4IS+4	5'-GCGAATTGGAGCTCCTAGGC -3'	Exón 2 +24 a +4	Diferenciación entre precursor y exones ligados en construcciones pKGEMA4
4PPE	5'-GAATTGGAGCTCCTAGGCCAGG -3'	Exón 2 +22 a +1	
oxi+4	5'-GCTGCTGGAGGAACAGCGTC -3'	Exón 2 +24 a +4	Diferenciación entre precursor y exones ligados en el contexto del gen <i>oxi1</i>

(a) Anotada en dirección 5'-3' del cebador, esto es en dirección 3'-5' de la secuencia donde hibrida.

(b) Las posiciones en el exón 2 se anotan como +X a +N haciendo alusión a que son posiciones aguas abajo de sitio de inserción del intrón (0).

## 12.5. RT-PCR

Este protocolo se utilizó con dos finalidades: la detección de moléculas circulares y la evaluación de los niveles de exones ligados. En ambos casos los procedimientos seguidos fueron similares, lo único que varió fueron los cebadores utilizados en cada caso. Para llevar a cabo la reverso transcripción se partió de 3.25 µg de ARN o RNPs y se incubó junto con 13 pmol oligo RT (nombre genérico) en presencia de 0,625 mM de cada dNTP a 90°C durante 2 min, para después mantenerlo 15 min a temperatura ambiente y 15 min en hielo. Se añadieron los componentes del tampón de extensión hasta un volumen final de reacción de 10 µl: tampón de la enzima 1x, DTT 10 mM, 15 U de inhibidor de ARNasas (GE healthcare) y 120 U de *SuperScript*®II RT (200 U de M-MLV RT en los ensayos para detección de exones ligados) (Invitrogen). Se mantuvo a 42°C durante 2 h, para después inactivar la enzima 15 min a 70°C.

Una vez tenemos el ADNc se procede con la amplificación por PCR. Se diluyen 1/30 las muestras de ADNc y se usa 1 µl como molde para el PCR. Un control paralelo fue poner una cantidad equivalente de esas mismas muestras no sometidas a RT, esto es, aproximadamente 10 ng. Las reacciones de amplificación se llevaron a cabo en un volumen final de 25 µl en presencia de dNTPs 1 mM, 15 pmol de cada uno de los cebadores y 2 U de *Taq* polimerasa. El programa de PCR se ajusta al estándar con una temperatura de anillamiento que depende de la pareja de oligonucleótidos utilizada y 35 ciclos de amplificación.





Este protocolo de PCR puede modificarse para conseguir una mayor sensibilidad en la detección, marcando por fosforilación con  $[\gamma\text{-}^{32}\text{P}]\text{ATP}$  y PNKinasa uno de los oligonucleótidos. Ésto nos permitió distinguir la diferencia de 6 nt que existe entre la molécula *lariat* y la circular en geles de poliácridamida desnaturalizantes al 6%.

El análisis y tratamiento de imágenes, así como la cuantificación de las bandas se realizó mediante el programa *Quantity One*® 4.6.2. (BioRad).

OLIGO	SECUENCIA	POSICIÓN	USO
Ect1	5'-CACCTGCTCGGATCTCGTC-3'	Complementario a una secuencia situada a 138 nt del extremo 5' del intrón	Cebador en la reverso transcripción. Muñoz-Adelantado, 2003
P	5'-TGAAAGCCGATCCCGGAG-3'	Complementario a una secuencia situada a 97 nt del extremo 5' del intrón	Oligo PCR en el extremo 5'. Muñoz-Adelantado <i>et al.</i> , 2003
LL	5'-GAGGTTACACGCACCGTTCTG-3'	Complementario a una secuencia situada a 59-40 nt del extremo 3' del intrón	Oligo PCR en el extremo 3'. Este trabajo
FMIV	4'-ACGAACACGTTTACCTGATG-3'	Complementario a una secuencia situada a 95-76 nt del extremo 3' del intrón	Oligo alternativo para PCR en el extremo 3'. Este trabajo
Int3	5'-GACGCTGGCCGAAATAGC-3'	Hibrida dentro de la secuencia de la IEP, entre las posiciones -288/-269 respecto del sitio de inserción del intrón	Cebador en el exón 1 para la amplificación de exones ligados/precursor. Martínez-Abarca <i>et al.</i> , 1998
4IS+4	5'-GCGAATTGGAGCTCCTAGGC -3'	Hibrida en el exón 2 de construcciones pKGEMA4 entre las posiciones +24 a +4	Cebador en la RT pero también el oligo en 3' (exón 2) para el PCR.
5'-RTSp	5'-CGCGCGTAATACGACTCACT-3'	Hibrida en el exón 2 de todas las construcciones pKG (secuencia del vector)	Este trabajo

### 12.5.1. cRT-PCR

Los experimentos de cRT-PCR se llevaron a cabo de manera similar a los que acabamos de explicar para la RT-PCR, pero las muestras son sometidas a ligación con T4 ARN ligasa para circular cualquier molécula lineal presente en nuestras preparaciones (Vogel *et al.*, 1997; Moore y Query, 2000). Para ello, 25  $\mu\text{g}$  de ARN o 0,625 U A260 de construcciones pKG2.5 se incubaron en una mezcla que contiene Tris-HCl 50 mM pH 7,8,  $\text{MgCl}_2$  10 mM, DTT 10 mM, ATP 1 mM y 50 U de T4 ARN ligasa (New England Biolabs) en un volumen final de 20  $\mu\text{l}$  durante 2 h a 37°C. La enzima se inactivó por calor y la muestra se limpió para reacciones posteriores (extracción con  $\Phi\text{CIA}$  y CIA;

finalmente, se precipita). Resuspendimos en 5  $\mu$ l y se usaron 2,5  $\mu$ l en la reverso transcripción. El cebador en la reverso transcripción fue Ect1 y la pareja de oligonucleótidos en el PCR, P y LL.

## 12.6. ENSAYOS DE PROTECCIÓN DEL ARN FRENTE A DIGESTIÓN CON ARNasas (RPA)

Este ensayo consiste en comprobar si una secuencia de ARN de interés está presente en nuestras preparaciones mediante la formación de un dúplex de ARN y la digestión con una mezcla de endorribonucleasas que digieren ARN de cadena simple (Sambrook *et al.*, 1989). En nuestro caso la técnica se utilizó para la detección de formas circulares escindidas en las preparaciones de ARN total y RNPs. Para llevar a cabo esta técnica se necesitó un ARN específico con el que hibridar; así, para poder visualizar la banda de protección se sintetizó un transcrito ARN *in vitro* marcado radiactivamente con [ $\alpha$ -<sup>32</sup>P]UTP (sonda) a partir de un molde que consistió en el fragmento amplificado por la pareja de oligos P/LL correspondiente a una molécula circular y/o *lariat* clonado bajo el promotor de la T7 ARN polimerasa en el vector comercial pGEM-T easy vector (Promega) (ver RT-PCR; pGT7CIR(AS) que se digirió con *Nde*I). La secuencia de este transcrito va a ser complementaria a aquella que queremos detectar. En el caso de que la sonda se corresponda con la secuencia circular, aproximadamente hacia la mitad del transcrito hay una región que no aparece con el *lariat*; por tanto, en preparaciones en las que estén presentes sólo moléculas de *lariat* la digestión con ARNasas nos debe generar dos bandas a partir de la sonda (si sólo hay circular debe aparecer una sola banda, y si están las dos formas, tres). Debemos tener la precaución de eliminar completamente el ADN molde utilizado en la transcripción *in vitro*.

Se secaron 10  $\mu$ g de ARN a vacío sin temperatura que luego se resuspendieron en 30  $\mu$ l de HB [80% formamida desionizada, 0,4 M NaCl, 1 mM EDTA, 40 mM PIPES pH 6,7] conteniendo 300.000 CPM de la sonda marcada. Se desnaturalizaron en un termobloque a 85°C durante 5 min y luego se dejaron anillar en estufa a 50°C durante la noche. Tras este tiempo, se procedió con la digestión con una mezcla de ARNasas A y T1; se añadieron 350  $\mu$ l de una solución de digestión que contuvo 10 mM Tris-HCl pH 7,5, 300 mM NaCl, 5 mM EDTA pH 8, 0,8 U ARNasa T1 (Roche) y 1 mg/ml ARNasa A<sup>41</sup>, y se incubó 60-90 min a 30°C. Para la inactivación de las ARNasas se añadieron 12,5  $\mu$ l de una mezcla compuesta por 16% SDS y 4 mg/ml proteinasa K (20-30 min a 37°C). Se limpian las muestras mediante fenolización y se precipitan usando 1 ml de etanol 100% + 2  $\mu$ l tRNA (1 mg/ml) + 50  $\mu$ l LiCl 8M. La mitad de la reacción se resolvió en un gel de poliacrilamida desnaturalizante (8 M UREA) al 8%.

---

<sup>41</sup> Esta enzima se vende liofilizada y es conveniente utilizarla preparada reciente. Así antes de comenzar se preparó una solución de ARNasa A 35 mg/ml como se indica en la nota al pie 33, apartado 5.2.1.



## 12.7. ENSAYO DE LIGACIÓN DE OLIGONUCLEÓTIDOS (OLA)

Este tipo de aproximación se utilizó con el mismo fin que el ensayo anterior, la detección directa de moléculas circulares. Este protocolo se fundamenta en que dos oligonucleótidos fosforilados en 5' y anillados contiguos en otra molécula molde pueden ser ligados por una ADN ligasa (Landegren *et al.*, 1988). Se diseñaron una pareja de oligonucleótidos complementarios a los extremos 5' y 3' de la secuencia del intrón. Estos oligonucleótidos se fosforilaron gracias a 10 U de PNKinasa en presencia de [ $\gamma$ - $^{32}$ P]ATP. Se realizó una reacción de anillamiento añadiendo 1 pmol de cada oligonucleótido marcado, el tampón específico de la enzima 1x y nuestro ARN molde (1-15  $\mu$ g) en un volumen final de 8,5  $\mu$ l e incubando a 85°C durante 3 min para después dejar bajar lentamente a temperatura ambiente hasta alcanzar 65°C. Se incubó a esta temperatura durante 30 min y luego se pasó a un termobloque a 37°C durante 20 min. Se añadieron a esta mezcla 30 U de T4 ADN ligasa (Promega) y se incubó 4 h a 37°C. Como control de la reacción se utilizó una ARN transcrito *in vitro* (2-400 ng) que coincide con la secuencia esperada para las moléculas en nuestras preparaciones. Se utilizaron además como controles parejas de oligonucleótidos que no hibridan contiguos (dejan entre ellos al anillar 3 nt) y que, por tanto, no deben ser ligados por la enzima.

Alternativamente, un protocolo idéntico al descrito se desarrolló para la detección de moléculas en lazo con la única diferencia en uno de los oligonucleótidos utilizados (el oligonucleótido que hibrida en la región 5' del intrón es igual en los dos tipos moleculares, ARN circular y *lariat*). En este caso tuvimos la precaución de diseñar dos oligonucleótidos para cubrir un posible reconocimiento erróneo de la A que forma el enlace modificado en la molécula escindida en forma *lariat*, que se diferenciaron en el nucleótido situado en el extremo 5'.

OLIGO	SECUENCIA	POSICIÓN <sup>a</sup>	USO
RmInt1 3'	5'-GGTGAGTAGGCCGGAGGAGTTTCAC-3'	Extremo 3' Posiciones 1860-1884	Detección de formas circulares
RmInt1 5'-3	5'-GAACTCCTTCCGTGCCTCTGCAGCA-3'	Extremo 5' desplazado Posiciones 4-29	Detección de formas circulares y en lazo
RmInt1 5'	5'-CTCCTTCCGTGCCTCTGCAGCACAC-3'	Extremo 5' Posiciones 1-25	
OLAL A	5'-TAGGCCGGAGGAGTTTCACCAGCCG-3'	Extremo 3' Posiciones 1878-1854	Detección de formas <i>lariat</i>
OLAL T	5'-AAGGCCGGAGGAGTTTCACCAGCCG-3'	Extremo 3' Posiciones 1878-1854	

(a) Todos los oligos hibridan en la secuencia del intrón.

## 12.8. AMPLIFICACIÓN DE LOS EXTREMOS DE UN ARN (RACE)

Este protocolo (Rapid Amplification of cDNA Ends, RACE) se ha usado para determinar los extremos del intrón escindido en nuestras preparaciones. Ciertos indicios aportados por otros experimentos inducen a pensar que estos extremos podrían no ser 5'-fosfato y 3'-OH; esto nos ha llevado a introducir una serie de modificaciones en los protocolos para intentar corroborar esta hipótesis.

### 12.8.1. RACE 5'

Este protocolo nos sirvió para la caracterización del extremo 5' de las moléculas lineales derivadas de intrón que están presentes en nuestras muestras de ARN total. Se utilizó el kit comercial *First Choice*® RLM-RACE (Ambion) siguiendo las instrucciones del proveedor. Este procedimiento se basa en la ligación de un adaptador al extremo 5' de las moléculas que presenten un extremo 5' monofosfato libre en la muestra (se usaron 2,2 µg de ARN total en un volumen final de 10 µl) mediante el uso de una T4 ARN ligasa (New England Biolabs). Tras 1 h a 37°C, se toma un quinto de la reacción de ligación y se somete a reverso transcripción con un cebador específico en la secuencia del intrón, Ect1 (ver RT-PCR). Aunque el kit recomienda el uso de *M-MLV* RT, trabajos previos en nuestro laboratorio aconsejan el uso de 15 U de *Thermoscript*™ RT (Invitrogen) que, al tener una temperatura de trabajo superior, soslaya mejor la formación de estructuras secundarias (del Val *et al.*, 2007). Las condiciones de incubación fueron modificadas en función de las condiciones aconsejadas por el proveedor de la enzima: 20 min a 55°C, 20 min a 60°C, 20 min a 65°C y 5 min a 85°C. Después se procede con la amplificación del ADNc generado mediante 2 rondas de PCR, en las que la segunda PCR sitúa los cebadores en posiciones interiores respecto a la primera. Las parejas de cebadores incluyen, en ambos casos, un cebador que hibrida en el adaptador (proporcionado en el kit) y un cebador complementario a la secuencia del extremo 5' del intrón (para la PCR externa se usó ECT1 y, para la interna, el oligonucleótido P, ambos descritos en el apartado 12.5. RT-PCR). Los fragmentos generados se visualizaron en gel de agarosa y también en geles de poliacrilamida no desnaturalizantes. Se aislaron de gel, fueron clonados en pGEM-T *ev* y una serie de colonias seleccionadas por PCR de colonia y digestión fueron enviadas a secuenciar.

### 12.8.2. RACE 3'

Utilizamos este protocolo para la determinación del extremo 3'-OH de las moléculas lineales que contienen la porción final de la secuencia de RmInt1 (al menos dIV-dVI) en las preparaciones de ARN total. En este caso, aunque se utilizaron determinados componentes del kit comercial *First Choice*® RLM-RACE (Ambion), se introdujeron algunas modificaciones ya que este kit está pensado para ARN transcritos de eucariotas con cola de poli A y los ARNm de bacterias carecen de estas modificaciones. Las



muestras (20  $\mu\text{g}$  de ARN en un  $V_{\text{r}} = 20 \mu\text{l}$ ) fueron sometidas a un tratamiento con fosfatasa alcalina (CIP, en el kit) durante 1 h a  $37^{\circ}\text{C}$  para eliminar los grupos 5' monofosfato de las moléculas lineales de manera que no puedan ligarse con el adaptador en 3'. 0,3  $\mu\text{g}$  de un adaptador diseñado con este fin (del Val *et al.*, 2007) se incubaron con 5  $\mu\text{l}$  del ARN previamente tratado en presencia de 10 U de T4 ARN ligasa (New England Biolabs) durante 1 h a  $37^{\circ}\text{C}$ . Los siguientes pasos fueron idénticos a los descritos en el apartado anterior; únicamente destacar que el cebador utilizado para la RT hibridaba en el adaptador (del Val *et al.*, 2007) y el doble PCR se realizó con parejas de cebadores uno complementario a la secuencia del adaptador y el otro específico de la secuencia en el extremo 3' de RmInt1. En concreto, los oligonucleótidos dentro de la secuencia del intrón fueron FMIV para la primera PCR y LL para la PCR interna (ver apartado 12.5. RT-PCR).

### 12.8.3. MODIFICACIONES EN RACE

La inconsistencia de los resultados nos indujo a pensar que podríamos estar encontrando otro tipo de extremos en las moléculas escindidas como intrón lineal. Se ha descrito para muchas ribozimas que su actividad da lugar a productos con extremos 3'-OH y 2',3'-fosfato cíclico. Para tratar de comprobar si nuestras moléculas lineales presentan este tipo de extremos, se hicieron una serie de modificaciones sobre los protocolos tanto de RACE 5' como 3'.

La posible ausencia de un extremo 5' fosfato libre se solventó mediante dos aproximaciones experimentales. La primera de ellas consistió en la fosforilación de nuestras muestras de ARN total previo a la ligación con el adaptador en 5'. 5  $\mu\text{g}$  de ARN total fueron tratadas con 10 U de T4 PNKinasa (New England Biolabs) en presencia de 2 mM de ATP durante 2 h a  $37^{\circ}\text{C}$ ; la mitad de esta reacción una vez fenolizada y precipitada fue usada en los siguientes pasos del protocolo tal y como se ha descrito anteriormente.

Existen otros protocolos para el RACE 5' que no están basados en la ligación de un adaptador sino en la extensión de un ADNc. Este tipo de ensayos podría evitar el problema ocasionado por la naturaleza del extremo de nuestro ARN escindido. Así, se trata de obtener un ADNc específico de nuestra muestra que una enzima transferasa terminal se va a encargar de extender por la adición en su extremo 3'OH de residuos de dATP (sintetizaríamos un ADN con una cola de poli A). La RT se realizó partiendo de 2  $\mu\text{g}$  de ARN total con 15 U de *Thermoscript*<sup>TM</sup> RT en un volumen final de 20  $\mu\text{l}$ . Para la limpieza del ADNc (eliminar resto de ARN) se usó el *CyScribe*<sup>TM</sup> *GFX*<sup>TM</sup> *Purification Kit* (GE healthcare) según las instrucciones del fabricante. La cola de poli A se añadió a un tercio del producto purificado mediante el uso de 160 U de transferasa terminal en las condiciones indicadas por el proveedor (Roche). Se realizaron dos rondas de PCR en las que uno de los cebadores era específico de la secuencia 5' del intrón (Ect1 en la 1ª PCR y

P en la segunda) y el otro contiene una secuencia de poli T complementaria a la cola que hemos añadido.

En cuanto a las modificaciones en el RACE 3', pensamos en la defosforilación del extremo 3'. Una de las especificaciones de la T4 PNKinasa es que en pH ácido es capaz de eliminar el grupo fosfato presente en extremos 3'. El tratamiento consistió en incubar 25 µg de ARN con 20 U de enzima en un tampón compuesto por 70 mM MOPS pH 6, 10 mM MgCl<sub>2</sub> y 5 mM DTT durante 2 h a 37°C (V<sub>F</sub>= 50 µl) previo al tratamiento con CIP y la ligación del adaptador en 3'.

Ninguna de las modificaciones proporcionó datos satisfactorios.

## 13. ACTIVIDADES BIOQUÍMICAS DE LAS RNPs

### 13.1. ACTIVIDAD RT EXÓGENA

Los ensayos realizados para estudiar la actividad reverso transcriptasa (RT) de RmInt1 fueron llevados a cabo básicamente como describen Moran *et al.* (1995) y Matsuura *et al.* (1997). La actividad RT de las RNPs van a usar un sustrato de poli(rA) hibridado a un oligo(dT)<sub>18</sub> para generar un ADNc de alto peso molecular marcado radiactivamente por la incorporación de [ $\alpha$ -<sup>32</sup>P]dTTP; este material de gran tamaño molecular queda retenido en papel Whatman DE-81 y la cuantificación de este producto nos va a permitir evaluar la actividad de nuestras RNPs. En la reacción se va a llevar un control paralelo que nos va a marcar el nivel de fondo del ensayo: la utilización de un sustrato poli(rA) sin cebador. El control positivo será la enzima comercial AMV RT (Roche). Las reacciones con el sustrato poli(rA)/oligo(dT)<sub>18</sub><sup>42</sup> se realizaron por triplicado y luego se obtuvo la media de los valores resultantes. Se prepara el medio de reacción conteniendo tampón RT 1x [KCl 10 mM, MgCl<sub>2</sub> 25 mM, Tris-HCl 50 mM pH 7,5, DTT 5 mM (preparar fresco)], 1 µg de sustrato poli(rA)/oligo(dT)<sub>18</sub> o poli(rA) y 10 µCi de [ $\alpha$ -<sup>32</sup>P]dTTP (800 Ci/mmol; GE healthcare). Se reparte en los correspondientes tubos y la reacción se inició con la adición de 0,025 U D.O.<sub>260nm</sub> de RNPs de manera que el volumen final de reacción fue de 10 µl. Se incubó 37°C 10 min y la reacción se para poniendo 8 µl en papel Whatman DE-81. Dejamos secar y se lavó 4 veces con 250 ml de 2xSSC para eliminar en el mayor grado posible el nucleótido marcado no incorporado. La radiactividad presente en el papel DE-81 se cuenta en un contador de centelleo (*Liquid Scintillation Analyzer Tri-Carb 1500*,  $\Delta$  Packard).

---

<sup>42</sup> Los sustratos se prepararon mezclando poli(rA) 1 mg/ml (Sigma) y oligo(dT)<sub>18</sub> 1 mg/ml (GE healthcare) en proporciones 9:1. Se hirvió 2 min y luego se pasa a hielo. Se hacen alícuotas y se congela. El oligo(dT) es sustituido por H<sub>2</sub>O en el caso de sustrato poli(rA). (Moran *et al.*, 1995).



## 13.2. CORTE DE SUSTRATOS ADN

Hemos hablado de que las RNPs producidas por el intrón son capaces de reconocer su secuencia diana e insertarse en una nueva localización. *In vivo*, este proceso se ha evaluado con los ensayos de *retroboming*. En este apartado se va a detallar cómo se caracteriza el proceso *in vitro*, desde la construcción y purificación de sustratos hasta la reacción final. Los protocolos que se describen a continuación son modificaciones ligeras de los detallados en Muñoz-Adelantado *et al.* (2003).

### 13.2.1. MARCAJE DE SUSTRATOS

Independientemente del tipo y del marcaje aplicado, todos los sustratos fueron purificados en gel no desnaturante de poliacrilamida al 6% de 3 mm de grosor (3 h 6W constantes). Para ello, una vez sometidos a electroforesis, se retiró uno de los cristales y se expuso a una película fotográfica durante 10 min (como indicadores de posición se utilizan unos marcadores fluorescentes). Usando como patrón la autoradiografía se recortaron las bandas del gel (dependiendo de la intensidad de la banda se divide en 2 o 3 tubos para mejorar la eficiencia de elución) y se trituraron en el interior de un tubo de 1,5 ml. Se sometieron a dos ciclos de congelación-descongelación (hielo seco-temperatura ambiente); justo antes del último ciclo se añadieron 700-800  $\mu$ l de tampón de elución [TE1x, 0,5 mM acetato amónico]. Se incubó a 37°C durante la noche. Se procedió con la eliminación de todos los restos acrilamida: se centrifugaron los tubos arrastrando al fondo el polímero; recogimos la máxima cantidad de sobrenadante que pudimos y se filtró mediante columnas *Spin X-100* (Costar). El eluato correspondiente a cada sustrato se repartió en tubos con un volumen aproximado de 300  $\mu$ l, se sometieron a extracción con 1 V de  $\Phi$ CIA y se precipitaron en etanol 100% en presencia de 0,2% de acrilamida lineal (Bio-Rad) y 25 mM acetato amónico. Se mide seco y se calcula el volumen necesario para que al resuspender quede a 300.000 CPM/ $\mu$ l.

#### 13.2.1.1. Marcaje de sustratos de ADN de cadena simple

##### 13.2.1.1.1. Marcaje en 5' con polinucleótido kinasa y [ $\gamma$ -<sup>32</sup>P]ATP

Como sustrato de cadena simple se utilizó un oligonucleótido sintético de 70 nt purificado por HPLC, WT (5'-AATTGATCCCGCCCGCCTCGTTTTCATCGATGAG ACCTGGACGAAGACGAACATGGCGCCGCTGCGGGGC-3'). Este cebador es simétrico respecto al sitio de inserción del intrón. Para el marcaje en 5' se utilizó [ $\gamma$ -<sup>32</sup>P]ATP y T4 polinucleótido kinasa. La cantidad de sustrato marcado varía dependiendo del número de muestras que se necesiten ensayar pero una reacción estándar para 20 muestras sería: 100 pmol de oligonucleótido WT, 100 U de enzima y 80  $\mu$ Ci [ $\gamma$ -<sup>32</sup>P]ATP (3.000 Ci/mmol, 10  $\mu$ Ci/ $\mu$ l; GE healthcare) en un volumen de 100  $\mu$ l. Se incubó 2h a 37°C y se retira el exceso de nucleótido marcado no incorporado por filtración en

columnas *MicroSpin™ G25* (GE healthcare). Se le añaden 20  $\mu$ l de un tampón de carga 6x [glicerol 50% más colorantes].

#### 13.2.1.1.2. Marcaje en 3' con terminal transferasa y [ $\alpha$ - $^{32}$ P]ddATP

Este mismo sustrato fue marcado en el extremo 3' para efectuar los ensayos de *splicing* reverso, con el *kit* comercial *DNA 3'-End Labeling kit* (Roche). 100 pmoles de oligonucleótido WT se incubaron, según las condiciones descritas por el fabricante, con 1600 U de transferasa terminal en presencia de 100  $\mu$ Ci [ $\alpha$ - $^{32}$ P]ddATP (3.000 Ci/mmol, 10  $\mu$ Ci/ $\mu$ l; GE healthcare). La reacción se paró añadiendo EDTA a una concentración final de 25 mM. El nucleótido marcado no incorporado se retiró por filtración en columnas *MicroSpin™ G25* (GE healthcare). Añadimos la correspondiente cantidad de tampón de carga 6x.

#### 13.2.1.2. Marcaje de sustratos de ADN de cadena doble

Los sustratos de cadena doble se sintetizaron mediante PCR. El molde que se usó fue el propio oligonucleótido WT que sirve de sustrato de cadena simple y se amplificó con la pareja de cebadores S70ds/UP (5'-AATTGATCCCGCCCGCCTC-3') y S70ds/DN (5'-GCCCCGCAGCGGCCCATGTT-3') que hibridan en los extremos de ese molde amplificando un sustrato de 70 bp. En la amplificación (25 ciclos con una temperatura de anillamiento de 60°C) se utilizaron 2,5 x 10<sup>-3</sup> pmol de WT y 60 pmol de cada cebador en presencia de 0,4 mM de dNTPs y 0,2 U de *Vent* polimerasa (New England Biolabs) ( $V_F$ = 50  $\mu$ l). La reacción se efectuó en un termociclador *Gene Amp PCR System 2400* (Perkin-Elmer).

##### 13.2.1.2.1. Marcaje en 5' con polinucleótido kinasa y [ $\gamma$ - $^{32}$ P]ATP

Consiste en el marcaje del extremo 5' de la cadena que va a invadir el intrón. Esto se consigue fosforilando radiactivamente el cebador S70ds/UP antes de hacer el PCR. Las condiciones de marcaje son similares a las descritas con anterioridad.

##### 13.2.1.2.2. Marcaje interno mediante PCR y [ $\alpha$ - $^{32}$ P]dCTP

Una opción para conseguir detectar todos los productos que se generen por la invasión del intrón (tanto banda de corte como *splicing* reverso completo y parcial) es el marcaje interno del sustrato por incorporación durante la amplificación de [ $\alpha$ - $^{32}$ P]dCTP. Este protocolo precisa de la adición en la reacción de una mezcla de dCTP frío/caliente (sin marcar/marcado en P  $\alpha$ ). Así, la mezcla de reacción contuvo 400  $\mu$ M de 3 dNTPs (dATP, dGTP, dTTP), 40  $\mu$ M de dCTP, 60  $\mu$ Ci [ $\alpha$ - $^{32}$ P]dCTP (3000 Ci/mmol; 10  $\mu$ Ci/ $\mu$ l), 50 pmol de cada cebador, 2,5 x 10<sup>-3</sup> pmol de WT y 0,2 U de *Vent* polimerasa (New





England Biolabs) en un volumen final de 50  $\mu$ l. Se limpió de nucleótido marcado no incorporado mediante el uso de columnas *MicroSpin™ G25* (GE healthcare).

### 13.2.2. REACCIÓN DE CORTE/*SPLICING* REVERSO

En realidad aunque se pueda definir de dos maneras posibles se trata de una misma reacción. Consiste en incubar las preparaciones enriquecidas en RNPs (0,2 U D.O.<sub>260</sub>) con 300.000 CPM del sustrato marcado en un volumen de reacción de 10  $\mu$ l. La composición del tampón de reacción se determinó experimentalmente (Muñoz-Adelantado, 2003): 10 mM KCl, 25 mM MgCl<sub>2</sub>, 50 mM Tris-HCl pH 7,5 y 5 mM DTT; el tiempo de incubación se prolongó 35 min a 37°C. La reacción se paró añadiendo 40  $\mu$ l TE, 0,2% acrilamida lineal, 0,3 M acetato sódico pH 5,2. Tras una extracción con 1V de  $\Phi$ CIA se precipita a -80°C con etanol 100%. La mitad de la muestra se resuelve en un gel de poliacrilamida desnaturizante al 6%.

El análisis y tratamiento de imágenes, así como la cuantificación de las bandas se realizó mediante el programa *Quantity One®* 4.6.2. (BioRad).

### 13.2.3. REACCIÓN DE UNIÓN DE LAS RNPs AL SUSTRATO ADN

Este tipo de ensayos se desarrollan de forma similar a los anteriores, usando sustratos de cadena simple y doble marcados en 5'. Se realizó un ensayo de retardo de movilidad en gel (Sambrook *et al.*, 1989; Buratowski y Chodosh, 1996) para comprobar si la unión de nuestros complejos retrasa la migración del sustrato de ADN en geles de poliacrilamida. 0,2 U D.O.<sub>260</sub> de RNPs se incubaron en presencia de distintos tampones con 15000 CPM de sustrato. En el medio también se incluyó poly(dI-dC) 20  $\mu$ g/ml y BSA 200  $\mu$ g/ml para minimizar uniones inespecíficas. La incubación se llevo a cabo a 30°C (salvo un punto que se incubó a 37°C) en termobloque durante 25 min. Transcurrido este tiempo las muestras se analizaron por electroforesis en minigeles de poliacrilamida no desnaturizantes al 4,5% en Tris-Glicina 375 mM pH 8,3 mediante el sistema *Mini-PROTEAN® II Electrophoresis Cell* (BioRad). Se utilizaron los siguientes tampones:

RT → 50 mM Tris-HCl pH 7,5, 25 mM MgCl<sub>2</sub>, 10 mM KCl, 5 mM DTT

A → 10 mM Tris-HCl pH 8, 5% glicerol, 0,5 mM EDTA, 1 mM DTT

B → 10 mM Tris-HCl pH 8, 10 mM MgCl<sub>2</sub>, 5% glicerol, 0,5 mM EDTA, 1 mM DTT

C → 50 mM Tris-acetato pH 8, 8 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 100 mM K(CH<sub>3</sub>COO), 27 mM NH<sub>4</sub>(CH<sub>3</sub>COO), 3,5% PEG, 1 mM DTT

D → 40 mM Tris-HCl pH 7,5, 5 mM MgCl<sub>2</sub>, 450 mM NaCl, 5% glicerol, 5 mM DTT, 1U/ $\mu$ l inhibidor ARNasas

E → 20 mM HEPES-KOH pH7,9, 100 mM KCl, 8% glicerol, 1 mM EDTA, 1 mM DTT

F → 20 mM HEPES-KOH pH7,9, 2 mM MgCl<sub>2</sub>, 50 mM KCl, 20 μM Zn(CH<sub>3</sub>COO)<sub>2</sub>, 10% glicerol, 4 mM espermidina, 0,5 mM DTT

G → 20 mM HEPES-KOH pH7,9, 5 mM MgCl<sub>2</sub>, 60 mM KCl, 4 mM Tris-HCl pH 8, 12% glicerol, 0,6 mM EDTA, 0,6 mM DTT

H → 10 mM Tris-HCl pH 8, 5% glicerol, 0,1 mM EDTA, 0,1 mM DTT

I → 50 mM Tris-HCl pH 8, 300 mM NaCl, 20 mM Imidazol, 10% glicerol

J → 25 mM Tris-HCl pH 8, 8 mM Mg(CH<sub>3</sub>COO)<sub>2</sub>, 10 mM KCl, 3,5% PEG, 1 mM DTT

K → 40 mM Tris-HCl pH 8, 100 mM NaCl, 10% glicerol, 0,1% EDTA, 1 mM DTT

#### 13.2.4. ESCALERAS DE SECUENCIACIÓN: MÉTODO DE SANGER

Dependiendo de las necesidades del ensayo se precisó una u otra combinación de molde y cebador, pero en cualquier caso la secuenciación que se usó como marcador de tamaño molecular en algunos geles de poli(acrilamida) desnaturizantes se realizó con el *T7 Sequencing Kit* (USB). Consistió en la desnaturización alcalina del molde (1,5-2 μg de ADN) para su posterior anillamiento con 10 pmol de cebador (5 min a 65°C, 10 min a 37°C, 5 min a temperatura ambiente) en un volumen final de 14 μl. La reacción de secuenciación se inició con la adición de la mezcla de marcaje (dATP, dGTP y dTTP 1,375 μM y 333,5 mM de NaCl), 5 μCi de [α-<sup>32</sup>P]dCTP (3.000 Ci/mmol; GE healthcare) y 2 μl de una dilución 1/4 de la T7 ADN polimerasa y se mantuvo 5 min a temperatura ambiente (la enzima es algo menos activa a esta temperatura con lo que nos aseguramos una buena lectura a lo largo de toda la escalera). Ponemos 4,5 μl de esta reacción en cada uno de los cuatro tubos que contienen los ddNTP (840 μM de 3 dNTPs y el cuarto se encuentra en proporción 93,5 μM dNTP/14 μM ddNTP en un tampón 40 mM Tris-HCl pH 7,5 y 50 mM de NaCl) y se incuban a 37°C. Se le añaden 5 μl de tampón desnaturizante 2x y se cargan 2 μl por carril en gel.

En concreto las escaleras realizadas como marcador para los ensayos de corte de las RNPs se llevaron a cabo con pGEM0.6 (pGEM-T que contiene 640 bp pertenecientes a *ISRm2011-2* e incluyen la secuencia de nuestros sustratos) y el cebador S70ds/UP.

## 14. CONSTRUCCIÓN DE MUTANTES

Se siguieron dos estrategias diferentes para conseguir las distintas mutaciones. El sistema más habitual fue la mutagénesis puntual mediante *Altered Sites® II* in vitro



*Mutagenesis Systems* (Promega), aunque hubo mutantes que por el mayor número de cambios o porque, por razones que desconocemos, no se pudieron obtener con este sistema; en estos casos puntuales se utilizó el PCR solapante para la introducción de las mutaciones.

#### 14.1. MUTAGÉNESIS MEDIANTE KIT COMERCIAL

El fundamento del protocolo es la introducción de mutaciones mediante hibridación con oligonucleótidos que contienen cambios respecto a la secuencia original. Debemos tener clonada nuestra secuencia a mutar en el vector pALTER-1® (Promega); en nuestro caso se utilizaron dos versiones: pAL2.5 (que contiene el intrón flanqueado por exones -174/+466; Muñoz-Adelantado *et al.*, 2003) para las mutaciones en la IEP y pALΔORF<sub>17</sub> (que contiene la ribozima en el contexto de pKGEMA4 con el promotor T7 clonado en el sitio *Xba*I) para los cambios en los dominios I y VI de la ribozima. Se anillan tres oligonucleótidos a la vez: uno de ellos fue nuestro oligonucleótido mutagénico y los otros dos van a servir para cambiar la resistencia que el plásmido confiere a la bacteria (además de darnos una idea de la probabilidad de que se haya introducido nuestra mutación). Originalmente el plásmido es resistente a tetraciclina pero sensible a ampicilina; si ambos oligonucleótidos anillan en su lugar correspondiente, tras la síntesis y reparación de las muescas, deberemos obtener un plásmido Ap<sup>R</sup> pero Tc<sup>R</sup>.

Una vez que hemos comprobado por secuenciación que la mutación es correcta los insertos se introducen por digestión en el vector definitivo. Así, las construcciones con el C-terminal de la IEP mutado se transfirieron como fragmentos *Bam*HI/*Spe*I al plásmido pKG0, mientras que los mutantes en la ribozima se introdujeron como *Sac*I en construcciones pKGEMA4<sub>17</sub> digeridas con el mismo enzima (extremos defosforilados). Los correspondientes mutantes se transfirieron por conjugación a *S. meliloti* RMO17 y *S. meliloti* RMO17: plásmido receptor.

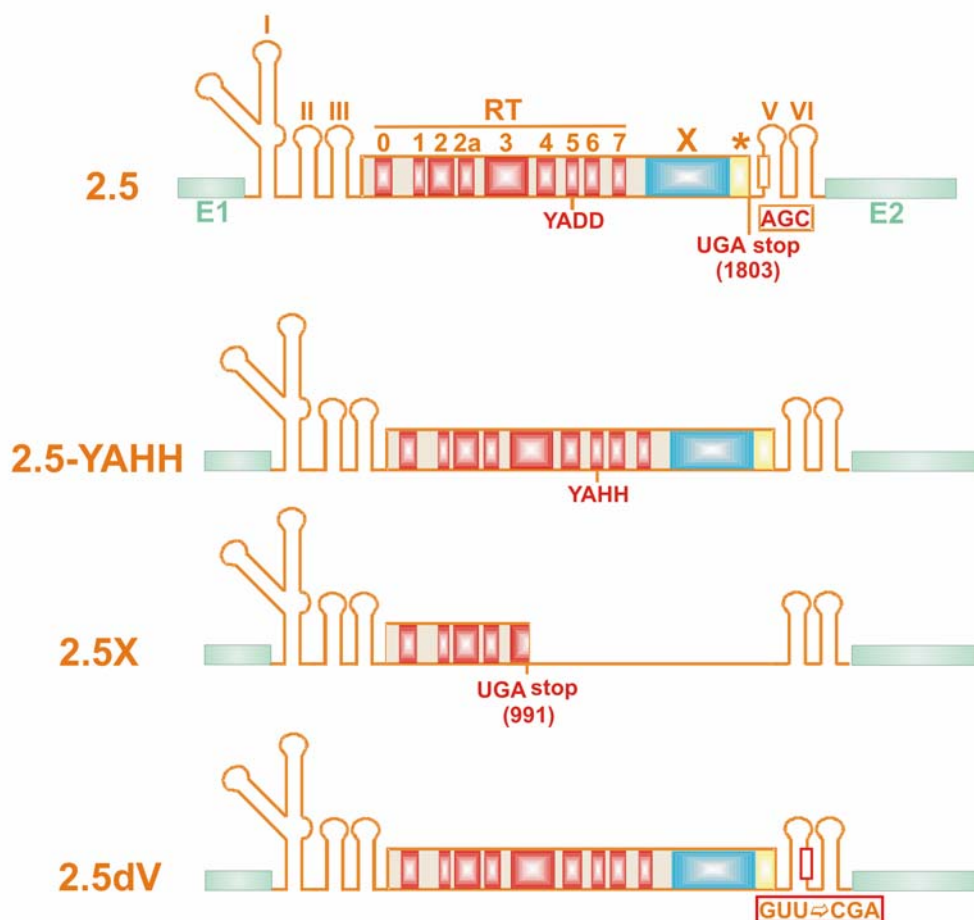
#### 14.2. INTRODUCCIÓN DE MUTACIONES POR PCR SOLAPANTE

Esta técnica es un sistema muy eficiente para la introducción de mutaciones descrito por Higuchi *et al.* (1988). Consiste en el diseño de dos parejas de oligonucleótidos que amplifican dos regiones que solapan en ~20 nt; es precisamente en esta región donde se encuentra la mutación. Una amplificación posterior utilizando como molde los dos fragmentos generados en la PCR anterior y los cebadores en los extremos no solapantes, va a dar lugar al fragmento mutado de longitud completa. Este fragmento será clonado en el vector definitivo y transferido a *S. meliloti* RMO17 con o sin plásmido receptor. Las condiciones de PCR fueron similares a las descritas en un protocolo estándar (60°C de temperatura de anillamiento y 30 ciclos): 5 ng de molde, 0,1 μM cada cebador, 0,1 μM dNTPs, pero la enzima usada fue *Eppendorf*® *Triple Master* en tampón “*high fidelity*” 1x (*Eppendorf*). Para la segunda PCR, el molde lo constituyen 3 μl de cada una de las anteriores.



# **Apéndices**



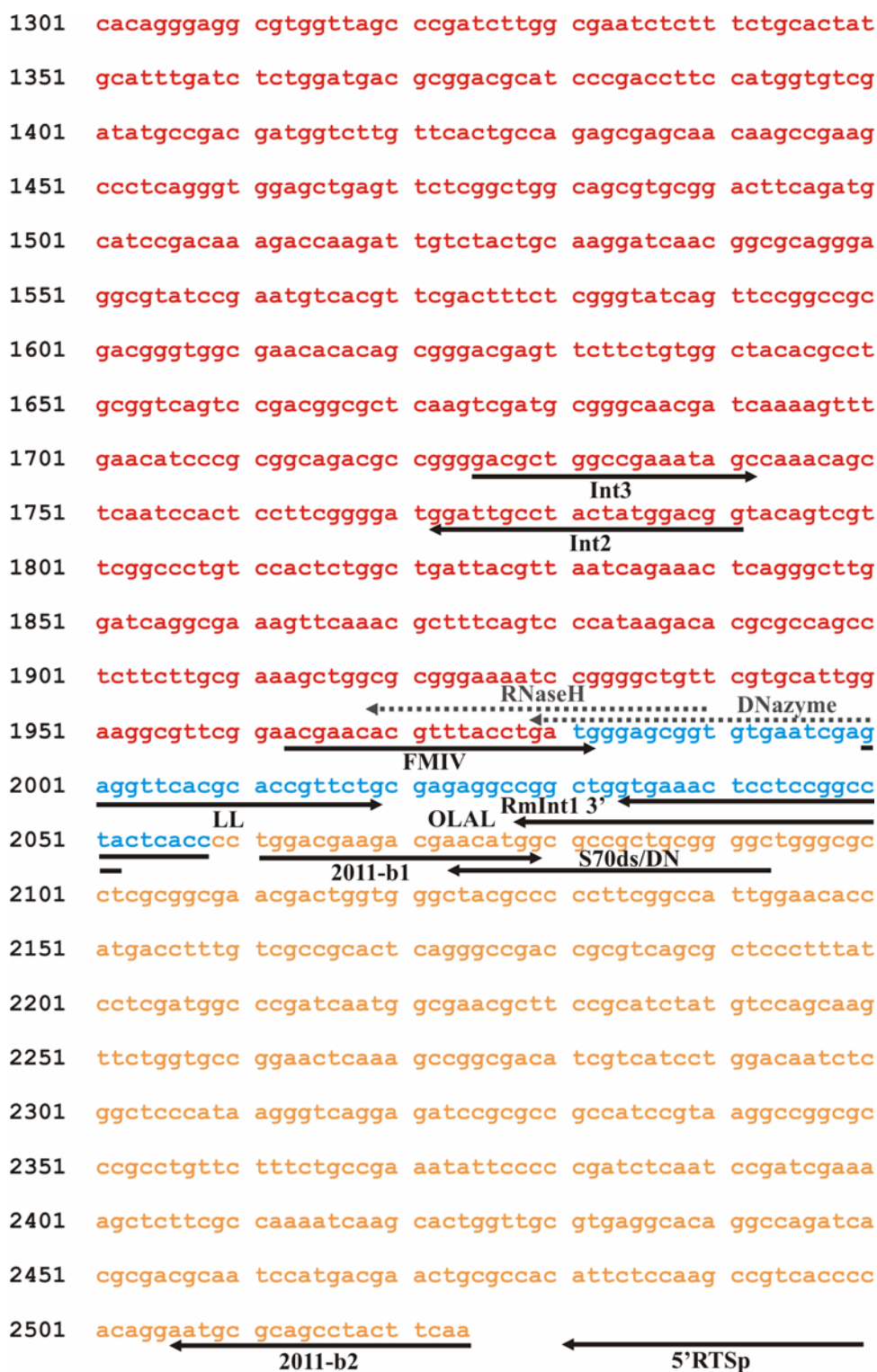


**Apéndice I.** Representación esquemática de diferentes mutantes que se utilizan como controles generales en muchos de los ensayos expuestos en este trabajo. En la parte superior se muestra la construcción silvestre, en la que se anotan las posiciones más importantes desde el punto de vista catalítico: en el dominio RT5 de la IEP se encuentra el centro activo para la actividad reverse transcriptasa (YADD), mientras en el dominio V de la ribozima se encuentra la triada catalítica AGC, residuos críticos para que el ARN del intrón sea activo. En orden descendente se representan los siguientes mutantes: mutante afectado en el dominio activo de la RT, donde los aspárticos en las posiciones 230 y 231 (DD) han sido cambiados a histidinas (HH) (Muñoz-Adelantado *et al.*, 2003); mutante 2.5X en el que la proteína se ha truncado por la generación de un codón de parada al cambiar la pauta de lectura (al eliminar un sitio EcoRI) después de la posición 140 de la proteína, aproximadamente hacia la mitad del dominio RT3 (Martínez-Abarca *et al.*, 2000); y finalmente, el mutante en el dominio V de la ribozima en el que los 3 nucleótidos enfrentados con los residuos de la triada catalítica se han cambiado de GUA a CGA, aboliendo la capacidad de maduración (Muñoz-Adelantado *et al.*, 2003). Los dominios se representan: en rojo, los dominios RT; en azul, el dominio madurasa; y la extensión C-terminal (distinguida con un asterisco), en amarillo. Las posiciones anotadas se refieren a la secuencia de nucleótidos de RmInt1.

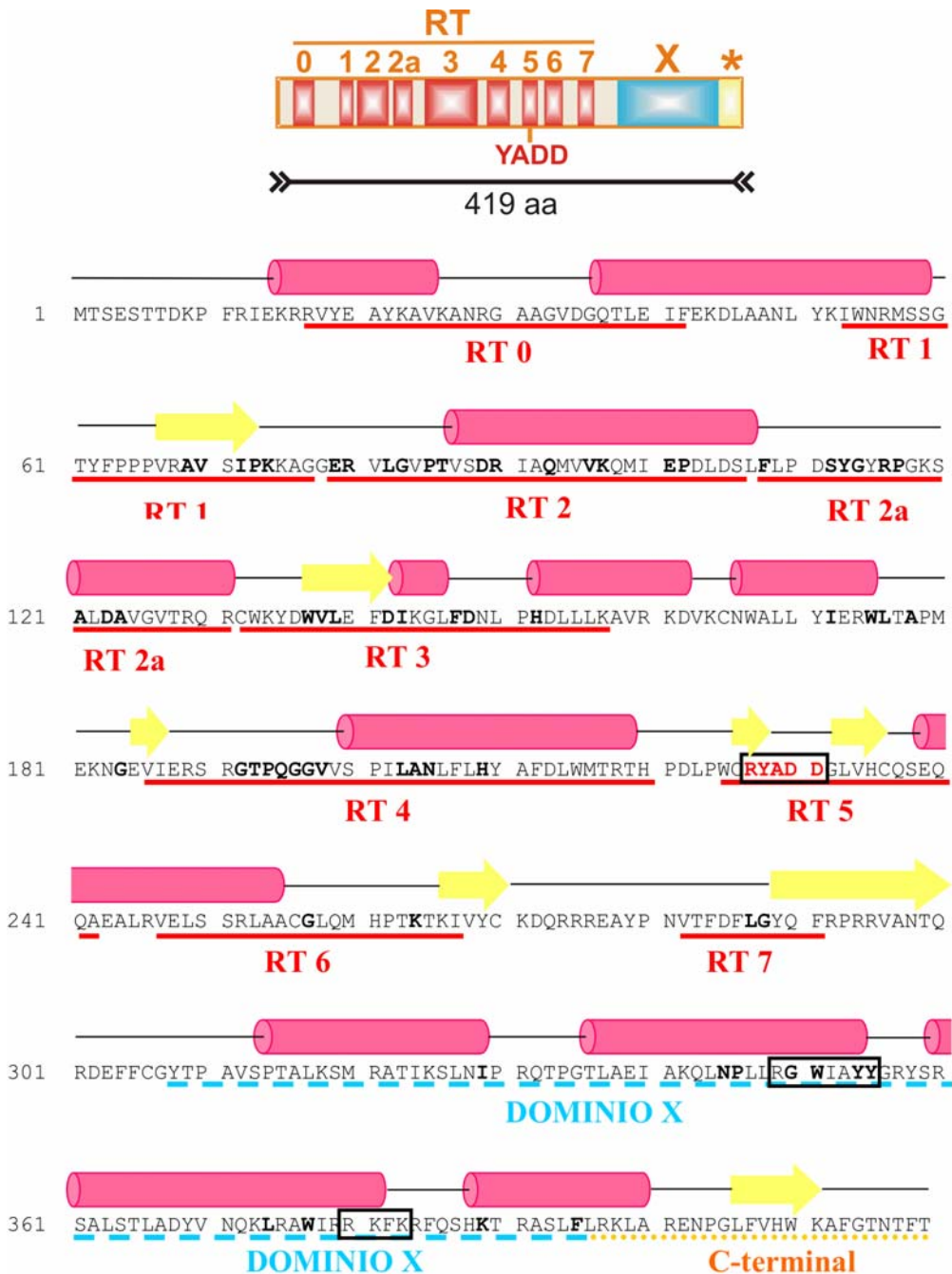
1 ctgaaggtgg attatcgcgc cgtctggacc ttcgtgcacg aagagggggt  
 51 gagttataaa aaaagacgct ggtcgccagc gaacgggagc ggccccgacgt  
 101 cgccccgccac cgggcacgat ggctgaagca ctgccccgga attgatcccc  
 151 cccgctcgt tttcatcgat gagagtgtgc tgcagaggca cggaaggagt  
 201 tcaacatgaa ctaagaccgt ggcgtaaagc tgcgtgaatg atgggggacg  
 251 gccctccggg atcggctttc aggagcgggt ctcaaaccag tccgagctgc  
 301 tgcggtaaag agccgtggtg gtgagcgtcg gatgaaacgt tccgacgaga  
 351 tccgagcagg tgcattgtcca aaagacgaac gaaagtgaac cctccgagga  
 401 cgcgtcgtta tgaacgtaag tgtcgtcga accaggaccg tttcgtcatc  
 451 ctgggacaag tccgccagat gcctgatgac cgggcgggcg gcgaccggcg  
 501 tagagggggc gtgagttgga cataggcttt cacgcggaac tgcaggaacc  
 551 aggctcctga tgtcaaggga gaagctcaag cggcgcaaac cgcaaggcga  
 601 gagtaccgat gcaggagact ggggcggatc gccccgtatg agcgtcgagg  
 651 accctgtaat ggggtcggag caaagggggc ggatcaggcc gtcgtattgt  
 701 ttgaaacaac tggaaacag atgacttcgg aaagtacgac agacaagccg  
 751 tttgaattg agaaacgtcg agtgtacgaa gcttaciaaag cggtcaaage  
 801 caaccgtggc gcggccgggg tggacgggca gacgctggag atatttgaga  
 851 aagaccttgc agcaaaccct tacaagatct ggaatcggat gtcctcggga  
 901 acctactttc cgccgccggt gcgcgccgct tccattcoga agaaggctgg  
 951 aggcgaaagg gttttgggtg tgcccacggt cagcgatcgg atcgcgcaga  
 1001 tgggtgtcaa gcagatgatc gagccggatt tggactccct ctttcttccg  
 1051 gactcctacg gttacaggcc gggaaaatcg gccctggatg ctgtcggagt  
 1101 gacgcgtcag cggtgctgga agtatgattg ggttttgaa ttcgacatca  
 1151 aagggtgttt tgacaatctt ccgcatgatc tcttgctgaa ggcggtcaga  
 1201 aaagacgtca aatgcaactg ggctctgctc tacatcgaaa gatggctgac  
 1251 tgcgcctatg gaaaagaacg gagaagtcac cgagcgggta cgcggtaccc

LCP2 ←  
 S70ds/UP ←  
 RmInt1 5' ←  
 LCP1 ←  
 P ←  
 ε1 ←  
 ECT1 ←  
 18R.0 ←





**Apéndice II.** Secuencia de nucleótidos de RmInt1 (en azul, la ribozima, y en rojo la proteína) flanqueada por exones -174/+466 (en naranja), como se encuentra en pKG2.5. Los cebadores se encuentran anotados con flechas negras; en trazo discontinuo gris, los oligos usados en las digestiones del ARN.



**Apéndice III.** Secuencia de aminoácidos de la proteína codificada por RmInt1. Bajo la secuencia se marcan los límites de los dominios identificados: en rojo, de los dominios RT0 al RT7 relacionados con la actividad reverso transcriptasa; en azul, el dominio X implicado en la interacción con el ARN del intrón necesaria para su maduración; y en amarillo, la extensión C-terminal de función desconocida. Enmarcados se señalan los residuos críticos de la actividad RT y los motivos mejor conservados en el dominio madurasa. Sobre la secuencia la predicción de la estructura secundaria (Jpred): en cilindros fucsia las hélices  $\alpha$  y en flechas amarillas las cadenas  $\beta$ .



## *RESULTADOS Y DISCUSIÓN*



# Capítulo I

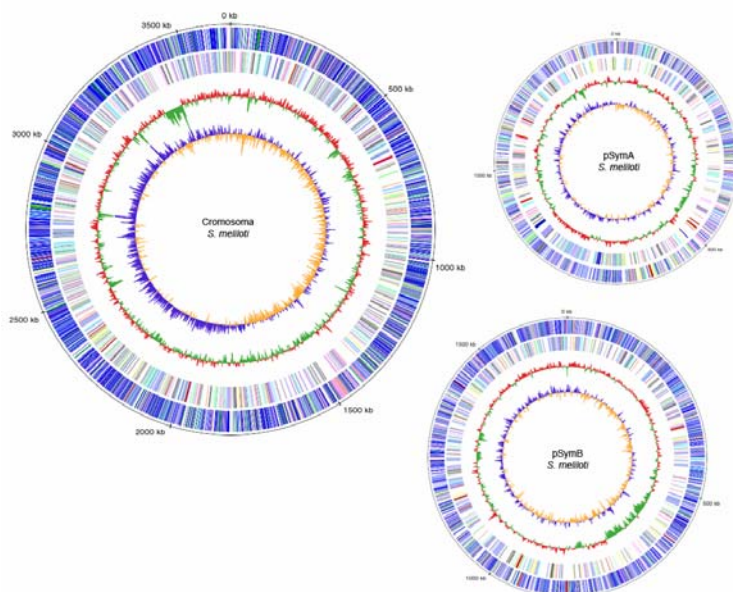


## I. NUEVOS RETROELEMENTOS CARACTERIZADOS EN *S. meliloti* GR4: INTRONES DE LA CLASE E

Las bacterias simbióticas, junto con los organismos fitopatogénicos, son los microorganismos de mayor interés agronómico; dentro de este grupo se encuentran los rizobios capaces de inducir la formación de nódulos en la raíz vegetal. Estos microorganismos, además de proporcionar un sistema ideal para el estudio de la interacción microbio-planta, juegan un papel fundamental en el ciclo global de nutrientes ya que son responsables del 50% del nitrógeno fijado por los organismos vivos (80 millones de toneladas). En relación con esto, se ha establecido un sistema modelo para el estudio de la simbiosis *Rhizobium*-leguminosa formado por la bacteria *Sinorhizobium meliloti* y una de las leguminosas mejor conocidas, *Medicago truncatula*.

*S. meliloti* es una bacteria gram-negativa perteneciente a la familia *Rhizobiaceae* dentro del orden Rhizobiales de las  $\alpha$ -Proteobacterias. Está presente tanto en el suelo como en la rizosfera induciendo la formación de nódulos indeterminados en raíces de un conjunto de especies de *Medicago*, incluyendo *M. sativa* (alfalfa), dentro de los cuales fija nitrógeno atmosférico. En los últimos años, en paralelo con la secuenciación de *M. truncatula*, un consorcio de investigadores se ha unido para secuenciar la cepa *S. meliloti* 1021 (Barloy-Hubler *et al.*, 2000a; 2000b; Barnett *et al.*, 2001; Capela *et al.*, 1999; 2001; Finan *et al.*, 2001; Galibert *et al.*, 2001), elegida de entre todos los aislados de esta bacteria como cepa modelo. Se ha descrito que su genoma (de un tamaño total de 6,7 Mb) está constituido por 3 grandes replicones (figura I.1): el cromosoma de 3,65 Mb y dos megaplásmidos, pSymA de 1,35 Mb y pSymB de 1,68 Mb, conocidos como plásmidos simbióticos (Burkhardt *et al.*, 1987; Sobral *et al.*, 1991). Es uno de los genomas bacterianos más grandes secuenciados hasta la fecha; a partir de los datos de secuencia se infiere que podría codificar del orden de 6204 proteínas de las cuales al 59,7% se le puede asignar una función por comparación con las secuencias disponibles en las bases de datos, mientras el 8,2% se consideran con función desconocida. Al contrario de lo que se esperaba por su tamaño, el genoma de *S. meliloti* no presenta secuencias altamente repetidas; se han descrito unos pocos genes duplicados (el 42% de los genes pertenecen a 548 familias parálogos), lo que demuestra que esta cepa no ha sufrido presión para la reducción del tamaño de su genoma durante su evolución facilitando así la adquisición de nuevas funciones adaptativas relacionadas con la vida en suelo y en simbiosis (además el 12% de los genes codifican sistemas transportadores). El 2,2% del genoma está compuesto por secuencias de inserción (IS) y secuencias derivadas de la invasión de fagos con distribución variable. Abundan especialmente en el pSymA, en las cercanías de los genes simbióticos, probablemente como consecuencia de reagrupamientos genómicos.



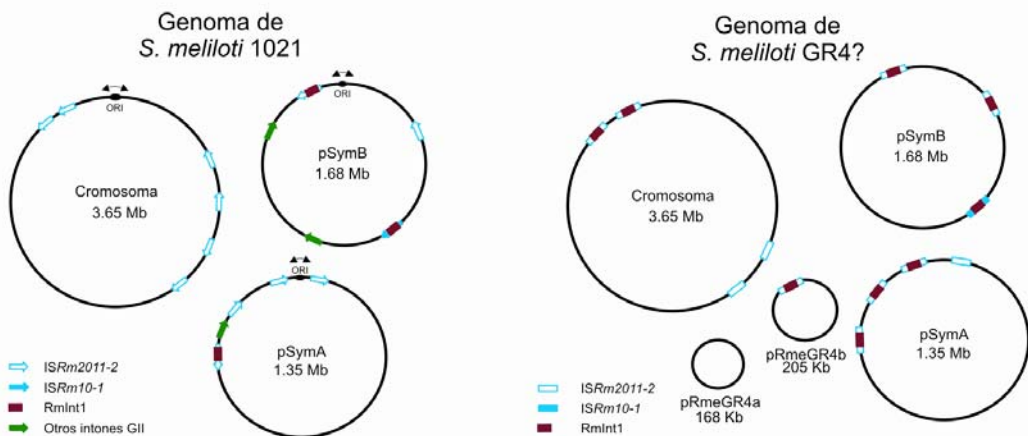


**Figura I.1.** Representación a escala del genoma completo de *S. meliloti* 1021. Los círculos más exteriores representan la distribución de genes codificados según el siguiente código de colores que representa categorías funcionales: en azul, los genes relacionados con el metabolismo de pequeñas moléculas; en verde, los implicados en el metabolismo de macromoléculas; en naranja, los genes que codifican elementos estructurales; en amarillo, los implicados en procesos celulares; los genes

que representan elementos de origen externo se muestran en rojo; en rosa, se marcan los reguladores no clasificados en otras categorías; y, por último, en gris se señalan los genes conservados o no de función desconocida. En el círculo central se representa el contenido GC a lo largo de los diferentes replicones (en rojo se muestran las zonas con más del 50% de GC y en verde, las de menos del 50%). Diagrama obtenido de la página [www.cebitec.uni-bielefeld.de/transcriptomics/sinogate.html](http://www.cebitec.uni-bielefeld.de/transcriptomics/sinogate.html)

En el cromosoma el contenido medio en GC ronda el 62,7%, pero si se hacen estudios pormenorizados a lo largo del cromosoma podemos encontrar regiones con menor contenido en GC, en concreto seis regiones distribuidas arbitrariamente en el genoma (Capela *et al.*, 2001). La región mayor (80 kb.) y más interesante agrupa un gran número de genes procedentes de elementos móviles, esto es, dos integrasas, 10 genes asociados a transposones y varios fragmentos de transposasas; todo esto indica que se trata de una región que ha sufrido una elevada tasa de reagrupamientos, muy posiblemente proceda de sucesivos eventos de transferencia horizontal. Se han descrito genes implicados en todos los procesos celulares: replicación, reparación, recombinación, división celular, transcripción, traducción, biosíntesis, metabolismo, transporte, catabolismo, protección a estrés oxidativo, transporte de hierro, biosíntesis del grupo hemo, genes relacionados con genes de virulencia, regulación y transducción de señales. En cuanto al pSymA se dice que es el verdadero plásmido simbiótico porque en él se encuentran la mayoría de los genes implicados en el proceso, tanto a nivel de nodulación como a nivel de fijación agrupados en lo que se conoce como la “región simbiótica” (Barnett *et al.*, 2001; Galibert *et al.*, 2001; Barloy-Hubler *et al.*, 2000a). Finalmente, el pSymB se considera como el plásmido con mayor implicación en la competencia saprofítica de esta bacteria en suelo (Finan *et al.*, 2001). Presenta una elevada densidad génica (1 ORF/1,1 kb.) relacionada con genes implicados en transporte, síntesis del polisacárido extracelular, reguladores transcripcionales, protección celular y otros enzimas

catabólicos; además contiene un gen indispensable para la viabilidad de la bacteria, el ARN<sub>t</sub><sub>CCG</sub> correspondiente a la arginina. En cuanto al contenido y distribución de intrones del grupo II podemos decir que *S. meliloti* 1021 presenta 3 copias de RmInt1, dos en el plásmido simbiótico pSymB (dentro de IS*Rm2011-2* y IS*Rm10-1*) y la tercera copia insertada en el pSymA (IS*Rm2011-2*), pero se encuentra ausente en el cromosoma (Toro *et al.*, 2003)<sup>43</sup>. La secuencia de ADN de estas tres copias de RmInt1 en 1021 son 99% idénticas a la descrita en el plásmido críptico pRmeGR4b de *S. meliloti* GR4 (Martínez-Abarca *et al.*, 1998); se trata de transiciones que no afectan a la estructura del ARN aunque uno de los cambios en el dIV conlleva un cambio de aminoácido en la secuencia de la proteína que se codifica (Toro *et al.*, 2003). Además de las tres copias de RmInt1 presentes en el genoma de la cepa 1021, se han descrito otras tres secuencias con similitud a reverse transcriptasas: SMb21167 (453 aa) y SMb21477 (392 aminoácidos), en el plásmido pSymB, y SMA1875 (505 aa), en el pSymA. Estudios filogenéticos y estructurales apuntan a que SMb21167/SMb21477 se corresponden con intrones bacterianos del grupo C, mientras que la clasificación del intrón SMA1875 es incierta (figura I.2; Toro *et al.*, 2003).



**Figura I.2.** Distribución de secuencias de inserción e intrones en el genoma de la bacteria *S. meliloti* 1021 (genoma secuenciado) en comparación con una representación hipotética propuesta para la cepa GR4 (no secuenciada). El número de copias de IS*Rm2011-2* (en azul, no relleno) es similar en ambas cepas, sin embargo el número de copias de RmInt1 (en rojo) es considerablemente superior en GR4 respecto a 1021 (9 frente a 3). En verde se representan las copias de intrones del grupo II que se han identificado en *S. meliloti* 1021. Aunque ninguno de los esquemas está hecho a escala, hemos tratado de reflejar de manera aproximada la posición relativa de los distintos elementos en los replicones de 1021 (Figura adaptada de Nisa-Martínez *et al.*, 2007)

<sup>43</sup> La secuencia de inserción IS*Rm2011-2* se ha definido como la diana natural donde se inserta RmInt1 (Martínez-Abarca *et al.*, 1998). Resulta interesante que aún estando presente en los tres replicones (6 copias en el cromosoma, 2 copias en el plásmido pSymB y 4 en el pSymA), el intrón sólo se encuentre invadiendo dos de ellas (Toro *et al.*, 2003).

Hasta ahora hemos descrito las características la cepa *S. meliloti* 1021, pero se ha visto que existen otras cepas bacterianas que presentan cualidades diferenciales con respecto a ésta. Es el caso de la cepa *S. meliloti* GR4 con la cual se lleva trabajando desde los años 70 en los distintos grupos de la Estación Experimental del Zaidín (Casadesús y Olivares, 1979). La peculiaridad de esta cepa es que presenta un elevado grado de competitividad cuando se compara con otras cepas del mismo género (Toro y Olivares, 1986). En relación directa con esto está el hecho de presentar dos plásmidos crípticos en su genoma además de los tres replicones ya descritos en el caso de 1021, pRmeGR4a (115 MDa) y pRmeGR4b (140 MDa) (Toro y Olivares, 1986). Pero estos plásmidos son dispensables en el proceso simbiótico, lo que se deduce de la existencia de cepas curadas. De las pocas cosas que se conocen acerca del plásmido pRmeGR4a es que se trata de un plásmido autotransmisible que posiblemente esté implicado en la cotransfección de pRmeGR4b (Mercado-Blanco *et al.*, 1993); el único elemento interesante descrito en este plásmido es un gen que muestra cierta homología con *rep C*, una proteína que se supone implicada en la replicación de los plásmidos *Ri* y *Ti* de *Agrobacterium* (Mercado-Blanco y Olivares, 1994). Si bien se ha determinado que no existen genes de especial relevancia en el pRmeGR4a, diferentes estudios han demostrado que la competitividad para la nodulación de la cepa GR4 radica en una región de ADN denominada *nfe*, localizada en el plásmido pRmeGR4b (Toro y Olivares, 1986; Sanjuan y Olivares, 1989; Soto *et al.*, 1993; García-Rodríguez y Toro, 2000). Además en este plásmido se han identificado y caracterizado los genes *dapA* y *dapB*, que codifican las enzimas dihidropicolinato sintasa y dihidropicolinato reductasa, respectivamente (García-Rodríguez *et al.*, 2000); estos dos genes se encuentran como copia única en el genoma de GR4, y siempre asociados a pRmeGR4b, que parece presentar una distribución clonal asociada a GR4 (Villadas *et al.*, 1995; Muñoz *et al.*, 2001). Una de las características que nos llamaron especialmente la atención en este plásmido fue su enorme contenido en elementos genéticos móviles. Se tienen datos muy parciales acerca de la secuencia de este plásmido (figura I.4), y aún así se han identificado y caracterizado una amplia variedad de secuencias de inserción: IS*Rm3* (Soto *et al.*, 1992a), IS*Rm4* (Soto *et al.*, 1992b), IS*Rm4-1* e IS*Rm9* (Zekrí *et al.*, 1998), IS*Rm6* (Zekrí y Toro, 1996) y una de las más abundantes (si no la que más) en *S. meliloti*, IS*Rm2011-2* (Selbitschka *et al.*, 1995). Junto a las ISs, se determinó la presencia de un intrón del grupo IIB de la clase bacteriana D, RmInt1 (Martínez-Abarca *et al.*, 1998), del que hemos hablado extensamente a lo largo de este trabajo. Así, la cepa GR4 contiene en su genoma al menos 9 copias del intrón del grupo II RmInt1, 8 de ellas asociadas a la secuencia de inserción IS*Rm2011-2* (Martínez-Abarca *et al.*, 1998) y la restante insertada en IS*Rm10-1* (Martínez-Abarca y Toro, 2000b). Únicamente presenta 3 copias de IS*Rm2011-2* no invadidas (figura I.2). Un estudio pormenorizado de la variación genética de la población de suelo de *S. meliloti* GR4 mostró 5 grupos de aislados en uno de los cuales (11/268) el intrón se encontraba insertado en el gen *oxi1* (transposición ectópica; Muñoz *et al.*, 2001).

Se han llevado a cabo numerosos estudios sobre la abundancia y distribución de RmInt1, tanto en cepas de *S. meliloti* como en cepas de otros microorganismos del orden

Rhizobiales (Martínez-Abarca *et al.*, 1998; Muñoz *et al.*, 2001; Toro *et al.*, 2003; Fernández-López *et al.*, 2005). RmInt1 es muy abundante entre los rizobios y en algunos casos pueden encontrarse hasta 9 copias en el genoma. Hasta la fecha, RmInt1 es uno de los pocos intrones del grupo II presente en alto número de copias en un genoma bacteriano, siendo un elemento ampliamente distribuido en diferentes cepas de esta especie. Esto le convierte en una herramienta muy útil para ser empleado como marcador genético en estudios ecológicos para analizar diversidad genómica en poblaciones bacterianas de *S. meliloti*.

Dado el elevado número de elementos genéticos móviles que se ha descrito en tan sólo ~21 kb secuenciadas del pRmeGR4b (un 10% de la secuencia total), es de suponer que la densidad de este tipo de elementos sea muy alta en este plásmido críptico. Por esto nos propusimos comprobar su abundancia relativa en la secuencia del replicón. Nos centraremos fundamentalmente en intrones del grupo II distintos de RmInt1 y trataremos de caracterizarlos desde el punto de vista filogenético y estructural.

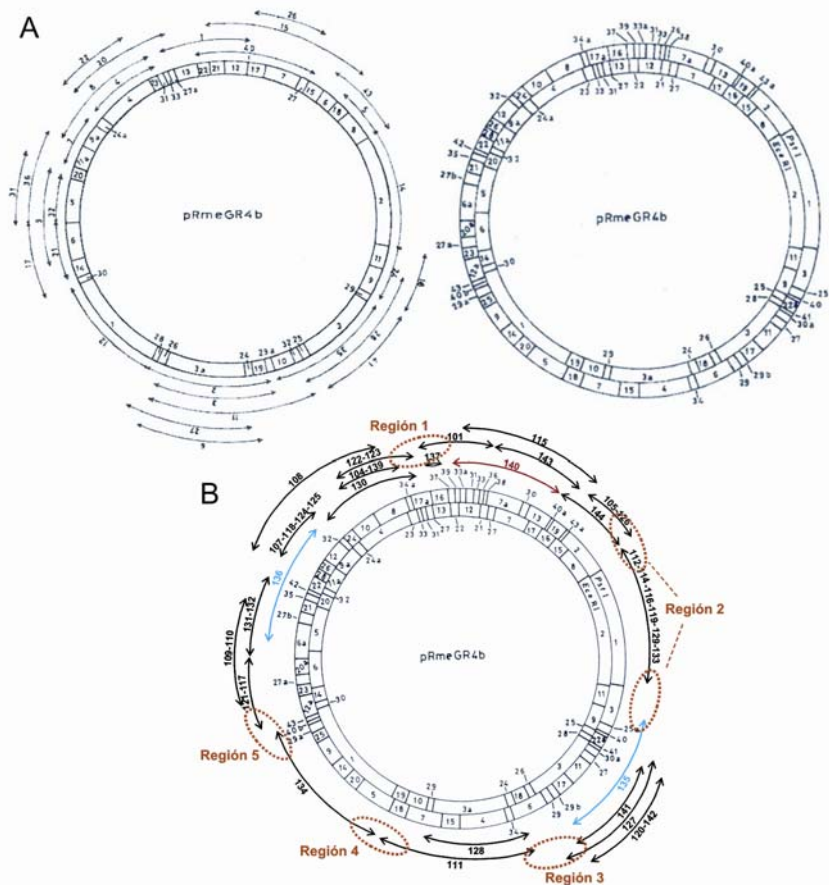
## I.1. BÚSQUEDA DE NUEVOS RETROELEMENTOS EN EL PLÁSMIDO CRÍPTICO pRmeGR4b

El grupo de investigación dispone de un banco de clones del plásmido pRmeGR4b en un vector de amplio rango de hospedador. Esta librería se encuentra en *E. coli* HB101 y fue construida por digestión parcial del plásmido purificado con los enzimas *EcoRI* y *HindIII*. Se clonaron fragmentos de aproximadamente 20 kb en el vector pRK290 que tiene resistencia a tetraciclina (Toro, 1985). Se dispone además de un mapa del plásmido inferido del patrón de digestión con dos enzimas de restricción, *EcoRI* y *PstI*, en el que se sitúan los diferentes clones de la librería (figura I.3A; Toro, 1985; Mercado-Blanco, 1993). A esto debemos sumarle que conocemos la secuencia completa de uno de los clones del banco, el pRmNT140, plásmido en el que se localizan los genes *nfe*, *dap* y la secuencia del intrón RmInt1 inicialmente descrita en *S. meliloti* GR4. La estrategia planteada fue tratar de obtener la secuencia de los extremos de los insertos presentes en algunos de los plásmidos que forman parte de la librería.

Antes de iniciar la secuenciación, quisimos verificar los fragmentos presentes en cada inserto mediante digestión con *EcoRI* y *PstI*. En la mayoría de los casos coincidió con los resultados de que disponíamos, aunque se pudieron apreciar algunas diferencias. En la tabla AI.1 del anexo se muestran la distribución de fragmentos que compone cada uno de los plásmidos del banco analizados. Los fragmentos se encuentran anotados en sentido 5'→3' según se encuentran en el plásmido y, en diferente estilo de fuente se destacan las diferencias respecto a los datos de los que se disponía hasta el momento<sup>44</sup>. El número

<sup>44</sup> En algún caso se trató de fragmentos mal identificados como por ejemplo en pRmNT130, en el que se pensó que incluía el fragmento *EcoRI* 27 pero según los datos obtenidos en este trabajo sería más bien el fragmento 24. En cambio, lo más habitual es que se trate de fragmentos adyacentes que no se han incluido en el patrón de digestión de un determinado inserto; es el caso de pRmNT143 en el que el tamaño similar

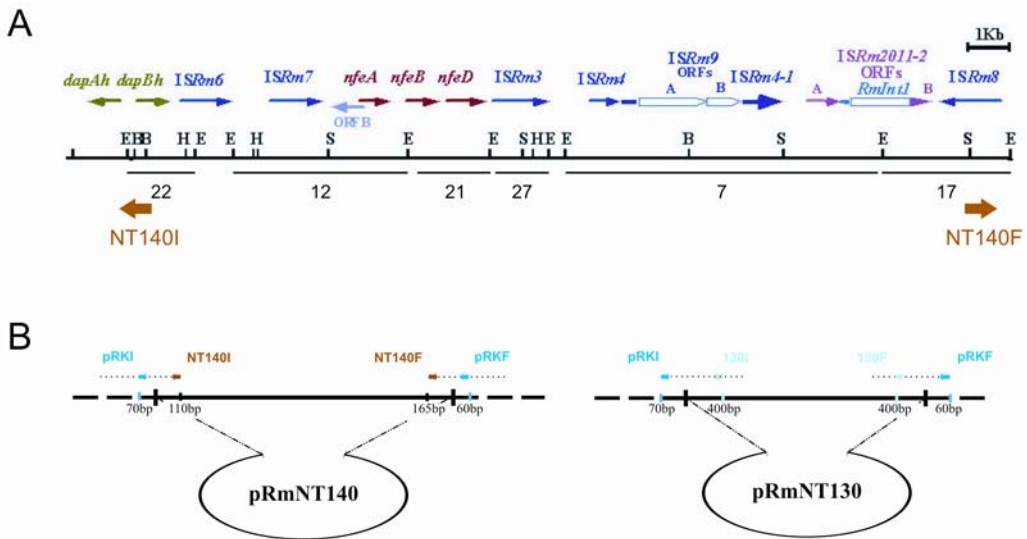
con el que se identifica cada fragmento está en relación con su peso molecular, que es diferente dependiendo del tipo de digestión de que se trate (Anexo, tablas AI.2 y AI.3). Todos estos análisis por digestión nos han permitido reducir el número de plásmidos con los que trabajar puesto que hemos llegado a la conclusión de que muchos de ellos son idénticos (figura I.3B). Igualmente, los patrones de digestión nos permiten establecer una nueva organización de los fragmentos en el mapa de plásmidos pRmNT, que como veremos más adelante va a ser determinante para conseguir la secuencia completa de algunos retroelementos (figura I.3B).



**Figura I.3.** Mapas de la distribución de los plásmidos la librería del pRmeGR4b. **(A)** Diagrama construido inicialmente basado en digestiones *EcoRI* y *PstI* (Toro, 1985; Mercado-Blanco, 1993). **(B)** Esquema final obtenido tras los estudios de digestión, hibridación ADN-ADN y secuenciación. En rojo se representa el plásmido pRmNT140 cuya secuencia conocemos. En azul, los dos plásmidos en uno de cuyos extremos se han localizado retroelementos. Las regiones no solapadas se marcan en naranja y se nombraron en el sentido de las agujas del reloj. Aunque la región 2 parezca dividida en 2, se trata de una única zona no solapada pues la secuencia central parece ser un artefacto.

de las bandas impidió determinar con precisión su composición de fragmentos, de modo que incluye un fragmento *EcoRI* más de los que se identificaron en los primeros estudios.

Como decíamos, se dispone de la secuencia completa del inserto en el plásmido RmNT140 (figura I.3B, en rojo; figura I.4A). La utilización de oligonucleótidos divergentes en los extremos de este fragmento de ~21 kb nos permitió conocer la secuencia del vector que flanquea los insertos en los plásmidos de la librería (figura I.4A). El diseño de cebadores convergentes en estas regiones del vector nos permitió la secuenciación de los extremos de los insertos de algunos de los plásmidos del banco (figura I.4B; oligonucleótidos en tabla AI.4). Siguiendo un criterio de continuidad (situación consecutiva en el mapa del plásmido), se seleccionaron una serie de clones para la posterior secuenciación de sus extremos: 101, 111, 112, 114, 116, 117, 119, 128, 129, 130, 133, 134, 135, 136, 137 y 143.



**Figura I.4.** Diseño experimental. (A) Distribución a escala de los genes y secuencias de inserción (IS) presentes en el plásmido RmNT140. En la barra central se representan los sitios de corte *EcoRI* (E), *BamHI* (B), *HindIII* (H) y *SalI* (S); debajo hemos dibujado la distribución de fragmentos identificados según el patrón de digestión en gel de agarosa. En la parte inferior se esquematiza la situación de los cebadores utilizados para conocer la secuencia de los extremos del vector pRK290 (NT140I y NT140F). (B) Estos esquemas explican cómo se diseñaron los cebadores (pRKI y pRKF) en los extremos del vector (panel izquierdo) y cómo se construyeron las sondas usadas en los experimentos de hibridación ADN-ADN (un ejemplo en el panel derecho). Las secuencias obtenidas a partir de los oligonucleótidos pRKI y pRKF (eliminando las 70 y 60 bp del vector, respectivamente) fueron sometidas a BLAST (x, n y p).

En paralelo, se trabajó para conseguir un esquema de distribución de los plásmidos de la librería mediante hibridación ADN-ADN. Como dijimos con anterioridad, en ocasiones, conocer la secuencia completa de algunos de los elementos presentes en los extremos de los insertos analizados nos va a exigir recurrir a otros plásmidos que contengan el elemento completo; en este sentido, un mapa de solapamientos de nuestro banco de clones será muy útil. En las membranas para hibridación se decidió incluir inicialmente el mínimo número de plásmidos que consideramos suficiente para los

solapamientos, pero finalmente vimos la conveniencia de incluirlos todos. Para comprobar los solapamientos se sintetizaron e hibridaron sondas de algunos de los extremos de algunos de estos insertos: 101 (Inicial y Final), 111 (I y F), 117 (I y F), 119 (I y F), 130 (I y F), 134 (I y F), 135 (I y F), 136 (I y F), 140 (I y F) y 143 (I). Esta aproximación nos permitió establecer la existencia de 6 regiones no solapadas y reestructurar la distribución de clones que se había establecido según los patrones de restricción (figura I.3B). Estos datos apuntan a que el banco de clones del plásmido críptico pRmeGR4b no está completo y que debemos buscar una alternativa para conocer la secuencia en las regiones no solapadas. Así pues, recurrimos a la hibridación en colonia sobre una genoteca de *S. meliloti* GR4<sup>45</sup>. Se hibridó con las sondas de las regiones no solapadas: 101I, región 1; 143I, región 2; 135F, región 3; 111F, región 4; 111I, región 5; y 117F, región 6. Los supuestos clones positivos mediante esta técnica se comprueban por PCR de lisados de colonia usando los cebadores con los que se sintetizaron las sondas. Dado el elevado número de falsos positivos que detectamos en la hibridación en colonia, decidimos analizar el patrón de digestión de los insertos, seguido de un nuevo análisis mediante hibridación ADN-ADN y secuenciación (Tabla AI.5, en el anexo). De los 228 clones positivos (228/1500) de los que partíamos, ninguno de ellos presenta ninguna de las regiones solapadas que queremos completar.

### I.1.1. ANÁLISIS DE SECUENCIAS

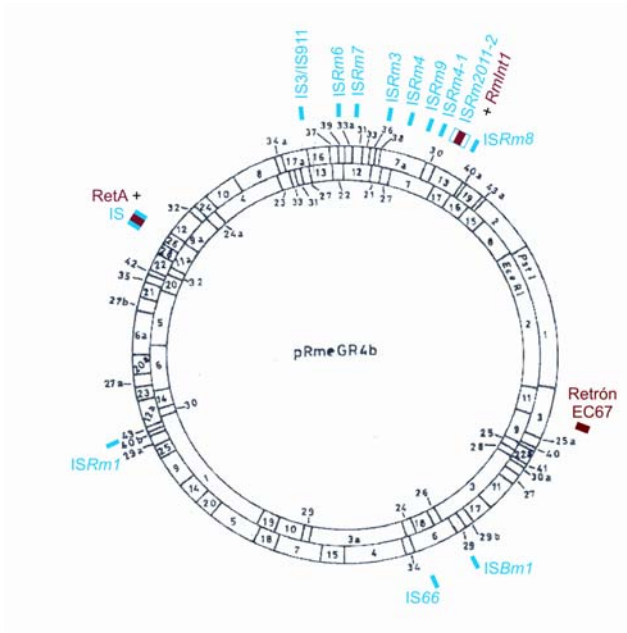
El objetivo planteado consistió en la evaluación de la abundancia de retroelementos y otras secuencias genéticas móviles en el plásmido críptico pRmeGR4b de *S. meliloti* GR4. Para ello disponemos de la secuencia de los extremos de algunos de los insertos del banco de clones construido a partir de la digestión parcial de éste. Se buscaron regiones de similitud entre nuestras secuencias y las anotaciones de las bases de datos mediante el uso de diferentes algoritmos del programa BLAST (x, n y p). En cuanto a las bases de datos utilizadas, quisimos hacer una búsqueda general en todos los genomas secuenciados pero también nos planteamos restringir nuestra comparativa con el genoma de *S. meliloti* 1021 (aunque esta cepa no contenga plásmidos crípticos). Todos los resultados se muestran en la tabla AI.6.

Los resultados más interesantes, además de revelarnos nuevos datos acerca de las regiones con ausencia de solapamiento, también nos proporcionaron información acerca de nuevos elementos genéticos algunos de los cuales estaban relacionados con intrones del grupo II. Cuando se analizaron las secuencias de ambos extremos de los plásmidos pRmNT119 y 129 comprobamos que mostraban identidad (100% de identidad con un 100% de cobertura de secuencia, valor esperado  $E = 0.0$ ) con secuencias de plásmidos RK2; en relación con esta observación también se secuenció uno de los extremos de otros insertos que se habían considerado idénticos a 119 y 129 (112, 114, 116 y 133)

---

<sup>45</sup> Se consideró, teniendo en cuenta el tamaño del genoma y el tamaño de los insertos en la librería, que un número de colonias representativo del genoma completo podría ser 1500.

confirmando este resultado. Si tenemos en cuenta que pRK290 es también un derivado del plásmido RK2 (Ditta *et al.*, 1985) y a esto le sumamos lo extraño que resulta que un fragmento tan grande (~20 kb) no tenga ninguna diana de restricción para *EcoRI* o *PstI*, parece claro que estamos hablando de que la doble banda que observamos en gel pudiera tratarse de formas mal digeridas del pRK290; otra posibilidad es que se trate de dos vectores ligados. Esta eventualidad nos reduciría las regiones no solapadas a 5.



**Figura I.5.** Patrón de distribución de secuencias de inserción y retroelementos de diversa naturaleza localizados a lo largo del pRmeGR4b gracias a la secuenciación de los extremos de algunos de los clones de la librería del plásmido críptico. En azul se muestran las diferentes secuencias de inserción encontradas a las que debemos sumar las múltiples presentes en el clon 140. *ISRm2011-2*, la diana natural de *RmInt1* se representa sin relleno e invadida por el intrón. Además se localizaron otros dos retroelementos: uno, presentaba identidad con *RetA* que, a su vez, muestra algunas semejanzas con intrones bacterianos del grupo II D; el otro, aunque no se ha conseguido secuenciar completo, parece corresponderse con un retrón similar a *EC67* de *E. coli*.

En cuanto a la localización de secuencias de inserción podemos decir que, además de las que ya se han descrito en el pRmNT140<sup>46</sup>, se encontraron regiones de identidad con otras 5 secuencias de inserción en diferentes localizaciones del pRmeGR4b (figura I.5). Uno de los aspectos que más interesantes en cuanto al origen de estas secuencias de inserción fue que más del 50% de las identidades (3/5) se correspondían con elementos de este tipo presentes en otras especies bacterianas, lo que podría ser indicativo de fenómenos de transferencia horizontal entre cepas relacionadas. La secuencia de los extremos 128F, 130I y 137I mostró similitud con tres ISs en diferentes bacterias: la secuencia 128F alineaba con un 98% de identidad con la transposasa de la *IS66* descrita en *S. medicae* WSM419 ( $E=0.0$ ); 130I presentó un 95% de similitud ( $E=4e-152$ ) con parte de la secuencia de inserción (*ISRle39a*) que flanquea un sistema de metilación-restricción en *R. leguminosarum*; por último, el 33% de la extensión de la secuencia del extremo I del

<sup>46</sup> Dos de los extremos secuenciados, 101F y 143F, presentaron identidad con *ISRm3*. Esto viene a demostrar que estos dos insertos se encuentran contiguos en mapa del plásmido críptico solapando con pRmNT140, ya que esta secuencia de inserción se describió en este plásmido (Soto *et al.*, 1992a).

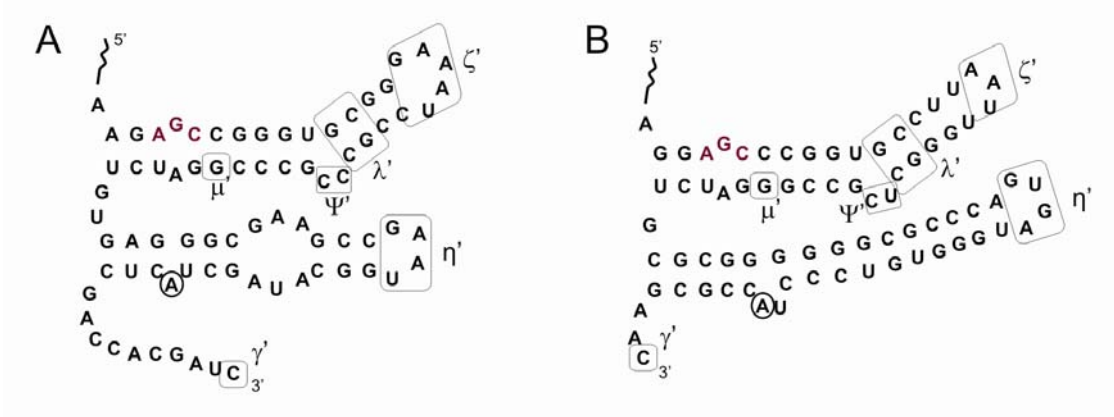


pRmNT137 se identificaba (76% similitud,  $E=2e-31$ ) con la transposasa IS3/IS911 de *Nitrobacter winogradskyi*. Por otro lado, los extremos 134I y 135I presentaron identidad con las transposasas ISRm1 de *S. meliloti* y ISBm1 descrita en el pRmeGR4b de *S. meliloti* GR4, respectivamente. Es conveniente resaltar que, si restamos las comprobaciones relacionadas con la identificación del supuesto fragmento EcoRI 2 (8) del total de extremos secuenciados (28), siete secuencias se correspondieron con elementos genéticos móviles (7/20) y en concreto 5 de ellas con ISs (5/7).

En relación con la presencia de retroelementos llamaron nuestra atención las secuencias obtenidas a partir de dos de los insertos. La identidad que nos proporcionaron los alineamientos locales se correspondía con dominios de ADN polimerasas dirigidas por ARN (reverso transcriptasas). En concreto, los extremos 136F y 135F mostraron un porcentaje de similitud significativo con RetA de *Serratia marcescens* (Kulaeva *et al.*, 1998) y el retrón EC67 de *E. coli* (Hsu *et al.*, 1990; Lim, 1991), respectivamente (figura I.5 y tabla AI.6). Como hemos dicho, en ambos casos fueron secuencias parciales y para englobarlos dentro de uno u otro tipo de retroelementos necesitamos conocer su secuencia completa.

#### **I.1.1.1. La secuencia completa del intrón**

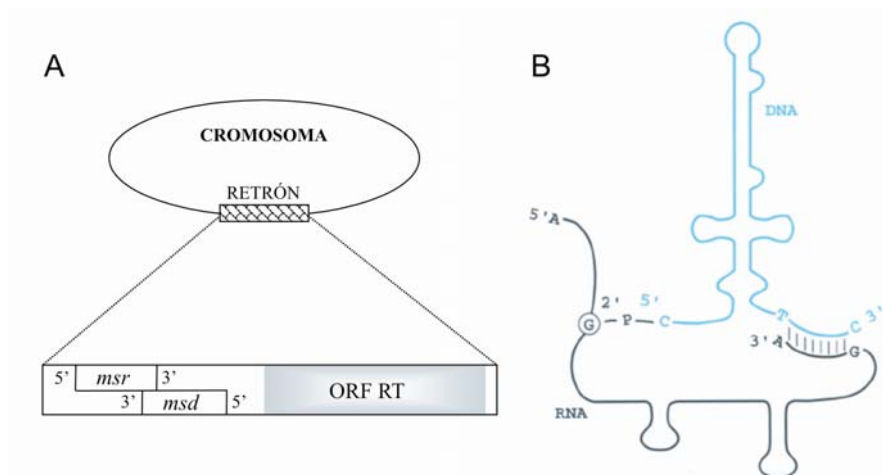
La obtención de la secuencia completa del elemento encontrado en el extremo del pRmNT136 se consiguió gracias al uso de un plásmido del banco de pRmeGR4b que solapaba en esta región, en concreto se seleccionó pRmNT107 (figura I.3B). Se trataba de diseñar oligonucleótidos divergentes sobre los fragmentos de secuencia que íbamos conociendo, con lo que pudimos obtener el elemento en toda su longitud (ver oligonucleótidos en tabla AI.7). Se secuenciaron 2714 bp y el análisis de la secuencia nos reveló la presencia de un intrón del grupo II de 1894 nt de longitud relacionado con los intrones del grupo II de la clase bacteriana D (al cual pertenece RmInt1). Esta identificación fue posible, entre otras razones, gracias a la presencia de dos de los dominios característicos de intrones del grupo II, el dV y el dVI (figura I.6). En virtud de la relación con intrones de la clase D y en relación con la identidad que nos proporcionaron los alineamientos BLAST, en su secuencia deberíamos identificar un marco abierto de lectura (ORF) que contenga los dominios RT y madurasa, pero que carezca de dominios D y En. Efectivamente, el análisis exhaustivo de la composición de aminoácidos en relación con otras RT permitió la identificación de todos los dominios RT así como del dominio X (figura I.9 y figura AI.1). Sin embargo, también reveló la existencia de dos cambios en la pauta de lectura que tenían lugar, uno en el dominio RT1 (fase +1 pasa a la +3) y otro en el dominio madurasa (que implica un cambio de la fase +3 a la +2). En el caso de que los ribosomas fuesen capaces de soslayar esta dificultad, generarían una proteína de 469 aa. Ambos cambios en la pauta de lectura han sido exhaustivamente analizados y secuenciados en varias ocasiones y en ambas direcciones para cerciorarnos de que no son artefactos generados durante la secuenciación.



**Figura I.6.** Predicción de la estructura secundaria de los dominios V y VI de los intrones *S.me14* (A) y *S.ma11* (B). En ambos casos, el dV presenta la estructura esperada con la particularidad de que el segundo residuo de la tríada catalítica (en rojo) se encuentra desapareado. Aunque la arquitectura del dVI varía entre los dos intrones, siguen manteniendo los elementos que caracterizan a este dominio: una adenosina desapareada hacia el extremo final del intrón (A dentro de un círculo) y una estructura en forma de tallo culminada con una secuencia de 4 residuos desapareados. Con letras griegas y recuadradas se indican las interacciones terciarias que estos dominios establecen con otras regiones del intrón.

### I.1.1.2. La secuencia del retrón

Los retrones son elementos genéticos relacionados con intrones del grupo II que codifican una RT estructuralmente similar a las polimerasas de retrovirus (Inouye e Inouye, 1996). Son secuencias de ADN que codifican una RT que se encarga de sintetizar cientos de copias por célula de un ADN satélite particular conocido como msADN (ADN multicopia de cadena simple). Este ADN extracromosómico está compuesto por un corto ADN de cadena simple cuyo extremo 5' se encuentra unido mediante un enlace 2'-5' a un residuo G interno de un molécula de ARN de cadena simple de pequeño tamaño (Inouye e Inouye, 1991a). Ambos forman una estructura bastante característica que funciona como cebador pero también como molde para la síntesis de más moléculas de msADN (figura I.7B). Los tres elementos [ARN (*msr*), ADN (*msd*) y RT] se encuentran codificados en la secuencia del retrón (figura I.7A). Su función es desconocida aunque se sospecha que podrían causar ciertos desequilibrios en las células que los contienen como, por ejemplo, incrementar la frecuencia de mutación espontánea (Lampson *et al.*, 2005). No se ha demostrado experimentalmente su capacidad de movimiento, pero su distribución en los organismos que los contienen hace sospechar que podríamos estar hablando de retroelementos móviles (Inouye e Inouye, 1993; Rice y Lampson, 1996). En su integración se ha visto que eliminan un fragmento (~3,5 kb) del ADN de su hospedador (Lim, 1995).

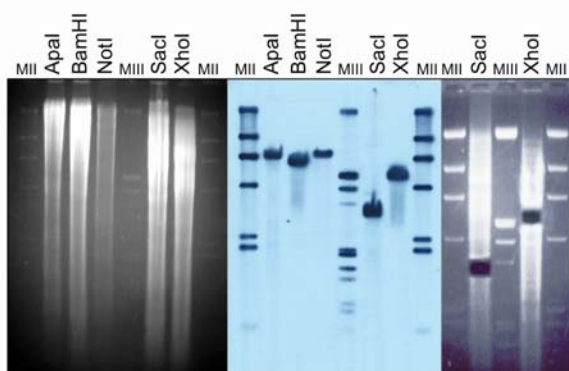


**Figura I.7.** Organización genómica de un retrón. **(A)** Se genera un único transcrito que codifica tanto el msADN como la RT. *msr* y *msd* se corresponden, respectivamente, con los componentes ARN y ADN del msADN. La fase abierta de lectura (ORF) que codifica la RT se encuentra siempre aguas abajo de la región *msr-msd* (figura adaptada de Rice y Lampson, 1996) **(B)** Estructura característica que adopta el msADN. El ADN de cadena simple se ilustra en azul y la molécula de RNA en negro. El ADN se une al RNA por un enlace fosfodiéster 2'-5' a través de un residuo de rG interno en el RNA. Los extremos 3' de ambas moléculas son complementarios formando un híbrido ADN-RNA (Inouye e Inouye, 1991b).

Uno de los principales problemas que nos encontramos a la hora de completar la secuencia de este retroelemento fue, precisamente, que se encontraba en uno de los extremos que las diferentes aproximaciones no nos han permitido solapar, la secuencia 135F (figura I.3B y I.5). Por esta razón, sólo pudimos avanzar en la secuenciación en una de las direcciones (ver oligonucleótidos en tabla AI.7). Se consiguió finalmente una secuencia con una longitud total de 1705 bp en la que se pudo identificar una fase abierta de lectura que se corresponde con el inicio del dominio RT1 (no completo) hasta el final de la proteína que caracteriza a los retrones, en concreto, a uno de los más estudiados, EC67 (figura AI.2). Aproximadamente, la primera mitad de la secuencia se corresponde con los siete dominios RT característicos de este tipo de proteínas (RT1-7), mientras el resto de la proteína parece estar relacionada con un dominio ARNasa H. Este tipo de organización se identifica en la RT de retrovirus y, por ende, se corresponde con la estructuración de la proteína presente en retrones. Este tipo de elementos se caracteriza por la presencia de las secuencias *msr* y *msd* aguas arriba del ORF (figura I.7A). En nuestro caso fue imposible identificar este tipo de secuencias por lo que no podríamos asegurar con exactitud que se trata de un retrón, aunque varios indicios apuntan en esta dirección: (1) la presencia de un dominio ARNasa H; (2) la longitud de secuencia proteica aguas abajo de los dominios RT; (3) alineamientos con la región de unión a ADN y ADN endonucleasa de intrones del grupo II no mostraron un porcentaje de identidad significativo con esta región de la proteína identificada (dato no mostrado). Por otro lado, si se tratase de un intrón del grupo II debíamos ser capaces de identificar los dominios V

y VI de la ribozima. El plegamiento de los 127 nt finales de la secuencia, aguas abajo del codón de parada de la proteína, no mostraba similitud con los dominios mencionados. Este hecho supone una nueva evidencia que apunta a que nos encontramos ante un retroelemento tipo retrón.

Resultados de hibridación ADN-ADN total demostró que se trata de un elemento que se encuentra en copia única en el genoma de la bacteria y que además se encuentra asociado al plásmido críptico pRmeGR4b (figura I.8). Se trabajó para tratar de conseguir la secuencia del elemento completo pero nuestros esfuerzos fueron infructuosos<sup>47</sup>, por lo que nos centramos en el análisis del intrón del grupo II asociado con el extremo 136F.



**Figura I.8.** Presencia de retrones en la secuencia del genoma de *S. meliloti* GR4. El primer panel muestra una misma preparación de ADN total digerido con varias enzimas de restricción. En todos los casos, la hibridación con una sonda específica del extremo F del plásmido NT135 dio una única banda (panel central). También se usó una cepa curada del plásmido críptico b (GRM6) comprobando una ausencia total de señal

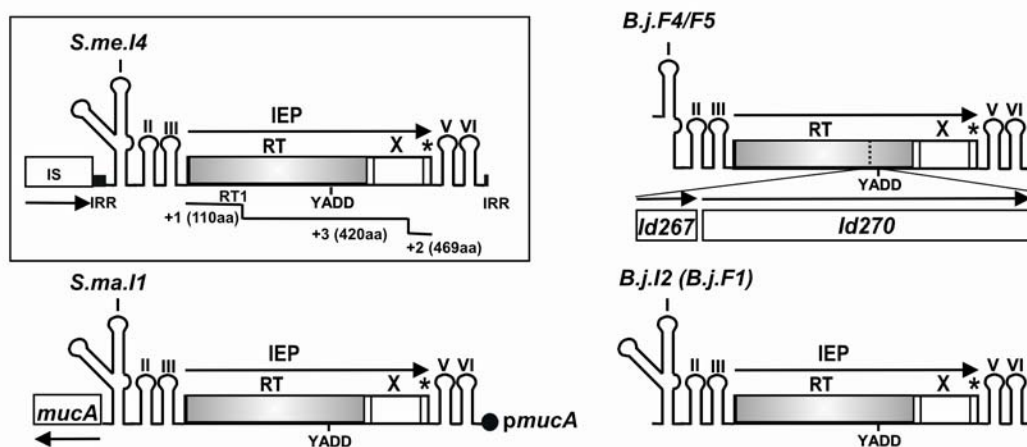
de hibridación (dato no mostrado). El tercer panel supone un ejemplo de la purificación de la banda de gel para su posterior clonaje y secuenciación, si procediese. Los marcadores II y III sirven de guía para el aislamiento de los fragmentos de gel.

## I.2. IDENTIFICACIÓN DE UNA NUEVA CLASE DE INTRONES DEL GRUPO II

Algunos de los intrones mejor caracterizados hasta la fecha, entre los que se encuentran los intrones mitocondriales de levaduras aI1 y aI2 y el intrón LLtrB de *L. lactis*, presentan en el extremo C-terminal de la proteína un dominio de unión a ADN y un dominio ADN endonucleasa. Sin embargo, muchos de los intrones bacterianos del grupo II descritos hasta la fecha carecen de estos dominios necesarios para la rotura de la cadena antisentido para que pueda tener lugar el *retrobombing* (Lambowitz *et al.*, 1999; Belfort *et al.*, 2002). Como ya hemos mencionado, RmInt1 es un intrón móvil que

<sup>47</sup> Con ayuda de diferentes marcadores de peso molecular y diversas digestiones se trató de aislar, en un gel de agarosa, la región correspondiente a la única señal observada cuando se hibridó con la sonda del 135F (figura I.8). Se eligieron enzimas que no cortasen nuestra secuencia, que no cortasen la secuencia del retrón EC67, pero también que fuesen únicas en el vector de clonaje utilizado (pKSBluescript II). Tras la ligación de los múltiples insertos en el vector, se comprobaba si el fragmento de interés se encontraba entre ellos por PCR de lisados de colonia. Como ya nos había ocurrido con la hibridación en colonia para esta región, este tipo de aproximación tampoco dio resultados satisfactorios.

presenta una IEP que contiene únicamente los dominios RT y X (Martínez-Abarca *et al.*, 2000). Este intrón bacteriano se ha clasificado dentro de la clase D, en la que también se incluyen, entre otros, *B.j.I1* de *Bradyrhizobium japonicum*, *E.c.I2* de *E. coli* y *R.e.I1* de *R. eutropha* (Martínez-Abarca y Toro, 2000a; Zimmerly *et al.*, 2001; Dai y Zimmerly, 2002). Con los nuevos avances en secuenciación, el número de intrones bacterianos descritos ha ido incrementándose. Las búsquedas en bases de datos mediante BLAST usando la secuencia de la IEP de RmInt1 ha permitido la identificación de nuevos miembros de la clase bacteriana del grupo D, como por ejemplo intrones en bacterias marinas no cultivadas (AAL78688; Podar *et al.*, 2002), en *S. flexneri* (S0042; Verkatesan *et al.*, 2001) y en otros muchos organismos pertenecientes tanto a eubacterias como a arqueas (como veremos en el capítulo III).



**Figura I.9.** Representación esquemática de los intrones bacterianos de la clase E identificados en este trabajo. La IEP se encuentra dentro del DIV de la ribozima y consta de dos dominios: reverso transcriptasa, RT, y madurasa, X. Con un asterisco se quiere hacer notar la presencia de una corta extensión aguas abajo del dominio X similar a la descrita en intrones de la clase D. El dominio activo para la actividad RT también se indica, YADD. Las flechas indican la orientación de las diferentes ORFs. En un recuadro se representa el intrón encontrado en el pRmeGR4b de *S. meliloti* GR4; asimismo se muestran los dos cambios en la pauta de lectura descritos en la secuencia de la IEP. IS y IRR se refieren a secuencia de inserción y repetición invertida terminal derecha, respectivamente. Debajo se muestra el intrón descrito en *S. marcescens* que se encuentra insertado en la región -10 del promotor del gen *mucaA*. Junto a estos se representan dos de los intrones de *B. japonicum*: el intrón completo y el intrón fragmentado (acompañado de los ORFs que lo interrumpen) del que sólo se ha conseguido plegar hasta parte del dI (ver texto).

Entre las IEPs que presentan similitud con la codificada en RmInt1 se encuentra RetA. Esta proteína se identificó en el plásmido R471a de *Serratia marcescens* (Kulaeva *et al.*, 1998) y forma parte del intrón de *S.ma.I1* (figura I.9; Martínez-Abarca y Toro, 2000a). Sin embargo, hasta el momento no se ha determinado la estructura secundaria de su ARN y no ha sido posible enmarcarla dentro de una determinada subclase bacteriana. Entre otras cosas, se ha sugerido que podríamos estar hablando de un intrón degenerado

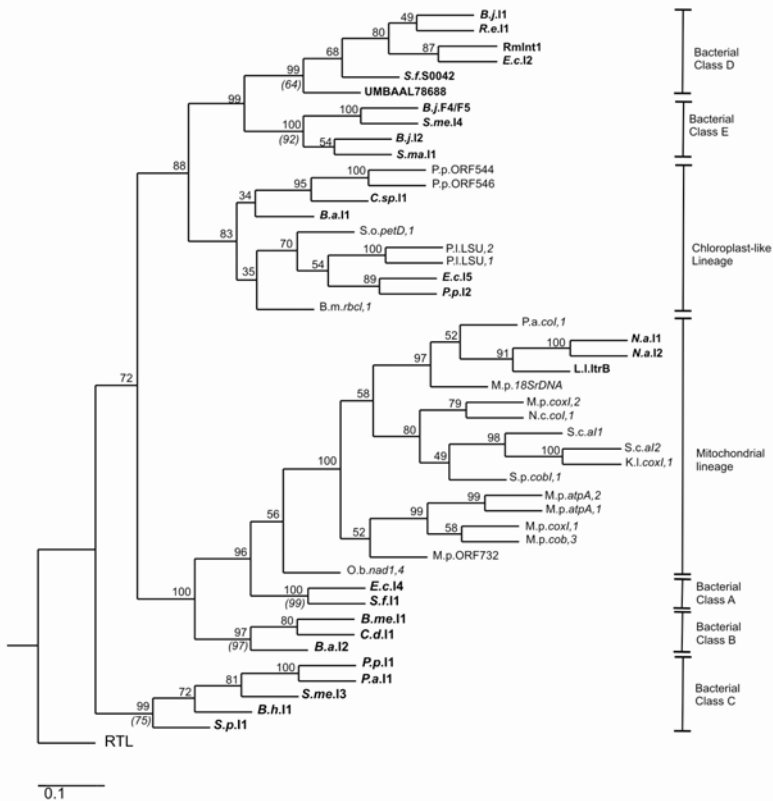
o que podría formar parte de una nueva subclase de intrones todavía por determinar (Dai y Zimmerly, 2002). Por otro lado, se ha descrito una cantidad importante de intrones fragmentados, cuyas IEP muestran un grado bastante elevado de similitud con *S.ma.I1* (figura I.9; Dai y Zimmerly, 2002). Entre ellos se podemos destacar varios intrones de *Bradyrhizobium japonicum* [*B.j.I2* (491 aa), *B.j.F4* (240 aa) y *B.j.F5* (237 aa)] (Göttfert *et al.*, 2001) y otros pertenecientes a *Pseudomonas putida* [*P.p.F3* (267 aa) y *P.p.F4* (226 aa)] (Kholodii *et al.*, 2000). Tras un detenido estudio basado en alineamientos de secuencia hemos llegado a la conclusión que los intrones fragmentados *B.j.F4* (240 aa) y *B.j.F5* (237 aa) podrían dar lugar a una única proteína que se encuentra interrumpida por otra proteína que codificaría una hipotética recombinasa específica de sitio (*id270*) junto a otro ORF de pequeño tamaño, *id267*; Göttfert *et al.*, 2001). En el primer fragmento (*B.j.F4*) se encontrarían los dominios RT del 0 al 4, mientras que en *B.j.F5* podemos identificar los dominios 5 al 7 junto con el dominio madurasa. Así pues, la reconstitución *in silico* de la secuencia daría lugar a una proteína de 460 aa que también manifiesta un elevado grado de similitud con la IEP de *S.ma.I1*; por esta razón pasamos a llamar al conjunto de los fragmentos como *B.j.F4/F5*. Además en nuestro estudio del plásmido críptico pRmeGR4b hemos identificado otra IEP con características muy similares a todas las anteriores a la que vamos a denominar *S.me.I4*. Como en el caso de *B.j.F4/F5*, debemos asumir que se corrigen los cambios en la pauta de lectura que se han determinado basándonos en la secuencia de nucleótidos.

En cuanto a los sitios de inserción de estos intrones (figura I.9), podemos decir que *S.ma.I1* se localiza siete nucleótidos aguas abajo de la región -10 del promotor del gen *mucA*. El análisis de la secuencia que circunda el intrón *S.me.I4* en su localización original (pRmeGR4b) muestra similitud con la repetición terminal invertida derecha de la secuencia de inserción *ISRtr1* (AF041379) de *Rhizobium tropici*. Por su parte, los intrones presentes en *B. japonicum* se sabe que están flanqueados por transposasas y ORFs aún sin caracterizar (Göttfert *et al.*, 2001).

### I.2.1. AGRUPAMIENTO FILOGENÉTICO DE ALGUNAS IEPs BACTERIANAS

Aunque sabemos que existen diferentes criterios a la hora de construir árboles filogenéticos en intrones del grupo II (ver introducción apartados 1.1.3. y 2.1.4.), lo más habitual es utilizar los siete bloques conservados del dominio RT de las IEPs (Martínez-Abarca y Toro, 2000a; Zimmerly *et al.*, 2001). En nuestro estudio hemos considerado algunas de las ORFs representativas de los linajes de intrones mitocondriales y de cloroplastos, así como de las clases bacterianas A, B, C y D. A este análisis le hemos sumado el nuevo intrón descrito en *S. melioli*, *S.me.I4*, junto con *S.ma.I1*, *B.j.I2* y *B.j.F4/F5* (todas las IEPs usadas se detallan en la tabla AI.8). Se construyeron filogenias basadas en distintos algoritmos computacionales (máxima parsimonia, MP, y *Neighbor-Joining*, NJ) que nos proporcionaron una topología de las ramas internas del árbol muy

similar, aunque la probabilidad de distribución de esas mismas ramas por el método NJ era menor (figura I.10). En cualquier caso, el árbol filogenético se encuentra dividido en 7 grupos principales, 6 de los cuales coinciden con las subclases de intrones previamente definidas mientras que la séptima rama incluiría a *S.ma.I1*, *B.j.I2* y *B.j. F4/F5* y *S.me.I4*. Las RT bacterianas que no se incluyen dentro de los linajes mitocondrial y cloroplástico forman 5 agrupamientos bien definidos (97-100% de probabilidad según MP). En particular, la rama que nos define el nuevo grupo de intrones bacterianos presenta un valor probabilístico del 100% que, aunque no coincide con el 92% proporcionado por el método NJ, en ambos casos es muy significativo. El hecho de que en este grupo existan intrones fragmentados (*B.j. F4/F5*) o con mutaciones que aparentemente los trunquen (*S.me.I4*), podría suponer una complicación a la hora de definir, basándonos únicamente en estos datos, una nueva clase de intrones del grupo II. Las restricciones no son tan importantes si tenemos en cuenta que en este grupo también se incluyen intrones completos (*S.ma.I1* y *B.j.I2*).



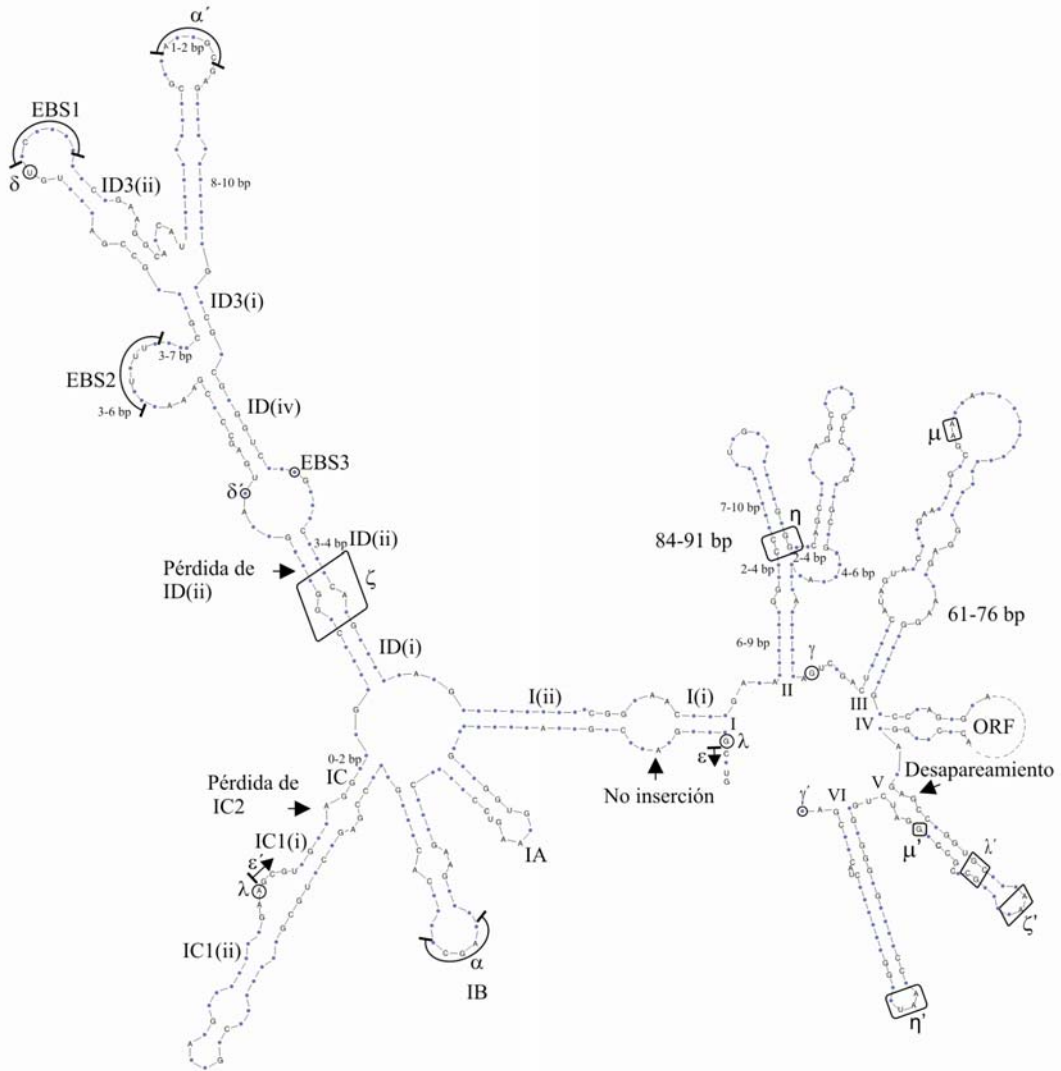
**Figura I.10.** Relaciones filogenéticas entre las RT codificadas por intrones del grupo II. Se muestra un árbol consenso construido por el método de máxima parsimonia; el número en cada nodo indica el valor de fiabilidad de la distribución sobre 100. En algunos casos también se muestra el valor equivalente obtenido por el método de *neighbor-joining*. La secuencia de la RT de una proteína no perteneciente pero relacionada con intrones (RTL, n° acceso P11660) nos sirvió para establecer la raíz del árbol (Xiong y Eickbush, 1990). La nomenclatura de las clases de intrones del grupo II se basa en la propuesta en Zimmerly *et al.*, 2001.

## I.2.2. ESTRUCTURA CONSENSO PARA LA RIBOZIMA DEL NUEVO GRUPO DE INTRONES

Se ha sugerido la existencia de una co-evolución entre las IEPs y la organización estructural del ARN en intrones del grupo II. De hecho, los agrupamientos filogenéticos basados en las ORFs coinciden con aquellos fundamentados en la clase estructural del ARN. En este estudio disponemos de la secuencia de cuatro intrones cuya estructura aún no ha sido determinada pero cuyo agrupamiento de acuerdo con los dominios RT de la IEP sugiere que forman un subgrupo independiente de todas las clases bacterianas descritas hasta el momento. Para verificar esta hipótesis intentamos generar un modelo consenso de plegamiento para esos cuatro intrones.

El alineamiento con Clustal X de las secuencias de ADN nos permitió localizar las regiones implicadas en interacciones estructurales dentro de los intrones. En cuanto a similitud, *B.j.I2* mostró un 65% de identidad respecto a *S.ma.I1*, mientras que *B.j.F4/F5* presentó un 83% de identidad respecto a *S.me.I4*. Además, se consiguieron identificar los seis dominios de la ribozima en tres de los cuatro intrones; la única excepción, *B.j.F4/F5*, sólo pudo mapearse hasta el tallo del subdominio ID3. El plegamiento de la estructura del ARN se mostró similar en los cuatro intrones, de acuerdo con lo previsto, lo que nos permitió establecer una estructura consenso (figura I.11) que se asemeja bastante con la descrita para los intrones de la clase bacteriana D (figura 12 en introducción; Toor *et al.*, 2001). Como en el caso de la clase D, presenta algunas características de los grupos IIA y IIB, pero también mostró algunas particularidades destacables. Dentro del dominio I, podemos señalar la desaparición de la inserción en la región inicial de este dominio (concretamente entre lo que se define como el tallo I(i) y I(ii)) y de las estructuras IC2 e ID(ii); quizá uno de los aspectos más destacables en relación con este dominio, sea la ausencia de la interacción  $\kappa$ - $\kappa'$  considerada como el principal punto de anclaje del dV al dI (receptor del dominio V). El dII presenta una estructura similar a la descrita para intrones bacterianos de la clase B (Toor *et al.*, 2001). Finalmente, los cuatro intrones muestran un desapareamiento en la región de la triada catalítica del dominio V de la ribozima no descrita hasta el momento (Peebles *et al.*, 1995), de modo que el apareamiento AGC-GUU se convierte en AGC-GAU.





**Figura I.11.** Estructura consenso para el ARN de los intrones bacterianos de la clase E. La nomenclatura de los dominios está de acuerdo con la usada en otros intrones por otros autores (Michel *et al.*, 1989; Qin y Pyle, 1998; Costa *et al.*, 2000). Se señalan las secuencias de reconocimiento de los exones predichas (EBS1, EBS2 y EBS3) así como los diferentes contactos terciarios a lo largo de toda la ribozima. Estos últimos se destacan con recuadros y se identifican con diferentes letras griegas. Con flechas se indican las principales características de la estructura de la ribozima de estos intrones. Los nucleótidos variables se representan con un círculo azul y las diferencias en la longitud de la secuencia entre los diferentes intrones del mismo grupo se indican en pares de bases (bp). Para la realización de este consenso, se han tenido en cuenta los siguientes intrones: *S.ma*.I1, 1909 bp (AF027768, cadena complementaria posición 657-2565); *B.j*.F1 (*B.j*.I2), 1912 bp (AF322013, posición 10857-12758), *S.me*.I4, 1894 bp (AJ496462, posición no definida); y, *B.j*.F4/F5, 1729 bp desde el dominio ID3 al dVI (AF332012, posiciones 109529-110496 y 112733-113493).

### I.3. DISCUSIÓN

El análisis de parte de la secuencia del plásmido críptico pRmeGR4b ha revelado aspectos muy importantes que podrían estar en relación con las ventajas simbióticas que en los últimos 30 años se han otorgado a *S. meliloti* GR4.

En cuanto a aspectos meramente metodológicos, debemos comentar que nuestro trabajo ha servido para determinar una nueva distribución de clones de la librería de pRmeGR4b y además ha permitido corroborar que el banco se encuentra incompleto, faltando cinco regiones por determinar (figura I.3B). Si analizamos nuestros resultados desde el punto de vista genético, parece evidente que nos encontramos ante un “replicón sumidero” para elementos transponibles de muy diversa naturaleza (figura I.5). Este dato no tendría especial relevancia si una de las características más destacables de la cepa GR4, su incrementada competitividad para la colonización de la raíz sobre otras bacterias de la misma especie, no se encontrase asociada con la presencia de este plásmido (Toro y Olivares, 1986; Sanjuan y Olivares, 1989; Soto *et al.*, 1993; García-Rodríguez y Toro, 2000). Se han descrito secuencias de inserción que muestran similitud con otras procedentes de especies y géneros relacionados ecológicamente con *S. meliloti* lo que apunta a fenómenos de transferencia horizontal del material genético, un fenómeno bastante extendido entre los procariotas. Podría resultar interesante determinar si, junto con estas ISs, también se transfirieron otros genes que pudieran ser los responsables de las ventajas ecológicas de GR4. Esta teoría podría ser verificada si se conociese la secuencia completa del plásmido, comprobando el contenido GC<sup>48</sup> y la naturaleza y origen de los genes que rodean a estas secuencias de inserción.

Siguiendo en la línea argumental anterior debemos destacar que la presencia de retrones se asocia habitualmente con la presencia de un profago integrado en el genoma de la bacteria (Lampson *et al.*, 2005). Además los profagos se consideran importantes agentes de transferencia genética horizontal (Frost *et al.*, 2005). Así pues, el supuesto retrón que hemos encontrado en copia única en este plásmido críptico podría tener alguna relación con la presencia de un fago latente cuyos genes virales podrían proporcionar a la bacteria ventajas en cuanto a infectividad. Este podría ser el caso de *S. meliloti* GR4 que coloniza la raíz más rápido que otros rizobios, presentando una tasa de fijación de nitrógeno similar (Toro y Olivares, 1986; Soto *et al.*, 1993). No es la primera vez que se ha descrito la presencia de este tipo de retroelementos en el genoma de un rizobio (Rice *et al.*, 1993). Los estudios de Rice y colaboradores pusieron de relieve la abundancia relativa de estos elementos en la población (16% de las cepas y especies analizadas, 10/63, presentaron msADN) y su variabilidad entre los diferentes rizobios (seis retrones diferentes en las diez cepas), llegando incluso a establecer parentescos

---

<sup>48</sup> Este parámetro se usa como medida de la homogeneidad del material genético en procariotas. Si una región presenta un contenido GC muy diferente del resto del genoma nos estaría indicando que ese fragmento en cuestión fue adquirido recientemente por el organismo. Se suele utilizar para definir eventos de transferencia horizontal.

filogenéticos. En muchos casos, parecía claro que habían sido transmitidos por herencia horizontal (mismo contexto genómico) aunque sugieren que podría tratarse de elementos genéticos móviles.

Finalmente, los estudios estructurales y filogenéticos realizados en el intrón localizado durante el desarrollo de este trabajo en la secuencia del plásmido pRmeGR4b, así como de otros tres intrones relacionados con éste, nos llevan a concluir que *S.me.I4*, *S.ma.I1*, *B.j.I2* y *B.j.F4/F5* forman una nueva clase filogenética dentro de los intrones del grupo II a la que designaremos como clase bacteriana E. Se ha conseguido una estructura consenso para esta nueva clase de intrones que reúne características de los intrones grupo II de la clase D, intrones IIB, algunas de los intrones IIA y un dII de intrones de la clase bacteriana B (figura I.11). Sin embargo, presentan algunas características propias que los diferencian de otros intrones en lo que se refiere al dominio V. En principio, sólo el último apareamiento de la triada estaría implicado en la interacción  $\kappa$ - $\kappa'$ , por lo que no podemos afirmar que exista una relación directa entre la falta de la región ID(ii) y el cambio en el nucleótido central opuesto a la triada catalítica, pero es relevante que estos dos cambios se produzcan simultáneamente. En la mayoría de los intrones tienen lugar tres interacciones que estabilizan la unión del dV con el dI y posicionan ambos dominios en el centro catalítico:  $\lambda$ - $\lambda'$  (región de la región desapareada interna del dI),  $\zeta$ - $\zeta'$  (zona del bucle terminal) y  $\kappa$ - $\kappa'$  (adyacente a la triada catalítica) (Introducción, figura 1). Teniendo en cuenta que ni la interacción  $\lambda$ - $\lambda'$  ni la  $\zeta$ - $\zeta'$  se encuentran afectadas podríamos pensar que la interacción entre ambos dominios no se ve comprometida. Sin embargo, se ha sugerido que el contacto  $\kappa$ - $\kappa'$  podría ser esencial a la hora de orientar el dV en el sitio activo de la ribozima (Boudvillain y Pyle, 1998). Además, estudios *in vitro* apuntan a la existencia de un centro de control del plegamiento de intrones del grupo II que implica tanto a la interacción  $\zeta$ - $\zeta'$  como a  $\kappa$ - $\kappa'$  e incluso podría extenderse hacia la región de coordinación (Waldsich y Pyle, 2007). Es probable que este grupo de ARNs de la clase E pudieran tener afectada su capacidad de compactación, ya que el elemento que desencadena el plegamiento se encuentra ligeramente alterado. Sin embargo, es interesante destacar que la región adyacente a la triada catalítica ( $\kappa'$ ) se encuentra conservada respecto al resto de intrones del grupo II, por lo que podría existir otra región en el dI que interaccionase con el dV reconstituyendo la interacción  $\kappa$ - $\kappa'$ .

El antepenúltimo nucleótido del tallo del dV se enfrenta a una de las posiciones más críticas, desde el punto de vista catalítico, identificadas en intrones del grupo II, la G central de la triada (figuras I.6 y I.11; Chanfreau y Jacquier, 1994; Peebles *et al.*, 1995). Estudios *in vitro* ponen de manifiesto una reducción de la capacidad de unión del dV a una molécula que representa el exón 1 unido a los dominios I-III en mutantes en los que la triada catalítica se compone de AGC-GAU (Chanfreau y Jacquier, 1994). Este leve defecto no parece ocasionar ninguna alteración de la capacidad de formar intrón escindido en lazo (Boulanger *et al.*, 1995) ni exones ligados (Chanfreau y Jacquier, 1994), por lo que podríamos afirmar que los intrones de la clase E no deberían presentar *a priori*

ningún defecto en su capacidad de escisión. Sin embargo, el entorno generado por el apareamiento G•U en esta región central se ha asociado con la coordinación de  $Mg^{2+}$  en el sitio catalítico de este ARN, en concreto con el posicionamiento del metal respecto de otros elementos del centro activo (Boulanger *et al.*, 1995) y se sabe que el  $Mg^{2+}$  es un elemento indispensable en el plegamiento y la actividad catalítica de las ribozimas.

Con todo, la caracterización funcional de los intrones bacterianos de la clase E requiere datos adicionales que permitan establecer si nos encontramos ante intrones activos o ante formas inactivas derivadas de un ancestro funcional.



# **Anexo capítulo I**



Tabla AI.1. Distribución de fragmentos entre los plásmidos del banco del pRmcGR4b.

PLÁSMIDO	DIGESTIÓN <i>EcoRI</i>	DIGESTIÓN <i>PstI</i>
	Nº fragmento	Nº fragmento
pRmNT101	27, 13, 22, 12, 21	16, 37, 39, 33a, 31, 33, 38
pRmNT102	26, 18, 24, 3a, 29	34, 4, 15
pRmNT103	18, 24, 3a	34, 4, 15
pRmNT104	4	10
pRmNT105	15, 8	43a
pRmNT106	24, 3a, 29	34, 4, 15
pRmNT107	11a, 9a	28, 26, 12, 32
pRmNT108	32, 11a, 9a, 24a, 4, 23	28, 26, 12, 32, 24, 10
pRmNT109	6, 5	27a, 20a, 6a, 27b
pRmNT110	6, 5	27a, 20a, 6a, 27b
pRmNT111	24,3a, 29, 10	34, 4, 15, 7
pRmNT112	2	1
pRmNT114	2	1
pRmNT115	12, 21, 27, 7, 17, 16	33a, 31, 33, 38, 7a, 30, 13, 40a
pRmNT116	2	1
pRmNT117	14, 6	23, 27a, 20a
pRmNT118	11a, 9a	28, 26, 12, 32
pRmNT119	2	1
pRmNT120	3, 26	41, 30a, 27, 11, 17, 29b, 29
pRmNT121	14, 6	23, 27a, 20a
pRmNT122	4, 23	10
pRmNT123	4, 23	10
pRmNT124	11a, 9a	28, 26, 12, 32
pRmNT125	11a, 9a	28, 26, 12, 32
pRmNT126	15, 8	43a
pRmNT127	28, 3,26, 18	41, 30a, 27, 11, 17, 29b, 29
pRmNT128	25, 28, 3, 26, 18, 24	22a, 41, 30a, 27, 11, 17, 29b, 29, 6
pRmNT129	2	1
pRmNT130	24a, 4, 23, 33, 31	24, 10, 8, 34a
pRmNT131	5,20,32	27b, 21, 35, 42
pRmNT132	5,20,32	27b, 21, 35, 42
pRmNT133	2	1
pRmNT134	1	1
pRmNT135	9, 25, 28, 3, 26	25a, 40, 22a, 41, 30a, 27, 11, 17, 29b, 29
pRmNT136	5, 20, 32, 11a	27b,21, 35, 42, 22, 28
pRmNT137	13	16



<b>pRmNT139</b>	4	10
<b>pRmNT140</b>	22, 12, 21, 27, 7, 17	39, 33a, 31, 33, 38, 7a, 30
<b>pRmNT141</b>	3	30a, 27, 11, 17, 29b, 29
<b>pRmNT142</b>	3, 26	41, 30a, 27, 11, 17, 29b, 29
<b>pRmNT143</b>	27, 7, 17, 16	7a, 30, 13, 40a, 19
<b>pRmNT144</b>	7, 17, 16	30, 13, 40a
<b>pRK290</b>	Se lineariza	

**Tabla AI.2.** Tamaño de los fragmentos *EcoRI* en kb.

<b>1</b>	<b>2</b>	<b>3 y 3a</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9 y 9a</b>	<b>10</b>
25,1	23,4	16,98	12,58	10	8,12	7,94	7,1	5,62	5,12
<b>11 y 11a</b>	<b>12</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17</b>	<b>18</b>	<b>19</b>	<b>20</b>
4,57	4,36	3,7	3,63	3,5	3,4	3,31	3,21	3,16	2,63
<b>21</b>	<b>22</b>	<b>23</b>	<b>24 y 24a</b>	<b>25</b>	<b>26</b>	<b>27 y 27a</b>	<b>28</b>	<b>29 y 29a</b>	<b>30</b>
2,08	1,9	1,86	1,65	1,54	1,44	1,38	1,17	1,09	1
<b>31</b>	<b>32</b>	<b>33</b>							
0,95	0,91	0,87							

**Tabla AI.3.** Tamaño de los fragmentos *PstI* en kb.

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6 y 6a</b>	<b>7 y 7a</b>	<b>8</b>	<b>9</b>
22,4	9,5	8,28	8,12	7,76	7,53	7,43	7,12	5,58
<b>10</b>	<b>11</b>	<b>12 y 12a</b>	<b>13</b>	<b>14</b>	<b>15</b>	<b>16</b>	<b>17 y 17a</b>	<b>18</b>
5,15	5,06	4,75	4,37	3,98	3,63	3,51	3,43	3,34
<b>19</b>	<b>20 y 20a</b>	<b>21</b>	<b>22 y 22a</b>	<b>23</b>	<b>24</b>	<b>25 y 25a</b>	<b>26</b>	<b>27, 27a y b</b>
3,21	2,97	2,60	2,24	2,15	2,04	1,97	1,88	1,84
<b>28</b>	<b>29, 29a y b</b>	<b>30 y 30a</b>	<b>31</b>	<b>32</b>	<b>33 y 33a</b>	<b>34 y 34a</b>	<b>35</b>	<b>36</b>
1,7	1,55	1,50	1,43	1,32	1,25	1,20	1,17	1,12
<b>37</b>	<b>38</b>	<b>39</b>	<b>40, 40a y b</b>	<b>41</b>	<b>42</b>	<b>43 y 43a</b>		
1,05	1	0,95	0,85	0,80	0,70	0,60		

Tabla AI.1. Oligonucleótidos usados para secuenciación y obtención de sondas.

Nº	CEBADOR	SECUENCIA	APLICACIÓN	REFERENCIA
1	NT40F	5' CGATCGCAACCAGTATGACA 3'	Secuenciación de los extremos de pRmNT140	Este trabajo
2	NT40I	5' CAACAACGAGGCGCATCTGA 3'	Secuenciación de los extremos de pRmNT140	Este trabajo
3	pRKI	5' GCATTACGGACTTCATG 3'	Secuenciación de los extremos del resto de los plámidos interés	Este trabajo
4	pRKF	5' CCCTTCACCAAGTTCGACGA 3'	Secuenciación de los extremos del resto de los plámidos interés	Este trabajo
5	111F	5' TAACCGCCATTCTTGTACGC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
6	111I	5' CTTCCGGTGTCAAGCGGTCTC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
7	117F	5' ATTCTACGCCGTTCTCCTGC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
8	117I	5' ACAGGCTGGCATTTCAGA 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
9	119F	5' CTGGCAGTCCTTGCCATT 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
10	119I	5' TTCCTGCTGGTACTGCTTCG 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
11	130F	5' GGCCAGAATATTCATCATGC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
12	130I	5' ATCCACATTGAGGATCGAGC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
13	134F	5' GTAACITGCGGATGCTGTCAC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
14	134I	5' GTCAATGCCCTTGGGAC 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo

15	135F	5' AGCATAGCGCGAGTATCGCA 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
16	135I	5' AGCTCTTGCGCACGTTTCAT 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
17	136F	5' GTTATGTGGCCTGCCGTAATAG 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
18	136I	5' CTTGGCTTCTCCGCATATGG 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
19	143I	5' GTTTCGGGTCTCATCGTG 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
20	101I	5' ATGGTTTGAGACCGGAGAGA 3'	Obtención de sonda por PCR para hibridación ADN-ADN	Este trabajo
21	128I	5' GCTCCTGTGCTTTCTTCATTG 3'	Obtención de sonda por PCR para hibridación ADN	Este trabajo
22	128F	5' TTTCCATTGGCAGGTTTCAG 3'	Obtención de sonda por PCR para hibridación ADN	Este trabajo
23	2-143I	5' GGAAACGAGCGTTCGATTTAC 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo
24	2-101I	5' ATTGAAGTCGATGGCTGACG 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo
25	1-135F	5' ATTCCAAAGAAGTCTGGCGG 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo
26	2-111F	5' TAGAGACGCAGAGCGAGAGC 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo
27	2-111I	5' CGCTATCCAGCAATAGGAGG 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo
28	2-117F	5' AGCCGCTCAAGAGATGAATG 3'	Obtención de sonda por PCR para hibridación en colonia	Este trabajo

**Tabla AI.2.** Sucesión de comprobaciones experimentales en la búsqueda de posibles clones de la genoteca de *S. meliloti* GR4 que nos permitan completar la secuencia del plásmido críptico pRmeGR4b.

SONDA	COLONIAS POSITIVAS POR HIBRID. EN COLONIA	COLONIAS POSITIVAS POR PCR	ANALISIS DE RESTRICCIÓN	HIBRIDACIÓN ADN-ADN	SECUENCIA
<b>101I</b>	31 colonias	8 lisados	7 muestras	0	-
<b>111I</b>	31 colonias	1 lisados	1 muestra	1 positivo	0
<b>111F</b>	83 colonias	10 lisados	6 muestras	0	-
<b>117F</b>	28 colonias	10 lisados	4 muestras	0	-
<b>135F</b>	15 colonias	6 lisados	0	0	-
<b>143I</b>	40 colonias	8 lisados	8 muestras	0	-

Tabla AI.1. Resultados de la búsqueda por alineamiento local en las bases de datos (BLAST).

<b>101I</b>						
Acceso	Descripción	Max score	Total score	Cobertura secuencia	E value	Max ident
ref ZP_02121104.1	ABC transport solute-binding protein [ <i>Methylobacterium nodulans</i> ORS 2060]	244	244	91%	1e-63	82%
ref ZP_01847183.1	Extracellular ligand-binding receptor [ <i>Methylobacterium</i> sp. 4-46]	235	235	89%	7e-61	77%
ref YP_001525916.1	ABC transport solute-binding protein [ <i>Azorhizobium caulinodans</i> ORS 571]	199	199	85%	5e-50	62%
ref ZP_02121737.1	ABC transport solute-binding protein [ <i>Methylobacterium nodulans</i> ORS 2060]	181	181	76%	2e-44	59%
<b>101F</b>						
X64613.1	<i>R.meliloti</i> plasmid pRmeGR4b OCD gene for ornithine cyclodeaminase	821	821	100%	0.0	100%
X63715.1	<i>R.meliloti</i> plasmid DNA for insertion sequence Rm3	268	268	32%	2e-68	100%
<b>111I</b>						
AE007274.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 80 of 121 of the complete plasmid sequence	31.9	31.9	5%	0.34	85%
AE007231.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 37 of 121 of the complete plasmid sequence	30.1	30.1	3%	1.2	94%
AE007220.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 26 of 121 of the complete plasmid sequence	30.1	30.1	4%	1.2	90%
AL591985.1	<i>Sinorhizobium meliloti</i> 1021 complete pSymB	30.1	260	36%	1.2	100%
<b>111F</b>						
ref YP_001372863.1	Phage integrase family protein [ <i>Ochrobactrum anthropi</i> ATCC 49188]	148.8	148.8	89%	1e-21	75%
ref YP_576128.1	Hypothetical protein Nham_0790 [ <i>Nitrobacter hamburgensis</i> X14]	100.9	100.9	86%	9e-15	69%
<b>112F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	609	609	100%	3e-171	100%
AY204475.1	Cloning vector pRK404, complete sequence	609	609	100%	3e-171	100%
AF327712.1	Cloning vector pRK310, complete sequence	609	609	100%	3e-171	100%
<b>114F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	609	609	100%	3e-171	100%
AY204475.1	Cloning vector pRK404, complete sequence	609	609	100%	3e-171	100%
AF327712.1	Cloning vector pRK310, complete sequence	609	609	100%	3e-171	100%
<b>116F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	609	609	100%	3e-171	100%
AY204475.1	Cloning vector pRK404, complete sequence	609	609	100%	3e-171	100%
AF327712.1	Cloning vector pRK310, complete sequence	609	609	100%	3e-171	100%
<b>117I</b>						
CP000740.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED02, complete genome	473	473	100%	2e-130	89%
CP000741.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED03, complete genome	275	343	85%	9e-71	85%
AL591985.1	<i>Sinorhizobium meliloti</i> 1021 complete pSymB	259	259	67%	7e-66	82%

<b>117F</b>						
Acceso	Descripción	Max score	Total score	Cobertura secuencia	E value	Max ident
AE007231.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 37 of 121 of the complete plasmid sequence	944	944	100%	0.0	98%
CP000740.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED02, complete genome	917	1460	100%	0.0	100%
DQ145546.1	<i>Sinorhizobium meliloti</i> plasmid pSmeSM11a, complete seq	866	866	100%	0.0	95%
ref[NP_43564 5.1]	Probable LysR-family transcriptional activator [ <i>Sinorhizobium meliloti</i> 1021]	331	331	99%	1e-89	99%
<b>119I</b>						
AY741093.1	Integration vector pJK202, complete sequence	881	881	99%	0.0	100%
AF327712.1	Cloning vector pRK310, complete sequence	881	881	99%	0.0	100%
AJ744860.1	uncultured bacterium pTB11 plasmid complete genome	881	881	99%	0.0	100%
M62426.1	Plasmid RK2 <i>kilA</i> locus encoding <i>klaA</i> , <i>klaB</i> , and <i>klaC</i> genes	881	881	99%	0.0	100%
<b>119F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	609	609	100%	3e-171	100%
AY204475.1	Cloning vector pRK404, complete sequence	609	609	100%	3e-171	100%
AF327712.1	Cloning vector pRK310, complete sequence	609	609	100%	3e-171	100%
<b>128I</b>						
AE007289.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 95 of 121 of the complete plasmid sequence	33.7	33.7	11%	0.089	75%
AL591985.1	<i>Sinorhizobium meliloti</i> 1021 complete pSymB	31.9	411	39%	0.31	100%
<b>128F</b>						
CP000741.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED03, complete genome	825	925	100%	0.0	98%
CP000740.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED02, complete genome	825	2207	100%	0.0	98%
CP000738.1	<i>Sinorhizobium medicae</i> WSM419, complete genome	825	825	100%	0.0	98%
<b>129I</b>						
AY741093.1	Integration vector pJK202, complete sequence	870	870	100%	0.0	99%
AF327712.1	Cloning vector pRK310, complete sequence	870	870	100%	0.0	99%
AJ744860.1	uncultured bacterium pTB11 plasmid complete genome	870	870	100%	0.0	99%
M62426.1	Plasmid RK2 <i>kilA</i> locus encoding <i>klaA</i> , <i>klaB</i> , and <i>klaC</i> genes	870	870	100%	0.0	99%
<b>129F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	574	574	100%	7e-161	97%
AY204475.1	Cloning vector pRK404, complete sequence	574	574	100%	7e-161	97%
AF327712.1	Cloning vector pRK310, complete sequence	574	574	100%	7e-161	97%
<b>130I</b>						
X99520.1	<i>R. leguminosarum</i> genes for restriction-methylation system, in a transposon-like structure	545	545	78%	4e-152	95%
Z68203.1	<i>Rhizobium</i> sp. plasmid NGR234a DNA	176	176	58%	8e-41	75%
U00090.1	<i>Rhizobium</i> sp. NGR234 plasmid pNGR234a, complete sequence	176	176	58%	8e-41	75%

<b>130F</b>						
Acceso	Descripción	Max score	Total score	Cobertura secuencia	E value	Max ident
ref YP_47294.1.1	Probable two-component sensor histidine kinase/response regulator hybrid protein [ <i>Rhizobium etli</i> CFN 42]	97.1	97.1	60%	4e-19	49%
AE007259.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 65 of 121 of the complete plasmid sequence	35.6	63.9	11%	0.025	100%
ref NP_77313.6.1	Hypothetical protein bll6496 [ <i>Bradyrhizobium japonicum</i> USDA 110]	71.2	71.2	62%	2e-11	41%
<b>133F</b>						
AY204476.1	Cloning vector pRK437, complete sequence	609	609	100%	3e-171	100%
AY204475.1	Cloning vector pRK404, complete sequence	609	609	100%	3e-171	100%
AF327712.1	Cloning vector pRK310, complete sequence	609	609	100%	3e-171	100%
<b>134I</b>						
AY914873.1	<i>Sinorhizobium meliloti</i> plasmid pMBA9a insertion sequence IS-like element putative transposase gene, complete cds; Rep(pMBA9a) (rep(pMBA9a)), TraI-like protein, RepA-like protein, and RepB-like protein genes, complete cds; and RepC-like protein gene, partial cds	333	333	54%	4e-88	97%
AF040724.1	<i>Sinorhizobium meliloti</i> NodD ( <i>nodD</i> ) gene, complete cds	329	329	54%	5e-87	97%
AE007216.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 22 of 121 of the complete plasmid sequence	264	264	45%	2e-67	95%
X56563.1	<i>R. meliloti</i> insertion sequence ISRM1 DNA	264	264	45%	2e-67	95%
<b>134F</b>						
AE007236.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 42 of 121 of the complete plasmid sequence	31.9	31.9	7%	0.20	90%
AL591985.1	<i>Sinorhizobium meliloti</i> 1021 complete pSymB	30.1	302	47%	0.70	100%
<b>135I</b>						
AY795967.1	<i>Sinorhizobium meliloti</i> plasmid pRmeGR4b insertion sequence ISBm1 transposase P2 gene, partial cds; RepC ( <i>repC</i> ) gene, complete cds; ExoI ( <i>exoI</i> ) gene, partial cds; and unknown gene	792	792	100%	0.0	99%
AY682089.1	<i>Sinorhizobium meliloti</i> strain GR4 plasmid pRmeGR4a RepC ( <i>repC</i> ) locus, partial sequence	212	212	37%	1e-51	88%
<b>135F</b>						
ref ZP_01312.827.1	RNA-directed DNA polymerase (Reverse transcriptase) [ <i>Desulfuromonas acetoxidans</i> DSM 684]	183		70%	5e-45	53%
sp P21325 RT67_ECOLI	Retron EC67 protein (ORF4-EC67 RT) [Includes: RNA-directed DNA polymerase (Reverse transcriptase); Ribonuclease H (RNase H)]	180		68%	4e-44	51%
ref ZP_00231.674.1	RNA-directed DNA polymerase from retron ec67 [ <i>Listeria monocytogenes</i> str. 4b H7858]	176		70%	5e-43	53%
<b>136I</b>						
CP000740.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED02, complete genome	765	1117	100%	0.0	94%
CP000741.1	<i>Sinorhizobium medicae</i> WSM419 plasmid pSMED03, complete genome	398	685	91%	2e-107	92%

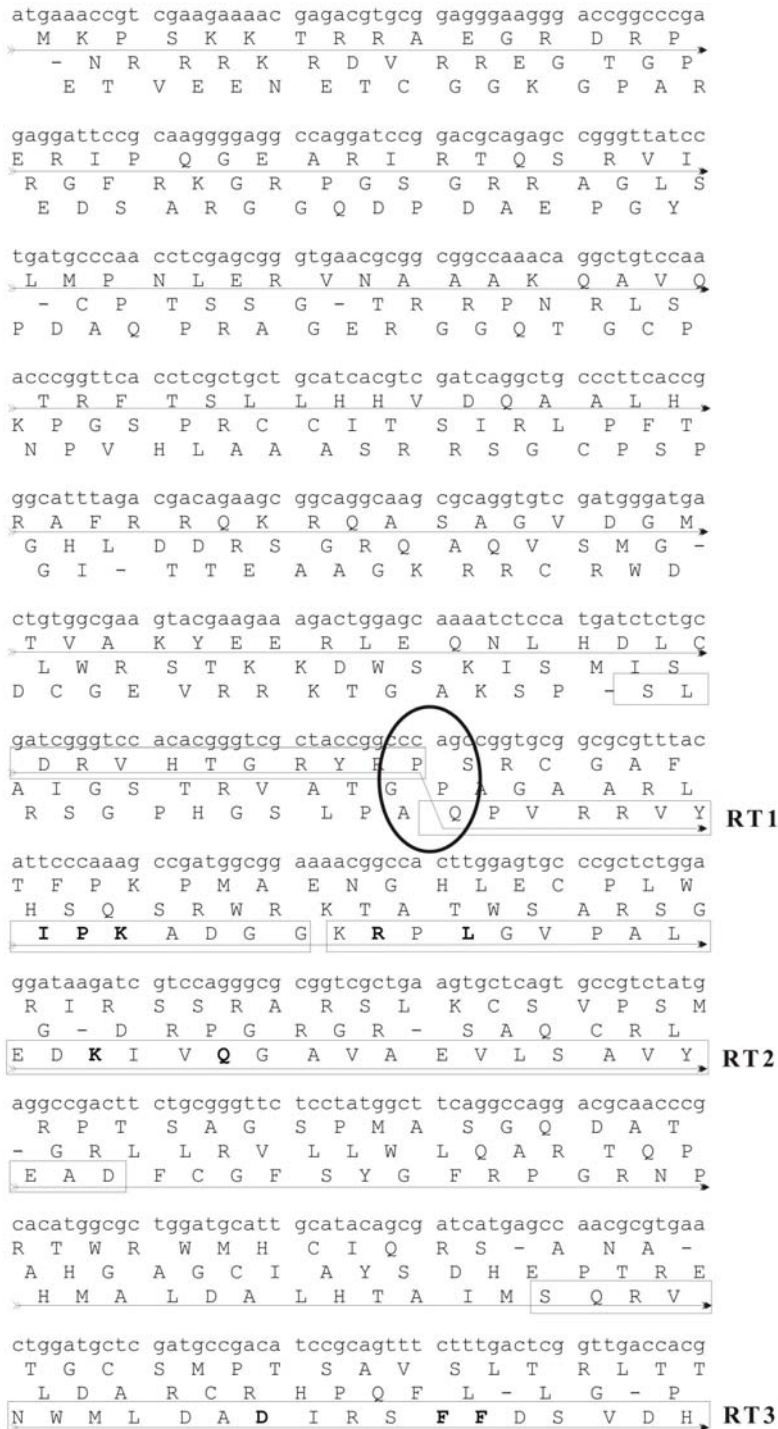
<b>136F</b>						
Acceso	Descripción	Max score	Total score	Cobertura secuencia	E value	Max ident
ref NP_94126.0.1	Putative reverse transcriptase/maturase [ <i>Serratia marcescens</i> ] RetA [IncL/M plasmid R471a]	216	216	71%	5e-55	54%
gb ABG56796.1	Putative <i>retA</i> reverse transcriptase [ <i>Klebsiella pneumoniae</i> ]	216	216	71%	5e-55	54%
ref NP_43716.4.1	Putative reverse transcriptasematurase protein [ <i>Sinorhizobium meliloti</i> 1021]	84.0	84.0	46%	9e-18	34%
ref NP_43623.9.1	Reverse transcriptase, maturase [ <i>Sinorhizobium meliloti</i> 1021] of intron RmInt1 protein	84.0	84.0	46%	1e-17	34%
<b>137I</b>						
CP000115.1	<i>Nitrobacter winogradskyi</i> Nb-255, complete genome transposase IS3/IS911	145	291	33%	2e-31	76%
<b>137F</b>						
AL591985.1	<i>Sinorhizobium meliloti</i> 1021 complete pSymB	31.9	179	20%	0.28	100%
AE007197.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 3 of 121 of the complete plasmid sequence	30.1	30.1	4%	0.97	100%
ref ZP_02857.143.1	FMN-binding negative transcriptional regulator [ <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i> WSM2304]	154	154	76%	2e-36	60%
<b>143I</b>						
CP000459.1	<i>Burkholderia cenocepacia</i> HI2424 chromosome 2, complete sequence	82.4	82.4	29%	2e-12	71%
AF010151.1	<i>Pseudomonas aeruginosa</i> PscN ( <i>pscN</i> ) gene, complete cds, and PscO ( <i>pscO</i> ) gene, partial cds	77.0	77.0	29%	8e-11	70%
AM236080.1	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> chromosome complete genome, strain 3841	73.4	116	28%	9e-10	96%
ref NP_65998.0.1	Type III secretion system ATP synthase protein [ <i>Rhizobium etli</i> CFN 42]	287	287	88%	2e-76	75%
ref YP_79154.8.1	Type III secretion system ATPase [ <i>Pseudomonas aeruginosa</i> UCBPP-PA14]	166	166	62%	7e-40	48%
<b>143F</b>						
X63715.1	<i>R. meliloti</i> plasmid DNA for insertion sequence Rm3	774	774	100%	0.0	99%
AE007261.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 67 of 121 of the complete plasmid sequence	756	756	100%	0.0	98%
AE007213.1	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA section 19 of 121 of the complete plasmid sequence	751	751	100%	0.0	98%



**Tabla AI.1.** Oligonucleótidos usados para obtener la secuencia completa de los retroelementos identificados en el pRmeGR4b.

<b>Nº</b>	<b>CEBADOR</b>	<b>SECUENCIA</b>	<b>APLICACIÓN</b>	<b>REFERENCIA</b>
<b>29</b>	<b>136FA</b>	5' ACGCCATTGATGTACCCAGA 3'	Obtener secuencia completa del elemento 136	Este trabajo
<b>30</b>	<b>2-136FB</b>	5' CTACCCAGCACAAGTGGCTT 3'	Obtener secuencia completa del elemento 136	Este trabajo
<b>31</b>	<b>136FABA</b>	5' CACCGGTGAACTGTAACG 3'	Obtener secuencia completa del elemento 136	Este trabajo
<b>32</b>	<b>STOP</b>	5' GTTATCCTGATGCCAACCT 3'	Obtener secuencia completa del elemento 136	Este trabajo
<b>33</b>	<b>107A</b>	5' CGATCTTATCCTCCAGAGCG 3'	Obtener secuencia completa del elemento 136	Este trabajo
<b>34</b>	<b>2-135F</b>	5' ATGCACAGGTCGTTGGCTA 3'	Obtener secuencia completa del elemento 135	Este trabajo
<b>35</b>	<b>135FB</b>	5' AGCATGGTCTGCGATACTCG 3'	Obtener secuencia completa del elemento 135	Este trabajo
<b>36</b>	<b>2-135FB</b>	5' TTAGACGTTGACTTCTGCG 3'	Obtener secuencia completa del elemento 135	Este trabajo
<b>37</b>	<b>136FB-bis</b>	5' AACTCTGGCAGTGCCTGGTT 3'	Obtener secuencia completa del elemento 135	Este trabajo
<b>38</b>	<b>ABA135F</b>	5' GGTCACGGATAACCTTGCTC 3'	Obtener secuencia completa del elemento 135	Este trabajo
<b>39</b>	<b>Aj135+</b>	5' CTCGACTACCTGCGAAACCTAG 3'	Obtener secuencia completa del elemento 135	Este trabajo

**Figura AI.1.** Secuencia de la RT identificada en el pRmNT136 con identidad a IEP de intrones del grupo II. Con flechas se indica la fase de lectura. Las cajas representan los dominios conservados de la RT y en negrita se muestran los residuos que caracterizan cada dominio. Los círculos marcan los cambios en la pauta de lectura.



agtggttgct gcagatggtg ggcgatagaa ttgccgatcc ccgatatcctg  
 S G C C R W W R I E L P I P V S  
 R V V A A D G G A - N C R S P Y P  
 E W L L Q M V A H R I A D P R I L **RT3**

caacttatta aactatggtc gcgggccgga attcttgaga gcgggcgagac  
 C N L L N Y G C G P E F L R A A R  
 A T Y - T M A A G R N S - E R R D  
 Q L I K L W L R A G I L E S G E

gtatgagact gacaggggga ctccgcaagg ggcgggcatc agtccgctcc  
 R M R L T G G L R K G R A S V R S  
 V - D - Q G D S A R G G H Q S A  
 T Y E T D R G T P Q G A G I S P L

tgcgcaacat ctttttgac tacatccttg atctctgggt acatcaatgg **RT4**  
 S P T S F C T T S L I S G Y I N  
 P R Q H L F A L H P - S L G T S M  
 L A N I F L H Y I L D L W V H Q W

cgctgctgcc atgcacgcg ggcgtattgtg atcgtgcgct atgccgacga  
 G V V A M H A A V L - S C A M P T  
 A S S P C T R P Y C D R A L C R R  
 R R R H A R G R I V I V R Y A D

cttcgtcatg ggcttcgaga agaaggatga tgcgcaggag atgctcttgg **RT5**  
 T S S W A S R R R M M R R R C S W  
 L R H G L R E E G - C A G D A L  
 D F V M G F E K K D D A Q E M L L

ccctcaagga acggctaggc gaatttggtc tggcgctcca tgaggggaag  
 P S R N G - A N L V W R S M R G  
 G P Q G T A R R I W S G A P - G E  
 A L K E R L G E F G L A L H E G K **RT6**

acgcggtctaa tcgaatttgg cgggttcgcg gccttatctc gacagcgggc  
 R R G - S N L A G S R P Y L D S G  
 D A A N R I W P V R G L I S T A A  
 T R L I E F G R F A A L S R Q R

tggtgagcga aaaccagaga ccttcgcctt cctgggcttc attcactatt  
 V V S E N Q R P S P S W A S F T I  
 W - A K T R D L R L P G L H S L  
 R G E R K P E T F A F L G F I H Y **RT7**

gtggattgac ccgatgatgc cggttcatcg tgaagcacia gacggaagg  
 V D - P V M A G S S - S T R R K  
 L W I D P - W P V H R E A Q D G R  
 C G L T R D G R F I V K H K T E G

aagcgctga cgcgaaagct caaggcgttg cgcgagggag catggcggca  
 G S A - R E S S R R C A R E H G G  
 E A P D A K A Q G V A R G S M A A  
 K R L T R K L K A L R E G A W R

catgcacagg tcggttgcta cccagcacia gtggcttgcc gctgtgctgc  
 T C T G R W L P S T S G L P L C C  
 H A Q V V G Y P A Q V A C R C A  
 H M H R S L A T Q H K W L A A V L

```

gtggacacta cggctattac ggcaggccac ataactatcc cgcgctcaac
  V D T  T A I  T A G H  I T I  P R S
A W T L  R L L  R Q A  T - L S  R A Q
» R G H  Y G Y Y  G R P  H N Y  P A L N
»
gggttcctatc agcaaacacg cgcatatggt tccgctgcct gaggcggcga
  T G S I  S K H  A H M  V P L P  E A A
  R V L  S A N T  R I W  F R C  L R R R
» G F Y / Q Q T  R A Y G  S A A  - G G
»
agtcagaaaa gccggcggat gagttgggtg gaatttgaaa tctgaccgc
  K S E  K P A D  E L V  G I - N P D R
» S Q K  S R R  M S W L  E F E  I L T
» E V R K  A G G  - V G  W N L K  S - P
»
acgcttcacg ctgccagtcc cacacattac tcgaacttgg ggcgaggctc
  T L H  A A S  P T H Y  S N L  G A G
A R F T  L P V  P H I  T R T W  A Q A
» H A S  R C Q S  H T L  L E L  G R R L
»
ggatatga
S D M
» R I -
  G Y

```

**Figura AI.1.** Secuencia de la RT identificada en el pRmNT135 con identidad al retrón EC67 de *E. coli*. En negro se muestra la secuencia de ADN y en verde su traducción a proteína. Las cajas verdes representan los residuos conservados de la proteína y en magenta se muestran los dominios identificados. No se han detectado cambios en la pauta de lectura aunque la presencia de nucleótidos no determinados en la secuencia de ADN (n) impide disponer de la secuencia traducida completa (?). La porción terminal carece de estructura.

```

1  gaattcgaaa ttccaagaag gtctggcggc aaacgggtca tcaagcccc gatcccaaaa ctcaaaaaa tccaactca tcttgcgcat
   E F E I P K K S G G K R V I K A P I P K L K K L Q T H L A H
91  gtctctacg catgcctcgc cgagattgaa caaacctccc acgcgaaatc cgtgtcgtat ggatttcgaa aaggcgcgg gatttagcggc
   V L Y A C L A E I E Q N L H A K S V S Y G F R K G R G I S G
181 aacgcaaaaa tccacaacag gaagcgtctc gtcctgaact tggatctcgc cgactttttc cccaccttca acttcggcgg agttcgcggc
   N A K I H K R K R F V L N L D L A D F F P T F N F G R V R G
271  tttttcctga aagacaattc cttcgcgctg acaaacgagg tagcaacaac cattgccag atagcgtgtg atgtaccgc ccttctcag
   F F L K D N S F A L T N E V A T T I A Q I A C D G T A L P Q
361  ggcagmccgt gctctccagt gatctcggaa ttgatctgcc acatattaga cgttcgact ctgcgcctcg caagaaagca tggctcgcga
   G ? P C S P V I S E L I C H I L D V R L L R L A R K H G L R
451  tactcgcgct atcgtgacga cattactctc tcaacgaacc gaaaagatt tccgtctgag attcgtcgc tgaacaaaat tgaccgcctc
   Y S R Y A D D I T L S T N R K E F P S E I A S V N K I D P S
541  acctgggtta ttggccagga actagcaagc agaatagacg atttgggtt cgctattaat ctgacgaaga cgcgtatgca ctgtcggcgt
   T W V I G Q E L A S R I D D C G F V I N L T K T R M H C R G
631  agccgtcaga tggtagcggg cctagttgtg aatgagaag tgaacattcg cagcgaatac tatcggaacg cccgtcggat gtgcgatagc
   S R Q M V T G L V V N E K V N I R S E Y Y R N A R A M C D S
721  cttttcaga agggtgccta ttttactctc caccagcgcg ccacagtgg tggagggtcg ccaattccaa aggtcagcga taaccttgc
   L F Q K G A Y F H S H Q P P T V G G G S P I P K V T D N L A
811  ccattggaag gcattctcaa ccacatctac tcagtaactc agaaggaaga gcggcgatcc gtacaagatc agcgtttgga gccacgcgcc
   P L E G I L N H I Y S V T Q K E E R R S V Q D Q R L E P R A
901  atccgagcac tctatcgtcg ttctcttttc tacaagcgtt tcattgcctc cgatgctccg ctgattgtta cggagggaaa gacagatccc
   I R A L Y R R F L F Y K R F I A L D A P L I V T E G K T D P
991  gtctaccttc gtgaggctat taagcacctc cctctttggc atccgctttt ggggcaacca acagcggatg gcttcaaatt ctcaactca
   V Y L R E A I K H L P S W H P L L G Q P T A D G F K F S L S
1081 tttttcaact acggcggaaa atcacatgag atcatggatc tcggtggcgg aacaggggac ctaaaaatca taccctcga ctacctcga
   F F N Y G G K S H E I M D L G G G T G D L K S I P L D Y L R
1171 aaactagatt catctaagtc tgacagaaag ccaatgctgc ataagccgat gaaattcccg gtcacatggt ttctcgataa cgatgatggg
   N L D S S K S D R K P M L H K P M K F P V I H V L D N D D G
1261 ttgngcgcgc taccggggac catcaagaan aactcaacg tggggatgtc tgtaacgagc acccaagact tctatcacgt ctgcgacaac
   L ? D V P G T I K ? N F N V G M S V T S T Q D F Y H V C D N
1351 ctttatgtgg ncaaaacgcc tmgggcagm ggcaaaaact cgatgaaag catgtttcct aaaagctgga ccacgatcgt gctcaatggc
   L Y V ? K T P ? A ? G K S S M E S M F P K S W T T I V L N G
1441 aaaaagttaa atccnagcaa caagattgac attacaaaag aanacgggaa gaaaattttt gccaccaacg ttgtnaaagc caacgcggcc
   K K F N ? S N K I D I T K E ? G K E I ? A T N V ? K A N A A
1531 aagatcgatt tcagcggctt taacctttg cttcgaacgg atcgtaaag tcaactcgatc attacgcgcc accagcggcc tgagcttttt
   K I D F S G F N P L L R T D R
1621 cgttcccag aggtcctaac acacgcccga aattttccc ttctnctt cctgtcatct eggcctgcg ttctggcact ctctt

```

## **Capítulo II**



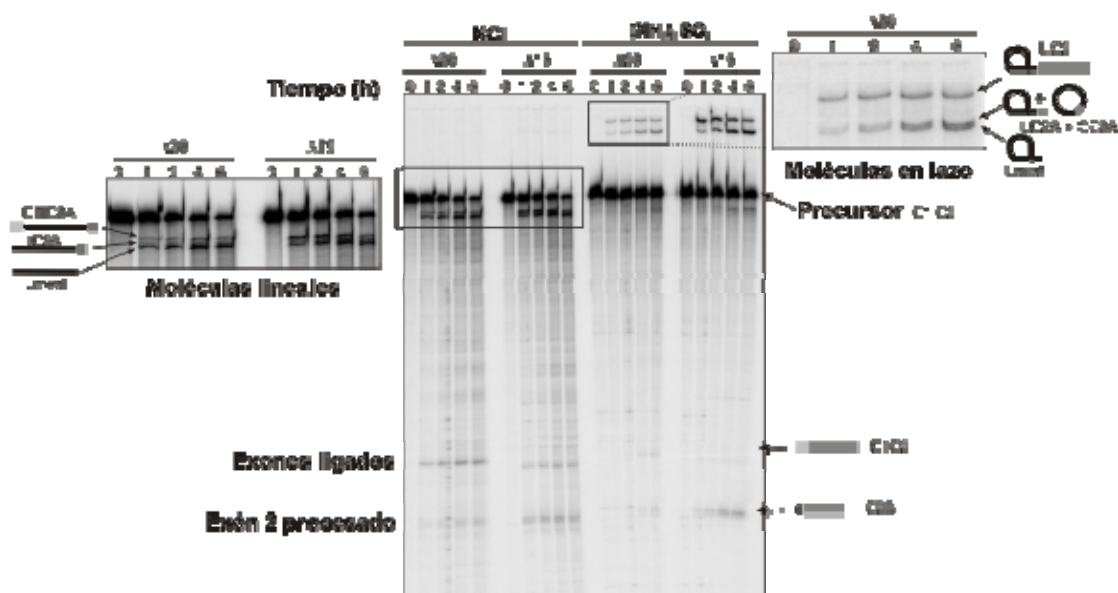
## II. FORMAS DE ESCISIÓN GENERADAS EN LA MADURACIÓN DE RmInt1

En trabajos previos se ha descrito la capacidad para madurar *in vivo* e *in vitro* de RmInt1. El proceso de escisión *in vivo* se ha caracterizado usando diferentes aproximaciones experimentales y se ha comprobado que depende de la IEP (Martínez-Abarca *et al.*, 1998; Muñoz-Adelantado *et al.*, 2003). Además se ha demostrado que construcciones modificadas respecto a la silvestre mejoran la eficiencia de la reacción de *splicing in vivo* hasta casi alcanzar el 100% (Nisa-Martínez *et al.*, 2007; este trabajo).

Por otro lado, los datos resultantes del análisis de los productos obtenidos de la reacción de *splicing in vitro* en ausencia de la IEP proporcionan una imagen más compleja del proceso (Figura II.1; Costa *et al.*, 2006a y b). En general, se considera una reacción considerablemente ineficiente y en la que se generan una serie de productos no esperados junto a los resultantes de la catálisis detectados en otros intrones del grupo II. Coincidiendo con los datos obtenidos *in vivo*, la reactividad de RmInt1 aumenta al acortar el exón 1. Se manejaron dos construcciones en las que se observó un comportamiento diferencial: una construcción denominada  $\Delta 15$  en la que el exón 1 queda reducido a los IBSs 1 y 2, y una construcción  $\Delta 20$  que contiene los 20 últimos nucleótidos del exón 1. Ambas construcciones se estudiaron en presencia de sales de amonio [1 M  $(\text{NH}_4)_2\text{SO}_4$ , que favorece la formación de *lariat*, y  $\text{NH}_4\text{Cl}$ , genera un 50% de ambos productos] y de sales de potasio [1 M  $\text{KCl}$ , que beneficia que se produzca la maduración por hidrólisis]. La primera construcción en ensayarse fue  $\Delta 15$ ; en  $(\text{NH}_4)_2\text{SO}_4$ , se observó una acumulación de intermediario *lariat* unido al exón 2 (LE2) así como una muy baja eficiencia de ligación de exones a pesar de que la proporción de moléculas de precursor reactivas alcanzó el 30% a 100 mM  $\text{MgCl}_2$ . Al incluir en la construcción los nucleótidos -20 a -16, que se sabe que son esenciales para obtener un *homing* eficiente *in vivo* (Jiménez-Zurdo *et al.*, 2003), la reactividad del ARN precursor disminuye alrededor de 3 veces (10% de moléculas reactivas) así como la velocidad de la primera reacción del *splicing* ( $0,12 \text{ h}^{-1}$  en  $\Delta 20$  frente a  $0,34 \text{ h}^{-1}$  en  $\Delta 15$ ) (Costa *et al.*, 2006a); sin embargo, aunque muy baja, la proporción de exones ligados es mayor en  $\Delta 20$  que en  $\Delta 15$  (Costa *et al.*, 2006a y b). El análisis de los productos de *splicing* mediante extensión a partir de cebador, RT-PCR y secuenciación revelaron la presencia de una serie de productos inesperados (LE2A, E1IE2A, IE2A y E2B) que podrían justificar la ineficiencia observada en la reacción (Costa *et al.*, 2006a). En todos los casos, se deduce la existencia de una posible interacción IBS1\* en el exón 2 que competiría con la original en el exón 1 (IBS1) y que conllevaría la hidrólisis del exón 2 en la posición +10 (+11 en casos puntuales). Además de estas moléculas, datos derivados de RT-PCR ponen de manifiesto la existencia de formas de intrón circular. Estas formas circulares contienen los diez primeros nucleótidos (CE2A) del intrón por lo que se considera que se han formado como consecuencia del ataque del



3'OH libre de la cola de este particular *lariat* (unido al E2A), al enlace 2'-5' del mismo. Este comportamiento *in vitro* podría traducirse *in vivo* en una estrategia para regular la expresión de su diana natural, la secuencia de inserción *ISRM2011-2*.



**Figura II.1.** *Splicing in vitro* de RmInt1 con construcciones  $\Delta$ ORF a 45°C con diferente sal monovalente: KCl y  $(\text{NH}_4)_2\text{SO}_4$ .  $\Delta$ 20 y  $\Delta$ 15 hace referencia a la longitud del exón 1 en el transcrito precursor. A ambos lados se marcan tanto con esquemas como con siglas la identificación de los productos en las correspondientes bandas. E1, exón 1 (cajas gris claro); E2, exón 2 (cajas gris oscuro); I, L y C, se corresponden respectivamente con intrón lineal, *lariat* y circular (trazo negro). Se resaltan algunos detalles del gel. Figura adaptada de Costa *et al.*, 2006a.

Con el fin de conseguir una imagen completa del proceso de maduración de RmInt1 *in vivo*, se ha llevado a cabo el análisis de los productos esperados en la reacción: intrón escindido y exones ligados. En resumen, podemos afirmar que este intrón se escinde dando lugar a ARN *lariat* (forma potencialmente invasiva), ARN circular, en mucha menor proporción, y ARN lineal. Además, nuestros esfuerzos encaminados a detectar exones ligados han sido poco productivos, en parte reflejando lo observado *in vitro*.

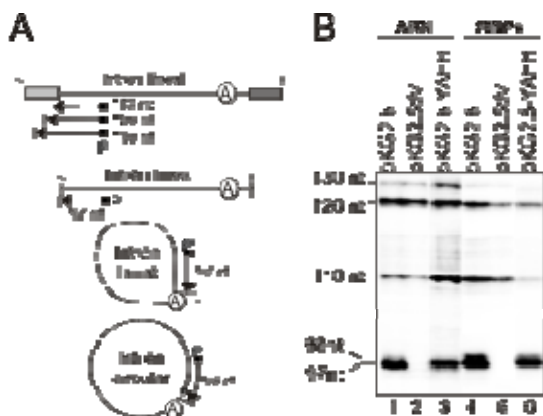
## II.1. RmInt1 FORMA CÍRCULOS *IN VIVO*

El planteamiento de este trabajo se divide en dos fases: una fase inicial de caracterización del proceso de escisión de RmInt1 donde identificamos moléculas en lazo y moléculas circulares cuya formación depende en gran medida de la presencia de la IEP, y una segunda fase que implica la caracterización fenotípica de distintos mutantes afectados en su capacidad para madurar.

## II.1.1. DETECCIÓN DE MOLÉCULAS CIRCULARES

### II.1.1.1. Escisión *in vivo* del intrón *RmInt1* determinada por extensión a partir de cebador

La primera aproximación para la detectar la presencia de intrón escindido en nuestras preparaciones de ARN y RNPs extraídas *in vivo* se realizó mediante la técnica de extensión (reverso transcripción) a partir de cebador. Para ello utilizamos el oligonucleótido P complementario a las posiciones 80-97 del extremo 5' del intrón (ver figura II.2A; Muñoz-Adelantado *et al.*, 2003). Coincidiendo con datos previos (Muñoz-Adelantado *et al.*, 2003), las muestras procedentes de células de *S. meliloti* RMO17 que expresan el intrón silvestre en construcciones pKG2.5 dan lugar a varios productos de extensión: una serie de bandas de longitud variable (110 nt, 126 nt y 130 nt) que se corresponden con la presencia de precursores y una banda mayoritaria de 97 nt que representa el/los productos de escisión del intrón (figura II.2B, carriles 1 y 4). Junto a este producto mayoritario, aparece una segunda banda de 98 nt que parece estar asociada también con la escisión del intrón. Ninguna de estas dos bandas se detectaron en muestras de ARN o RNPs aisladas de bacterias que portan pKG2.5dV, un plásmido que expresa un intrón mutado en el dominio V de la ribozima (las posiciones G<sub>1840</sub>U<sub>1841</sub>U<sub>1842</sub> se han modificado a GCA; Muñoz-Adelantado *et al.*, 2003) (figura II.2B, carriles 2 y 5). Sin embargo, comprobamos que un intrón mutado que codifica una proteína carente de actividad RT, pKG2.5-YAHH [los dos residuos aspártico (D) conservados en el motivo catalítico del dominio RT5 de la proteína se han cambiado por dos histidinas (H)], ve disminuida ligeramente su capacidad de maduración (~80% respecto del WT; figura II.2B, carriles 3 y 6).



**Figura II.2.** Análisis por extensión de cebador del *splicing in vivo* de *RmInt1*. (A) Ilustración esquemática de los productos esperados en el ensayo tanto a partir de las moléculas de precursor como de las posibles formas de escisión del intrón. La adenosina del dominio VI se representa como una A dentro de un círculo. (B) Resultado de la RT usando como molde tanto ARN total como RNPs extraídos de *S. meliloti* RMO17 que expresa el intrón desde las diferentes construcciones indicadas y como cebador el oligonucleótido P.

Es evidente que esta técnica no revela la naturaleza de los productos escindidos pues el producto de extensión a partir de una molécula lineal es idéntico al que se obtendría de la reverso-transcripción de una molécula *lariat*. Se ha establecido que la mayoría de las reverso transcriptasas (entre ellas la AMV RT usada en este ensayo) no son capaces de leer más allá de los enlaces modificados 2'-5' en moléculas con forma de lazo y se paran

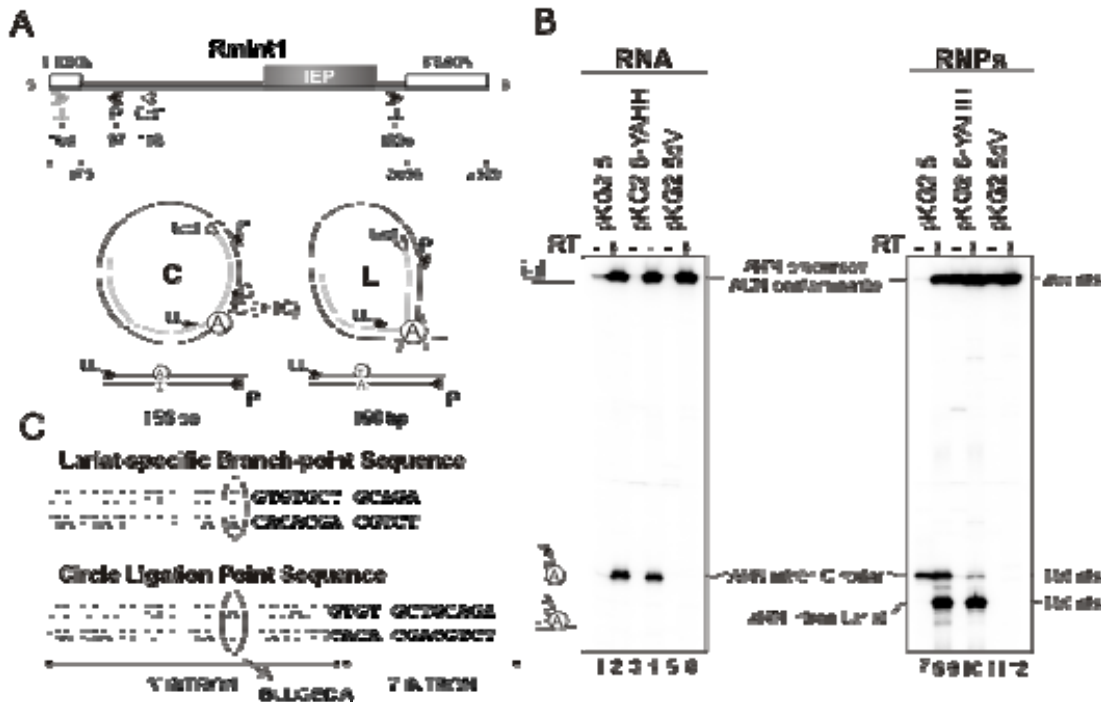
justo antes (Murray *et al.*, 2001); por tanto, el producto de extensión derivado de la molécula *lariat* en nuestro caso debe tener 97 nt. Por el contrario, si este enlace 2'-5' se encuentra en una molécula lineal (o no *lariat*), la pausa tiene lugar una vez se ha pasado el enlace (Lorsch *et al.*, 1995). De este modo, podríamos interpretar la aparición de la banda de 98 nt como la presencia en nuestras preparaciones de moléculas derivadas de la escisión del intrón en las que existe un enlace 5'-2' y en las que el 3'OH no se encuentra unido a ningún otro nucleótido o cadena de nucleótidos. Así pues, en función de las rutas descritas para la maduración de intrones del grupo II, podríamos pensar que se trata de moléculas circulares en las que el primer nucleótido del intrón se encuentra unido al último.

### II.1.1.2. Detección de dos productos de escisión diferentes por RT-PCR

Nuestro siguiente paso fue tratar de confirmar esa hipótesis mediante RT-PCR. Sometimos nuestras preparaciones de ARN total y RNPs a reverso-transcripción usando un oligonucleótido que hibrida a 118 nt del inicio del intrón (Ect1) y una RT que se supone que es capaz de atravesar el enlace modificado 5'-2' sin desengancharse (SuperScript II®, Invitrogen), aunque no es 100% eficiente. A continuación, el ADNc resultante se sometió a amplificación mediante PCR, usando los cebadores divergentes P (complementario a las posiciones 80-97 del intrón y que amplifica en dirección al extremo 5' del intrón) y LL (hibrida entre los nucleótidos 1826-1845 del intrón y permite la polimerización en dirección hacia el final del intrón) (figura II.3A). El marcaje radiactivo en 5' del oligonucleótido P nos permite la identificación de los productos de PCR una vez se han separado por electroforesis en geles de poliacrilamida desnaturizantes al 6% (UREA 8M) (figura II.3B)<sup>49</sup>. Al analizar los productos de amplificación de las muestras de ARN procedentes de células que expresan el intrón silvestre (pKG2.5) o mutado en el dominio RT5 (pKG2.5-YAHH), se observó la aparición de dos bandas: una de 265 nt y otra de 156 nt (figura II.3B, carriles 2 y 4). La banda de 265 nt también aparecía en el mutante pKG2.5dV, lo que nos estaba indicando que esta banda no deriva de intrón escindido, sin embargo, en este mutante defectuoso en *splicing* no fuimos capaces de detectar la banda de 156 nt (figura II.3B, carril 6), cuyo tamaño coincide con el esperado para moléculas resultantes de la amplificación de un intrón circular en el que el primer (G) y el último (C) nucleótido se encuentran formando un enlace (figura II.3A). La banda de 156 nt tampoco aparece en ausencia de RT previa al PCR, lo que nos indica que deriva de una molécula ARN (figura II.3B, carriles 1, 3 y 5).

---

<sup>49</sup> Se trató de separar los productos en geles de agarosa o en geles de poliacrilamida no desnaturizantes pero la capacidad resolutive en ambos casos fue insuficiente. Mientras la banda de 265 bp se puede distinguir fácilmente, las bandas de 150 y 156 bp fue imposible separarlas. Se usó agarosa de bajo punto de fusión que permite aumentar el porcentaje del gel hasta el 5% y varios porcentajes de poliacrilamida, con escaso éxito (datos no mostrados).



**Figura II.3.** Caracterización de los productos de escisión mediante RT-PCR. (A) Representación esquemática del ensayo indicando la localización de los cebadores en la secuencia de *RmInt1*. Ect1 fue el oligonucleótido elegido para la RT (flecha blanca), mientras que P y LL se usaron en el PCR (flechas negras). También se indica el tamaño de los productos de amplificación esperados de cada uno de los tipos moleculares, círculo y *lariat*. Con una flecha gris se indica un apareamiento inespecífico del cebador LL como consecuencia del cual se amplifica la banda de 265 bp. (B) Análisis en gel de los productos de RT-PCR de ARN y RNPs de *RmInt1* (pKG2.5) y mutantes derivados (pKG2.5dV y pKG2.5YAHH). Se muestran resultados del PCR con (+) y sin (-) previa RT. En los laterales de los geles se indican el tamaño y la naturaleza de los productos identificados. (C) Se muestran las regiones que rodean al punto de ligación de la molécula circular, así como en torno al enlace 2'-5' del *lariat*, tal y como aparecen en la secuenciación de los productos de PCR.

El análisis de los amplificadores obtenidos a partir de muestras enriquecidas en RNPs nos deparó algunas sorpresas. De modo similar a los resultados obtenidos con ARN, las RNPs procedentes de cepas con construcciones pKG2.5 permitieron la amplificación de una banda de 156 nt, ausente en bacterias que expresan el intrón mutado en el dV de la ribozima (pKG2.5dV) (figura II.3B, carriles 8 y 12). Es importante destacar, que este producto se encuentra considerablemente reducido en cepas que expresan el intrón a partir de la construcción pKG2.5-YAHH (figura II.3B, carril 10). A diferencia de lo descrito para las muestras de ARN, en el caso de las preparaciones de RNPs, tras la amplificación del ADNc, nos aparece una banda claramente mayoritaria de 150 nt. Este tamaño se corresponde con el esperado para un sustrato ARN derivado del anclaje mediante enlace 5'-2' del primer nucleótido del intrón (G) con la adenosina desapareada del dominio VI de la ribozima (A). Esta banda aparece en bacterias con construcciones pKG2.5 y pKG2.5-YAHH, pero está ausente en aquellas que portan el plásmido

pKG2.5dV (figura II.3B, carriles 8, 10 y 12). Así pues, los datos de RT-PCR parecen demostrar la existencia de dos formas de escisión para RmInt1: una, aparentemente mayoritaria en RNPs, que se manifiesta como un amplificado de 150 nt y que no se amplifica en ARN total, y otra que se traduce en una banda de 156 nt que se amplifica en todas las muestras con cierto nivel de *splicing* (ya sea ARN o RNPs).

Llamó nuestra atención la presencia de la banda de 156 nt en muestras de RNPs obtenidas a partir de cepas que expresan el intrón WT que no fueron sometidas a RT (por lo que el cebador Ect1 no se adiciona) (figura II.3B, carril 7). Este dato nos estaría indicando que se trata de amplificadas procedentes de moléculas de ADN presentes en nuestra preparación y que coinciden en tamaño con las moléculas derivadas de intrón circular. Así pues, nos preguntamos si existe la posibilidad de que se estén formando moléculas de ADN circular a partir de moléculas de ARN de intrón escindidas en forma circular. Este tipo de amplificado no aparece en cepas con construcciones con ribozima o la RT IEP mutada (figura II.3B, carriles 9 y 11).

### II.1.1.3. RmInt1 se escinde *in vivo* dando lugar a intrón *ariat* y circular

Con el fin de confirmar la identidad de los productos de RT-PCR se procede a su secuenciación. Para ello, los amplificadas se extraen de gel de agarosa al 2% y se clonan.

Los 19 clones resultantes de la banda de 265 bp en muestras de ARN y RNPs procedentes de pKG2.5 y pKG2.5-YAHH, presentaban una secuencia que se corresponde con 168 nt del extremo 3' del exón 1 unidos a 97 nt del extremo 5' del intrón. Este análisis coincide con la secuencia resultante de 9 clones de la misma banda obtenida en ausencia de RT en preparaciones de RNPs del mutante en el dominio RT5 (pKG2.5-YAHH). Todos los resultados en referencia a esta banda apuntan a que nos encontramos ante un producto de amplificación derivado de contaminación en las muestras con ADN (por eso aparece en ausencia de RT) o, en el caso de los carriles +RT, procedente de moléculas de ARN precursor. Aún así, se trataría de moléculas lineales que una pareja de cebadores divergentes no pueden amplificar. Efectivamente, este producto deriva de amplificación con los oligonucleótidos P/LL, pero comprobamos la existencia de un apareamiento inespecífico del cebador LL entre las posiciones -168 a -149 de la secuencia de inserción aguas arriba del intrón (figura II.3A).

La secuenciación de plásmidos en los que se clonó la banda de menor tamaño en las muestras de ARN en geles de agarosa (156 bp), vino a confirmar nuestras sospechas y nos desveló algunos datos adicionales que hasta el momento se habían pasado por alto. Cinco de los 9 clones que se secuenciaron de la amplificación de los productos de escisión del intrón WT se corresponden exactamente con la secuencia que se esperaría del PCR, previa reverso-transcripción, de una molécula de ARN intrón circular en la que los nucleótidos inicial y final están ligados (figura II.3C). Los 4 clones restantes mostraron una citosina (C) extra justo en el punto en el que se unen los dos extremos del intrón. Los productos de estas formas con extra C son visibles en los geles de poliacrilamida, aunque

como bandas minoritarias (figura II.3B, carriles 2, 4 y 8). La proporción de uno u otro tipo de moléculas fue manifiestamente diferente en muestras de ARN de cepas con la IEP mutada en la actividad RT: 10/11 se correspondieron con círculos perfectos, mientras 1/11 contenía la extra C.

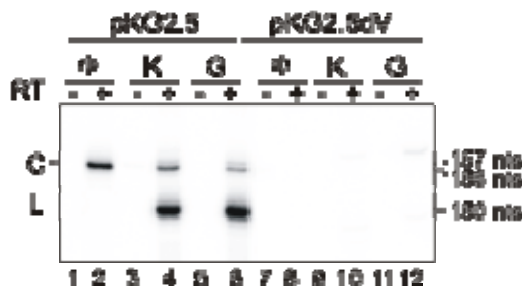
Por otro lado, en el caso de las muestras procedentes de RT-PCR de preparaciones de RNPs también se aisló de gel de agarosa la banda de menor tamaño molecular. Como ya hemos mencionado, en agarosa al 2% fuimos incapaces de discriminar entre el amplificado de 150 y el de 156 bp, por lo que la proporción relativa de secuencias debe corresponderse con la cantidad presente de cada uno de los amplificados en las diferentes muestras. Para la construcción pKG2.5, 11 de los 13 clones analizados contenían la secuencia esperada en la amplificación de una molécula de intrón escindido en forma *lariat*, es decir, la secuencia del final del intrón hasta la A desapareada del dVI unida al primer nucleótido del intrón (figura II.3C). Además, la A no se manifiesta en el producto de PCR como tal, sino como una timidina (T); se ha descrito que la RT no reconoce correctamente la A como consecuencia del enlace modificado que presenta, por lo que suele enfrentarla a otra A en lugar de a una T (Vogel *et al.*, 1997; Vogel y Bönen, 2002). Los dos clones restantes se corresponden con la amplificación de intrón circular. De acuerdo con la alta proporción de moléculas de 150 nt que se observa en el carril 10 de la figura II.3B respecto a la de 156 nt, las únicas formas que se secuenciaron fueron las derivadas de intrón escindido *lariat* (10 clones de 10). La conclusión más plausible de todos los datos expuestos hasta el momento es que RmInt1 genera en su escisión, al menos, dos tipos de moléculas: intrón *lariat* e intrón circular.

No nos olvidamos de secuenciar el producto de PCR de 156 bp obtenido en muestras de RNPs extraídas de cepas que expresan el intrón WT (pKG2.5) en ausencia de RT (figura II.3B, carril 7). Los seis clones analizados contuvieron la misma secuencia, equivalente a la identificación de la banda de 156 bp en el resto de muestras. Así pues, se demuestra la existencia de círculos de ADN que contienen el intrón, probablemente generados a partir de ARN circular.

#### II.1.1.4. ¿Contradicciones en los resultados?

Con todo, nuestros datos revelan una extraordinaria contradicción: la presencia de banda derivada de la amplificación de *lariat* en preparaciones de RNPs pero su ausencia en muestras de ARN total. Esto además pone en entredicho los resultados de extensión a partir de cebador en muestras de ARN, de donde se deduce que el intrón lleva a cabo, al menos, la primera reacción de maduración (probablemente por el mecanismo de *branching*). Nuestra primera actuación fue analizar qué hace a ambas muestras diferentes y la conclusión fue casi inmediata: mientras en que en la extracción de RNPs se trata de un enriquecimiento progresivo por sedimentación en gradiente de sacarosa, el adecuado aislamiento de ARN precisa de la eliminación del ADN pero también de las proteínas presentes en la preparación mediante fenolización. Debemos recordar que la estabilidad

del ARN escindido es mayor cuando cohabita con la proteína, por lo que parece lógico que el ARN *lariat* se encuentre mayoritariamente formando complejos con la IEP. Es probable que para otros fines, la disociación de estos complejos previa a la fenolización no sea importante, pero da la impresión de que el rendimiento final de *lariat* en las preparaciones de ARN total se ve dramáticamente afectado. En relación con esto se plantearon dos alternativas para la extracción de ARN total: una mediada por la digestión con proteinasa K previa a la fenolización y otra en la que se añade tiocianato de guanidina, un fuerte agente desnaturante de proteínas, a la mezcla de fenolización. El dibujo descrito hasta el momento derivado de los datos de RT-PCR cambió radicalmente (figura II.4). Mientras que en la muestra obtenida con el protocolo tradicional ( $\Phi$ ) sólo se observó la banda de mayor tamaño (156 nt) tras la electroforesis en poliacrilamida desnaturante, en las muestras extraídas mediante cualquiera de los otros dos protocolos (K, proteinasa K, y G, tiocianato de guanidina) pasamos a ver las dos bandas (150 y 156 nt). Claramente, la asunción que podríamos haber lanzado (las moléculas de intrón escindido circular es mayoritaria en las muestras de ARN) da paso a otra realidad bien distinta, demostrándose que el producto mayoritario tanto en ARN como en RNPs es el ARN escindido en forma *lariat*.



**Figura II.4.** Determinación de la influencia que tiene el protocolo de extracción del ARN en la recuperación de *lariat*. Para ello, ARNs extraídos según 3 protocolos diferentes se sometieron a RT-PCR.  $\Phi$ , representa el protocolo básico, mientras que K se corresponde con la adición de proteinasa K antes de la fenolización y en G hemos añadido tiocianato de guanidina a la mezcla de fenolización.

Este resultado también nos invita a la reflexión desde el punto de vista cuantitativo: cuando en la muestra sólo existe una banda susceptible de ser amplificada, se magnifica la señal obtenida cuando se compara con muestras en las que más de una especie molecular puede servir de molde a los cebadores (figura II.4, comparar la intensidad de banda de 156 nt en los carriles 2, 4 y 6). *A priori* la cantidad de intrón circular no debe haberse modificado y si lo ha hecho, debería ser en un sentido positivo similar al *lariat* (se han roto las posibles interacciones que estableciese con la IEP, por tanto, se mejora el rendimiento de recuperación en la muestra), sin embargo, la intensidad de la banda se ve disminuida. Otro indicio que nos pone de manifiesto esta modificación en los protocolos de extracción del ARN total, es que la eficiencia con la que la RT supera el enlace modificado es considerablemente baja: mientras que por extensión a partir de cebador somos capaces de detectar una cantidad aceptable de intrón escindido en muestras de ARN tratadas con fenol por el protocolo tradicional, la amplificación del ADNc a partir de esas mismas muestras no refleja este resultado (ausencia de la banda *lariat*). Aunque no se muestra, la intensidad de la banda de 97 nt en la reverso-transcripción con el oligonucleótido P es similar en los tres casos (ligeramente superior para las muestras

extraídas con los dos nuevos respecto al protocolo tradicional). Del conjunto de datos, deducimos que la síntesis de ADNc a partir de cebador refleja de manera más fiel que la RT-PCR (entre otras cosas porque no está mediada por amplificación) la eficiencia de la reacción de *splicing*, aunque no es lo suficientemente resolutiva. Por esto, los datos se complementan.

Por consiguiente, este tipo de aproximación nos está generando un cierto sesgo que hace imposible establecer la proporción absoluta de intrón escindido en forma circular respecto al intrón que madura formando *lariat*.

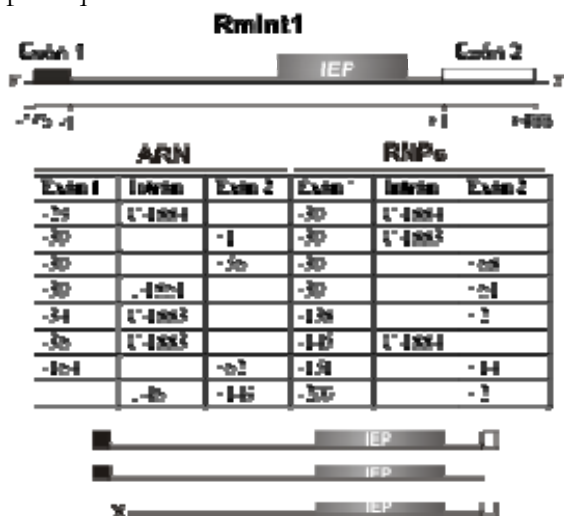
#### II.1.1.5. Identificación de moléculas de ARN generadas por cRT-PCR

Esta técnica es muy similar a la RT-PCR descrita con anterioridad pero está precedida de una reacción de ligación con T4 ARN ligasa (cRT-PCR). Esta ligación se realiza en unas determinadas condiciones de manera que se trata de producir circularización intramolecular, no ligación entre más de una molécula. El uso de esta técnica nos permitiría dar respuesta a dos cuestiones. La primera y original, era la detección de moléculas generadas como consecuencia de una reacción de maduración por hidrólisis; estas moléculas lineales se convertirían en moléculas circulares con un enlace 5'-3' (aparentemente más fáciles de sintetizar por la RT, con lo que serían muy abundantes en las muestras de PCR) que comparando su proporción respecto a muestras no tratadas con ARN ligasa, podríamos aseverar la presencia de moléculas lineales en nuestra preparación (Vogel y Börner, 2002). Por otro lado, podría servirnos para tratar de determinar el mecanismo por el cual se generan los círculos; en el apartado 1.2.3. de la introducción decíamos que el mecanismo propuesto para la formación de moléculas de ARN circular en el intrón de mitocondrias de levadura *ai5γ* implicaba la liberación inicial del exón 2, atacado por exón 1 libre presente en el medio (Murray *et al.*, 2001). En ambos casos, el estudio de la población de moléculas y de intermediarios de reacción puede proporcionarnos una idea de lo que está sucediendo *in vivo*.

Se usaron muestras de ARN y RNPs extraídas de bacterias que expresaban el intrón a partir de pKG2.5. Tras la amplificación fue imposible distinguir una banda clara en geles de agarosa porque nos aparecía una mancha continua a lo largo de todo el carril, de manera que decidimos no aislar de banda y clonar directamente todos los productos generados en el PCR. Bien es cierto que después del clonaje hubo una selección por tamaño y se eligieron aquellos clones que se encontraban en el rango aproximado del tipo de moléculas que se buscaban (insertos que en cualquier caso no excediesen de 400 bp). El análisis de 16 clones (8 procedentes de muestras de ARN y 8 de RNPs) puso de manifiesto la existencia de una mayoría de moléculas en las que el exón 1 se encontraba exactamente unido al final del intrón ( $C_{1884}$ ) o muy cerca del sitio de *splicing* 3' (al penúltimo nucleótido del intrón o al primer o segundo nucleótido del exón 2) (9 de los 16 clones analizados, ver tabla en figura II.5). Cinco de 16 contenían secuencia de ambos exones, por lo que perfectamente podría tratarse de moléculas de ARN precursor.



Finalmente, los dos clones restantes coincidieron con moléculas de intrón degradadas en el 5' y en el 3' (un ejemplo de cada, en una que se pierden 46 nt del extremo 5' del intrón y en la otra 22 nt del final del intrón). Además se observó una tendencia al procesamiento de nuestros precursores en torno a la posición -30 dentro del exón 1 lo que coincide con previas observaciones realizadas por extensión a partir de cebador (banda de 126-127 nt, ver figura II.2). Estos datos apoyan la sugerencia realizada por Murray y colaboradores que proponen que el primer paso en la formación de círculos podría ser la liberación del exón 2 (Murray *et al.*, 2001), puesto que observamos moléculas que carecen de este exón pero que aún conservan el exón 1.



**Figura II.5.** Detección de moléculas de ARN derivadas del procesamiento del precursor por cRT-PCR. En la parte superior se muestra una representación esquemática de la región que contiene el intrón con sus correspondientes exones en la construcción pKG2.5. En la tabla se detallan los resultados obtenidos del análisis de la secuenciación de clones obtenidos por cRT-PCR a partir de ARN y RNPs con intrón WT. Los nucleótidos en el intrón están señalados como el residuo y la posición que ocupa en la secuencia, así la C-1884 es el último nucleótido del intrón. En la parte inferior se ilustran los tipos moleculares encontrados en nuestras preparaciones de manera esquemática.

### II.1.1.6. Detección directa de ARN circular

El sesgo introducido por la RT a la hora de sintetizar el ADNc no nos permite establecer una comparación cuantitativa de la cantidad de cada una de las especies moleculares que se generan durante la maduración de RmInt1. Se han tratado de realizar estimaciones ayudándonos de protocolos de PCR semicuantitativa<sup>50</sup>, que tampoco permitió la eliminación del sesgo puesto que el problema se arrastra desde la reverso-transcripción (datos no mostrados). Lo lógico era pensar en técnicas que nos permitieran la detección directa de las moléculas circulares generadas y, a poder ser, que nos sirviesen para compararlas con las moléculas *lariat* presentes en nuestras muestras de ARN o RNPs. Se pensó en una amplia variedad de alternativas (todas detalladas en el material y métodos apartado 12), pero lo cierto es que ninguna proporcionó un resultado concluyente. Todas apuntaban a que la proporción de círculos en nuestras preparaciones podía considerarse

<sup>50</sup> Se trata de analizar de manera exhaustiva los productos generados desde ciclos iniciales del PCR, esto es, retirar muestras cada pocos ciclos y no permitir que la reacción se sature. La aparición y cantidad relativa nos permitiría estimar la cantidad inicial de cada muestra en la preparación. En nuestro caso, partimos del ciclo 12, se tomaron muestras cada 2 ciclos y se llegó hasta el ciclo 28. La muestra inicial, como en el caso de la RT-PCR ya descrita fue el ADNc sintetizado por la RT.

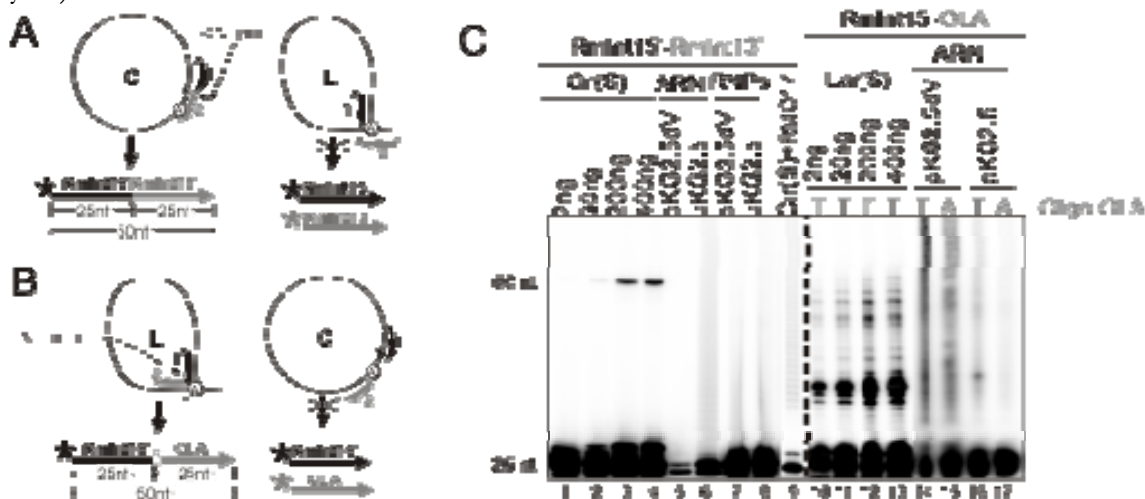
despreciable en comparación con la cantidad de *lariat* presente. A continuación haremos una breve descripción de las observaciones que podemos extrapolar con las diferentes técnicas utilizadas.

#### II.1.1.6.1. Ensayo de ligación de oligonucleótidos (OLA)

Por analogía con los experimentos llevados a cabo por Murray y colaboradores (Murray *et al.*, 2001) para la detección de moléculas circulares *in vitro*, se diseña un protocolo de ligación de oligonucleótidos. Esta técnica está basada en el uso de una pareja de oligonucleótidos monofosforilados que hibriden contiguos en una molécula de interés, de manera que la adición de una ADN ligasa generaría un nuevo producto que se correspondería con la unión de los dos oligonucleótidos (Landegren *et al.*, 1988). En una primera aproximación, nuestro principal objetivo fue la detección de moléculas circulares, por lo que el diseño experimental se centró en verificar la existencia de productos derivados de la ligación de oligonucleótidos que hibridan justo en el punto de circularización del intrón (figura II.6A). Los controles de reacción fueron transcritos *in vitro* que se corresponden con esta región de la molécula circular o con la unión de la molécula *lariat*, dependiendo del planteamiento. Para simplificar las posibles interpretaciones de los resultados, se usaron únicamente muestras de ARN y RNPs donde ya hemos confirmado por RT-PCR la existencia de ARN circular, pKG2.5, contrastadas con muestras de pKG2.5dV donde sabemos que no existe escisión (figura II.6C, panel izquierdo). Mientras los controles nos auguraban una sensibilidad del ensayo en torno a los 2 ng de ARN transcrito *in vitro* (figura II.6C, carril 1), en nuestras muestras apareció un bandeo difuminado a lo largo de todo el carril donde fue imposible discriminar la aparición de una banda mayoritaria (figura II.6C, carriles 6 y 8). Una posible explicación al resultado podría ser que las muestras contienen algún contaminante que impide que la reacción se lleve a cabo, pero no tiene mucho sentido si tenemos en cuenta que los ARNs están resuspendidos en agua. Otra alternativa es que se trate de muestras muy complejas que dificulten enormemente el proceso como consecuencia de hibridaciones inespecíficas de los oligonucleótidos, aunque la opción más plausible es que la proporción de moléculas de intrón escindido circular sea tan ínfima que no somos capaces de conseguir una reacción eficiente.

Por otro lado, aunque es posible que un enlace modificado 5'-2' no pueda servir como puente para que tenga lugar la ligación (Murray *et al.*, 2001), intentamos el mismo protocolo (variando uno de los oligonucleótidos) para la detección de *lariat*, una forma que suponemos más abundante (figura II.6B). Como no estábamos seguros de la pauta de reconocimiento/hibridación de la A desapareada como consecuencia del enlace, se usaron dos oligonucleótidos que variaban su posición 5' entre T y A con el fin de abarcar las dos posibilidades. La eficiencia de ligación en muestras control fue mucho menor a la que se obtuvo para el circular, además de ir acompañada de una serie de bandas inespecíficas y mayoritarias, particularmente cuando usábamos el oligonucleótido OLA T

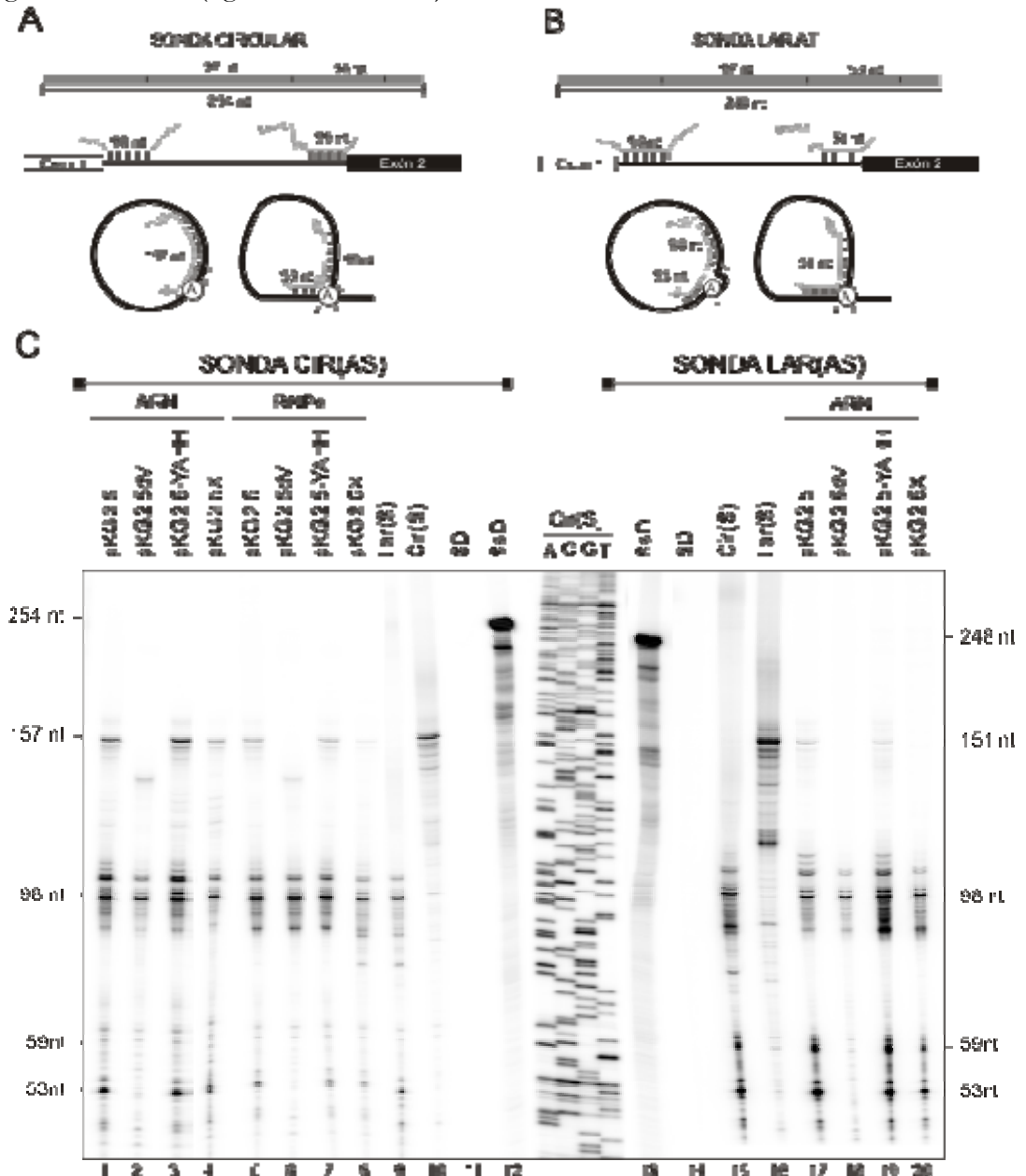
(figura II.6C, carriles 10 al 13). En cuanto a las muestras problema se repitió lo que ya observamos en los experimentos sobre circular, una mancha a lo largo del carril que no permite la discriminación de la banda del tamaño esperado, 50 nt (figura II.6C, carriles 16 y 17).



**Figura II.6.** Detección de las moléculas de escisión de RmInt1 mediante ligación de oligonucleótidos (OLA). (A) Diseño experimental para la detección de círculos en la que se usaron dos oligonucleótidos (RmInt15'/RmInt13') de 25 nt que al unirse generarían un producto de 50 nt; en ningún caso, deberían ligarse en la molécula *lariat*. (B) Diseño de los oligonucleótidos para que se ligan usando como molde la molécula en lazo. Uno de los oligonucleótidos coincide con el esquema en A (RmInt15'); sin embargo, en el extremo 3' del intrón se han diseñado dos oligonucleótidos pensando en el reconocimiento erróneo de la A en el *lariat* observada en los ensayos de RT-PCR. Ambos oligonucleótidos (OLA) carecen de la cola del *lariat* pero en su extremo 5' puede contener una A o una T. (C) Resultado del ensayo con los dos abordajes experimentales descritos en A y B. El panel izquierdo se corresponde con la detección de moléculas circulares en la que se usaron muestras de ARN y RNPs de construcciones con intrón silvestre y mutado en la ribozima. El panel derecho representa el ensayo con ARN de intrón silvestre y mutado testado con las dos parejas de oligonucleótidos. En ambas aproximaciones se usaron transcritos *in vitro* como controles de reacción (Cir(S) y Lar(S)).

En estas condiciones no se puede extraer un resultado concluyente. Desconocemos el tipo de enlace que une los extremos del intrón en la molécula circular, de manera que si es un enlace 5'-2' podría suponer un problema para la efectividad de la T4 ADN ligasa. Una alternativa a tener en cuenta podría ser el diseño de los oligonucleótidos de manera que se evite el punto de ligación (enlace modificado) como puente donde actúe la ligasa. Sin embargo, se abandona este tipo de aproximación porque consideramos que es muy probable que el principal escollo que nos estamos encontrando es la tremenda complejidad de nuestras muestras (gran número de especies de ARN, mayoritariamente ARNr), lo que impediría la mejora del rendimiento de la ligación de oligonucleótidos. Esto viene refrendado por el hecho de que, cuando se lleva a cabo la ligación usando como molde transcritos *in vitro* control circular en presencia de ARN total extraído de *S.*

*meliloti* RMO17 (recordar que esta cepa carece de intrón en su genoma), la reacción de ligación se inhibe (figura II.6C, carril 9).



**Figura II.7.** Detección de las moléculas de escisión de *RmInt1* gracias a protección frente a digestión con ARNasas A y T1 (RPA). (A y B) Naturaleza de la sonda usada y patrón de hibridación sobre las distintas moléculas que contienen intrón, según si el ARN sonda se diseña para proteger el círculo (A) o el *lariat* (B). (C) SsD representa la sonda sin digerir mientras en los carriles SD sufre digestión sin previa hibridación. Se usaron preparaciones de ARN y RNPs de distintas construcciones que contienen intrón WT o mutado. Como controles se usaron ARNs transcritos *in vitro* y el tamaño de las bandas viene marcado por una escalera de secuenciación con ddNTPs a partir del cebador LL, y un molde que representa la unión de la molécula circular.

### II.1.1.6.2. Ensayos de protección frente a digestión con ARNasas (RPA)

Otra aproximación posible para conseguir discriminar entre moléculas de intrón circular e intrón *lariat* en nuestras preparaciones es la técnica RPA. Se fundamenta en la eliminación de ARN de cadena simple presente en nuestra muestra una vez hibridado con otro ARN sonda, complementario a las regiones que queremos proteger. Para ello se suelen utilizar una mezcla de endoribonucleasas con distintas especificidades de sustrato: ARNasa A produce cortes tras pirimidinas y ARNasa T1 que corta en el extremo 3' de guaninas (Vogel *et al.*, 1997). En principio, la mejor ARN sonda para la detección y posterior cuantificación de todas las especies que este protocolo nos revela se corresponde con la región complementaria del amplificado de 156 bp de la molécula circular, CIR(AS) (comparar tamaño de productos entre paneles A y B en figura II.7). La sonda marcada sintetizada *in vitro* es algo mayor (254 nt) como consecuencia de las secuencias arrastradas del vector. El uso de esta sonda nos generaría 3 fragmentos que nos identifican cada una de las formas de intrón: como precursor, 59 nt (además de otro fragmento de 98 nt que tiene un tamaño muy similar a una de las bandas que aparecen en la molécula *lariat*); como *lariat*, 53 nt (los 6 nt que forman la cola no hibridarían); y como circular, 157 nt (figura II.7A). Como indicadores del tamaño esperado para algunas de las bandas se usaron ARN transcritos *in vitro* control que se corresponden con la secuencia de la región de ligación de la molécula de intrón circular y *lariat*. La primera conclusión que se puede extraer de los resultados es que el protocolo se está desarrollando en unas condiciones adecuadas pues la digestión de la sonda cuando no se encuentra hibridada es completa (figura II.7C, carriles 12 y 11). Si atendemos a los resultados obtenidos para los transcritos control circular, podemos observar la presencia de una banda mayoritaria del tamaño esperado acompañada de otra serie de bandas de menor intensidad que pueden corresponderse con degradación (o síntesis) parcial de nuestra sonda (figura II.7C, carril 10). Sin embargo, en el ARN control *lariat*, la sonda parece haberse digerido generando varias bandas entre las que se distingue una mayoritaria de aproximadamente 98 nt. La identificación de la otra banda que debe generarse resultó más complicada, puesto que el tamaño que nos indica la secuencia no se corresponde con el de la otra banda mayoritaria que aparece en el carril (figura II.7C, carril 9). A la altura esperada aparece una banda de escasa intensidad que no se encuentra en relación molar con la banda de 98 nt, por lo que es difícil señalar sin lugar a dudas la banda inferior de origen en el *lariat* (57 nt)<sup>51</sup>. El requisito para discriminar entre bandas pasaba por la comparación entre ARNs problema

---

<sup>51</sup> Las bandas en los controles no coinciden exactamente con las bandas que deben aparecer en las muestras, son ligeramente más grandes como consecuencia de la identidad de las secuencias flanqueantes (la sonda y el ARN control proceden del mismo inserto en orientación inversa respecto al promotor). Concretamente, en los controles esperábamos encontrar una única banda de ~163 nt en el carril donde el ensayamos el ARN transcrito *in vitro* procedente de unión circular, mientras que en el de las moléculas *lariat* debía dividirse en dos fragmentos: uno de ~100 nt correspondiente a la región 5' de la molécula de sonda y el otro de ~57 nt.

en los que el intrón estuviera presente y otros en los que no. Por ello, se usaron muestras de ARN y RNPs procedentes de cepas en las que hemos conseguido detectar *lariat* (pKG2.5 y pKG2.5-YAHH, carriles 1, 3, 5 y 7), y los comparamos con otras en las que no hay escisión (pKG2.5dV y pKG2.5X, carriles 2, 4, 6 y 8). El análisis exhaustivo de las bandas no nos permite extraer ninguna conclusión. En cuanto a la banda derivada de la protección de la forma circular, las diferencias parecen más patentes en el caso de las RNPs, aunque nos llama la atención la presencia en ambos tipos de muestras de una banda a esa altura en la construcción pKG2.5X y ligeramente más pequeña en el caso de pKG2.5dV. En concreto, hemos podido determinar que esta banda de mayor tamaño en pKG2.5dV se identifica precisamente con la no protección en la región del dV donde se localiza la mutación CGA que impide el *splicing* a este mutante.

Todos estos detalles nos llevan a pensar que, de nuevo, la heterogeneidad de moléculas existentes en nuestra preparación enmascara los resultados esperables. Con intención de comprobar si se trataba de un problema particular del planteamiento o de la sonda que habíamos elegido, intentamos el mismo tipo de ensayo pero desde la perspectiva contraria: sintetizar la sonda a partir de un molde que se correspondiese con la secuencia complementaria de la molécula *lariat*, LAR(AS). Lamentablemente, la imagen obtenida fue similar (figura II.7C, carriles del 13 al 20), muchas bandas inespecíficas y, lo que es más importante, la aparición de bandas de protección en mutantes en los que otras técnicas nos mostraron diferente resultado.

Tras muchos intentos de mejorar la técnica, sin resultado alguno, se pensó en la secuenciación de las bandas para definir su identidad. Se usaron muestras procedentes de hibridación con sonda CIR(AS), se aislaron de gel las bandas mayoritarias y se trataron de secuenciar. La secuenciación reversa por el método de Sanger (terminación con ddNTPs) generó muchas complicaciones en unos casos por la falta de ARN molde para que la RT pudiese actuar y cuando lo hacía, su ineficiencia a la hora de incorporar ddNTPs sólo nos permitió distinguir un producto de extensión final que se corresponde con el tamaño de la banda aislada. En el caso de las bandas de menor tamaño se observó un alto grado de degradación o de mezcla con otras especies de tamaño similar. Basándonos en esta observación desistimos de una secuenciación química por el método de Maxam-Gilbert (es muy probable que nos muestre mucho ruido de fondo y nos impida la correcta interpretación de la lectura de bandas). En el caso de la banda derivada del circular (157 nt), disponemos de los oligonucleótidos P y LL que nos permiten la RT-PCR y posterior secuenciación de la banda de protección. Efectivamente, se confirma la naturaleza de esta banda que protege muestras de ARN procedentes de cepas que expresan las construcciones pKG2.5 y pKG2.5X. Otras técnicas considerablemente sensibles como la RT-PCR ponen en duda este resultado, por lo que consideramos que esta aproximación tampoco nos va a ayudar a la identificación de las formas circulares que existen en nuestras preparaciones.

### II.1.1.6.3. Extensión de cebador modificada por la adición de un ddNTP

Nuestra siguiente opción fue tratar de establecer diferencias entre los dos tipos moleculares en función de su secuencia. En teoría, esto puede conseguirse gracias a una técnica que se conoce con el nombre de extensión de cebador envenenada (PPE, Poisoned Primer Extension). Se trata de la síntesis de un ADN complementario a nuestra molécula a partir de un cebador, con la particularidad de que la extensión se para gracias a la adición en la mezcla de reacción de un ddNTP. Dado que el oligonucleótido elegido para la reverso-transcripción hibrida en todos tipos moleculares que investigamos (ARN precursor, intrón *lariat* e intrón circular), podremos cuantificar la cantidad de cada una de esas moléculas presentes en nuestra preparación en función de la longitud del fragmento generado después de la extensión en presencia de un determinado ddNTP. Debemos elegir meticulosamente el ddNTP para que no nos coincida el tamaño de los ADNc en función de la secuencia de nuestras moléculas y de la posición de partida de nuestra reverso-transcripción. Por otro lado, los productos de extensión pueden identificarse marcando la incorporación de un determinado  $[\alpha\text{-}^{32}\text{P}]\text{dNTP}$  durante la polimerización o gracias al marcaje en 5' del cebador previo a la extensión, la elección de un sistema u otro va a depender de las circunstancias del ensayo<sup>52</sup>. Es importante resaltar que utilizaremos una RT que sea capaz de superar el enlace modificado 5'-2' que, al menos, sabemos que existe en el *lariat*. En la figura II.8A y B se muestra un ejemplo de las combinaciones posibles de cebador, ddNTP y técnica de marcaje. La extensión del oligonucleótido LCP1, que hibrida a cinco nucleótidos del inicio del intrón, en presencia de ddTTP y  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$  debe generar, al menos, 3 fragmentos: uno de 36 nt que se corresponde con el precursor, otro de 39 nt que equivale a la molécula circular y un tercero de 49 nt correspondiente a la extensión del *lariat*. En el caso que el enlace en la molécula circular sea 5'-2', o que la RT tenga dificultades para sobrepasar el enlace modificado en el *lariat*, ambas moléculas darían lugar a un producto cuya extensión sería de 35 nt que coinciden con el inicio del intrón (figura II.8A). Nuevamente como controles de reacción se han usado ARNs transcritos *in vitro* que reproducen la unión circular [Cir(S)] y/o *lariat* [Lar(S)] pero en los que no existe ningún enlace modificado. El análisis de estos controles nos reveló, en primer lugar, que la técnica funciona puesto que aparecen productos del tamaño esperado. Sin embargo, también podemos observar bandas de mayor peso molecular que se corresponden con nuevas paradas por la incorporación de ddTTP en virtud de la secuencia (figura II.8B, carriles 1 y 2). Esto sólo puede indicar dos cosas, o que la RT prefiere introducir cualquier otro dNTP en lugar de ddTTP (algo que ya habíamos observado con anterioridad), o que la pureza de la preparación de ddTTP que proporciona la casa comercial (99,7%) no es tal. Con este nuevo hándicap, la interpretación de las señales en las muestras problema se simplifica. Si atendemos a los

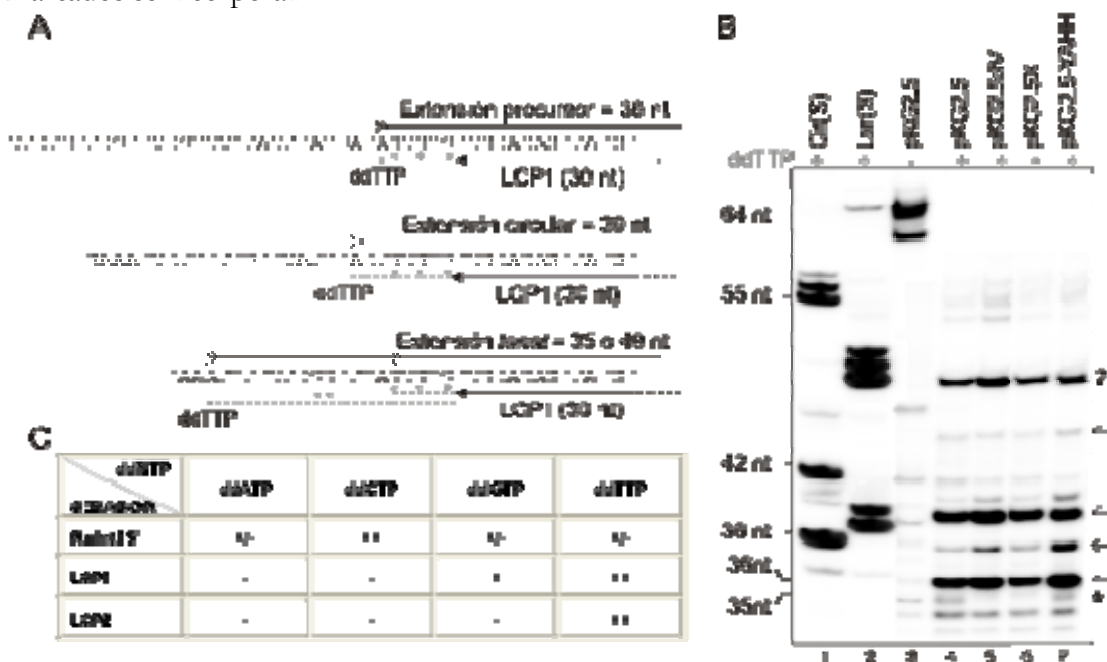
---

<sup>52</sup> La primera opción tiene la desventaja de no ser cuantitativa porque depende del número de nucleótidos marcados que se incorporen, pero se evita la enorme mancha del oligonucleótido marcado, lo que es muy importante si nuestra banda queda muy cerca del tamaño del cebador.

productos de extensión que se generan en ausencia de ddTTP podemos observar una serie de bandas que podrían corresponderse con moléculas en distinto grado de procesamiento (figura II.8B, carril 3). De entre todas quizá sería conveniente resaltar dos: la de mayor tamaño molecular que se corresponde con una de las formas precursoras mayoritarias identificadas mediante otras técnicas (extensión de cebador con el oligonucleótido P o cRT-PCR) y una banda de 35 nt marcada con un asterisco que, como después veremos, podría corresponderse con la parada en el enlace modificado. A grandes rasgos, el patrón de bandas en las muestras de ARN total extraídas de bacterias que portan los plásmidos pKG2.5, pKG2.5dV, pKG2.5X y pKG2.5-YAHH se mostró similar (figura II.8B, carriles del 4 al 7). Las bandas señaladas con una flecha en el lateral se identificaron por tamaño con moléculas de ARN precursor. Destacar la presencia de una banda marcada con una (?) que es dependiente de la presencia de ddTTP y que muestra intensidad parecida en todos los carriles; probablemente derive de una hibridación inespecífica del oligonucleótido que no hemos sido capaces de determinar. Lamentablemente, el que no tenga lugar la parada de la síntesis del ADNc desde el precursor nos enmascara la detección de la extensión en moléculas circulares (se produce la coincidencia en la banda de 38 nt). Sin embargo, observamos algo que ya vaticinábamos: en muestras en las sabemos que se produce escisión nos aparece un producto de 35 nt que está ausente en los dos mutantes *splicing* (figura II.8B, carriles 4 y 7 frente a 5 y 6, banda marcada con \*). Como antes adelantamos, al coincidir además con una banda independiente de adición de ddTTP, probablemente se trate de la parada en el enlace 5'-2' presente en el *lariat* (desconocemos si también en el circular porque, aunque se especula que puede tratarse de un enlace de este tipo, aún no se ha demostrado experimentalmente). Curiosamente, no se identificó ninguna banda consecuencia de la incorporación de ddTTP por la extensión de las moléculas de *lariat* más allá del enlace modificado. Este dato resulta cuanto menos extraño si atendemos a la intensidad de banda derivada de intrón escindido respecto de precursores en extensiones de cebador realizadas con el oligonucleótido P (ver figura II.2B, carril 1) y la relación de intensidades obtenida en este ensayo (comparar con figura II.8B, carril 3, o aún más dramático con carril 4). Lo esperable sería, en el caso de que la RT no sea capaz de leer a través del enlace, la misma relación de intensidades pero, si lo atravesase, debíamos haber conseguido identificar la banda. Así pues, debemos asumir que la RT no está superando el enlace con la eficiencia deseada, de manera que somos incapaces de detectar esos eventos. Esto debe ir asociado con algún otro fenómeno como por ejemplo impedimentos estéricos que no permitan a la RT iniciar la síntesis del ADNc tan próxima al punto de unión. Se han usado varios posibles cebadores para intentar descartar problemas con la hibridación, pero lo cierto es que la propia naturaleza del ensayo nos limita enormemente el diseño de los mismos. Por otro lado, se han usado combinaciones alternativas de ddNTPs para legitimar los resultados (figura II.8C). También se ha usado varias RT para evitar los inconvenientes intrínsecos de cada una de ellas (la escasa procesividad o la baja eficiencia en la incorporación de ddNTPs) pero el resultado fue



similar en todos los casos aunque la SuperScript II® se reveló como una de las más eficientes (datos no mostrados). En relación con la diferencia de intensidades, puede deberse en parte al tipo de marcaje: cuanto más grande es la molécula, más nucleótidos marcados se incorporan.



**Figura II.8.** Detección de los productos de la escisión de RmInt1 *in vivo* por extensión de cebador envenenada (PPE). (A) Ejemplo de una de las posibles combinaciones de las que hemos usado para el diseño experimental. El cebador (LCP1) se indica con una flecha negra y el producto de extensión se corresponde con la línea discontinua gris. Sobre la secuencia la longitud de los productos de extensión esperados con la adición de ddTTP. Los asteriscos marcan las posibles incorporaciones de  $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$  en el producto de extensión. (B) Detalle de los productos obtenidos en el ensayo señalado en A. Se usaron muestras de ARN procedentes de bacterias que expresan el intrón WT pero también varios derivados que contienen mutaciones en la IEP o la ribozima, a las que se adicionó (+) o no (-) ddTTP. Como controles se usaron ARNs transcritos *in vitro*. Los tamaños de los productos se señalan en el lateral izquierdo del gel. Las señales en el lado derecho se corresponden con: moléculas derivadas de precursor ( $\leftarrow$ ), parada en el enlace modificado 2'-5' (\*) y productos de extensión de origen desconocido (?). (C) Todas las combinaciones que se han usado con más o menos éxito: tres oligonucleótidos en la misma región y los 4 ddNTP. Con (-) nos referimos a combinaciones no factibles para la diferenciación de las diferentes moléculas (por supuesto, no se han ensayado); (+/-) significa que permite la discriminación entre dos de los 3 tipos de moléculas; con (++) señalamos las mejores combinaciones para ensayar porque generan la mayor diferencia de tamaño entre productos, mientras que (+) dan productos de tamaño similar.

De este ensayo podemos deducir que la RT pasa muy ineficientemente el enlace modificado 5'-2'. Por otro lado, no hemos conseguido detectar extensión del producto circular en ninguna de las combinaciones ensayadas por lo que es probable que su cantidad sea extremadamente reducida. La otra posibilidad que explicaría este hecho es que el intrón circular también se encuentre ligado por un enlace 5'-2'.

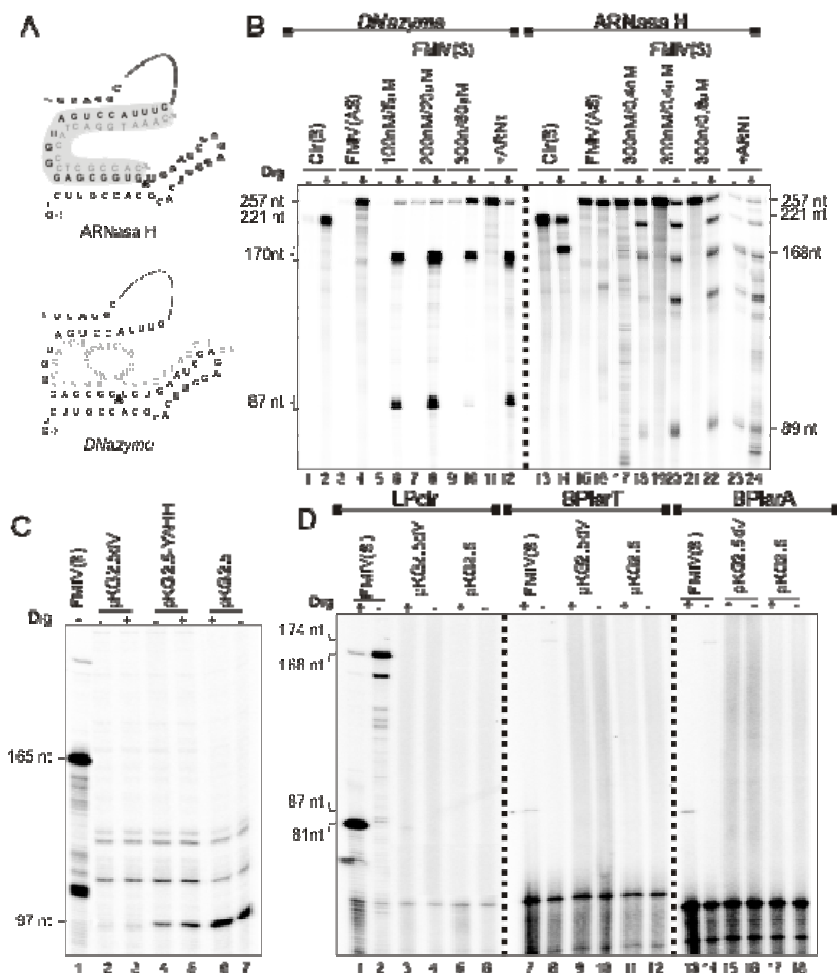
#### II.1.1.6.4. Digestión mediante el uso de ARNasa H o de DNazymes seguida de extensión de cebador

Tomando como referencia el ensayo anterior, quisimos verificar las conclusiones con un nuevo procedimiento de detección directa de moléculas circulares. En esta ocasión la síntesis del ADNc a partir del cebador P va precedida de la digestión de las moléculas que contienen intrón, entre ellas las moléculas *lariat* y circular. Según los datos del experimento anterior, es probable que en el *lariat* no exista producto de extensión más allá del inicio del intrón (enlace modificado). Sin embargo, si observásemos la banda derivada de la extensión del circular hasta el punto de corte establecido, confirmaríamos por un lado, la cantidad de este tipo de moléculas presentes en nuestras preparaciones y, por otro, podríamos apuntar el tipo de enlace que tiene lugar en el intrón circular. En caso de que esta banda esté ausente, la conclusión sería similar a la del ensayo anterior.

La forma de generar un producto lineal a partir de la molécula circular, o en Y si atendemos al producto formado a partir del ARN *lariat*, fue la digestión en un punto concreto mediante el uso de una *DNazyme* tipo 10-23 o bien la digestión del ARN en regiones híbridas ADN-ARN mediante el uso de ARNasa H (ver material y métodos, apartado 12.3). Ambos oligonucleótidos se han diseñado para que la disrupción de las moléculas tenga lugar en la misma región del intrón, dentro del dominio V de la ribozima (figura II.9A). En esta ocasión, la técnica precisó de una puesta punto con transcritos *in vitro* marcados internamente para comprobar la eficiencia del corte. En estos ensayos se usaron diferentes condiciones, concretamente en las digestiones con *DNazyme* se variaron las concentraciones relativas de uno y otro componente (transcrito *in vitro* y oligonucleótido), la concentración de  $Mg^{2+}$  en el medio de reacción, así como la temperatura. Las reacciones con ARNasa H permiten menos versatilidad por lo que nos limitamos a cambiar las proporciones relativas de ambos componentes (figura II.9B). La conclusión más evidente de esta primera fase fue que la digestión de los híbridos es bastante más ineficiente que la actuación de la *DNazyme* (aparición de bandas inespecíficas y baja eficiencia de corte), por lo que se descartó este abordaje para sucesivos ensayos. Aunque no se llegó a conseguir una eficiencia de corte del 100% con la *DNazyme*, las condiciones óptimas seleccionadas fueron 20  $\mu$ M de oligonucleótido, 100 mM  $MgCl_2$  y 2 h. a 37°C (figura II.9B, carril 8). La adición de ARN total procedente de *S. meliloti* RMO17 no alteraba la capacidad de corte de los transcritos *in vitro* (figura II.9B, carriles 12 y 24).

El siguiente paso fue comprobar qué ocurría si realizábamos el mismo tipo de digestión (*DNazyme*) en nuestras muestras problema y luego eran sometidas a extensión de cebador. Un transcrito circular sintetizado *in vitro* nos sirvió como control de reacción; se puede observar que, tras la digestión, se genera un producto de 165 nt que nos marca el punto donde debemos esperar nuestra banda derivada de intrón circular (figura II.9C, carril 1). Ratificando resultados anteriores, la RT no supera el enlace 5'-2' en el *lariat* y da lugar a una banda de 97 nt además de las consabidas bandas de precursor;

independientemente de si las muestras se han sometido o no a digestión el resultado es idéntico (figura II.9C, comparar carriles 4/5 y 7/6). En consonancia con los datos extrapolables de todas las técnicas de detección directa de moléculas circulares ensayadas hasta el momento, no se consigue detectar ninguna molécula derivada de intrón circular. Podemos dar dos explicaciones recurrentes en nuestra argumentación, o hay muy poco o la unión de los extremos 5' y 3' del intrón se produce por un enlace 5'-2'.



**Figura II.9.** Ensayos de extensión de cebador previa digestión con ARNasa H o una DNazyme en la región del dominio V. **(A)** Esquema del diseño tanto del ADN oligonucleótido que nos servirá para la digestión con ARNasa H así como de la DNazyme 10-23 (ADN con capacidad catalítica). Las flechas marcan el punto de digestión en el caso de la DNazyme o el final de la digestión de la ARNasa H. **(B)** Puesta a punto del corte y selección de estrategia mediante el uso de ARNs marcados con  $[\alpha\text{-}^{32}\text{P}]\text{UTP}$  sintetizados *in vitro*: Cir(S) no contiene la secuencia a ser cortada aunque representa la unión de la molécula de intrón circular; FMIV(S) es 37

nt más larga que Cir(S) y contiene la secuencia que se corta; finalmente, FMIV(AS) genera el transcrito complementario a FMIV(S) y no debe cortarse. Los carriles se marcan con la concentración del transcrito (T) frente a la de oligonucleótido (O) en reacción (T/O). En los laterales se marca el tamaño de los productos obtenidos tras la digestión; 257 nt es la longitud del transcrito FMIV(S) sin digerir. **(C)** Extensión de ADNc a partir del cebador P usando como moldes ARNs de distintas construcciones previamente digeridas con la DNazyme. El control para la extensión del ADNc en una molécula circular fue un transcrito *in vitro*, FMIV(S). **(D)** Ensayo similar al presentado en C pero situando el cebador justo encima del punto de ligación en la molécula circular (LPcir) o en el lariat (BPlarT o BPlarA dependiendo del nucleótido que se enfrenta a la A con el enlace modificado). Los signos + y - sobre los geles significan muestras sometidas o no a digestión, respectivamente.

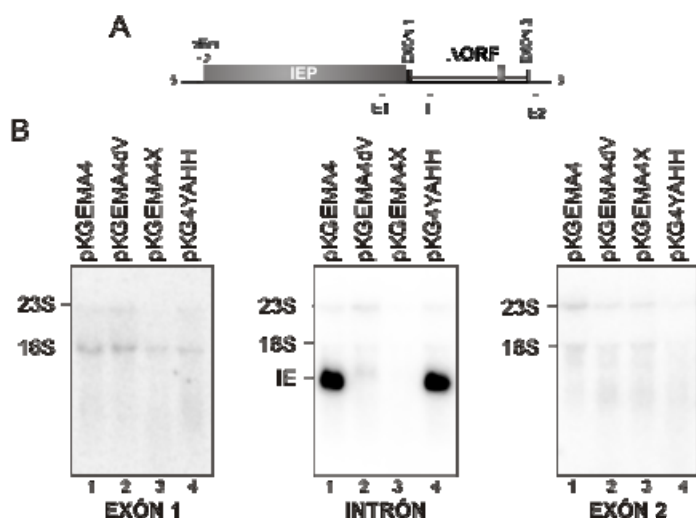
Si el problema es el enlace modificado, ¿podríamos soslayarlo? La respuesta es afirmativa pero sólo si se consigue el diseño de una serie de cebadores que hibriden en el punto de ligación de las moléculas circular y *lariat*, pero que sean lo suficientemente específicos para no dar problemas de hibridación cruzada. Nuevamente, queda patente la fiabilidad del ensayo si atendemos al resultado en los transcritos control (figura II.9D, carriles 1/2, 7/8 y 13/14) pero no tenemos una explicación para lo que acontece en nuestras preparaciones. El resultado más visual es que probablemente el oligonucleótido diseñado sobre el punto de ligación en la molécula circular, LPcir, hibrida de una manera más eficiente que cualquiera de los otros dos localizados en el punto de unión de la molécula *lariat* (BPlar). Independientemente de este hecho, no se han conseguido detectar bandas específicas en nuestras muestras problema, por lo que esta modificación del ensayo no ha solucionado nuestro problema inicial.

#### II.1.1.6.5. Northern blot

En nuestro afán por cuantificar la proporción relativa de productos de escisión, recurrimos a una nueva aproximación basada en la hibridación de una sonda de ADN sobre una membrana a la que habremos transferido el ARN total extraído de bacterias que portan el intrón silvestre en construcciones pKGEMA4 o los diferentes intrones mutados (Northern blot). Se utilizaron este tipo de construcciones porque el intrón escindido es de menor tamaño pero también porque la secuencias de los exones que lo flanquean reducen las posibilidades de hibridación cruzada<sup>53</sup>. Para la separación electroforética de las muestras se utilizaron tanto geles de agarosa desnaturizante (1,85%(v/v) formaldehído) como geles de poliacrilamida desnaturizante (8M urea). En la hibridación se usaron varias sondas para la identificación de los diferentes productos de escisión así como de los posibles intermediarios de las reacciones y las moléculas de precursor, aunque lamentablemente no se obtuvieron los resultados esperados (figura II.10A). Los geles de agarosa no se mostraron lo suficientemente resolutivos para la identificación de los productos de escisión pero revelaron que sólo íbamos a ser capaces de detectar productos de un tamaño muy definido derivados de la escisión de intrón. Con esto queremos hacer referencia a la aparición de una banda específica únicamente en construcciones que expresaban un intrón con capacidad de escisión, pKGEMA4 o pKG4YAAH (figura II.10B, panel central, carriles 1 y 4), mientras que no fuimos capaces de apreciar la banda en construcciones carentes de *splicing*, pKGEMA4x o pKGEMA4dV

<sup>53</sup> Nuestra experiencia con la manipulación y revelado de productos *in vitro* obtenidos a partir del intrón completo (1884 nt) aconsejan el uso de estas construcciones (740 nt) pues de lo contrario corremos el riesgo de que la banda no sea capaz de entrar en el gel. Por otro lado, los exones se corresponden con la secuencia de inserción *ISRm2011-2* de la que se encuentran hasta 12 copias en el genoma de *S. meliloti* RMO17. En las construcciones pKG2.5 la longitud de estos exones es de -175/+466, pero en construcciones pKGEMA4 se reduce a -20/+5 de manera que la elección de una sonda fuera de esta región pero todavía muy próxima al intrón es más factible.

(carriles 2 y 3). Además cuando hibridamos esas membranas con sondas específicas de cada uno de los exones el resultado fue la aparición de una pareja de bandas inespecíficas, pero en ningún caso detectamos, como cabría esperar, moléculas de precursor y mucho menos de exones ligados o libres (figura II.10B, paneles laterales); un análisis posterior nos permitió determinar que esas bandas inespecíficas se correspondían con los ARNs ribosómicos de mayor tamaño (23S y 16S). Del exón 2 se usaron incluso dos sondas distintas ante la posibilidad de que se encontrase procesado, pero el resultado fue el mismo en los dos ensayos (resultado no mostrado). Una posible explicación para estos resultados podría ser: en primer lugar, la amplia variedad de moléculas procesadas en su extremo 5' que hemos encontrado en los diferentes ensayos llevados a cabo (extensión de cebador y cRT-PCR e incluso alguno más que veremos más adelante, 5'-RACE); por otro lado, en nuestras construcciones no existe ningún tipo de terminadores de transcripción, de modo que la longitud de las moléculas precursoras no está perfectamente definida.

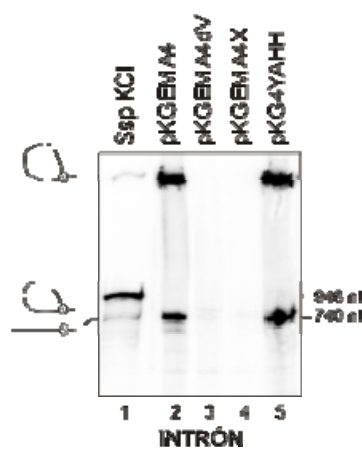


**Figura II.10.** Detección de los productos derivados de RmInt1 presentes en muestras de ARN mediante Northern blot. (A) Esquema que representa la distribución proporcionada de las diferentes partes del intrón en construcciones pKGEMA4. Sobre la secuencia se indica la posición de las sondas utilizadas en los distintos ensayos: E1 significa que se corresponde con el exón 1, I con intrón y E2 con el exón 2. (B) Resultado obtenido al hibridar con cada una de las sondas indicadas arriba cuando la separación

electroforética se llevó a cabo en geles de agarosa desnaturizante. Se utilizaron ARNs de cepas que expresan intrón con capacidad de madurar y otros en los que no se escinde. Se señala la posición de migración de los ARNr 23S y 16S, así como la banda correspondiente a las formas escindidas del intrón (IE). Aunque no se muestra, los resultados en RNPs fueron similares.

A la vista de los resultados decidimos mejorar la capacidad resolutoria del ensayo gracias a la utilización de geles de poliacrilamida 6% desnaturizantes (8M urea) para la separación de las muestras. Por experimentos anteriores de *splicing in vitro* (Costa *et al.*, 2006a), sabemos que las moléculas en lazo, por su estructura más compleja, retrasan su migración respecto a las moléculas lineales. Sin embargo, parece ser que las moléculas circulares deberían co-migrar con las moléculas *lariat*; al menos, esto es lo que se desprende de los estudios de Tabak y colaboradores que afirman que cuando la cola del *lariat* es corta (en nuestro caso es de 6 nt) es muy complicado separarlo de las moléculas circulares de similar tamaño molecular (Tabak *et al.*, 1988). En cualquier caso, tomando como referencia los productos de maduración generados en una reacción *in vitro* a partir de un

transcrito precursor que contiene un intrón de la misma longitud que el de nuestras construcciones (figura II.11, carril 1), pudimos distinguir la presencia de dos bandas en aquellos mutantes que nos habían mostrado previamente su capacidad de maduración (carriles 2 y 5). La banda de mayor tamaño migraba a la misma altura que la molécula *lariat* en la reacción de *splicing in vitro*, mientras que la banda que migró más rápido lo hizo en paralelo con el producto de escisión lineal en las preparaciones *in vitro*. De nuevo, estas bandas estuvieron ausentes en las muestras de ARN total que expresaban un intrón defectuoso en la capacidad de maduración, pKGEMA4x o pKGEMA4dV (carriles 3 y 4). Si asumimos que la banda de intrón circular migró junto al intrón escindido en lazo, claramente no podremos cuantificar las proporciones, pero hemos descubierto la presencia en nuestras preparaciones de moléculas con un tamaño preciso que migran a la altura del intrón escindido lineal. Es la primera evidencia de este tipo que se tiene en referencia a la escisión *in vivo* del intrón RmInt1.



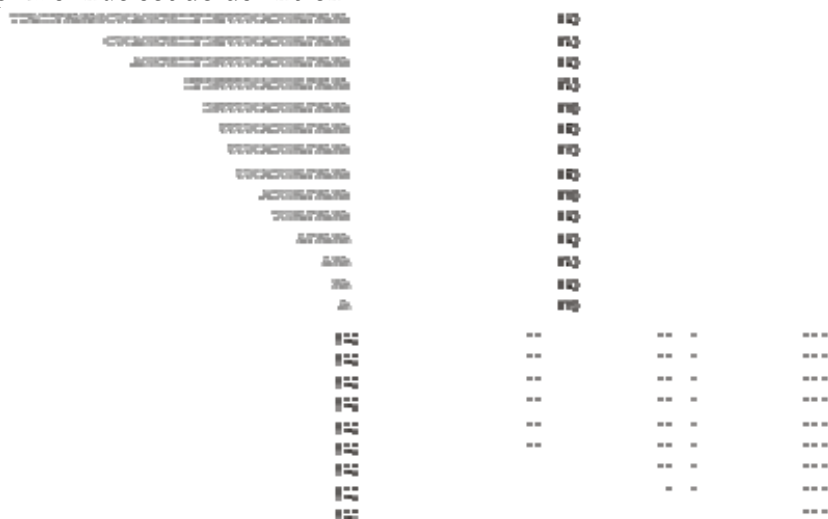
**Figura II.11.** Búsqueda de mayor resolución en los ensayos de Northern blot. De la misma manera que se hizo en agarosa, se utilizaron geles de poliacrilamida al 6% desnaturalizante para separar en paralelo muestras de ARN en las que el intrón se encuentra activo y otras en las que no es capaz de madurar. Como marcador, además de uno comercial que no se muestra, se utilizó una reacción de *splicing in vitro* realizada según las condiciones descritas en Costa *et al.*, 2006a. El tamaño de algunas de las moléculas se marca al margen. A la izquierda se puede ver una ilustración que se corresponde con la identificación teórica de los productos resultantes de la hibridación con una sonda específica del intrón. La hibridación también se hizo con una sonda del E2 pero no se vio ninguna banda destacable (dato no mostrado). La distorsión de las bandas se debe a que migran junto a una de las especies moleculares más abundantes en las preparaciones de ARN

total, los ARNr de mayor tamaño (23S y 16S).

#### II.1.1.6.5.1. ¿Son realmente moléculas lineales?

En los últimos tiempos, nuestro grupo de investigación ha usado con bastante éxito la técnica del 5'-RACE (amplificación rápida de extremos 5' de ADNc) para la caracterización de pequeños ARNs reguladores, riborreguladores (del Val *et al.*, 2007). Así nos planteamos su aplicación para la identificación del extremo 5' de esas supuestas moléculas lineales (ver fundamentos en material y métodos). Se habla en estos términos pues podría tratarse de moléculas de intrón en lazo degradadas, lo que se conoce como *broken lariat*. Para tratar de apoyar nuestra primera hipótesis (que se trata de moléculas lineales), entre otras precauciones, los clones se seleccionaron de acuerdo al tamaño del inserto que contenían sus plásmidos; así, restringimos nuestro análisis a amplificados con no más de 150 bp que se corresponde con el tamaño esperado para la amplificación del adaptador 5' y los primeros 97 nt de nuestro intrón. Se analizaron 228 clones por PCR de colonia de los que se seleccionaron 44 para enviar a secuenciar. Se detectaron una amplia variedad de moléculas de precursor, así como moléculas resultantes de degradación

(figura II.12), pero no fuimos capaces de encontrar ningún clon cuya secuencia se iniciase con el primer nucleótido del intrón.



**Figura II.12.** Secuencias obtenidas de los análisis de 5'RACE. En gris resaltado como negrita se señalan las secuencias de los exones presentes en los clones y en negro normal la secuencia del intrón. Entre paréntesis en negrita al lado de cada secuencia se especifica el número de clones secuenciados que se corresponden con cada tipo de inserto. En la parte inferior podemos ver clones en los que el intrón se encuentra procesado en mayor o menor grado.

En vista de la imposibilidad de determinar el extremo 5' por las presencia de moléculas precursoras de un tamaño muy similar, se pensó una aproximación semejante (3'-RACE) para el mapeo del extremo 3' de nuestras moléculas (ver detalles en material y métodos, apartado 12.8). En ninguna de las construcciones ensayadas pudimos adivinar la presencia de la banda del tamaño esperado por lo que no nos pareció oportuno proceder con la secuenciación. Tanto los resultados obtenidos con el 5'-RACE como la ausencia de ellos en el 3'-RACE llamó nuestra atención, pues no podíamos explicar la aparición de las bandas de migración lineal detectadas por *Northern blot*. Así que nos planteamos la posibilidad de que estas moléculas lineales no mostrasen en sus extremos el 5'-fosfato y el 3'-OH esperados, sino que se tratase, como se ha descrito para otras ribozimas, de extremos modificados 5'-OH y 2'-3'-fosfato cíclico. La presencia de cualquiera de estas modificaciones impediría la ligación del adaptador en 5' o 3', de manera que no podríamos poner de manifiesto la presencia de intrón escindido mediante el uso de la técnica de RACE convencional. Entonces, nuestros esfuerzos se dirigieron al diseño de protocolos de RACE alternativos con el fin de poder detectar ambos tipos de extremos modificados. Para el mapeo de un posible 5'-OH se plantearon dos alternativas: fosforilación de las preparaciones de ARN total previa a la ligación o hacer el 5'-RACE usando como material de partida el ADNc que se correspondiese con la región 5' de RmInt1. La única posibilidad que contemplamos para la modificación del protocolo de 3'-RACE fue utilizar T4 polinucleótido kinasa, que según datos del fabricante a pH ácido desfosforila extremos 3'. En ninguna de las modificaciones descritas se obtuvo resultado

alguno. Aunque no podemos descartar la posibilidad de que se estén generando moléculas lineales en nuestras preparaciones los datos de que disponemos hasta el momento, parecen apuntar que las bandas de menor tamaño derivan de la rotura de moléculas en lazo. Otra posibilidad es que se correspondan con la linearización de las moléculas circulares.

Así pues, si unificamos las conclusiones que los distintos experimentos de detección directa nos han permitido extraer podemos sugerir que la cantidad de moléculas de intrón escindido en forma circular es mínima respecto a otras especies moleculares de intrón (precursor o *lariat*) y que sólo una técnica que permita su amplificación nos posibilita su identificación. Por otro lado, cabe la posibilidad de que hayamos sido capaces de detectar moléculas de intrón escindido en forma lineal (aunque requiere de posterior comprobación), siendo la primera vez que se describe la capacidad de maduración *in vivo* por hidrólisis del intrón RmInt1.

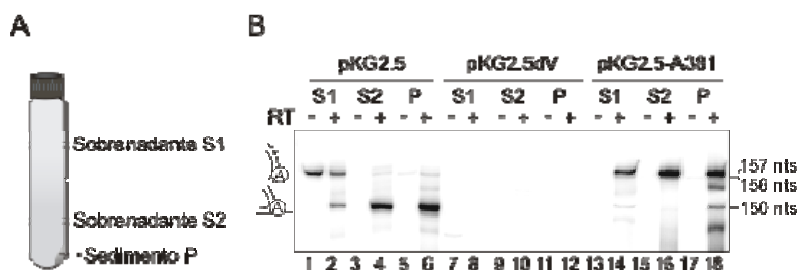
#### II.1.2. LA ACTIVIDAD MADURASA DE LA IEP DE RmInt1 ES NECESARIA EN LA ESCISIÓN TANTO *LARIAT* COMO CIRCULAR.

En las IEPs de intrones del grupo II podemos distinguir varias actividades desarrolladas por distintos dominios y, entre ellas, nos interesa destacar la asistencia en la maduración *in vivo* que presenta el dominio X. Quisimos verificar este hecho en RmInt1, además de comprobar el grado de implicación de este dominio en la escisión del intrón como círculos. Para ello, creamos el mutante YY→AA en el que se sustituyeron dos tirosinas muy conservadas en todos los dominios madurasa (Y<sub>354</sub> e Y<sub>355</sub>, ver figura III.4) por dos residuos alanina. A éste se le sumó otro mutante en el mismo dominio en el que un residuo lisina también conservado (K<sub>381</sub>, ver figura III.4) se cambió nuevamente por alanina. Los estudios de extensión del cebador, RT-PCR y *Northern blot* llevados a cabo con preparaciones de ARN y RNPs del mutante YY→AA mostraron niveles de *splicing* indetectables (figura II.13A, carriles 2 y 5; C, carriles 4 y 8; D, carril 3). Para contrastar el dato obtenido con RNPs extraídas de bacterias que contienen el plásmido pKG2.5-A354A355, se ensayó su actividad RT *in vitro* mediante el uso de un sustrato poli(rA)/oligo(dT)<sub>18</sub>; como cabía suponer a raíz del fenotipo observado en *splicing*, estas RNPs tenían la actividad RT exógena fuertemente inhibida, a niveles similares al mutante YAAH en el dominio activo de la RT. Podríamos decir pues, que este mutante no forma RNPs activas, resultado coherente con su incapacidad de invasión de nuevas dianas (dato mostrado en el capítulo de resultados siguiente, figura III.8). Todas estas observaciones confirman que el dominio X de RmInt1 se requiere para el *splicing* del intrón *in vivo*, así como para su escisión en forma circular.





cuando se comparaba con los niveles detectados en preparaciones de ARN y RNPs de intrón silvestre (figura II.13C, comparar carriles 4/6 y 10/12). A esto debemos sumarle la observación de que las moléculas circulares amplificadas en este mutante parecían contener un nucleótido adicional respecto al WT, lo que se deduce de la diferencia de movilidad del producto de amplificación del círculo. Este hecho fue confirmado gracias a la secuenciación del producto de RT-PCR que nos mostraba la presencia de una citosina justo en el punto de ligación de los dos extremos del intrón. La aparición de bandas en el Northern blot coincidió con las observaciones hechas por extensión de cebador, es decir, este mutante mostró una reducción (no evaluada) de la capacidad de *splicing* que se tradujo en un descenso de la intensidad de banda de migración más lenta (figura II.13D, carril 4). Aunque es difícil de precisar, la intensidad de la banda inferior no parece verse afectada. Quisimos comprobar si la mayor intensidad de banda derivada de circular que observábamos en el mutante K381A por RT-PCR podría deberse a que la mutación estaba afectando la afinidad de unión de la IEP al intrón escindido *lariat*. Ante la imposibilidad, hasta el momento, para demostrar esta hipótesis partiendo de los componentes purificados *in vitro*, se realizó el análisis mediante RT-PCR de las distintas fracciones resultantes del proceso de enriquecimiento en RNPs para comprobar posibles variaciones en la proporción *lariat*/circular a lo largo del gradiente de sacarosa. En este proceso, además del sedimento de RNPs (P), nos encontramos dos fracciones de sobrenadante con diferente densidad (S1 y S2) (figura II.14A). El estudio de los productos de RT-PCR en muestras con intrón silvestre (pKG2.5, figura II.14B), puso de manifiesto que el ARN *lariat* era el producto mayoritario tanto en la fracción enriquecida en RNPs (P, carril 6) como en el sobrenadante más denso (S2, carril 4), mientras que los niveles de ambos amplificados estaban balanceados en la fracción de sobrenadante más laxa (S1, carril 2). Este análisis contrasta con el patrón de distribución de bandas en el mutante pKG2.5-K381A, en el que el producto de amplificación mayoritario en todas las fracciones (P, S1 y S2; figura II.14B, carriles 14, 16 y 18) fue el derivado de moléculas circulares, reproduciéndose además la aparición del nucleótido extra. Únicamente en la fracción P, pudimos detectar niveles reducidos de amplificación de la molécula *lariat* (carril 18). Así pues, ya que no hemos encontrado *lariat* (no unido a la proteína) en las fracciones de sobrenadante del mutante, cabe afirmar que se está produciendo un descenso en la producción de *lariat* asociada con una mutación en un residuo lisina conservado en el dominio X de la IEP de RmInt1. En paralelo, existe un incremento en la intensidad de banda de amplificación derivada del circular, sobre todo entre el sobrenadante más laxo y el resto de fracciones (carril 14 en comparación con 16 y 18), lo que podría estar indicando que el intrón circular también se asocia a la IEP (probablemente en menor medida que el *lariat*).



**Figura II.14.** Detección mediante RT-PCR de las formas de intrón escindido presentes en las distintas fracciones obtenidas durante el protocolo de extracción de RNPs. **(A)** Diagrama esquemático de las tres fracciones ensayadas: los dos sobrenadantes, laxo (S1) y denso (S2), y el sedimento (P). **(B)** Se muestran los productos resultantes de la RT-PCR del intrón silvestre (construcciones pKG2.5) en comparación con el mutante en el dominio madurasa (pKG2.5-A381). Como control ensayamos en paralelo el mutante en el dominio V de la ribozima (pKG2.5dV). Se muestra el resultado del PCR precedido (+) o no (-) de RT. Los tamaños así como su equivalencia en un diagrama se muestran en los laterales.

**Nota adicional:** Todos los ensayos anteriores se han realizado en paralelo con construcciones pKGEMA4 (versiones  $\Delta$ ORF con exones -20/+5 y la IEP expresada en cis aguas arriba de la ribozima) y mutantes derivados (pKGEMA4X, pKGEMA4dV, pKG4YAHH, pKG4A381 y pKG4A354A355), salvo el Northern blot que sólo se ha llevado a cabo con construcciones  $\Delta$ ORF. No se ha considerado el mostrar los resultados puesto que vienen a confirmar las conclusiones arrojadas por las versiones pKG2.5.

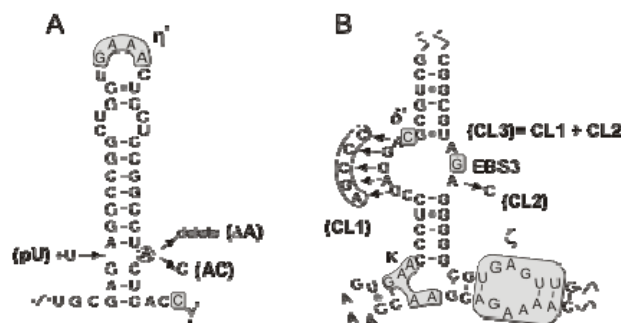
### II.1.3. ESTRUCTURAS DE LA RIBOZIMA QUE INTERVIENEN EN LA ESCISIÓN DE RmInt1 *IN VIVO*

Hasta el momento, hemos podido demostrar que RmInt1 se escinde dando lugar a moléculas *lariat* pero también a moléculas circulares. Para tratar de determinar el papel biológico de estas formas circulares nos pareció relevante obtener mutantes en los que la vía de transesterificación esté completamente inhibida, de manera que una cantidad mínima de *lariat* no pudiera enmascarar las conclusiones resultantes. Por ello, se pensó inicialmente en el estudio de mutantes en la adenosina desapareada (A) del dVI de la ribozima. El papel de este residuo en *splicing* ha sido ampliamente estudiado *in vivo* e *in vitro* en intrones mitocondriales de levaduras (fundamentalmente en ai5y), comprobándose que determinadas mutaciones que implican a esta A impiden la formación de *lariat*, aunque no parecen afectar a la escisión del intrón por la ruta hidrolítica (Liu *et al.*, 1997; Chu *et al.*, 1998; Podar *et al.*, 1998b; Chu *et al.*, 2001). Los datos *in vitro* mostraban que la sustitución de la A por G o U disminuye la capacidad de formación de *lariat* en ~100 veces, en cambio, la sustitución de este residuo por una

citosina impide que tenga lugar la primera transesterificación en reacciones de *splicing* en *cis* (Liu *et al.*, 1997). Se observó este mismo efecto cuando la adenosina era deletada ( $\Delta A$ ) (Schmelzer y Mueller, 1987; Chin y Pyle, 1995; Chu *et al.*, 1998) o cambiada por un deoxinucleótido (dA o dG) (Liu *et al.*, 1997). Estos resultados coinciden con observaciones filogenéticas que demuestran que la A puede encontrarse en la naturaleza cambiada a G o U (aunque no se ha determinado el mecanismo de *splicing in vivo* en estos intrones) (Michel *et al.*, 1989). Por otro lado, mutaciones en las que la A se encontraba apareada mostraron un comportamiento diferencial en función de fuerza de la interacción: pA-G, pA-A y pA-C mostraban niveles de *branching* silvestres (Chu *et al.*, 1998), mientras el apareamiento con U (pA-U) redujo drásticamente la cantidad de *lariat* detectado (van der Veen *et al.*, 1987a; Schmelzer y Mueller, 1987; Chu *et al.*, 1998). En general, los datos obtenidos *in vivo* refrendan los resultados *in vitro* aunque un mutante pA-G, resultado de la reversión *in vivo* de la mutación pA-U, manifestaba un ligero defecto en la segunda reacción de transesterificación que se tradujo en la acumulación de intrón escindido *lariat* unido al exón 2 (Podar *et al.*, 1998b). Estudios realizados en intrones del grupo II en cloroplastos de cebada, muestran que la delección o ausencia de la A en el sitio de *branching* da lugar a escisión lineal *in vivo* pero no son capaces de la detección de moléculas de intrón *lariat* (Vogel y Börner, 2002).

A pesar del papel esencial que esta A desapareada tiene en la reacción de maduración en lazo (Schmelzer y Schweyen, 1986; Chin y Pyle, 1995), se ha investigado poco su capacidad para establecer interacciones de tipo funcional o catalítico con otras regiones del intrón. Los datos más significativos se han obtenido recientemente *in vitro* combinando datos procedentes de entrecruzamiento químico por proximidad (*UV-crosslinking*) con el análisis de mutaciones puntuales en el intrón ai5 $\gamma$  (Hamill y Pyle, 2006). Se han investigado cuatro mutantes de los cuales tres tienen un efecto significativo en la maduración de ai5 $\gamma$  *in vitro* en una reacción de *trans-splicing*. El mutante puntual A370G mostró la inhibición más suave (~20% WT), seguido de un mutante que afecta a los residuos 223 a 228 en el lado 5' de la región de coordinación (UUAUAA $\rightarrow$ AGGCC) que presentó una inhibición del *branching* en torno a 35 veces respecto al intrón silvestre. El efecto más dramático se obtuvo en el doble mutante, de manera que la formación de *lariat* se redujo hasta alcanzar unos niveles 130 veces por debajo del WT. La conclusión extraída a partir de todos estos datos parece ser que el sitio de *branching* en el dVI, incluidos los dos apareamientos no canónicos G•U que flanquean la A, pudiera interaccionar con la región de coordinación identificada en el dI de la ribozima con una polaridad específica. Esta hipótesis se apoya además en la importancia que siempre se ha concedido a la distancia entre los dominios V y VI del intrón (Boulanger *et al.*, 1996; Podar y Perlman, 1999; Chu *et al.*, 2001), ya que la separación de esta región de coordinación con respecto al receptor del dV (definido por la interacción  $\kappa$ - $\kappa'$ ) es un tallo corto y rígido. Así, ambos receptores podrían funcionar en conjunto para orientar todos los residuos catalíticos hacia el centro de reacción (Zhang y Doudna, 2002).

Con estos antecedentes, decidimos elegir aquellas mutaciones cuyo efecto fue más drástico sobre la primera reacción de transesterificación en la ruta de *branching*. Según este criterio, se llevaron a cabo tres mutaciones en la A desapareada del dVI (figura II.15A). Una de ellas supuso la sustitución de este residuo por una C (AC); otra consistió en la delección de la A ( $\Delta A$ ); y una tercera que resultó del apareamiento de esta A765 con un residuo U que introdujimos en la posición 736 de la ribozima (pU). En lo que concierne a la región de coordinación, también se construyeron tres mutantes que implicaron a residuos en uno, otro o ambos lados de esta región desapareada (Figura II.15B). El mutante en el lado 5', denominado CL1, contienen cambios a lo largo de los residuos 222 al 226, de manera que la secuencia GAGGA presente en el intrón silvestre ha sido sustituida por AGCCC en nuestro mutante. En el lado 3', con menor número de nucleótidos desapareados, se construyó un mutante puntual en la posición 330 en el que la A original fue cambiada por una C (mutante CL2). El tercer mutante, CL3, contenía las dos mutaciones simultáneamente, lo que nos permitiría establecer el papel cooperativo de ambas regiones de la región desapareada en el reconocimiento de la A. Por supuesto, no se alteraron las interacciones  $\delta$ - $\delta'$ , importante para el apareamiento de bases entre el exón 5' y el intrón, y EBS3-IBS3, esencial para una eficiente ligación de los exones (Costa *et al.*, 2000) ( $\delta$  se encuentra en la posición 227 y EBS3 se localiza en la 329). Al tratarse de la misma interacción, los mutantes en la A desapareada y los obtenidos en la región de coordinación deberían mostrar similares características. Todos estos mutantes deben tener un fenotipo similar en relación a la reacción de *branching*: la producción de *lariat* debe estar disminuida según los resultados descritos en ai5 $\gamma$ .



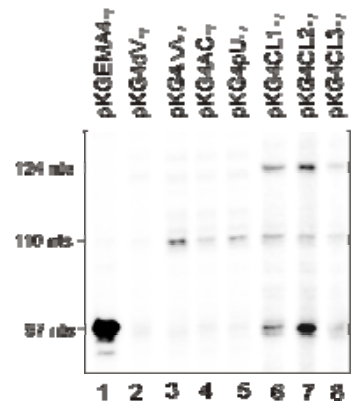
**Figura II.15.** Estudio por mutagénesis de las regiones implicadas en la escisión de RmInt1. (A) Diagrama del dominio VI de la ribozima donde se representan el conjunto de mutantes analizados: delección de la A desapareada ( $\Delta A$ ), transversión de A a C (AC) y apareamiento de la A con un U (pU). (B) Se representa la región de coordinación y los mutantes diseñados en ella: mutante en el lado 5' del bucle (CL1), en el 3' (CL2) y el

doble mutante (CL3). La parte inferior de este esquema es lo que se considera el receptor del dominio V. En letras griegas, resaltadas como zonas sombreadas, se marcan las interacciones terciarias descritas en la zona.

Los mutantes de la A desapareada y la región de coordinación se construyeron en el sistema pKGEMA4<sub>T7</sub> con el fin de conseguir la máxima sensibilidad a la hora de la evaluación de su capacidad de movimiento (ver descripción en material y métodos, apartados). Brevemente decir que son construcciones en las que la IEP se expresa aguas arriba del ARN del intrón en versión  $\Delta ORF$  flanqueado por exones -20/+5.

### II.1.3.1. Determinación de la eficiencia de escisión *in vivo* de los mutantes de RmInt1 mediante extensión de cebador

En una primera aproximación para la caracterización de los productos de escisión, se puso de manifiesto la eficiencia de la primera reacción de *splicing* mediante la técnica de extensión a partir de cebador descrita previamente (ver apartado II.1.1.1.). Así, los ensayos de extensión de cebador llevados a cabo con ARN total de *S. meliloti* RMO17 que expresa el intrón silvestre en pKGEMA4<sub>17</sub>, pusieron de manifiesto un producto de extensión mayoritario de 97 nt que se corresponde con el tamaño esperado para el intrón escindido. Esta banda no se detectó en el ARN de células que portaban la construcción con la ribozima mutada en el dV, pKG4dV<sub>17</sub> (figura II.16, carriles 1 y 2). Sin embargo, el patrón de bandas en las construcciones mutadas en el dVI o el dI fue bastante diferente al de pKGEMA4<sub>17</sub> y dependía de la región afectada. Como apreciación general, podemos decir que se observó una mayor disminución de la escisión en los mutantes en el sitio de *branching* que en los de la región de coordinación. Aunque se detectaban productos de extensión de mayor longitud derivados de precursores no escindidos, no se pudo apreciar ninguna banda derivada de intrón escindido en los mutantes de la A del dVI (carriles 3, 4 y 5), si bien más tiempo de exposición nos reveló la presencia de bandas de intensidad débil en la región donde se esperaba el producto de 97 nt (dato no mostrado). Por otro lado, en los mutantes del dI se pudieron identificar tanto bandas derivadas de la presencia de ARNs precursores de intrón no escindido como el producto de extensión del intrón maduro (carriles 6, 7 y 8). La reducción menor en la eficiencia de la reacción (~12% del WT) se observó en el mutante puntual que afectaba la región derecha del bucle, pKG4CL2<sub>17</sub>, seguido del mutante en lado 5' de esta región desapareada, pKG4CL1<sub>17</sub>, cuya reducción alcanzó el 5% respecto de la ribozima silvestre. Como cabría esperar de un efecto acumulativo de los dos cambios anteriores, la intensidad de la banda de escisión en el mutante pKG4CL3<sub>17</sub> fue casi inapreciable. Por tanto, de estos datos podemos concluir que la región de coordinación juega algún papel en la escisión del intrón. Se puede considerar que ambos lados de la región desapareada están implicados pero el efecto es más patente cuando se muta la parte izquierda del bucle, quizá por la mayor extensión de la región afectada por la mutación. En función de los datos previos obtenidos con mutantes en la A desapareada de otros intrones del grupo II, podemos concluir que nuestras mutaciones corroboran los efectos observados

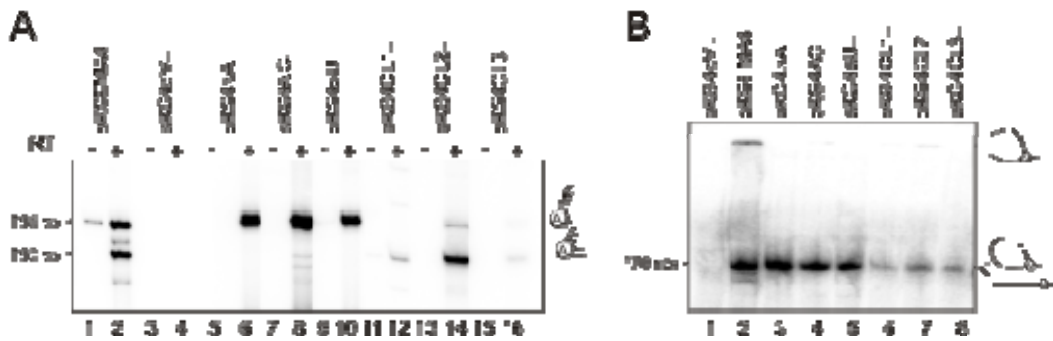


**Figura II.16.** Estudio de la escisión de RmInt1 *in vivo* por extensión de cebador. El molde usado fue ARN total extraído de *S. meliloti* RMO17 que porta los plásmidos señalados y el cebador, el oligonucleótido P. Se indica el ADNc de 97 nt derivado de intrón escindido, así como bandas de migración más lenta que se corresponden con precursores no maduros.

por el resto de investigadores, esto es, que determinados cambios que afectan a la A del dVI derivan en una fuerte reducción de la eficiencia de maduración. Además, basándonos en la descripción de productos dada para la escisión de esos otros intrones, estaríamos en condiciones de afirmar que los ensayos de extensión de cebador ponen de manifiesto mayoritariamente los productos de *splicing* originados por la ruta de *branching*, es decir, intrón escindido *lariat*.

### II.1.3.2. Naturaleza de los productos generados en la escisión *in vivo* de los mutantes de RmInt1

Con anterioridad hemos demostrado que RmInt1 madura principalmente dando lugar a un ARN *lariat* aunque un ensayo de RT-PCR nos permite la detección de moléculas de escisión circular. Recurrimos a esta técnica para tratar de determinar la identidad de los productos generados en el caso de los mutantes que nos ocupan (figura II.17A). De manera similar a lo que ya describimos para el ARN derivado de la construcción pKG2.5, el análisis del perfil de bandas obtenido con ARN extraído de células que expresan el intrón silvestre a partir de construcciones pKGEMA4<sub>T7</sub>, nos reveló la presencia de dos bandas, una de 156 nt asociada con la amplificación de la molécula de intrón circular y otra de 150 nt derivada del ARN *lariat* (carril 2). Ambas bandas estaban ausentes en el mutante control de *splicing* pKG4dV<sub>T7</sub> (carril 4). De nuevo, la región mutada marcó el diferencial de productos obtenidos. Mientras los mutantes del dVI se escindieron dando lugar a intrón circular como producto mayoritario, los mutantes de la región de coordinación apenas si generaron intrón circular aunque en estos tres mutantes pudo identificarse el producto de amplificación del ARN *lariat*. Si consideramos que la existencia de dos productos de PCR puede convertirse en un proceso competitivo por los cebadores, la mayor intensidad de la banda de migración lenta en los mutantes de la A desapareada puede considerarse un artefacto de la técnica. De modo que podríamos pensar que la cantidad de ARN intrónico circular generado en estos mutantes no se altera respecto a las condiciones WT (figura II.17A, carriles 6, 8 y 10). En consonancia con la presunción anterior, las banda derivada de *lariat* no se detecta en los mutantes  $\Delta A$  y pU (carriles 6 y 10), aunque un mayor tiempo de exposición de los geles nos permitió identificar un producto de amplificación que migraba justamente a 150 nt en el mutante AC (carril 8). Si atendemos a las mutaciones en el dI, podemos observar que el mutante CL2 produjo formas en lazo pero sólo trazas de intrón circular (carril 14); mientras, los dos mutantes restantes mostraron escisión como intrón *lariat* en concordancia con las proporciones relativas observadas en los ensayos de extensión de cebador, pero no fuimos capaces de detectar amplificados derivados de la escisión como círculos (carriles 12 y 16). El hecho de que la RT-PCR no permita la cuantificación de los productos nos impide precisar el grado de implicación de cada región en los distintos procesos de escisión.



**Figura II.17.** Caracterización de los productos de escisión *in vivo* de los mutantes en el dVI y el dI de RmInt1. **(A)** Determinación de la presencia o ausencia de formas circulares y en lazo por RT-PCR. La amplificación se realizó, con (+) o sin (-) RT previa, en muestras de ARN extraídas de bacterias que expresan el intrón silvestre (pKGEA4<sub>T7</sub>), mutado en el dV de la ribozima (pKG4dV<sub>T7</sub>), así como en los mutantes de los dominios I y VI. El tamaño de las bandas se marca en el lateral junto a un esquema de los amplificados donde la A desapareada se representa dentro de un círculo. **(B)** Análisis directo de los productos de escisión del intrón mediante Northern blot. La separación de las muestras de ARN se llevó a cabo en poliacrilamida desnaturalizante; tras la electrotransferencia, la membrana fue hibridada con una sonda específica de intrón. Aunque no se muestra, nos servimos de marcadores comerciales y una reacción de *splicing in vitro* para la identificación de los diferentes productos. Los supuestos productos se esquematizan en el lado derecho de la autorradiografía.

Nuestros estudios para la caracterización de los productos de escisión, se completaron con datos de Northern blot. Para simplificar trataremos los mutantes en dos bloques. Si comenzamos con el análisis de los productos detectados en los mutantes de la A desapareada, podemos concluir que en el caso de los mutantes  $\Delta A$  y pU no somos capaces de detectar más que la banda de mayor migración lo que estaría de acuerdo con observaciones realizadas en el caso de otros intrones del grupo II ( $\alpha 5\gamma$ ) en los que se ha descrito la existencia de productos generados por una primera reacción de hidrólisis durante el *splicing in vivo* (Podar *et al.*, 1998b). Llama nuestra atención que en estos mutantes en los que los datos de RT-PCR revelan la total ausencia de *lariat* pero la presencia de cantidades WT de moléculas circulares, no se aprecie ninguna banda de movilidad retrasada (recordar que el intrón circular debería migrar en paralelo al *lariat*). Esta observación podría ser la confirmación de que la cantidad de moléculas circulares presentes en nuestras preparaciones se encuentra por debajo del límite de detección de la hibridación y que la RT-PCR está sobreestimando la cantidad de círculos. Por su parte, de acuerdo a la pequeña proporción de intrón en lazo que se vislumbró en los ensayos de RT-PCR, en el mutante AC se observa una tenue banda de migración retrasada. También se observaba una correlación entre el patrón de bandas de Northern blot y los datos de extensión de cebador y RT-PCR con respecto a la banda de mayor tamaño en los mutantes del dI: la intensidad de la banda fue mayor en CL2, seguida de CL1 y CL3. En este caso, llama nuestra atención que la intensidad de banda de migración rápida es menor con respecto a los mutantes del dVI.

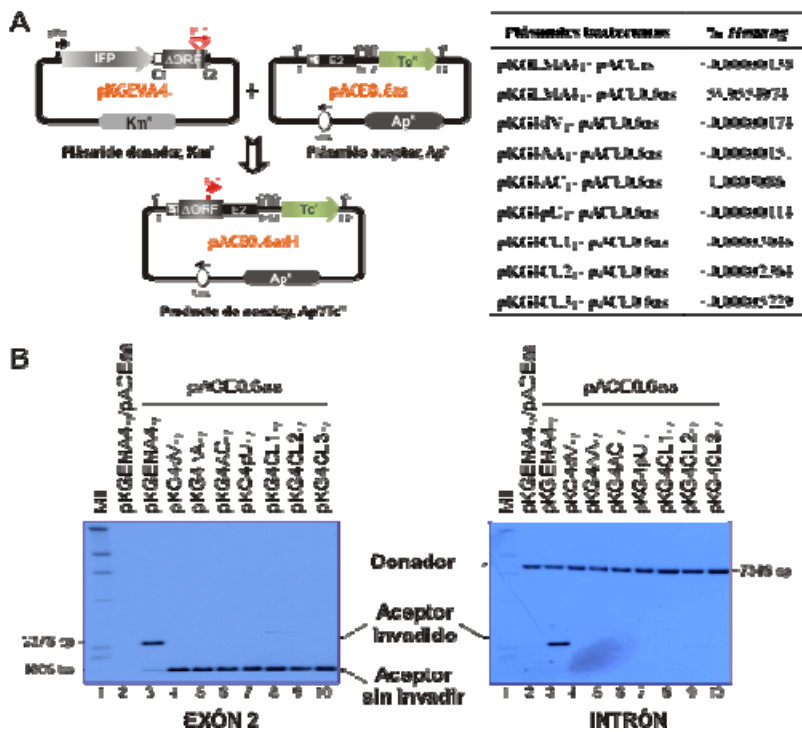


A la vista de los resultados, parece claro que mientras la adenosina en el sitio de *branching* es casi esencial para que el intrón se escinda en forma *lariat*, la alteración de esta región del dominio VI no parece afectar a la formación de intrón circular ni, en el caso de que finalmente se determine, tampoco a la escisión lineal. Por el contrario, aunque queda patente el papel de la región de coordinación en la primera reacción de transesterificación, su implicación en la formación de moléculas escindidas en forma circular, inclusive de supuestas moléculas lineales, parece mucho más crítica. La impresión general que parece extraerse de todos nuestros resultados podría resumirse diciendo que mientras los mutantes en el dVI sólo parecen ser importantes para la ruta de *branching*, una distorsión de la región de coordinación provoca una alteración de la maduración del intrón por cualquiera de sus vías.

### II.1.3.3. Movilidad de los mutantes en la ribozima

La capacidad de *retrobombing in vivo* de estos mutantes fue evaluada gracias al uso de un sistema de doble plásmido que se basa en la activación de la resistencia a tetraciclina ( $Tc^R$ ) cuando se produce un evento de salto del intrón sobre su secuencia diana (Karberg *et al.*, 2001; Barrientos-Durán, 2008). El sistema consiste en un plásmido donador (pKGEMA4<sub>T7</sub> o alguno de los mutantes) que contiene clonado en su dIV el promotor para la ARN polimerasa del fago T7. Por otro lado, el plásmido aceptor presenta el sitio de inserción del intrón rodeado por varios terminadores, todo seguido de un gen de  $Tc^R$  carente de promotor; el inserto se cierra con un terminador para la T7 ARN polimerasa (figura II.18A). Ambos plásmidos se introducen en *S. meliloti* RMO17 mediante conjugación para luego evaluar la eficiencia de invasión en *E. coli* HMS174 (DE3); la ARN polimerasa de T7 no funciona en *S. meliloti* por lo que se aisló el conjunto de plásmidos de varias colonias del rizobio y se electrotransformó en la cepa de *E. coli* *recA* y que expresa la enzima. De este modo, cuando el intrón invade su diana se activa la transcripción desde el promotor que porta el  $\Delta ORF$  y se establece la proporción de colonias resistentes (diana invadida) con respecto a aquellas que contienen el plásmido aceptor, invadido o sin invadir, ( $Ap^R + Tc^R / Ap^R$ ). Como se muestra en la tabla de la figura II.18, la eficiencia de invasión evaluada para la construcción silvestre, pKGEMA4<sub>T7</sub>, rondó el 56%, mientras que si la cuantificación se realizaba con un plásmido donador mutado en la ribozima o un plásmido aceptor que carece de la diana para la inserción del intrón este porcentaje se estimó cercano a  $10^{-6}$ . A pesar de que algunos mutantes retuvieron la capacidad de producir *lariat*, considerada por algunos autores como la única forma de intrón con capacidad de invasión (Zimmerly *et al.*, 1995a; Eskes *et al.*, 1997), la mayoría de los mutantes perdieron su capacidad de movimiento. El único mutante que retuvo, aunque a niveles muy bajos ( $3 \times 10^{-4}$ ), su capacidad de movimiento fue aquel en que la A se sustituyó por C en el dVI. Algunos de estos raros eventos de invasión fueron secuenciados para comprobar la veracidad de la inserción y pudimos identificar la mutación en la ribozima (dato no mostrado). Una de las cosas que más llamó nuestra

atención fue que mutantes que tenían niveles similares (CL3) o incluso mayores de intrón *ariat* (CL1 y CL2), no retenían su capacidad de movimiento. Tratando de buscar una razón para estos resultados, pensamos que quizá la inestabilidad o la baja proporción de RNPs que podrían generarse en los mutantes en la región de coordinación explicarían los niveles de *homing*. Hemos usado la capacidad RT exógena de estas RNPs *in vitro* como medida de la fracción de RNPs activas presentes en nuestras muestras (ver tabla II.19). Si tenemos presente que esta actividad fue extraordinariamente baja en todos los mutantes (cercana al valor mostrado por el mutante en el ribozima pKG4dV<sub>T7</sub>, pero superior a la del mutante en el centro activo de la actividad RT, dominio RT5 de la IEP, pKG4YAAH<sub>T7</sub>) y que las RNPs extraídas de los mutantes del dI mostraron una actividad ligeramente superior a la del mutante *homing*<sup>+</sup>, no podemos decir que la ausencia de movilidad en estos últimos se deba a una inestabilidad de las RNPs. Más bien podríamos pensar en un problema en la organización del centro activo la ribozima que impide el correcto reconocimiento del ADN diana.



**Figura II.18.** Análisis de los niveles de *retrohoming in vivo* de los diferentes mutantes de RmInt1. (A) Diagrama que representa el conjunto de plásmidos que intervienen en el ensayo de movilidad entre plásmidos. Encontramos un plásmido donador de intrón (silvestre o mutado) que es resistente a kanamicina. Por otro lado, el plásmido aceptor contiene la diana de invasión rodeada por terminadores de transcripción (T1, terminando tanto la transcripción llevada a cabo por la ARN polimerasa de *E. coli* como la del fago T7; T2, hace que termine la ARN polimerasa

de *E. coli* pero no la de T7) seguido de un gen de resistencia a tetraciclina que carece de promotor de manera que este plásmido sólo va a ser resistente a ampicilina; cerrando el conjunto se encuentra un terminador de la T7 ARN polimerasa. Cuando se produce la invasión de este plásmido aceptor con el ΔORF, se activa la transcripción del gen Tc<sup>R</sup> permitiendo seleccionar los clones en los que ha tenido lugar el *homing*. Junto a este esquema se muestra una tabla con los datos de movilidad de los mutantes anotados como el porcentaje de invasión de un plásmido aceptor que contiene el ADN diana en la misma orientación que la horquilla de replicación, pACE0.6as. Como control se usó pACEas que carece de la diana aunque el resto del plásmido es similar. (B) Niveles de invasión detectados por hibridación con una sonda del exón 2 que nos permite establecer la eficiencia de *homing* y con una sonda específica del intrón. El tamaño y la identidad de las bandas se indican a los laterales.

Los resultados de movilidad fueron confirmados por hibridación (Southern blot) del ADN plasmídico con una sonda del intrón y también con una sonda del exón 2. Como se puede ver en la figura II.18B, esta técnica no fue lo suficientemente sensible para detectar los bajos niveles de *retrobombing* del mutante AC.

**Figura II.19.** Se representa la capacidad de síntesis de un ADNc sobre un sustrato exógeno de los mutantes respecto del WT (en porcentaje). Los datos proceden de al menos 3 medidas y dos ensayos independientes y se expresan como la media  $\pm$  la desviación estándar.

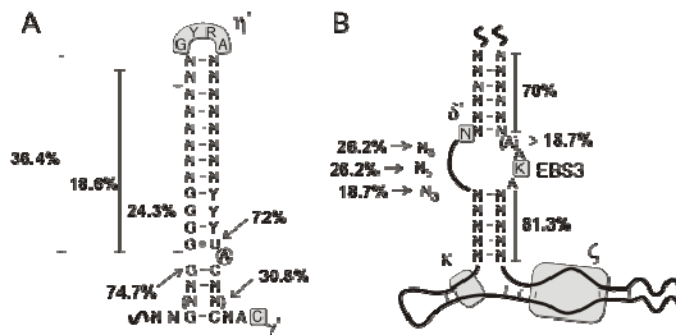
Mutantes	Actividad RT exógena
pKGENA-T7	100
pKG-dIVT7	1.65 $\pm$ 1.34
pKG-YAIIIIT7	7.32 $\pm$ 0.07
pKG-dVI-dA17	3.13 $\pm$ 3.13
pKG-dVI-ACT7	9.76 $\pm$ 0.51
pKG-dVI-pCT7	8.21 $\pm$ 3.96
pKG-dCL1T7	3.36 $\pm$ 4.98
pKG-dCL2T7	21.99 $\pm$ 1.35
pKG-dCL3T7	17.88 $\pm$ 1.54

#### II.1.3.4. Estudio estructural de los dominios I y VI de la ribozima de los intrones del grupo IIB

RmInt1 pertenece a la clase IIB de intrones del grupo II junto con otros intrones como ai5 $\gamma$ , aunque RmInt1 se encuadra dentro de la clase IIB3 y ai5 $\gamma$ , que carece de IEP en el dIV, se enmarca en la IIB1. Resulta curioso que estos dos intrones, a pesar de sus diferencias, conservan una estructura similar, y al parecer característica de este grupo de intrones, que coordina la conformación del sitio catalítico de la ribozima. En nuestro estudio de las características estructurales de los intrones del grupo IIB, hemos usado la base de datos de intrones bacterianos creada por Zimmerly y colaboradores (Dai *et al.*, 2003). En esta colección (actualizada en agosto de 2007) se incluyen 184 intrones bacterianos que contienen IEP y 19 intrones de arqueas, así como una completa colección de intrones fragmentados o incompletos. Nosotros hemos seleccionado 117 secuencias con estructura perfectamente definida correspondientes a intrones completos de diferentes microorganismos. Los intrones que son idénticos descritos en más de un organismo o los que están presentes en múltiples localizaciones se tienen en cuenta sólo una vez, de manera que finalmente nos quedamos con 107 secuencias entre las que se incluyen 41 intrones de la clase IIB1, 21 del grupo IIB2, 23 IIB3, 14 IIB4 y 8 IIB5.

El análisis de la estructura asociada a la secuencia del dVI de estas ribozimas nos reveló algunas características interesantes (figura II.20A). Si nos fijamos en la distancia que separa la A del final del intrón, encontramos que la más común suele ser 7 nt (61,7%) seguida por un 32,7% de los intrones en los que esta distancia se incrementa en un nucleótido. Uno de las cuatro características estructurales implicadas en el reconocimiento del sitio de *branching* se ha sugerido que es el tallo basal de 4 bp del dVI (Chu *et al.*, 2001). Según nuestros datos, sólo el 30,8% de los intrones presenta un tallo de esta longitud frente al 59,8% en los que esta región en doble hélice es de 3 bp; en un 7,4% de los casos el tallo es de 5 bp. Varios autores consideran importantes los pares de bases que flanquean la A desapareada (Chu *et al.*, 1998; Podar *et al.*, 1998b; Chu *et al.*, 2001; Schlatterer *et al.*, 2006; Erat *et al.*, 2007). Una amplia mayoría de intrones (74,7%) muestran un apareamiento G-C precediendo al nucleótido desapareado, seguido de un 13% con apareamiento G•U y un 9,3% con A-U. Si pensamos en el apareamiento que

acontece una vez pasado este residuo, el 72% presentan un apareamiento G•U aunque también pueden estar presentes otros apareamientos con menor frecuencia (A-U→ 11,2% o G-C→ 6,5%); es curioso que la purina suele encontrarse en la cadena opuesta donde se localiza la A. Es importante destacar que hay un 8,4% de los intrones IIB en los que los nucleótidos por encima de la A no se encuentran apareados. En un 79,3% de los casos, la distancia entre la A y la región desapareada terminal del dVI es 10-12 bp (10 bp→ 24,3%; 11 bp→ 18,6%; y 12 bp→ 36,4%); por lo general, esta doble hélice de 12 bp no muestra ningún desapareamiento. Esta distancia puede alcanzar los 16 bp, pero cuanto mayor es, más probable es la existencia de regiones desapareadas internas. Otra región importante en el dVI es el bucle final; lo más habitual es que presente 4 nt (65,4%) aunque pueden darse otras posibilidades (7 nt→ 20,5%, 5 nt→ 8,4% y excepcionalmente 3, 6 o 10 nt). En los que la región comprende 4 nt, su composición es G (94,3%) Y (94%) R (93%) A (96%). Finalmente, analizamos la presencia en el dVI de una región rica en purinas, una característica conservada incluida en estudios filogenéticos previos (Michel *et al.*, 1989; Chu *et al.*, 2001); la mayoría de los intrones estudiados (89,7%) presentan una región de 5-8 nt enriquecida en purinas que mayoritariamente suele estar compuesto de G (88,5%) y enfrentado con la A desapareada.



**Figura 20.** Análisis estructural de la región de coordinación y el dominio VI en los intrones bacterianos del grupo IIB. (A) Se destacan las características más conservadas entre las secuencias de 107 intrones estudiados. Se analiza la distancia entre la A desapareada (dentro de un círculo) y el final del intrón, la longitud de la horquilla por encima de la A, la identidad de los apareamientos

que rodean el nucleótido de *branching* y la longitud y composición del bucle terminal del dominio. (B) Estructura consenso de la región de coordinación en intrones IIB. Se ha evaluado la distancia con estructuras por encima y por debajo de la región desapareada, así como el grado de asimetría del bucle; se destaca la importancia de la secuencia AKA en el lado derecho del bucle que comprende la interacción EBS3. Las interacciones terciarias aparecen sombreadas e identificadas con letras griegas.

El análisis estructural de la región de coordinación fue algo más complicado (figura II.20B). La longitud de la zona 5' de esta región desapareada es bastante variable aunque en el 52,4% de los intrones IIB suele ser de 6-7 nt; también es posible una longitud de 8 nt (15,9%), 4 nt (18,7%), 5 nt (7,5%) y 2 nt (1,8%). En cuanto al lado derecho del bucle se observa una cierta tendencia: en el 68,2% de los intrones su longitud es de 3 nt. Siete de los 8 intrones IIB5 incluidos en este estudio muestran 8 nt en esta zona 3'. En cuanto a su composición, hemos comprobado que hay una secuencia que se repite en 100 de los 107 intrones analizados: A (90%) K (72%) A (88%), correspondiéndose K la mitad de las veces con una G. En cuanto a la distancia que separa esta región desapareada respecto de otras estructuras por arriba y por abajo, en el 92,4% de los casos suele encontrarse a 6 bp

de la región ID(iv) y en el 81,3% de estos intrones, la separación con la región receptora del dV es de 5 bp. Nuevamente, los intrones del grupo IIB5 rompen esta tendencia, ya que esta última distancia puede rondar los 10-14 bp. Sólo dos de los 107 intrones IIB analizados no muestran ningún vestigio que nos permita la descripción de una región de coordinación, fueron el intrón Sh.fr.I1 de *Shewanella frigidimarina* y M.a.I4-3 de *Methanosarcina acetivorans*; en ambos casos se encontró una G desapareada en su lugar.

## II.2. LA LIGACIÓN DE EXONES ES UN PROCESO INEFICIENTE EN RmInt1

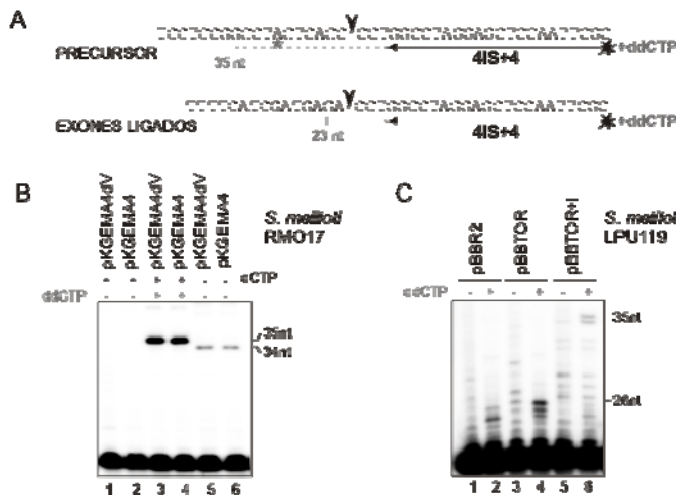
Hablábamos al principio de este capítulo de que la reacción de *splicing in vitro* de RmInt1 era bastante ineficiente, y esto es especialmente acusado si atendemos al porcentaje de exones ligados que se detectaron (Costa *et al.*, 2006a). Además, es bastante interesante que, dependiendo de la longitud del exón 1, esta proporción varía, siendo más efectiva la unión de exones en construcciones  $\Delta 20$  que en  $\Delta 15$  (Costa *et al.*, 2006a y b). Esto podría implicar que en este intrón y en otros relacionados exista una interacción adicional que involucre a los nucleótidos situados entre las posiciones -16 y -20 en el exón 1 (Costa *et al.*, 2006b). Se especula con que pueda estar relacionada con cambios en el sitio catalítico en la transición entre la primera y la segunda reacción de *splicing*.

En este apartado se van a describir una serie de resultados preliminares relacionados con la capacidad de RmInt1 de producir exones ligados *in vivo*. De entrada estos experimentos ponen de manifiesto un defecto en RmInt1 a la hora de unir los exones una vez que el intrón se ha escindido. Por supuesto, estos resultados deben ser confirmados mediante otras técnicas, a poder ser más sensibles, que permitan establecer el porcentaje real de ARNm maduro que este intrón permite.

### II.2.1. DETECCIÓN DE EXONES LIGADOS POR PPE

Nos hemos referido a esta técnica cuando tratábamos de discriminar entre formas en lazo y circulares (apartado II.1.1.6.3). El fundamento es el mismo, extensión de cebador frenada por la presencia de ddNTPs, pero el diseño experimental varía. En este caso el cebador a partir del cual la RT va a extender el ADNc se encuentra situado en el exón 2 y nos debe permitir discriminar entre precursor, intermediario (intrón unido a exón 2, si existe) y exones ligados. Debido a que los exones en construcciones pKG2.5 coinciden con ISRm2011-2 y teniendo en cuenta que *S. meliloti* RMO17 presenta 12 copias de este elemento en su genoma, nos llevó a usar para estos estudios la construcción pKGEMA4 que nos va a permitir el diseño de cebadores que no presenten hibridación cruzada con la secuencia de inserción. Nuestra primera tentativa fue el uso del oligonucleótido 4IS+4 combinado con la presencia de ddCTP (figura II.21A y 21B, carriles 3 y 4). En paralelo, sumamos a estos experimentos una variante que consistió en añadir 3 dNTP pero no adicionar dCTP de manera que el producto de extensión debía ser un nucleótido más corto (figura II.21B, carriles 5 y 6). Como se puede observar la adición de ddCTP o la

ausencia de dCTP nos marca la presencia de moléculas de precursor (comparar carriles 1 y 2 con el resto en figura II.21B), pero en ningún caso esta técnica nos permitió señalar la presencia de exones ligados o de intermediarios de reacción. Se probaron muchas variables: enzimas, temperaturas, proporción relativa de ddNTP, condiciones de anillamiento..., pero el resultado no mejoró. En muchos de estos intrones se ha observado apareamiento de los exones formando una horquilla perfecta (F. Michel, comunicación personal), de manera que esto podría dificultar la hibridación del oligonucleótido; de acuerdo a esto se diseñó un nuevo oligonucleótido, 4PPE, pero proporcionó resultados similares (dato no mostrado).



**Figura II.21.** Estudio de la ligación de exones mediante extensión de cebador adicionada de ddNTP. (A) Diagrama que representa una de las aproximaciones usadas en estos ensayos para el ARN procedente de construcciones pKGEMA4. En el se representa el oligonucleótido (4IS+4) marcado radiactivamente, hibridado a su correspondiente secuencia en el exón 2 (flecha). Con líneas discontinuas se señalan los productos de extensión generados si a la mezcla de reacción se le añade ddCTP tanto en las moléculas de precursor como de exones ligados. (B) Resultado del experimento en

construcciones con intrón silvestre en comparación de otras con la ribozima mutada en *S. meliloti* RMO17. Se probaron dos protocolos uno basado en la parada por la incorporación de ddCTP y otro en el que no se adicionó dCTP a la mezcla de dNTP. El tamaño de los productos de extensión se marca en el lateral. (C) Análisis de la ligación de los exones del gen *oxi1* en *S. meliloti* LPU119 (copia de *oxi1* genómica invadida por RmInt1). De manera similar al ensayo llevado a cabo en *S. meliloti* RMO17, se trataron de identificar los ADNc derivados de la extensión del oligonucleótido oxi+4 cuando se adicionaba ddCTP en el medio. En esta estudio se usaron 3 plásmidos: uno que se correspondía con el vector (pBBR2), otro fue ese mismo vector conteniendo la diana *oxi1* (pBBTOR) y, finalmente, aquel que contenía la diana invadida por el intrón (pBBTOR+I). En todos los casos, se llevó un control paralelo en el que se adicionaban los cuatro dNTPs y ningún ddNTP (-).

Esta contrariedad nos llevó a intentar refrendar los resultados mediante el uso de otras construcciones en las que el intrón estaba presente pero el contexto de exones es ligeramente diferente; son construcciones derivadas de la transposición ectópica de RmInt1 sobre el gen *oxi1*. En particular, decidimos usar una cepa de *S. meliloti* en la que la copia genómica de *oxi1* se encuentre invadida para eliminar ruido de fondo, LPU119 (figura II.21C). Pudimos comprobar que cuando se usaba como molde el ARN de la bacteria que expresa sólo el vector, pBBR2, se produjeron una serie de paradas inespecíficas no relevantes (carriles 1 y 2). En cambio, usando como molde ARN de células que portan el plásmido que contiene los exones ligados del *oxi1*, pBBTOR,

pudimos observar una parada dependiente de la adición de ddCTP a la altura de lo que consideramos exones ligados (comparar carriles 3 y 4). Finalmente, cuando el ARN se extrajo de bacterias con el plásmido anterior invadido por el intrón, pBBTOR+I, fuimos capaces de detectar una banda dependiente de ddCTP que migraba a la altura esperada para la molécula de precursor. Nuevamente, la banda de un posible intermediario o la esperada para los exones ligados estaba ausente. Como en el caso anterior, se probaron distintas condiciones pero también distintos fondos genéticos: *oxi1*<sup>-</sup> como *E. coli* DH5 $\alpha$ , *oxi1*<sup>+</sup> no invadido en el genoma como *S. meliloti* RMO17 y en cepas tipo GR4 con el *oxi1* invadido como *S. meliloti* GR46D6 (Muñoz *et al.*, 2001). En ningún caso hubo diferencias respecto al resultado descrito en *S. meliloti* LPU119 (dato no mostrado).

### II.2.1. DETECCIÓN DE EXONES LIGADOS POR RT-PCR

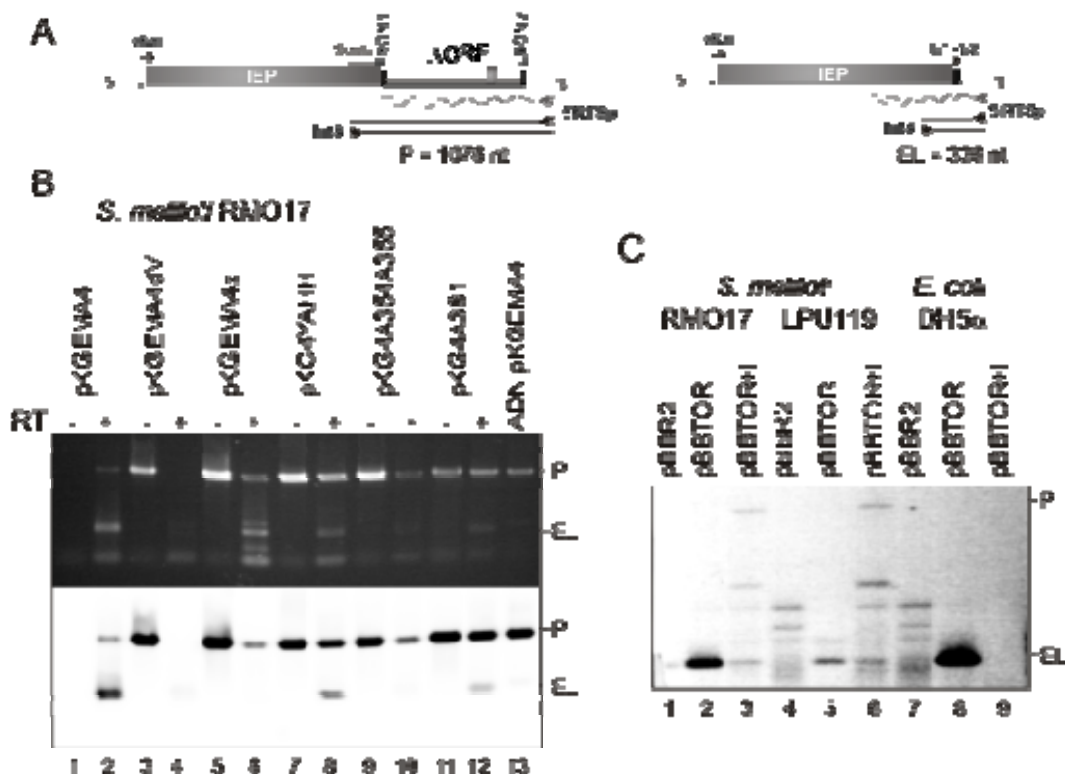
Ante la imposibilidad de poner de manifiesto la presencia de exones ligados por PPE, pensamos en intentarlo por RT-PCR, de manera que si existe poca cantidad podamos ser capaces de detectarla al magnificarla por amplificación. Siguiendo el mismo razonamiento que el caso anterior, decidimos usar tanto construcciones pKGEMA4 como aquellas otras derivadas de la invasión del gen *oxi1*. En el primer caso, se usó una pareja de cebadores (Int3/5'RTSp) que amplificaban 336 bp si se trataba de los exones ligados frente a 1076 bp si nos encontramos con ADNc derivado de ARN precursor (figura II.22A); el cebador usado para la síntesis del ADNc fue el propio 5'RTSp. Como era de esperar, la banda de precursor se detectó tanto para intrón WT (pKGEMA4) como para todas las variaciones de intrón mutado ensayadas (figura II.22B); de hecho, parece observarse cierta contaminación con ADN plasmídico en los carriles -RT. Nuestra sorpresa llegó al descubrir una banda que aparecía a la altura de los exones ligados en los geles de agarosa teñidos con bromuro de etidio (figura II.22B, parte superior); no sería sorpresa si apareciese únicamente en la muestra con intrón silvestre (carril 2), pero también era visible los carriles +RT de algunos mutantes defectuosos en *splicing* (carril 6 y, curiosamente más débil en los carriles 4 y 10). De igual modo, mutantes que sabemos que retienen en mayor o menor medida la capacidad de madurar según los datos de extensión de cebador mostraron una banda a la misma altura que el intrón WT (carriles 8 y 12); además parece existir una aparente correlación entre la cantidad de exones ligados obtenidos con esta técnica y la valoración del producto escindido por extensión de cebador. Se trató de comprobar si se trataba de una banda inespecífica en el mutante en comparación con el silvestre mediante RT-PCR semicuantitativa (ciclos 15 al 35), pero no obtuvimos un resultado concluyente (dato no mostrado). Finalmente, decidimos determinar sin lugar a dudas la naturaleza de la banda de migración a la altura de exones ligados. La separación del producto de PCR en un gel de poliacrilamida al 6% parece apuntar a que las bandas en el intrón silvestre o en construcciones en la que la IEP se encuentra mutada (pKGEMA4x y pKG4-A354A355) migran a diferente altura (dato no mostrado). Aún así decidimos confirmar esta apreciación mediante hibridación con una

sonda específica del exón 1 (esta sonda también hibrida con el precursor, figura II.22A). Se corrieron en paralelo muestras de PCR realizado sobre preparaciones de ARN, sometidas o no a previa RT, extraídas de células de *S. meliloti* RMO17 que expresan el intrón en sus diferentes variantes [*splicing*<sup>+</sup>: WT, pKGEMA4; con la IEP mutada en el dominio activo para la RT (RT5), pKG4-YAHH; IEP mutada en el dominio madurasa, pKG4-A381 (30% *splicing*); y mutantes *splicing*<sup>-</sup>: ribozima mutada en el dV, pKGEMA4dV; con una rotura de la pauta de lectura en la IEP, pKGEMA4x; otro de los mutantes en la madurasa de la IEP, pKG4-A354A355]. Tras transferir los productos separados por electroforesis en gel de agarosa al 1% a una membrana de nylon, se hibridaron con una sonda de ~260 bp que se corresponde con el extremo 3' de la secuencia de la IEP (figura II.22B, parte inferior). El resultado no dejó lugar a dudas: en aquellos mutantes en los que la mutación inhibe el *splicing* no somos capaces de detectar exones ligados (carriles 4, 6, y 10); en cambio, mutantes que retienen en mayor o menor medida su capacidad de maduración así como el intrón WT mostraron señal de hibridación (carriles 2, 8, y 12). Con todo, decidimos confirmar todos los resultados anteriores secuenciando la banda. Se aisló la banda de gel de agarosa, se clonó y se secuenció; efectivamente, la secuenciación demostró que, mediante RT-PCR, podíamos detectar exones ligados.

Se hicieron comprobaciones paralelas mediante ensayos de RT-PCR en el entorno *oxi1*, pero el resultado vino a confirmar las observaciones que ya habíamos hecho con la PPE. En este caso, el resultado fue, si cabe, más difícil de interpretar debido a la presencia de una serie de bandas *a priori* inespecíficas que enmascararon los resultados tanto en *S. meliloti* LPU119 como en un fondo genético en el que no se detecta la presencia de *oxi1* como es *E. coli* DH5 $\alpha$  (figura II.22C). El resultado tampoco fue mejor en RMO17. El propio vector (pBBR2) mostraba una serie de bandas atribuibles a artefactos generados por la técnica (carriles 1, 4 y 7). En los tres casos, a pesar del ruido de fondo que proporcionaba el vector solo, la amplificación de un ADNc obtenido para ARNs de bacterias que contienen la unión de los exones (pBBTOR) da lugar a una banda de intensidad considerable que marca la altura a la que deben aparecer los exones ligados (carriles 2, 5 y 8). En cambio, la inserción del intrón en su versión silvestre (pBBTOR+I) no modificada (esto es, no tipo  $\Delta$ ORF, sino ribozima con la IEP codificada en su dIV) vuelve a generar una serie de bandas inespecíficas entre las que parece distinguirse la banda de ligación de exones en las dos cepas de *S. meliloti* (carriles 3 y 6); también podemos observar una banda a la altura del precursor. No es el caso de *E. coli*, donde no somos capaces de detectar ningún producto de amplificación (carril 9)

En este punto, dadas las enormes dificultades que nos encontramos a la hora de la detección de exones ligados, decidimos abandonar esta línea hasta encontrar un nuevo sistema en el que no se planteen semejantes problemas. En cualquier caso, lo que parece deducirse de estos ensayos es que la proporción de ligación de exones es extraordinariamente baja; tan baja que es imposible detectarlos por extensión de cebador con ddNTP, aunque la RT-PCR revela su existencia. Ahora, el nuevo reto que se plantea es evaluar cuánto de baja es esta proporción.





**Figura II.22.** Detección de exones ligados mediante RT-PCR. (A) Estrategia de RT-PCR planteada: la RT se lleva a cabo con el cebador 5'RTSp complementario al exón 2, seguida de PCR con ese mismo oligonucleótido y el cebador Int3 que hibrida en el exón 1. Se muestra el producto generado tanto para el precursor (P) como para la ligación de exones (EL). Sobre el precursor se señala la zona de hibridación de la sonda usada en el Southern blot. (B) Resultado obtenido con la RT-PCR en muestras de ARN procedentes de *S. meliloti* RMO17 que portan los diferentes plásmidos que se indican, que a su vez contienen el intrón silvestre versión  $\Delta$ ORF (pKGEMA4) o distintos mutantes afectados en mayor o menor medida en su capacidad de maduración. En la parte superior se muestra la electroforesis en gel de agarosa al 1% de los productos del PCR, previa (+) o no (-) RT, teñida con bromuro de etidio. En la parte inferior, se presenta ese mismo gel transferido e hibridado con una sonda complementaria al exón 1. Como control que nos marca el tamaño de la banda de precursor usamos de molde en el PCR ADN de pKGEMA4. (C) RT-PCR para detección de exones ligados *oxi1*. Se muestra el negativo de un gel de agarosa teñido con bromuro de etidio (permite una mejor apreciación de las bandas) en el que se incluyen muestras de ARN de dos cepas de *S. meliloti* (RMO17 y LPU119) y la cepa *E. coli* DH5 $\alpha$  que pueden contener 3 plásmidos distintos (explicado en el texto).

## II.3. DISCUSIÓN

En este capítulo hemos estudiado los distintos productos que se generan durante la escisión *in vivo* del intrón RmInt1. Las distintas técnicas manejadas nos han permitido determinar la existencia de moléculas de escisión en lazo y circulares, aunque hay indicios que apuntan a la presencia de formas derivadas de una primera reacción de hidrólisis durante la liberación del intrón. Además, hemos puesto de relieve algunos de los requerimientos para el proceso de escisión, en cuanto a actividad maduradora de la IEP y determinadas estructuras de la ribozima. Por otro lado, datos preliminares acerca del porcentaje de ligación de exones llevada a cabo por este intrón revelan un escaso rendimiento en esta reacción.

### II.3.1. LAS FORMAS GENERADAS POR RmInt1 EN SU ESCISIÓN

Diversas técnicas utilizadas en el estudio del tipo de moléculas que se producen a partir del intrón durante la escisión, nos permiten afirmar sin lugar a dudas que RmInt1 se escinde generando ARN *lariat*. La cantidad relativa es difícil de establecer, puesto que siempre se presentan limitaciones técnicas que impiden la correcta cuantificación de los productos de escisión, aunque la mayoría de nuestros estudios apuntan a que se trata del producto mayoritario generado durante la escisión del intrón silvestre.

Además, por RT-PCR hemos conseguido poner de manifiesto la presencia en nuestras preparaciones de moléculas de ARN circular derivado de intrón. Muy probablemente, la proporción de este tipo de moléculas esté sobreestimada debido al sesgo que, como consecuencia del enlace 5'-2' en el *lariat*, genera la RT. Desconocemos la naturaleza del enlace en la molécula circular, de cualquier modo parece ser que la RT se comporta de un modo diferencial dependiendo si se trata de una molécula con (*lariat*) o sin (circular) residuos unidos al 3' del nucleótido que establece el enlace modificado (en nuestro caso, la A) (Lorsch *et al.*, 1995; Murray *et al.*, 2001). Estas formas son detectadas también en mutantes en los que la capacidad de movimiento se encuentra impedida (K381A o YAAH), por lo que podemos descartar que los amplificadores resulten de dímeros de intrón cabeza-cola generados durante la movilidad del intrón. A esto debemos sumarle, que el intrón circular generado por RmInt1 también se detecta en fondos genéticos RecA<sup>-</sup>, lo que impide reorganizaciones genómicas que nos generen artefactos en la RT-PCR. Los amplificadores de moléculas circulares obtenidos en ausencia de RT en algunas muestras nos indican la presencia de ADN intrón circular; aunque podría tratarse de moléculas sintetizadas por la propia RT del intrón, debemos profundizar más en este aspecto. Estas moléculas de ADN circular se describieron en *Podospora anserina* asociadas con senescencia celular (Osiewacz y Esser, 1984; Schmidt *et al.*, 1994; Begel *et al.*, 1999). Sin embargo, este no parece ser el caso de las moléculas de ADN circular de RmInt1 puesto que no se observa acumulación de este tipo de especies moleculares en curvas de crecimiento hasta saturación de *S. meliloti* RMO17: pKG2.5.

Por otro lado, el análisis del sobrenadante y el sedimento obtenidos durante el protocolo de aislamiento de RNPs sugiere que el *lariat* es el tipo molecular se encuentra asociado con la IEP en este tipo de complejos y es muy probable que el ARN circular presente diferentes propiedades hidrodinámicas que pudieran debilitar su interacción con la proteína. Este comportamiento en solución puede deberse entre otras razones al tipo de plegamiento que adoptan las formas escindidas o a diferencias en las interacciones (reconocimiento) de ambas formas con la IEP.

Finalmente, disponemos de indicios que apuntan a la existencia en RmInt1 de una ruta de maduración por hidrólisis *in vivo*. Bien es cierto que en nuestros intentos por determinar el extremo de este tipo de moléculas nos hemos encontrado con muchos escollos, no consiguiendo resultados positivos, pero trabajos posteriores tratarán de aclarar este aspecto. Como punto de partida, se piensa en tratar de optimizar la técnica del 3'-RACE y mejorar las condiciones del 5'-RACE a partir de ADNc gracias al uso de un kit comercial (5'/3'-RACE Kit, 2nd Generation, Roche Applied Science).

El cambio radical en el patrón de productos de escisión que una técnica de detección directa (Northern blot) pone de manifiesto en relación con el protocolo indirecto (RT-PCR) nos ha dado mucho que pensar y algunas de las justificaciones que se nos han ocurrido es que: (i) todos sean artefactos de la RT; (ii) que el *lariat* WT sea más resistente a degradación que cualquiera de los mutados; (iii) que el proceso de escisión no ocurra completo, de manera que nuestro intrón presente una mayoría de formas intermedias de *lariat* unido al exón 2 que nosotros no podemos detectar con el Northern blot (recordar que no conocemos el tamaño de nuestro exón 2). Esta última hipótesis queda descartada porque las muestras de ARN usadas en el Northern han sido sometidas a extensión de cebador usando un oligonucleótido que aparee en el exón 2 que debía habernos generado, al menos un producto de extensión que se correspondiese con la parada del enlace modificado del *lariat* y no fue el caso (dato no mostrado pero que se intuye de los datos de PPE para la detección de exones ligados).

Con todo, una de las principales conclusiones que podemos extraer de nuestros estudios es que RmInt1, como otros intrones del grupo II, usa rutas alternativas para la escisión del intrón tratando de prevenir los posibles efectos deletéreos que una interrupción de la maduración de un ARNm podría ocasionar. Se trata de no suponer nunca una carga para el organismo hospedador, de modo que éste no se vea en la obligación de tratar de expulsarlo de su genoma (o impedir su movilidad). En este sentido nos llama especialmente la atención el bajo rendimiento de exones ligados que hemos sido capaces de detectar; y por qué, con esta supuesta alta tasa de interrupción del proceso de traducción de la transposasa de la secuencia de inserción *ISRm2011-2*, es un elemento con tanto éxito entre los miembros del orden Rhizobiales. Este comportamiento, más propio de un retrotransposón que de un intrón, que muestra RmInt1 podría explicar el alto grado de truncamiento que se observa en distintas poblaciones de rizobios (Fernández-López *et al.*, 2005). Derivado de su comportamiento *in vitro*, se ha propuesto que RmInt1 podría estar siendo utilizado por su hospedador para mantener controlado al otro

elemento móvil, la secuencia de inserción (Costa *et al.*, 2006a). Claramente, esta hipótesis podría estar refrendada por los datos preliminares obtenidos *in vivo* que se presentan en este trabajo.

### II.3.1.1. El mecanismo de escisión de moléculas circulares en RmInt1

El análisis de los productos obtenidos en ensayos de RT-PCR en los que las preparaciones de ARN total habían sido tratadas con T4 ARN ligasa pusieron de manifiesto la presencia mayoritaria de intermediarios de intrón en los que pudimos señalar un sitio de rotura que coincide exactamente con el sitio donde se unen el extremo 3' del intrón con el exón 2 o bien distintas zonas de corte situadas en las proximidades de este sitio de maduración 3' (tanto dentro del intrón como en el exón 2). Estos resultados apoyan el mecanismo de escisión propuesto por Murray y colaboradores (Murray *et al.*, 2001). Así, una vez que el exón 2 se libera de la molécula de precursor, la C situada al final del intrón podría atacar el sitio de maduración 5' para dar lugar a intrón circular. Dependiendo de la naturaleza del nucleófilo que produzca la liberación del exón 2, podríamos discutir si en el proceso se estarían generando exones ligados o no. Desde luego, la influencia de esta ruta de escisión en cualquiera de los dos casos debe ser minoritaria dado el mínimo porcentaje de exones ligados que parece generar RmInt1.

### II.3.2. IMPLICACIÓN DE LA IEP EN LA ESCISIÓN DE RmInt1

Los resultados obtenidos por las diversas aproximaciones abordadas en este capítulo apuntan a que el dominio X es fundamental para la escisión *in vivo* de RmInt1 por cualquiera de las posibles rutas. Tanto es así que un mutante, A354A355, afectado en dos tirosinas muy conservadas dentro del dominio madurasa pierde la capacidad de maduración. Esta observación puede deberse a alteración de alguno de los contactos que la IEP mantiene con la ribozima aunque también puede estar ocasionada por una desestabilización de la proteína que conduce a su degradación. Sin embargo, una mutación que afecta también a una lisina conservada en el mismo dominio, no tuvo un efecto tan drástico. Este intrón retuvo parte de su capacidad para la formación de *lariat*, lo que nos lleva a pensar que es probable que de algún modo se esté afectando la interacción de la IEP con el ARN del intrón. Lamentablemente, hasta que el grupo de investigación no disponga de la proteína purificada no podremos determinar este aspecto. De los datos de RT-PCR sólo podemos extraer la conclusión de que se está produciendo una disminución en la generación de moléculas en lazo pero no que esté aumentando la cantidad de especies circulares debido al sesgo introducido por la RT.

Como dijimos en la introducción de este capítulo, RmInt1 es capaz de madurar *in vitro* en ausencia de la IEP dando lugar a moléculas circulares. Sin embargo, estos círculos debidos a la presencia de un sitio de corte inespecífico, IBS1\*, dentro del exón 2, no han sido encontrados *in vivo*, lo que es indicativo de que la IEP promueve el correcto reconocimiento de los exones. Hemos detectado un incremento muy llamativo de un

intrón circular que contiene una C extra en el punto de unión de los dos extremos del intrón cuando la IEP se encuentra mutada en el dominio madurasa, K381A. En la literatura se describe la presencia de nucleótidos extra en moléculas circulares generadas *in vitro* a partir del intrón ai5γ, que se consideran introducidos por la RT cuando se encuentra con el enlace fosfodiéster modificado 2'-5' (Murray *et al.*, 2001). En nuestro caso, llaman la atención dos aspectos: el primero es que siempre se trata de una C, pero más llamativo si cabe es que el primer nucleótido del exón 2 (IBS3) es justamente una C. Por tanto, es bastante probable que este tipo de moléculas se generen como consecuencia de un mal reconocimiento de la posición IBS3, de lo que podríamos concluir que el dominio X de RmInt1 determina la escisión precisa en el sitio de *splicing* 3' durante la formación de intrón circular. Probablemente su influencia en la formación de intrón lineal no sea tan evidente. Así, podríamos decir que, de algún modo, la IEP controla el balance entre intrón circular y *lariat*.

### II.3.3. EL PAPEL DE LA (A) DESAPAREADA DEL dVI Y LA REGIÓN DE COORDINACIÓN EN LA ESCISIÓN *IN VIVO*

La formación de moléculas en lazo durante la maduración ha sido un proceso ampliamente estudiado *in vitro* en intrones nucleares eucariotas pero también en intrones del grupo II, pero las bases moleculares que rigen el proceso *in vivo* no han podido ser establecidas. Nosotros hemos desarrollado una aproximación *in vivo* basada en el estudio de determinados mutantes implicados en la formación de *lariat* en el intrón bacteriano del grupo II RmInt1. Nuestros datos refuerzan estudios anteriores que pusieron de relieve la relevancia de la A en el dVI para la conformación catalíticamente activa que desencadena la primera reacción de transesterificación en el *splicing*. Sin embargo, este residuo parece ser prescindible en las otras rutas de escisión que operan en RmInt1. Además, hemos estudiado el hipotético papel que se ha atribuido a la región de coordinación como el receptor del dominio VI durante el *splicing*. Nuestros resultados apuntan a que se trata de una región absolutamente crítica en la conformación final del centro catalítico de la ribozima, por lo que cualquier modificación altera de manera significativa todas las rutas de escisión catalizadas por la ribozima de RmInt1. Estudios filogenéticos proporcionan indicios que nos permiten especular la posibilidad de que esta región pudiera estar implicada en la coordinación de metales en el sitio catalítico.

#### II.3.3.1. Papel de la A en la escisión de RmInt1

Aunque la secuencia primaria del dVI no esté conservada, suele presentar una estructura secundaria característica que consiste en un corta estructura en doble hélice con un nucleótido desapareado, generalmente una A, situada a 7-8 nt del final del intrón. En este trabajo hemos evaluado la capacidad de escisión y de invasión que muestran 3 mutantes diferentes en este residuo: por un lado, se delecionó la A para conseguir el efecto más dramático de los posibles; también se introdujo una U en la cadena opuesta que nos

permitiese introducir la A dentro de la doble hélice por apareamiento de bases; por último, cambiamos esta A por una C, un cambio que había dado lugar a niveles indetectables de transesterificación durante la reacción de *splicing in vitro* en otros intrones del grupo II. De acuerdo con observaciones anteriores (Liu *et al.*, 1997; Chu *et al.*, 1998; Podar *et al.*, 1998b), los tres mutantes se comportaron de modo similar: mientras  $\Delta A$  y pU parecen incapaces de producir intrón escindido *lariat*, una C en el sitio de *branching* no inhibe completamente una primera reacción transesterificación (aunque está muy afectada). Además este último mutante retiene una mínima capacidad de *homologation* no detectada por hibridación. No es ninguna novedad esta alta ineficiencia de la C como nucleótido de *branching* si lo comparamos con otras sustituciones en esa misma posición como G o U (Query *et al.*, 1996; Liu *et al.*, 1997; Podar *et al.*, 1998b), sin embargo, trabajos previos ya describieron la presencia de cantidades muy pequeñas de *lariat* en este tipo de mutante (Jacquier y Rosbash, 1986; Liu *et al.*, 1997). Esto contrasta con los resultados obtenidos en estudios de maduración pre-ARNm eucariotas en los que se establece que una C en el sitio de *branching* provoca una reducción en la velocidad de la primera reacción de *splicing* cercana al 70%, pero que la segunda reacción tiene lugar incluso más rápido que en el intrón silvestre (Query *et al.*, 1996). Además, el análisis de la resistencia a degradación (mediante el uso de enzimas que atacan los enlaces 5'-2') de una molécula en lazo con una C en el punto de ramificación obtenida *in vitro*, muestra que se trata de moléculas extraordinariamente estables que se acumulan en el interior celular (Jacquier y Rosbash, 1986). Ninguna de las dos observaciones anteriores parece ser el caso de este tipo de mutante en RmInt1: la reducción en *splicing* supera el 70% y no observamos acumulación en el ARN de la bacteria.

En referencia directa a lo que suele ocurrir en intrones espliceosómicos en los que la ausencia de la A desapareada suele inducir al uso de sitios críticos para la formación del *lariat*, RmInt1 parece que se comporta de manera similar al resto de intrones del grupo II; los datos de RT-PCR indican que la escisión tiene lugar en la posición correcta y no en otra A cercana. Este “mal hábito” de los intrones espliceosómicos provoca que se produzca un bloqueo tras la primera reacción (Query *et al.*, 1996; Liu *et al.*, 1997; Gaur *et al.*, 1997), algo que los intrones del grupo II no pueden permitirse; sobre todo si tenemos en cuenta que se trata de elementos genéticos móviles que muy probablemente necesiten una estructura *lariat* perfecta para llevar a cabo el *splicing* reverso.

Por otro lado, nuestros datos *in vivo* podrían apoyar la idea de que cualquier nucleótido que pueda mantener expuesto de la manera adecuada su 2'-OH es válido para que se forme el intrón en lazo. Claramente, la mejor elección para esta posición es una A desapareada puesto que si la emparejamos con una U inhibimos completamente la capacidad de *branching in vivo* (Podar *et al.*, 1998b; este trabajo). Sin embargo, una transversión de A a C permite niveles mínimos de formación de *lariat*. Se han propuesto varias teorías para explicar la conformación que la A podría adoptar en el entorno del dVI basadas en la estructura del intrón ai5γ. Algunas sugieren que existe un equilibrio conformacional entre dos estados en los que la A se encuentra alineada dentro de la

hélice o expuesta fuera de la misma, siendo esta última configuración el estado activo para que tenga lugar la transesterificación (Liu *et al.*, 1997; Chu *et al.*, 1998; Zhang y Doudna, 2002; Schlatterer *et al.*, 2006). De hecho, estos autores consideran que la ineficiencia de la C como nucleótido para la ramificación puede radicar en interacción de este residuo con guanosinas adyacentes. La otra hipótesis es que la A permanece situada dentro de la hélice y sólo cambia su configuración cuando los apareamientos G•U que la flanquean coordinan iones  $Mg^{2+}$  provocando una apertura del surco menor en la hélice (Erat *et al.*, 2007). En este caso, la explicación para el comportamiento de la C podría ser justificado por tratarse de una base con menor capacidad de apilamiento que el resto. Aunque los estados conformacionales de la A no es un aspecto que se enmarque dentro de los objetivos de este trabajo, nuestros resultados podrían proporcionar evidencias que apoyen la primera hipótesis. En RmInt1, la A se encuentra rodeada de un apareamiento G-C y otro U-A (ver figura II.15A), lo que podría dificultar la unión de metales en esta región. Este hecho también nos permite explicar por qué el mutante AC puede formar *lariat*: una C en el sitio de formación del enlace modificado en el *lariat* difícilmente puede establecer interacciones lo suficientemente fuertes con los residuos adyacentes situados en la cadena opuesta, de manera que los apareamientos que flanquean esta posición desapareada no se alteran.

A pesar de la fuerte inhibición de la ruta de *branching* en los mutantes del dVI, todos parecen retener una capacidad WT para la escisión *in vivo* como formas lineales y también como moléculas circulares. La cantidad de estos dos productos de escisión es similar a la que se produce en el intrón silvestre, por lo que es sencillo deducir que la A es requerida para la formación del *lariat* pero prescindible en cualquiera de las otras rutas de escisión *in vivo*. La formación de moléculas lineales está en consonancia con observaciones previas extraídas de múltiples estudios *in vitro* relacionados con la estructura del dominio (Schmidt *et al.*, 1993; Chu *et al.*, 2001), así como con el análisis mutacional de esta región (Liu *et al.*, 1997; Chu *et al.*, 1998; Podar *et al.*, 1998b), donde se demuestra que cualquier mutante que muestra una limitación en la maduración en forma de *lariat*, es capaz de escindir-se mediante una primera reacción de hidrólisis. En el caso de estos mutantes, es absurdo pensar que la banda de mayor migración en el Northern blot se corresponde con *lariat* roto puesto que no hay *lariat*. ¿Podría tratarse de formas circulares rotas? Quizá, pero entonces la pregunta sería por qué no se observa nada de forma circular intacta. La discusión en este sentido se ve extraordinariamente limitada por la nula información publicada en torno a la formación de moléculas circulares en mutantes de este tipo.

### II.3.3.2. Relevancia funcional de la región de coordinación en la conformación del sitio catalítico de la ribozima de RmInt1

La región de coordinación ha sido considerada como uno de los elementos estructurales más importantes en los intrones del grupo IIB, ya que esta estructura ensambla la mayoría de los elementos que participan en el *splicing* (Costa *et al.*, 2000; de Lencastre *et al.*, 2005;

Hamill y Pyle, 2006). Cuando modificamos esta región en RmInt1 se produce una alteración paralela de todas las rutas de escisión *in vivo*. Un estudio previo de esta región en ai5 $\gamma$  enfocado en su implicación en la reacción de *branching* (Hamill y Pyle, 2006), nos limita la comparación de las observaciones *in vitro* con nuestros resultados *in vivo*. A grandes rasgos, se observa una correlación en los datos a pesar de que se trata de intrones diferentes (aunque pertenecientes al mismo grupo estructural, IIB) y sistemas completamente distintos (los estudios *in vivo* en RmInt1 se han realizado en *cis*, mientras que el trabajo *in vitro* en ai5 $\gamma$  se ha llevado a cabo en *trans*). Así, el mutante CL2 en el dI de RmInt1 conserva el porcentaje más alto de *branching* (~8 veces menor que el WT frente a las ~20 veces en ai5 $\gamma$ ), mientras que CL1 retiene sólo el 5% de la actividad WT (~35 veces en ai5 $\gamma$ ). El doble mutante muestra unos niveles de formación de *lariat* difíciles de cuantificar lo que se correlaciona con la disminución de ~130 veces que se describe para ai5 $\gamma$ . Por otro lado, tanto la formación de moléculas lineales como circulares se ve severamente afectada de acuerdo con la reducción de transesterificación descrita arriba. Uno de los primeros aspectos que llama nuestra atención es que la reducción del *branching in vivo* en RmInt1 es menos acusada que la observada *in vitro* para ai5 $\gamma$ . Esto puede deberse a otros factores implicados en el éxito de la reacción *in vivo*, como por ejemplo, la IEP. Además, observamos un comportamiento diferencial entre ambos lados de esta región desapareada: la región derecha mostró una inhibición menor de formación de *lariat* que el lado 5'. Este efecto puede deberse a dos posibles causas: (i) que la mayor extensión de la mutación en el lado izquierdo induce problemas conformacionales más importantes que una mutación puntual en la región 3'; y (ii), que la región 5' juegue un papel directo en cualquier interacción establecida con otras estructuras dentro del intrón determinante para la escisión. Estudios que permitan un mayor detalle acerca de estas posibles interacciones podrían demostrar esta hipótesis.

Por otro lado, muchos autores consideran que el dominio I de los intrones del grupo II es una de las primeras estructuras en plegarse de manera que supone el andamiaje en el que el resto del intrón se colapsa junto con los exones (Qin y Pyle, 1997; Su *et al.*, 2005; Waldsich y Pyle, 2007). Además, los datos obtenidos *in vitro* apuntan a que el sitio catalítico que va a llevar a cabo las dos reacciones de transesterificación en la ruta de *branching* se forma antes de que tenga lugar la primera reacción (Hamill y Pyle, 2006), lo que indica que la región de coordinación podría ser esencial para la adecuada conformación del centro activo de la ribozima. Por tanto, independientemente de cuál sea el nucleófilo en la primera reacción del *splicing*, cualquier alteración de esta región conducirá a un más o menos acusado descenso de la escisión.

Si atendemos a la capacidad de movimiento, es curioso las diferencias entre los diferentes mutantes de la ribozima. Mientras el mutante AC retiene una mínima capacidad de invasión, mutantes como CL2, que parecen formar mayor cantidad de *lariat*, no son móviles. Este comportamiento puede ser un fiel reflejo de los problemas conformacionales que tienen lugar en los mutantes en la región de coordinación. Este efecto podría estar en relación con la interacción con la IEP o también con el



reconocimiento del ADN diana para la invasión. Es un hecho demostrado que las RNPs son la forma activa para el movimiento (Saldanha *et al.*, 1999; Zimmerly *et al.*, 1999), y que el *lariat* es la principal forma que se asocia a la IEP en estas RNPs (este trabajo), por lo que un mutante que no se escinde como *lariat*, no debe retener capacidad de invasión. Pero, ¿qué pasa con nuestros mutantes en la región de coordinación? Una posibilidad es que aunque se esté produciendo proteína tenga dificultades para asociarse con el *lariat* alterado. Estudios de extensión de cebador llevados a cabo en paralelo en muestras de ARN y RNPs muestran que la reducción de la intensidad del producto de escisión es similar en los dos casos (dato no mostrado) y nuestra experiencia con otros mutantes nos dice que no es suficiente para inhibir el *homing*. Además, la actividad de RT exógena ensayada *in vitro* para RNPs de los mutantes en la región de coordinación supera ligeramente los niveles del mutante AC, lo que en cualquier caso podría ser indicativo de una mayor estabilidad de las primeras respecto a las segundas RNPs. Por eso, pensamos que es poco probable que la interrupción de contactos entre el ARN del intrón y la IEP sea la causa principal de que los mutantes en el dI no tengan *homing*. La otra posibilidad planteada es que se altere la interacción con el ADN diana. Como decíamos al principio, esta región está implicada en el posicionamiento de los exones durante el *splicing* (Costa *et al.*, 2000), por lo que también debe estarlo durante el reconocimiento de la diana previo a la movilidad. La consecuencia inmediata de una alteración que provoque la rotura de las interacciones EBS-IBS es un defecto en el *splicing* pero también en el *homing*. De acuerdo con estas asunciones, no hemos podido detectar actividad ADN endonucleasa *in vitro* (corte en la cadena sentido) cuando ensayamos estas RNPs mutadas frente a un sustrato de ADNsc de 70 nt de longitud (dato no mostrado).

De todos los datos expuestos podemos sacar dos conclusiones principales. La primera nos dice que, los más que probables problemas conformacionales que hemos inducido en los mutantes de la región de coordinación, son suficientes para inhibir completamente la movilidad aunque todavía permitan que tenga lugar la reacción de *splicing* tanto por transesterificación como por hidrólisis. En segundo lugar, estamos en condiciones de afirmar que la invasión del ADN diana no sólo requiere la presencia de intrón *lariat* escindido sino que este *lariat* debe mantener una adecuada configuración del centro activo.

### II.3.3.3. Conservación filogenética de estas estructuras en intrones IIB

Los datos obtenidos en el análisis de estas estructuras en intrones bacterianos del grupo IIB muestran que la región que rodea el sitio de *branching* está bastante conservada, sin embargo, esta conclusión para la región de coordinación es más arriesgada. El dominio VI, junto con el dominio V, es una de las regiones mejor conservadas en la ribozima desde el punto de vista estructural. Suele representarse como una horquilla de longitud variable que contiene un nt desapareado a 3-4 bp de la base; se ha establecido que la región desapareada que presenta al final de la horquilla interacciona con la región distal

del dII (figura II.20A). Con el fin de definir los determinantes moleculares implicados en la selección del sitio de *branching* se realizó un estudio bastante completo de este dominio en el intrón de mitocondria de levadura  $\alpha 5\gamma$  (Chu *et al.*, 2001). Las principales conclusiones extraídas de nuestro particular análisis de los intrones bacterianos IIB coinciden con parte de las observaciones derivadas del mencionado trabajo: la base del dominio suele estar formada por 4 apareamientos W-C, existe una región rica en purinas alrededor de esta región y los apareamientos que rodean a la A suelen ser, por arriba, R-K y, por abajo, G-Y. Entonces, podemos asumir que estas características conservadas podrían tener una implicación real en el reconocimiento del sitio para la formación del *ariat*, pero también en la selección del sitio de *splicing* 3' aunque sean procesos independientes (Chu *et al.*, 2001; Hamill y Pyle, 2006).

Por otro lado, hemos sido capaces de identificar una estructura que podría ser considerada la región de coordinación a pesar de las diferencias estructurales en casi todos los intrones IIB. La mayoría de las excepciones se han detectado en los intrones IIB5; su región de coordinación suele ser de mayor tamaño que en el resto que a su vez se encuentra considerablemente separada de las otras estructuras situadas por abajo. En relación con esto, es importante resaltar la ausencia en este tipo de intrones de un receptor para el dV bien definido (interacción  $\kappa$ - $\kappa'$ ; este trabajo). Dicho esto, podría ser razonable pensar que existe una relación directa entre la distancia del receptor del dV respecto a la región de coordinación y la distancia constante observada entre el dV y el dVI descrita por varios autores (Boulanger *et al.*, 1996; Podar y Perlman, 1999; Chu *et al.*, 2001). Los intrones del grupo IIC a pesar de ser uno de los grupos más alejados filogenéticamente hablando, mantiene esta región de coordinación. Por el contrario, los intrones IIA presentan una región desapareada en esta misma posición pero claramente diferenciado respecto al grupo IIB; esto puede estar en relación con determinadas interacciones que distinguen ambos tipos de intrones. En particular nos referimos a la interacción  $\delta$ - $\delta'$  y al apareamiento EBS3-IBS3 determinantes en el proceso de *homing* e importantes en *splicing*; en el caso de los intrones IIA la posición IBS3 coincide con la  $\delta'$  por lo que no necesitan una interacción extra que aproxime los dos exones.

Aunque no existe un total consenso en cuanto a la existencia de una relación evolutiva entre los intrones espliceosómicos y los intrones del grupo II, podemos establecer muchos paralelismos estructurales y mecánicos entre ambos (Sharp, 1985; Cech, 1986; Madhani y Guthrie, 1992; Sontheimer y Steitz, 1993; Query *et al.*, 1994; Gaur *et al.*, 1997; Zhang y Doudna, 2002; Hamill y Pyle, 2006). Así podemos considerar que los distintos ARN pequeños nucleares que forman parte del espliceosoma son piezas ensambladas de intrones del grupo II (Sharp, 1991), por ejemplo: el apareamiento U6/sitio de *splicing* 5' en el espliceosoma se corresponde con la interacción  $\epsilon$ - $\epsilon'$  en los intrones del grupo II; la región I de la hélice entre los ARNs U2/U6 con el dominio V; la hélice entre U2 y la A desapareada sería el dominio VI; o el apareamiento de los exones con U5 que podría considerarse equivalente a las EBSs-IBSs de intrones del grupo II (para revisiones, Michel

y Ferat, 1995; Nielsen, 1998; pero también, Parker *et al.*, 1987; Madhani *et al.*, 1990; Fabrizio y Abelson, 1990; Newman y Norman, 1991; 1992; Madhani y Guthrie, 1992; McPheeters y Abelson, 1992; Hetzer *et al.*, 1997a). De acuerdo con esto, el papel de la región de coordinación podría asemejarse con U5, ya que es la región que expone al sitio catalítico los dos exones. Una de las cosas que no conseguimos explicarnos es que si nuestras mutaciones no alteran ni la interacción EBS3-IBS3, ni la  $\delta$ - $\delta'$ , ¿por qué se inhibe la escisión? La respuesta a esta pregunta parece sencilla, esta región además debe estar involucrada en cualquier otro tipo de proceso que es requerido para el *splicing*. Tres resultados experimentales proporcionan evidencias de un posible papel de la región de coordinación en la unión de  $Mg^{2+}$ . La primera es que para que tenga lugar la interacción entre la A del dVI y la región de coordinación *in vitro* es esencial la presencia de  $Mg^{2+}$  en el medio (Hamill y Pyle, 2006). La segunda está basada en la necesidad de un ión metálico con capacidad catalítica dentro del sitio activo durante el *splicing* que active los nucleófilos y estabilice los grupos reactivos en los estados de transición (Steitz y Steitz, 1993; Sontheimer *et al.*, 1999). Según el modelo propuesto (ver figura 5 en las generalidades), el ión  $Mg^{2+}$  coordinaría el 3'-oxianión de los nucleótidos reactivos en los sitios 1, 2 y 3. Los sitios 2 y 3 varían dependiendo de la reacción pero el sitio 1 permanece constante a lo largo de las dos reacciones de *splicing*; este último, está situado en el dominio I en la proximidades de las regiones EBSs, por lo que no sería descabellado pensar que la región de coordinación puede representar este sitio de unión a metales. Finalmente, la tercera evidencia puede estar relacionada con identidad de secuencia con motivos de unión a metales; en U6 se ha sugerido que la penúltima A del motivo conservado ACAGAG podría estar implicada en la unión de  $Mg^{2+}$  (Sontheimer y Steitz, 1993; Sontheimer *et al.*, 1997). Conviene recordar en este punto que los residuos más conservados de la región 3' del bucle de coordinación son AKA, similares a los que se observan en la parte central del hexanucleótido. Este podría ser uno de los motivos que explicasen por qué un cambio puntual como es el A330 a C (AKA) reduce considerablemente la capacidad de escisión.

## **Anexo capítulo II**



**Tabla AII.1.** Oligonucleótidos usados para construir los mutantes en la ribozima de RmInt1.

NOMBRE	SECUENCIA	APLICACIÓN	REFERENCIA
dVI-ΔA	5'-CCTCCGGCCTCTCACCCCTG-3'	Construcción del mutante delecionado de la A <sub>764</sub> (a)	Este trabajo
dVI-AC	5'-CCTCCGGCCTCCTCACCCCTG-3'	Construcción del mutante en el que la A <sub>764</sub> se cambia por C	
dVI-pU	5'-GTTCTGCGAGTAGGCCGGCTG-3'	Construcción del mutante la A <sub>765</sub> se encuentra apareada con una U (736)	
A330C	5'-ACCGGCCTAGCGGGGGCGTGA-3'	Construcción del mutante CL2	
CL/UP	5'-GTGAACCCTCCAGCCCCGCGTCGTTATGAACGTAA	Cebadores internos solapantes para la construcción del mutante pKG4CL1 <sub>17</sub>	
CL/DN	5'-ATAACGACGCGGGGCTGGAGGGTTCACTTTCGTTC		
SacBbr	5'-GGGAGCTCACGTGCCTCGTTTTCATCGATGAGA-3'	Cebadores externos para la construcción del mutante pKG4CL1 <sub>17</sub> .	Nisa-Martínez <i>et al.</i> , 2007
pT7R	5'-TCGACCCTATAGTGAGTCGTATTACTCGAG-3'	Amplifican parte del ΔORF	Barrientos-Durán, 2008

Las posiciones que se marcan en negrita tanto en esta tabla como en la anterior representan las mutaciones introducidas.

(a) La posición enumerada se corresponde con la secuencia del ΔORF



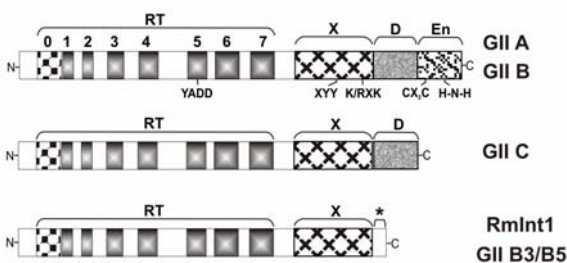
## **Capítulo III**





### III. CARACTERIZACIÓN DE LA REGIÓN C-terminal DE LA PROTEÍNA CODIFICADA POR RmInt1

Inicialmente, el trabajo con intrones mitocondriales de levaduras hizo suponer que las proteínas codificadas por intrones del grupo II presentaban tres dominios bien diferenciados: dominio RT, X y ADN endonucleasa (Mohr *et al.*, 1993; Moran *et al.*, 1994; Zimmerly *et al.*, 1995a; 1995b). Sin embargo pronto se pudo apreciar que esta no era la regla general, que el avance en el conocimiento de nuevos retrointrones revelaba una considerable variedad no sólo de estructuras secundarias para la ribozima sino también de composición en dominios para la IEP. Así, teniendo en cuenta ambos criterios, se habla de tres grandes grupos filogenéticos de intrones del grupo II: IIA, IIB y IIC, que en algunos casos presentan subdivisiones (ver figura 6 en la introducción). De este modo, se considera que las IEPs de los intrones presentes en orgánulos eucariotas (mitocondria y cloroplasto) y los derivados de éstos (intrones IIA, IIB1, IIB2 y IIB4) se componen de cuatro dominios: un dominio reverso transcriptasa (RT) usado por el intrón para copiar su propio ARN en ADNc durante el mecanismo de movilidad; a continuación encontramos un dominio madurasa (X) que los intrones han adaptado para asistirlos durante la maduración disminuyendo la energía libre necesaria en su plegamiento activo; y finalmente, en la región C-terminal se suelen encontrar dos dominios funcionalmente diferenciados, un dominio de unión a ADN (D), usado para el reconocimiento del sitio de inserción en su ADN diana durante la movilidad, y un dominio ADN endonucleasa (En) que sirve al intrón para producir el corte en la cadena de abajo y generar el cebador que necesita la RT para copiar el ARN (figura III.1). Por otro lado, las IEPs de los intrones bacterianos de la clase D y E (o lo que es lo mismo, los intrones IIB3 y IIB5; ver figura 6), entre los que se encuentra RmInt1, presentan únicamente los dominios RT y X, mostrando en lugar de los dominios D y En una pequeña extensión de función desconocida (figura III.1). En una situación intermedia se encuentran los intrones del grupo IIC, cuyas IEPs se componen de dominios RT, X y D, pero carecen de dominio ADN endonucleasa (figura III.1).



**Figura III.1.** Representación esquemática de las IEPs en los distintos grupos de intrones del grupo II. La primera mitad de estas proteínas está compuesta por una serie de dominios RT (RT0-7) seguido de un dominio de unión a ARN conocido como dominio madurasa (X). Algunos grupos presentan además un par de dominios en la región C-terminal: el primero de unión a ADN (D) y un segundo con función ADN endonucleasa (En).

Estos dos últimos dominios están ausentes en los intrones de las clases IIB3 y IIB5 que presentan en su lugar una corta extensión señalada con un asterisco. Las IEPs de intrones de la clase IIC mantienen el dominio D. Bajo el primer esquema se representan los aminoácidos más importantes en cada dominio.

Desde el punto de vista funcional, se ha demostrado que tanto la IEP como el ARN son necesarios para la maduración y movilidad de los intrones del grupo II (Lambowitz *et al.*, 1999; Belfort *et al.*, 2002; Lambowitz y Zimmerly, 2004). Estudios llevados a cabo con intrones del grupo IIA ponen de manifiesto que estos intrones precisan los dominios RT y madurasa para su maduración pero no los dominios D o ADN endonucleasa (Moran *et al.*, 1994; Zimmerly *et al.*, 1995b; Cui *et al.*, 2004). Por otro lado, la delección del dominio En inhibe completamente el corte en la cadena de abajo pero permite la integración del ARN del intrón en la cadena sentido (tanto en sustratos de cadena doble como ADN de cadena simple). La eliminación de ambos dominios, D y En, además impide la unión estable de la proteína al ADN diana de modo que inhibe el *splicing* reverso de la ribozima en ADNdc aunque retiene un ~10% de capacidad de integración en sustratos ADN de cadena simple (Zimmerly *et al.*, 1995b; Matsuura *et al.*, 1997; Guo *et al.*, 1997; Singh y Lambowitz, 2001; SanFilippo y Lambowitz, 2002). A pesar de carecer de dominios D y En de forma natural, RmInt1 es un retroelemento con una capacidad de movimiento comparable con la de otros intrones como Ll.LtrB, aI1 o aI2 (Martínez-Abarca *et al.*, 2000). Además, nuestro grupo ha demostrado que las RNPs de RmInt1, aunque no son capaces de producir corte en la cadena antisentido, se insertan eficientemente en sustratos de ADNsc y de ADNdc *in vitro* (Muñoz-Adelantado *et al.*, 2003).

Según el mecanismo de movilidad descrito para Ll.ltrB (TPRT, ver apartado 2.2.3.1. en la introducción), las carencias funcionales de la IEP de RmInt1 no deberían permitir el *retrohoming* a este intrón, pues carece de dominios que le permitan el reconocimiento del ADN diana y la generación de un cebador para la síntesis del ADNc. Por esto se ha sugerido que la extensión C-terminal característica de este grupo de intrones podría estar implicada de algún modo en la unión al ADN diana (San Filippo y Lambowitz, 2002). Estudios llevados a cabo en nuestro laboratorio en relación con el reconocimiento del sitio de inserción de RmInt1 ponen de manifiesto que ciertas posiciones (T-15 y G+4) son críticas para su movilidad (ver figura 10 en la introducción); en estos estudios no se pudo adscribir a una posible interacción con el ARN del intrón por apareamiento de bases (como ocurre con los contactos EBS/IBS 1, 2 y 3), pudiendo suponer residuos reconocidos por la proteína, posiblemente por su extremo C-terminal (Jiménez-Zurdo *et al.*, 2003). Sin embargo, estudios recientes de maduración *in vitro* de este intrón apuntan a la existencia de una posible interacción EBS2/IBS2<sup>1</sup> alternativa (EBS2/IBS2\*) que tiene lugar entre las posiciones -13 a -18 y que incluiría la posición -15 (Costa *et al.*, 2006b). Por tanto, la contribución de la extensión C-terminal de la IEP de RmInt1 en la unión y reconocimiento del ADN diana vuelve a ser una incógnita. Con el fin de arrojar luz en

---

<sup>1</sup> La interacción EBS1/IBS1 comprende desde la posición -1 a la posición -7 en el ADN diana; junto con ésta, la interacción EBS2/IBS2 (entre las posiciones -9 a -13) contribuye al reconocimiento del exón 1. Por otro lado, el reconocimiento del exón 2 se lleva a cabo por la interacción EBS3/IBS3 que implica a la posición +1 en la diana y que además es característica de intrones IIB (figura 10; Jiménez-Zurdo *et al.*, 2003; Costa *et al.*, 2000).



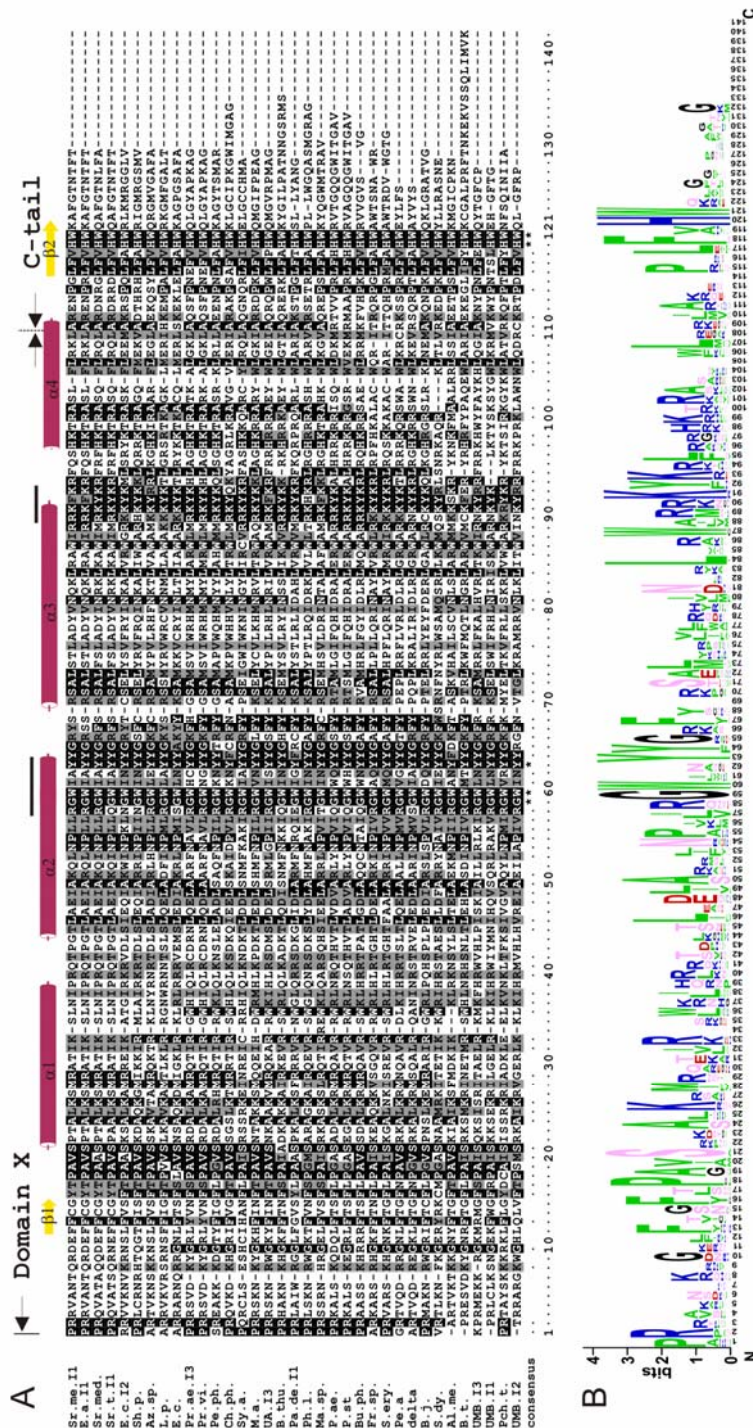
Con el fin de determinar el grado de conservación de esta región en los intrones del grupo IIB3, se ha realizado un alineamiento de las secuencias de la región C-terminal de varias IEPs pertenecientes a este subgrupo (figura III.3A). En este alineamiento se han comparado tanto la secuencia de la madurasa como de la extensión C-terminal de función desconocida. Ambas regiones se encuentran muy conservadas en todas las proteínas utilizadas en este estudio. Destacar el alto grado de invariabilidad encontrado en dos motivos conservados en el dominio X de intrones del grupo II: RGWXNYY (posiciones 58 a 64 en la figura III.3B; posiciones 349-355 en la IEP de RmInt1) y R(K/R)XK (posiciones 90 a 93 en la figura III.3B y 380-383 en la IEP de RmInt1). La estructura secundaria predicha de manera bioinformática para este dominio en la mayoría de los intrones del grupo II consiste en cuatro hélices  $\alpha$  flanqueadas por 2 cadenas  $\beta$  de pequeña longitud (Blocker *et al.*, 2005; figura III.3A); los dos motivos conservados mencionados se situarían en o cerca de la región final de las hélices  $\alpha_2$  y  $\alpha_3$ , respectivamente (figura III.3). Estableciendo una comparativa estructural con la RT del HIV-1<sup>2</sup>, las hélices  $\alpha_1$ ,  $\alpha_2$  y  $\alpha_3$  de intrones del grupo II se corresponderían con las hélices  $\alpha_H$ ,  $\alpha_I$  y  $\alpha_J$  en la región del pulgar de la RT retroviral (Blocker *et al.*, 2005; ver apartado 2.1.3. en la introducción). Sin embargo, la última hélice  $\alpha$  en intrones del grupo IIB3,  $\alpha_4$ , no tiene una estructura equivalente en la RT de HIV-1 pero tampoco en la proteína LtrA del intrón de *L. lactis*. En su lugar, en ambas proteínas se describe la presencia de tres cadenas  $\beta$  relacionadas con dimerización (Blocker *et al.*, 2005). La mutación de una lisina conservada en esta región en la proteína LtrA (K483) afecta su actividad madurasa (Cui *et al.*, 2004). La posición equivalente en intrones IIB3 sería una leucina extraordinariamente conservada en las IEPs estudiadas (posición 107 en figura III.3B) y que se correspondería con la leucina en la posición 396 de la IEP de RmInt1.

Como se puede observar en el alineamiento, la región aguas abajo del dominio X es muy similar entre los intrones IIB3 (figura III.3). Cabe destacar especialmente el motivo LFXHW en las posiciones 117-121 (figura III.3B; posiciones 406-410 en la IEP de RmInt1) cuya predicción de estructura secundaria coincide con una corta cadena  $\beta$  denominada  $\beta_2$  en el alineamiento (figura III.3A). A diferencia de lo que se había establecido en estudios anteriores<sup>3</sup>, la información que se desprende de nuestro análisis de la porción C-terminal de la secuencia proteica de los intrones de la subclase IIB3 indicaría que el dominio X finaliza en la posición 106 (figura III.3), de modo que la extensión C-terminal podría definirse por el motivo conservado LX<sub>3</sub>AX<sub>3</sub>PXLF(V/A)HW (que se corresponde con las posiciones 396 a 410 en la proteína de RmInt1). El grado de conservación que caracteriza a esta región en las IEPs de intrones IIB3 podría significar que este fragmento es estructural y funcionalmente importante para esas RTs.

---

<sup>2</sup> Recordar que las IEPs de intrones del grupo II se consideran emparentadas con las RTs de retrovirus y de elementos retrotransponibles sin repeticiones terminales invertidas (no-LTR) (Xiong y Eickbush, 1990).

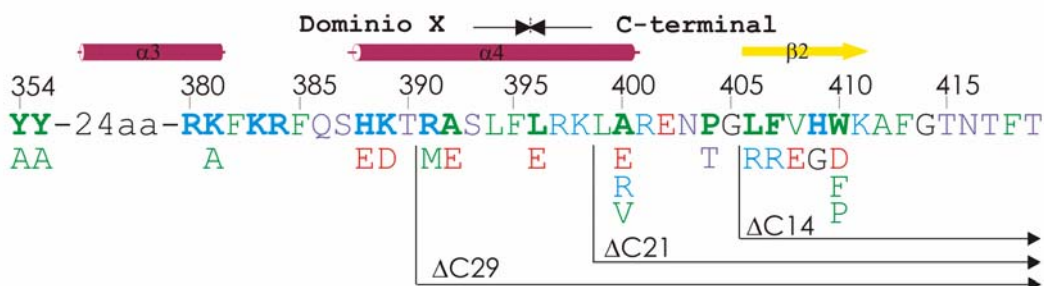
<sup>3</sup> San Filippo y Lambowitz (2002) consideraron que el dominio X se extiende hasta la posición A400 en RmInt1, siendo la longitud de la extensión C-terminal de 20aa.



**Figura III.3.** (A) Alineamiento múltiple de la secuencia de la región C- terminal de la IEP de RmInt1 con otras proteínas codificadas por intrones del grupo IIB3 usando Clustal W. Los aminoácidos resaltados en negro presentan > 75% de identidad, mientras que los sombreados en gris tienen > 50% de identidad. Para el sombreado se usó el programa Boxshade (bioweb.pasteur.fr/sequal/interfaces/boxshade.html). Para la predicción de la estructura secundaria En la parte superior del alineamiento se muestra la predicción de estructura secundaria realizada con el programa JPred: en fucsia se representan las hélices  $\alpha$  y en amarillo, las cadenas  $\beta$ . Los residuos idénticos son señalados con un asterisco y los motivos conservados con barras sobre la estructura predicha. Los intrones se han nombrado de acuerdo con la nomenclatura propuesta por Zimmerly ([www.fp.ucalgary.ca/group2/introns/](http://www.fp.ucalgary.ca/group2/introns/)) y se muestran en la tabla AIII.1 en el anexo. (B) Secuencia logo para las IEPs de los intrones IIB3 (http://weblogo.berkeley.edu). Está basado en el alineamiento mostrado en la parte A, asignando un máximo de 4 bits cuando la posición es idéntica en todas las secuencias. Los aa están coloreados según sus propiedades: básicos, en azul (K, R y H); ácidos en rojo (D y E); hidrofóbicos, verde (P, L, I, V, M, F, w, Y y A); polares, en púrpura (N, Q, S y T); y neutros, en negro (G y C).

### III. 2. ANÁLISIS DE LAS MUTACIONES EN EL C-terminal DE LA IEP DE RmInt1

Se construyeron una serie de mutantes en posiciones conservadas de la región C-terminal de la IEP de RmInt1 que nos permitiesen determinar características funcionales importantes en esta zona de la proteína (figura III.4; ver oligonucleótidos en tabla AIII.2 AIII.3). En concreto, se analizaron tres deleciones que se corresponden con motivos estructurales en esta región. Más en detalle se analizan posiciones puntuales en las que el aminoácido original es sustituido por otro de naturaleza diferente, con el fin de observar un efecto dramático. Además, quisimos comprobar la importancia del residuo por el que se sustituye el aminoácido original en dos posiciones en concreto, realizando hasta tres mutantes distintos para un mismo residuo.



**Figura III.4.** Diseño de las mutaciones realizadas en el extremo C-terminal de la proteína codificada por RmInt1. En el centro se muestra la secuencia correspondiente a la región final del dominio madurasa y la extensión C-terminal, acotadas por flechas. Los residuos más conservados se resaltan en negrita. Sobre la secuencia, las posiciones correspondientes en la proteína y la predicción de estructura secundaria. En la parte inferior de la secuencia, y bajo su correspondiente posición, se indican los cambios realizados. En el caso de las posiciones 400 y 410 se construyeron tres mutantes distintos, que se representan uno debajo de otro. Las deleciones practicadas se señalan con flechas acodadas. El código de colores es el mismo que el que encontramos en la secuencia logo (ver figura III.3).

A su vez, aunque los estudios se iniciaron en construcciones del tipo RmInt1 (pKG2.5), la mejora de las prestaciones bioquímicas con construcciones pKGEMA4 nos llevaron a construir algunos de los mutantes en este sistema. En concreto, este tipo de construcciones en las que la ribozima es de menor longitud ( $\Delta$ ORF presenta 740 nt frente a los 1884 nt de RmInt1) nos permitieron llevar a cabo ensayos de *splicing* reverso. En general, los resultados obtenidos con nuestros mutantes en construcciones derivadas de pKGEMA4 fueron similares a los que se obtuvieron con pKG2.5; si bien, debemos puntualizar que, debido al mayor rendimiento tanto en procesos *in vivo* (*splicing* y *retroboming*) como en actividades *in vitro* (RT y corte en la cadena de arriba) (figura AIII.1 en anexo), los valores relativos para todos los procesos muestran niveles superiores en mutantes construidos en plásmidos derivados de pKGEMA4 respecto de pKG2.5. Los datos referentes a construcciones pKGEMA4 se incluyen en el anexo (tabla AIII.4 y en

las figuras AIII.2 y 3). Además, la evaluación de la eficiencia de movilidad se estudió mediante el sistema de selección de colonias tetraciclina resistentes en la que el  $\Delta$ ORF en el plásmido donador contiene en el promotor de la ARN polimerasa del fago T7 y el plásmido receptor presenta un gen de resistencia a tetraciclina que carece de promotor. Este sistema permite la detección de eventos de *homing* muy por debajo de los niveles revelados por el Southern blot (sistema en el que vamos a basar nuestras observaciones) que no resultan significativos a la hora de explicar los efectos observados por los mutantes (por esta razón, se incluyen en el anexo, tabla AIII.5).

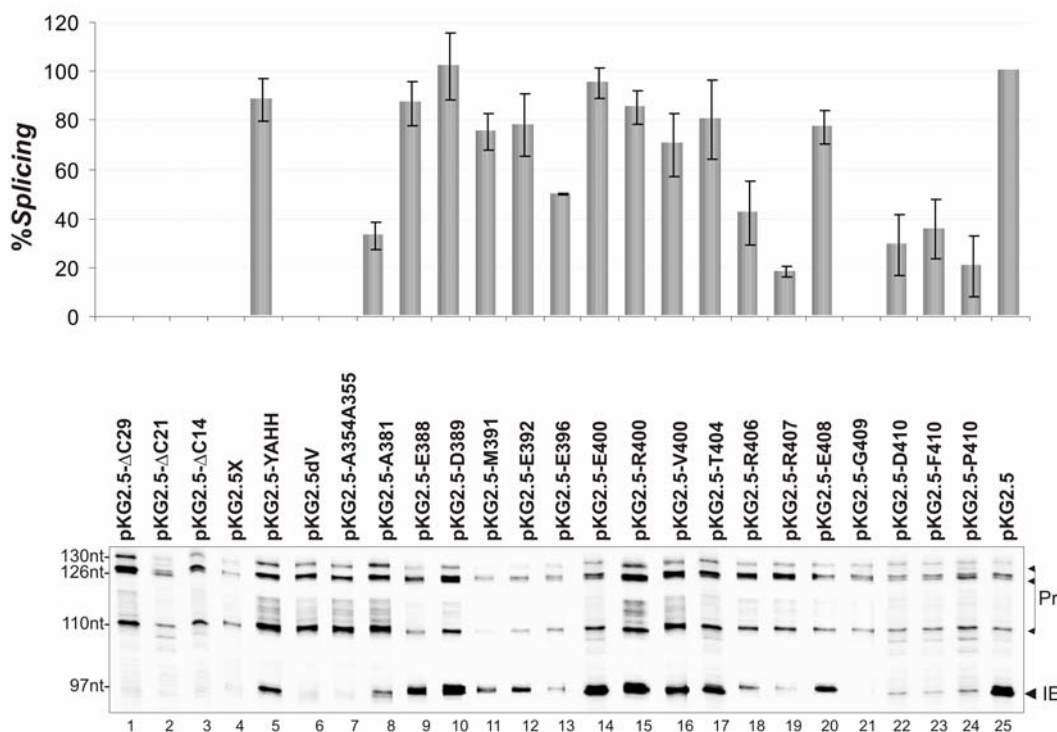
### III. 2.1. EFECTO DE LAS DIFERENTES DELECCIONES

La deleción de mayor longitud ( $\Delta$ C29) en la que la IEP pierde los últimos 29 aa, entre los que se encuentran los residuos finales del dominio X, no permitió la detección de formas derivadas de maduración (figura III.5, carril 1). Del mismo modo, mutantes con deleciones más cortas que no se encuentran dentro del dominio madurasa que pierden la totalidad o parte de la extensión C-terminal ( $\Delta$ C21 y  $\Delta$ C14, respectivamente) tampoco mostraron niveles de *splicing* detectables por extensión de cebador (figura III.5, carriles 2 y 3). Se obtuvieron resultados similares al ensayar el mutante YY $\rightarrow$ AA que contiene la doble sustitución de las tirosinas consecutivas (Y354 e Y355) del motivo conservado RGWXNYY del dominio madurasa por dos alaninas (figura III.5, carril 7). Las cuatro mutaciones presentaron unos niveles de actividad RT muy disminuidos ( $\leq 5\%$  de los niveles WT) junto con niveles indetectables de actividad de corte de sustratos de ADN (Tabla III.1 y figura III.6). De acuerdo con todo lo expuesto ninguno de los mutantes presentaba la capacidad de transponerse a su diana en el plásmido receptor (figura III.8, carriles 1, 2, 3 y 7). Otros mutantes usados como controles presentaron un fenotipo similar; tal es el caso de los mutantes pKG2.5X (presenta un cambio en la fase de lectura hacia el primer tercio de la secuencia de la IEP por lo que se genera una proteína truncada; carril 4 en figuras III.5 y III.8 y carriles 3 y 25 en figura III.6) o pKG2.5dV (mutado en la región de la triada catalítica en el dominio V de la ribozima; carril 6 en figuras III.5 y III.8 y carriles 2 y 24 en figura III.6). En cambio, un mutante afectado en el centro catalítico de la actividad RT (pKG2.5-YAHH, en el que dos aspárticos (DD) son sustituidos por dos histidinas en el dominio RT5) no muestra movilidad pero lleva a cabo el resto de actividades de manera ligeramente reducida o similar respecto del intrón silvestre (figuras III.5 y III.8, carril 5; figura III.6, carriles 4 y 26).

Contrastando con el importante efecto sobre la actividad del intrón que tiene la mutación YY $\rightarrow$ AA, se comprobó que la mutación de la última lisina (K381  $\rightarrow$ A) en el otro motivo conservado presente al final de la hélice  $\alpha 3$  del dominio X, R(K/R)XK, aún conserva ciertos niveles de *splicing* (33% del nivel WT; figura III.5, carril 8). Sin embargo, este mutante tiene muy afectados los niveles de actividad RT (6% WT, tabla III.1) lo que probablemente contribuya a la ausencia de movilidad (figura III.8, carril 8). Tampoco fuimos capaces de detectar corte en dianas de ADN de cadena simple ni doble (figura III.6, carriles 6 y 28, respectivamente). Por tanto, el patrón de inhibición de actividades



que muestran los mutantes por delección y el mutante YY→AA coincide con alteraciones estructurales importantes de la IEP, mientras que el residuo K381 podría estar implicado en alguna interacción puntual de la proteína con el ARN del intrón necesaria para que tenga lugar el *splicing* de manera eficiente. Esta mutación inhibe los niveles de *retroboming* debido probablemente a la inestabilidad de las RNPs que se generan lo que coincide con los mínimos niveles de actividades bioquímicas de estos complejos medidas *in vitro*. Del análisis de estas mutaciones podemos concluir que la integridad de esta región es importante no sólo desde el punto de vista estructural (resulta esencial para el mantenimiento de la estructura activa de la proteína) sino también funcional (el defecto en movilidad supera los efectos en maduración).



**Figura III.5.** Análisis de la eficiencia del *splicing in vivo* en RmInt1 por extensión de cebador. En el panel inferior se muestra una composición con una selección de carriles representativos de cada mutante. Se corresponde con los productos resultantes de la síntesis de ADNc a partir del cebador P en muestras de ARN total extraído de células de *S. meliloti* RMO17 transformada con los distintos plásmidos donadores de intrón mutado. En los laterales se señalan el tamaño de las bandas y su identificación correspondiente (Pr, precursor, e IE, intrón escindido). En la parte superior representamos las eficiencias de maduración de los diferentes mutantes como porcentaje del intrón silvestre, pKG2.5. La eficiencia absoluta se calculó según la razón producto escindido (IE)/todos los productos generados (Pr + IE). Los datos son el promedio de los valores procedentes de al menos 2 determinaciones con dos muestras extraídas de manera independiente; las barras de error se corresponden con la desviación estándar de la media. Además de los mutantes de la región C-terminal en el ensayo se usaron diferentes controles explicados en el texto.

**Tabla III.1.** Actividades bioquímicas de las RNPs del intrón silvestre y de los diferentes mutantes.

Construcciones	Actividad RT exógena <sup>a</sup>	Corte de sustrato ADNsc 70 nt <sup>b</sup>	Corte de sustrato ADNdc 70 bp <sup>c</sup>
pKG2.5-ΔC29	5 ± 2	ND	ND
pKG2.5-ΔC21	2 ± 0	ND	ND
pKG2.5-ΔC14	3 ± 1	ND	ND
pKG2.5D5-CGA	11 ± 2	ND	ND
pKG2.5X	5 ± 3	ND	ND
pKG2.5-YAHH	4 ± 3	46 ± 8	88
pKG2.5-A354A355	4 ± 3	ND	ND
pKG2.5-A381	6 ± 2	ND	ND
pKG2.5-E388	51 ± 16	24 ± 4	118
pKG2.5-D389	57 ± 17	26 ± 9	89 ± 10
pKG2.5-M391	14 ± 2	47 ± 22	42
pKG2.5-E392	21 ± 3	107 ± 30	48
pKG2.5-E396	8 ± 3	18 ± 9	ND
pKG2.5-E400	57 ± 24	117 ± 0	51 ± 23
pKG2.5-R400	30 ± 12	65 ± 14	55
pKG2.5-V400	12 ± 3	35 ± 1	48
pKG2.5-T404	10 ± 4	47 ± 5	38 ± 13
pKG2.5-R406	8 ± 2	ND	ND
pKG2.5-R407	8 ± 2	ND	ND
pKG2.5-E408	9 ± 2	172 ± 16	60
pKG2.5-G409	6 ± 1	ND	ND
pKG2.5-D410	5 ± 2	ND	ND
pKG2.5-F410	10 ± 4	ND	ND
pKG2.5-P410	7 ± 2	ND	ND
pKG2.5	100	100	100

(a) La actividad RT se estableció en función de la capacidad de incorporación de [ $\alpha$ -<sup>32</sup>P]dTTP por preparaciones enriquecidas en RNPs extraídas de bacterias con las construcciones indicadas usando un sustrato exógeno, poli(rA)/oligo(dT)<sub>18</sub>.

(b) El corte de sustratos de ADN de cadena simple resultó de la incubación de las preparaciones de RNPs arriba indicadas con un oligonucleótido de 70 nt marcado radiactivamente en 5' que contiene la secuencia de reconocimiento de RmInt1.

(c) El corte de la cadena de arriba en sustratos de ADNdc se realizó con productos de PCR de 70 bp con la cadena superior marcada en 5'

En los tres casos las actividades se expresan como porcentaje del comportamiento en pKG2.5. Los valores se corresponden de los resultados obtenidos en al menos dos ensayos independientes usando un mínimo de dos preparaciones de RNPs diferentes y se representan como la media ± S.D. ND significa no detectado.

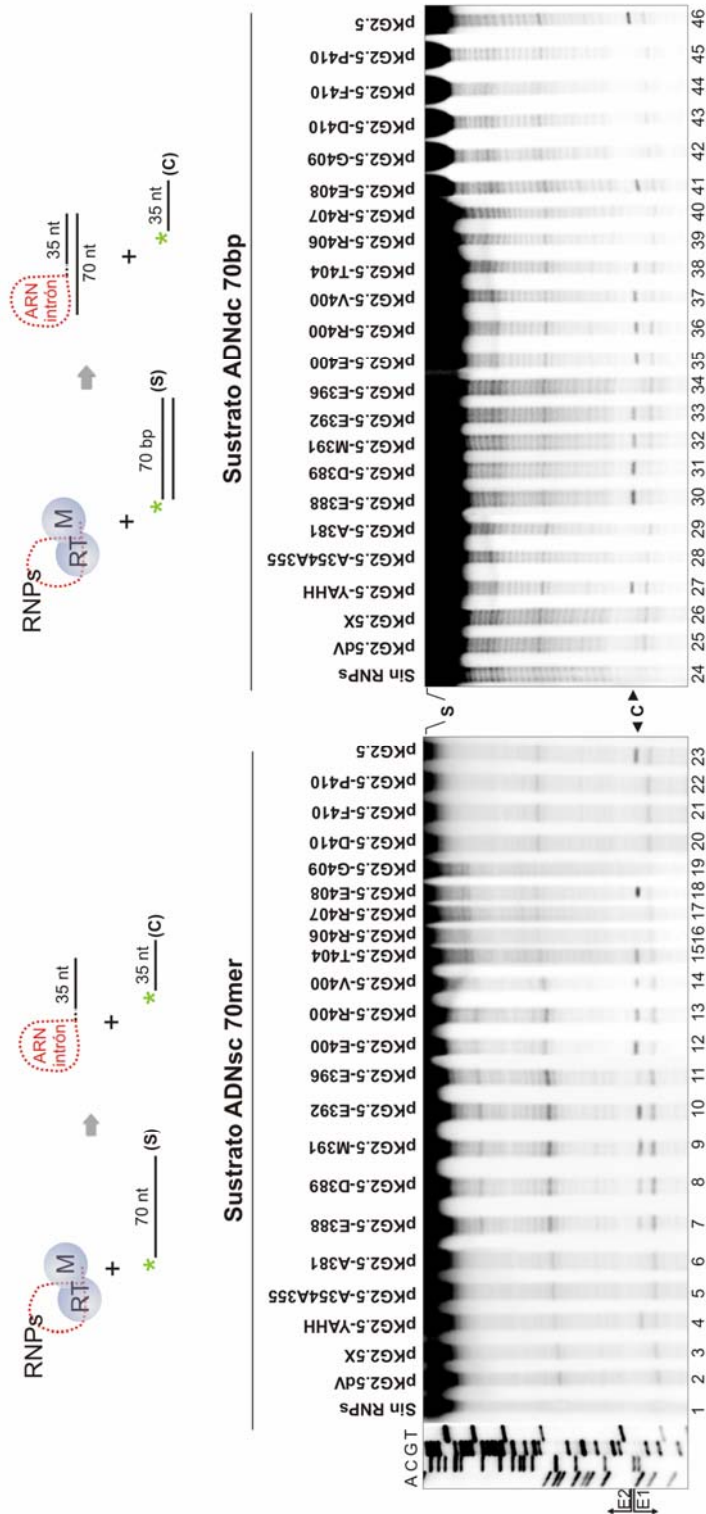
### III. 2.2. ANÁLISIS DE MUTACIONES EN RESIDUOS CONSERVADOS DE LA HÉLICE $\alpha 4$

Hacia el final del dominio madurasa las IEPs de los intrones IIB3 presentan la secuencia conservada HKXRA compuesta de aminoácidos básicos situados en la porción N-terminal de una hélice  $\alpha$  (posiciones 98-102 en figura III.3). Algunos de los residuos de este motivo, así como la estructura secundaria predicha, también se encuentran conservados en las IEPs de otros intrones del grupo II (Blocker *et al.*, 2005). Dentro de la misma estructura pero formando parte de la porción C-terminal encontramos otro par de residuos hidrofóbicos muy conservados entre los intrones de la clase IIB3 pero que ya no muestran correspondencia con la misma región en otras RT: una leucina en posición 107 del alineamiento (L396 en RmInt1) y una alanina en posición 111 (A400 en RmInt1) (figura III.3). A diferencia de lo que habíamos observado en anteriores mutantes afectados en dominios característicos del dominio X (YY $\rightarrow$ AA y K381A), mutaciones en los residuos H388, K389, R391, A392 y L396 mostraron una inhibición moderada de los niveles de *splicing* (87%, 100%, 75%, 78% y 50% del nivel WT, respectivamente; figura III.5, carriles del 9 al 13), a pesar de que en todos los casos los aminoácidos fueron reemplazados por otros con propiedades y estructura muy diferente (figura III.4). Una de las primeras conclusiones que se podrían extraer a partir de estos datos es que esta región no parece estar implicada en la escisión del intrón. Los mutantes H388E y K389D retienen unos niveles sustanciales de actividad RT (51% y 57% WT), mientras que las mutaciones R391M, A392E y L396E mostraron un mayor nivel de inhibición (14%, 21% y 8%, respectivamente; tabla III.1). De igual modo, la mayoría los mutantes mantuvieron la capacidad de rotura del ADN en sustratos de cadena simple y doble (tabla III.1 y figura III.6, carriles 7-10 y 29-32), aunque el mutante L396E sólo retuvo una actividad residual sobre sustratos de ADNsc (18% del nivel silvestre; tabla III.1 y figura III.6, carril 33 comparado con carril 11). Estas observaciones sugieren que la conformación de la proteína no debe encontrarse especialmente alterada en estos mutantes ya que se están produciendo RNPs activas. Sin embargo, cuando analizamos la eficiencia de *retroboming* en esos mutantes pudimos observar que en todos, excepto en aquel en la posición R391, se producía una inhibición del proceso hasta niveles indetectables independientemente de la orientación del sitio de inserción respecto a la dirección de la replicación<sup>4</sup> (figura III.8, carriles 9, 10, 12 y 13). Así, la movilidad del mutante R391M se restringió a dianas

---

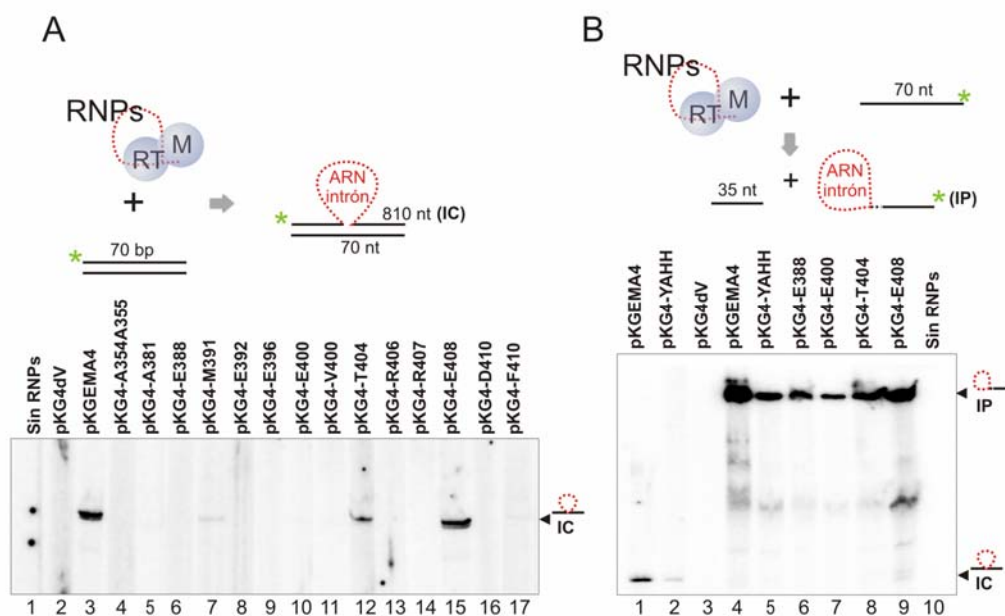
<sup>4</sup> Recordar que este tipo de intrones que carecen de los dominios D y En presentan una ruta mayoritaria de *retroboming* asociada con la replicación en la que utilizan como cebador para la reverso transcripción bien los cebadores de ARN generados por la primasa o los propios fragmentos de Okazaki que sintetiza la ADN polimerasa celular durante la síntesis de la hebra retardada. Por tanto, el intrón se insertaría en la hebra que sirve como molde para la síntesis de la cadena retardada durante la replicación. Por esta razón hemos adoptado la terminología LAG (del inglés *lagging*) y LEAD (del término *leading*) para designar la orientación de la diana respecto del sentido de la replicación (Martínez-Abarca *et al.*, 2004). Ver apartado 2.2.3.2. en la introducción.

situadas en la misma dirección que la horquilla de replicación en el plásmido pJB0.6 (pJB0.6LAG), alcanzando el 34 % de eficiencia de *retroboming*.



**Figura III.6.** Efecto de las diferentes mutaciones en la región C-terminal de la IEP de RmInt1 en la rotura de sustratos de ADN en el sitio de inserción del intrón. En el panel izquierdo se muestran los resultados obtenidos cuando las RNPs son incubadas (condiciones en Material y Métodos) con un oligonucleótido de 70 nt de longitud marcado en 5' con [ $\gamma$ - $^{32}$ P]; este sustrato es simétrico, esto es, el sitio de inserción del intrón se encuentra hacia la mitad del oligonucleótido dando lugar a una banda de corte de 35 nt como se esquematiza en la parte superior del gel. Como marcador se ha utilizado una escalera de secuenciación sintetizada a partir de un cebador complementario al extremo 5' del sustrato, S70ds/UP. En el panel derecho se muestra una composición representativa de los resultados obtenidos para el corte de la cadena sentido (marcada radiactivamente en su extremo 5') de un sustrato de ADN de cadena doble. El producto de corte (C) tiene el mismo tamaño que en el ensayo anterior, como se puede comprobar en el diagrama de la parte superior. S hace referencia al sustrato que en un caso es ADNsc y en el otro ADNdc.

Este fenotipo en el *retrohoming* podría ser explicado si atendemos a la capacidad de integración del ARN del intrón una de las hebras del ADN en sustratos de doble cadena (estudios de *splicing* reverso llevados a cabo con construcciones pKGEMA4). Los niveles de *splicing* reverso completo en los mutantes H388E, A392E y L396E fueron indetectables, en cambio se pudo apreciar una reducción del 83% en los niveles de integración completa del ARN del intrón (figura III.7A, carriles 6, 8 y 9). En cambio, la cuantificación del *splicing* reverso parcial en sustratos de ADN de cadena sencilla marcados radiactivamente en el extremo 3', permitió apreciar niveles similares de integración a los niveles de corte de sustratos en la cadena sencilla que se habían observado en este mismo mutante (figura III.7B, carril 6). Es conveniente destacar en este punto que fuimos incapaces de detectar *splicing* reverso completo con sustratos ADNsc probablemente debido a que, en este tipo de sustratos *in vitro*, las interacciones entre el ARN del intrón y su diana de cadena simple (EBSs/IBSs) sean más inestables que cuando en la reacción está presente la otra hebra complementaria; esta cadena se encargaría de mantener próximos ambos exones una vez que se ha producido el corte en



**Figura III.7.** Comprobación de la capacidad de integración del ARN de algunos de los mutantes en los sustratos correspondientes, ensayos de *splicing* reverso. (A) Se incubaron RNPs de construcciones derivadas de pKGEMA4 con un sustrato ADNdc 70 bp. Con este tipo de aproximación únicamente se pueden detectar eventos de invasión completa generándose un producto de 810 nt (35 nt + 740 nt + 35 nt) (ver esquema en la parte superior). (B) Se trata del mismo tipo de ensayo pero en esta ocasión se utilizó un sustrato ADNsc 70 nt marcado en su extremo 3' para poder detectar los fenómenos de invasión parcial de la diana. Este tipo de marcaje también debería revelarnos los eventos de invasión completa sin embargo, aunque tuvimos la precaución de correr en paralelo muestras del ensayo mostrado en el panel A, no fuimos capaces de detectar niveles significativos *splicing* reverso completo. El retraso en la movilidad de las formas parcialmente integradas es debido a la estructura en lazo del intrón.

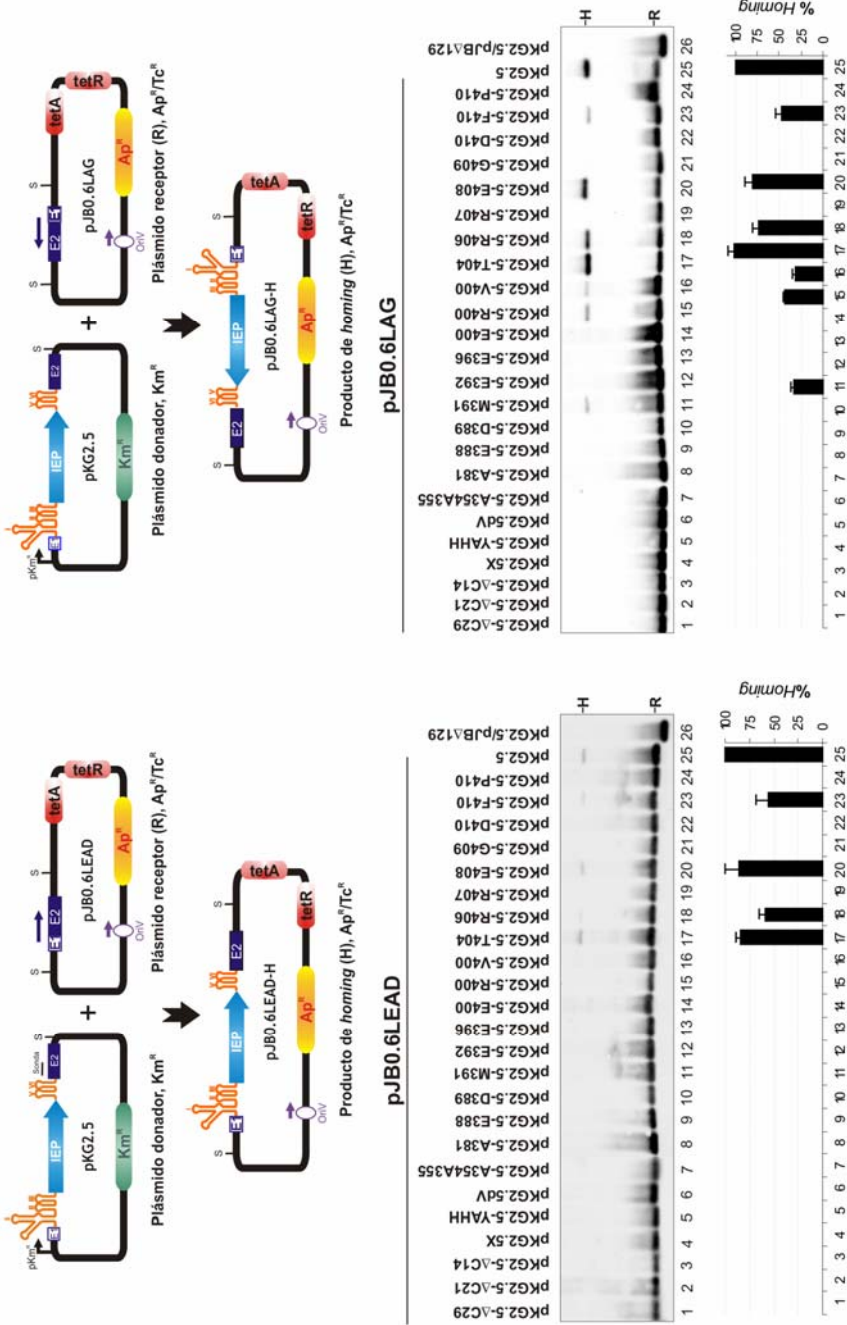
la cadena de arriba facilitando así que se complete la reacción<sup>5</sup>.

En cuanto al mutante A400E nos recuerda al comportamiento de otros residuos mutados aguas arriba en la misma hélice  $\alpha 4$  (niveles de *splicing* similares al WT, 95%, pero efecto deletéreo en *homing*) (figura III.5 y III.8, carril 14), por lo que podríamos considerar que forma parte de la misma región funcional en la proteína de acuerdo a su correspondencia estructural. Sin embargo, pudimos observar cambios en el fenotipo de los mutantes en esta posición dependiendo del cambio aminoacídico elegido para la mutagénesis. Esta alanina fue cambiada a otro aminoácido hidrofóbico, valina (V), y también se construyó un mutante que presentaba una arginina (R, aminoácido básico) en esta posición; ambas sustituciones tienen lugar en otros intrones del grupo IIB3 (figura III.3, posición 111). De manera similar al comportamiento observado en el mutante A400E, ambos cambios permitieron a estos mutantes madurar *in vivo* (70% y 85% de los niveles WT, respectivamente; figura III.5, carriles 15 y 16). La capacidad de las RNPs para cortar sustratos de ADNdc *in vitro* fue similar en los tres mutantes (tabla III.1 y figura III.6, carriles 34-36), pero diferían en el corte de ADN cadena sencilla mostrando niveles superiores la mutación A400E respecto de la otras dos (tabla III.1 y figura III.6, carriles 12-14). Similar a este último fue el efecto detectado en la actividad RT que pasa de un 52% en el mutante A400E a un 30% y un 12% en los mutantes A400R y A400V, respectivamente. Estas diferencias no son suficientes para explicar el singular comportamiento que estas mutaciones mostraron en *retrohoming*; al igual que el mutante R391M, las sustituciones A400V y A400R mostraron un 31% y un 44% de la actividad WT cuando se transformaban con plásmidos receptores en los que la diana se encuentra clonada en el sentido de la horquilla de replicación (figura III.8, carriles 15 y 16). Un dato que podría arrojar luz a este respecto es que el mutante A400E no muestra niveles detectables de *splicing* reverso completo (figura III.7A, carril 10) aunque fuimos capaces de detectar niveles considerables de *splicing* reverso parcial (figura III.7B, carril 7).

En resumen, podemos considerar que los residuos presentes en la hélice  $\alpha 4$  podrían estar implicados en el mecanismo de movilidad del intrón, en concreto podrían ser requeridos para el *retrohoming* en dianas que precisen la utilización de la hebra conductora como cebador para la síntesis del ADN complementario al ARN del intrón.

---

<sup>5</sup> Es evidente que el marcaje en 5' de la cadena de arriba en los sustratos ADNdc no permite la detección de *splicing* reverso parcial. La solución podría ser el marcaje interno de la molécula durante la polimerización pero hemos comprobado que la eficiencia de incorporación de nucleótidos marcados en sustratos tan pequeños es bastante ineficiente por lo que no obtenemos suficiente sustrato marcado que nos permita realizar estos ensayos (dato no mostrado). En estudios anteriores de este tipo de actividad se usaron sustratos de mayor longitud (Muñoz-Adelantado *et al.*, 2003).



**Figura III.8.** Eficiencia de movilidad *in vivo* del intrón silvestre pKG2.5 y los diferentes mutantes en la región C-terminal de la IEP. *S. meliloti* RMO17 fue cotransformada con una pareja de plásmidos: un plásmido donador de intrón (silvestre o mutado, pKG2.5) y un plásmido receptor que contiene la diana de inserción de RmlInt1 clonada en la misma (pJB0.6LAG) o en orientación opuesta (pJB0.6LEAD) a la dirección de la horquilla de replicación (ver los esquemas correspondientes en la parte superior del gel). El conjunto de plásmidos presentes en las bacterias fue analizado por digestión *Sma*I (S), electroforesis e hibridación ADN-ADN con una sonda específica del exón 2 (ver esquema en la parte superior). El porcentaje de invasión del ADN diana *in vivo* se representa en gráficos de barras en la parte inferior de los gels relativo al intrón silvestre. Los datos se corresponden con la media de los plásmidos extraídos de al menos 4 transconjugantes en dos ensayos independientes; mientras, las barras de error se corresponden con la S.D.

### III. 2.3. RESULTADO DE LA MUTACIÓN DE RESIDUOS CONSERVADOS EN LA CADENA $\beta 2$

Los siguientes mutantes en ser analizados comprendían desde la posición 404 a la 410 dentro del motivo conservado PXL $F(V/A)$ HW (posiciones 115-121 en figura III.3); esta región se encuentra dentro de la considerada como extensión C-terminal independientemente de la definición del dominio madurasa. En esta región dos mutaciones dan lugar a fenotipos similares, P404T y V408E. En ambos casos mantienen niveles considerablemente altos de *splicing* (80% y 77% del WT, respectivamente; figura III.5, carriles 17 y 20). La actividad de sus RNPs *in vitro* es desigual; mientras la capacidad para producir el corte en sustratos de ADN $_{sc}$  y ADN $_{dc}$  se reduce aproximadamente a la mitad (salvo el corte de V408E en ADN $_{sc}$  que se mantiene e incluso supera los niveles silvestres; figura III.6, carriles 15, 18, 38 y 41), la actividad RT en ambos mutantes se reduce considerablemente (tabla III.1). Las actividades de corte se ven refrendadas por la presencia de dianas invadidas completa y parcialmente por el intrón  $\Delta$ ORF en ambos mutantes (figura III.8). La particularidad que las distingue de otros mutantes en la hélice  $\alpha 4$  es que son capaces de invadir su diana independientemente de la orientación con la que se clone respecto a la dirección de la horquilla de replicación con una eficiencia similar al intrón silvestre (en torno al 80% en pJB0.6LEAD y entre el 80 y el 100% en pJB0.6LAG; figura III.8, carriles 17 y 20).

Otras posiciones situadas en la cadena  $\beta 2$  mostraron un efecto más importante en sus niveles de *splicing*; mientras que la reducción detectada en los mutantes L406R, F407R y W410D alcanzó en algún caso el 80% (eficiencias de *splicing* del 42%, 18% y 29%, respectivamente; figura III.5, carriles 18, 19 y 22), la sensibilidad de la técnica no nos permitió detectar *splicing* en el mutante H409G (figura III.5, carril 21). En todos los mutantes la actividad RT se encontraba por debajo del 8% de los niveles silvestres (tabla III.1) y no se detectó capacidad de corte de sustratos ADN (figura III.6, carriles 16, 17, 19, 20, 39, 40, 42 y 43). Estos datos coinciden con los resultados de *splicing* reverso testado en alguno de estos mutantes, esto es, ni L406R, ni F407R ni W410D invadían el ADN diana (figura III.7A, carriles 13, 14 y 16). Este comportamiento induce a pensar que estos residuos podrían intervenir de alguna manera preservando la estructura activa de la proteína. Es razonable pensar, por tanto, que estos mutantes deberían de tener abolida su capacidad de movimiento; esta afirmación es cierta pero sólo para los mutantes en las posiciones 407, 409 y 410D. La sustitución en la posición 406 de una leucina por una arginina sigue manteniendo niveles sustanciales de *retroboming* en ambas orientaciones de la diana (59% en pJB0.6LEAD y 73% en pJB0.6LAG; figura III.8, carril 19). Estos datos se ven refrendados con los resultados obtenidos tanto por el sistema pKGEMA4 como en estudios de movilidad por selección de colonias resistentes a tetraciclina (ver tablas AIII.3 y AIII.4). Los resultados obtenidos en este mutante inducen a pensar que, aunque se deben estar produciendo RNPs activas *in vivo* que soportan los niveles de *retroboming* descritos, su comportamiento *in vitro* sólo puede ser atribuible a un descenso



en los niveles de proteína activa que se traduce en importantes reducciones en sus actividades bioquímicas.

De manera similar a la comprobación realizada en la posición 400 de la proteína, se seleccionó la posición W410 para comprobar el efecto que distintas mutaciones puede ocasionar sobre las actividades *in vivo* e *in vitro* de RmInt1. Así, un triptófano (W) conservado en todas las IEPs de intrones IIB3 estudiadas (figura III.3B, posición 121) se sustituyó por una prolina (P), que distorsionaría la posible estructura en cadena  $\beta$  predicha en esta región, y por una fenilalanina (F), un aminoácido aromático de naturaleza similar (figura III.4). El cambio a prolina mostró un fenotipo similar al que hemos descrito en el caso del mutante W410D. Sin embargo, la mutación W410F mostraba, en general, niveles superiores (no espectaculares) en todas las actividades siendo el dato más llamativo que es capaz de moverse con una eficiencia del 55% y el 47% a dianas en pJB0.6LEAD y LAG, respectivamente (figura III.8, carril 23).

Todos los datos mostrados parecen demostrar que los residuos de la región  $\beta$ 2 forman una parte de la IEP distinta estructural y funcionalmente del resto, siendo el aminoácido H409 el residuo más crítico de esta región conservada de acuerdo a los estudios filogenéticos realizados al principio de este capítulo (figura III.3B, posición 120). Esta región contribuiría a la estabilización de la conformación activa de la proteína.

### III. 3. DISCUSIÓN

En este trabajo hemos tratado de caracterizar la parte C-terminal del dominio X de la IEP de RmInt1. Los alineamientos de la secuencia de esta región en intrones de la subclase bacteriana IIB3 ponen de manifiesto un elevado grado de conservación, definiendo el motivo  $LX_3AX_3PXL(V/A)HW$  como una secuencia característica de este grupo de retrointrones. Aunque es probable que esta extensión forme parte del dominio madurasa ya que no se le ha podido adscribir una función concreta independiente de éste (por ejemplo, el mutante H409G se encuentra muy afectado en *splicing*), la longitud que se le asignó en los primeros trabajos (San Filippo y Lambowitz, 2002) pudiera incrementarse en 4 aminoácidos de acuerdo a los datos aportados en este trabajo.

El comportamiento de mutantes puntuales dentro de dos de las regiones conservadas entre el dominio madurasa de todos los intrones del grupo II fue claramente distinto; mientras el mutante  $YY \rightarrow AA$  presentaba una inhibición total de la capacidad de maduración, el mutante K381A retuvo un 30% de los niveles WT (figura III.5, carriles 7 y 8). El residuo equivalente a la K381 de RmInt1 descrito en LtrA (K470) se encuentra en una de las regiones con menor tasa de mutaciones definida en ensayos de evolución unigénica (Cui *et al.*, 2004); al igual que ocurre en nuestro caso, el mutante K470A presenta unas actividades bioquímicas muy pobres. Estudios previos de escisión del intrón RmInt1 desarrollados en este trabajo apuntan a que este residuo podría estar implicado en una posible interacción con la ribozima. Se han descrito dos mecanismos de *splicing* para el intrón de *L. lactis*, uno de ellos dependiente de la unión al dIVa y otro en

el que esta estructura no se requiere. En ausencia del sitio específico de unión de la IEP al ARN del intrón en la región proximal del dIV de la ribozima, se retiene una capacidad residual de maduración (6-10%) por la interacción directa de la madurasa y parte de los dominios RT con determinadas regiones conservadas en el centro catalítico. La delección de regiones periféricas de este dominio (dIVb1/b2) reduce el *splicing in vivo* en un ~30%, mientras que la delección del dIVb completo inhibe su capacidad de maduración hasta alcanzar el 37% (Cui *et al.*, 2004). En cambio, la delección del dominio IVa en el intrón de levaduras aI2 da lugar a una reducción del *splicing* en sólo un ~30% (Huang *et al.*, 2003). En cualquiera de los dos intrones mencionados, la ausencia del dIVa tiene un efecto más acusado en la movilidad probablemente porque sea crítico en el posicionamiento de la IEP para la reverso transcripción (Wank *et al.*, 1999; Huang *et al.*, 2003). Por otro lado, el análisis de delecciones en la región N-terminal de LtrA sugiere que parte de los residuos presentes en esta zona, heredada de retrotransposones no-LTR, se requieren para la interacción con el dIVa. La delecciones más cortas afectan principalmente a la actividad de *splicing*, en cambio, cuando el número de aminoácidos eliminados aumenta también se alteran la actividad RT y la afinidad de unión de LtrA por el ARN de Ll.ltrB (Cui *et al.*, 2004). Curiosamente, la región N-terminal aguas arriba del dominio RT0 de la IEP de los intrones bacterianos de la subclase IIB3 es más corta que en el resto de linajes incluyendo las RTs de retrotransposones no-LTR (Blocker *et al.*, 2005). Es posible que la principal interacción que el ARN de este grupo de intrones establece con la proteína sea más débil que el resto, de modo que los contactos secundarios que los dominios X y RT establecen con el centro activo de la ribozima sean más críticos. Podría ser que la lisina en posición 381 estuviera involucrada en alguna de estas interacciones entre la IEP y el ARN del intrón necesarias para un *splicing* eficiente. La mayor implicación de este residuo en el *retroboming* puede estar relacionada con que la ausencia de esta interacción desestabilice la conformación activa de las RNPs tanto, que el nivel de complejos activos presentes en la bacteria sea insuficiente para desencadenar eventos de invasión.

La región cuya estructura predicha es la hélice  $\alpha 4$  parecer ser una parte del dominio madurasa con una función particular. Todas las posiciones conservadas en esta región estructural (H388, K389, R391, A392, L396 y A400) retienen una eficiencia de *splicing* sustancial (50-95%), pero su capacidad de movimiento se ve seriamente comprometida. Es más, el *retroboming*, cuando se detecta, sólo se produce a sitios diana clonados en la misma dirección de la horquilla de replicación de modo que la RT usaría como cebador la hebra retardada que se está sintetizando durante la replicación para copiar el ARN del intrón; esta es la ruta más favorecida en la movilidad de RmInt1 y probablemente la que menos restricciones presente en intrones que carecen de dominio ADN endonucleasa (Martínez-Abarca *et al.*, 2004; Nisa-Martínez *et al.*, 2007). Según nuestras observaciones pudiera ser que en esta región estén presentes residuos que son específicamente requeridos para el *boming* a sitios diana que se encuentran en dirección opuesta a la dirección de la horquilla de replicación. Se ha sugerido que esta ruta de *retroboming* minoritaria podría implicar la integración del ARN del intrón tanto en dianas que se

encuentren en doble cadena como en regiones del ADN que, de forma transitoria, se expongan como cadena sencilla. Entretanto, el cebador usado por la IEP para la síntesis del ADNc podría incluir muescas al azar producidas en la cadena complementaria en las cercanías del sitio de inserción del intrón, la cadena continua que se está generando durante la replicación (si pensamos que esta ruta de invasión también es dependiente de replicación) o iniciación *de novo* como se ha demostrado para algunas RTs (Martínez-Abarca *et al.*, 2004). Es un hecho que la mayoría de los mutantes mantienen su capacidad para producir el corte en la cadena de integración cuando se usan sustratos ADNsc y ADNdc, si bien es cierto que parece no producirse la integración completa del ARN en esos sustratos. La excepción es el mutante R391M que sufre *splicing* reverso completo pero que mantiene su tendencia a invadir dianas clonadas en pJB0.6LAG (figura III.8). Podríamos pensar que algunos de estos residuos podrían estar implicados en interacciones requeridas para el posicionamiento de la IEP una vez que ha tenido lugar la inserción del intrón para iniciar la reverso transcripción. Es probable que los condicionamientos para una u otra ruta de *retroboming* sean diferentes y también las interacciones ARN-proteína que intervengan en cada vía. Estudios más detallados sobre los residuos de la IEP implicados en la elección del cebador durante la movilidad podrían corroborar nuestras hipótesis. En cualquier caso, no es contradictorio que algunos de estos mutantes presenten una actividad RT que supere el 50% ya que en nuestra evaluación de la capacidad de polimerización de las preparaciones enriquecidas en RNPs usamos un sustrato exógeno que contiene su propio cebador (poli(rA)/oligo(dT)<sub>18</sub>).

Tanto la hélice  $\alpha 4$  como la cadena  $\beta 2$  de la IEP de RmInt1 se encuentran dentro de lo que se considera el dominio de conexión de la RT del HIV-1 (región que conecta los dominios RT y ARNasa H); la estructura obtenida para esta región comienza con 3 cadenas  $\beta$  que forman parte de la superficie de dimerización de esta proteína. Sin embargo, en los intrones IIB3 y, en general, en los intrones del grupo II, la predicción de estructura en esta región de la IEP consiste en una hélice  $\alpha$  (figura III.3A), aunque existen excepciones como la de la proteína LtrA en la que se predice una cadena  $\beta$  (Blocker *et al.*, 2005). LtrA es un homodímero que se forma en paralelo con el reconocimiento del ARN del intrón (Saldanha *et al.*, 1999; Rambo y Doudna, 2004); estudios de modelización de la estructura de LtrA comparativamente con la RT de HIV-1 apuntan que los dominios D y E de la primera reemplazan casi por completo al dominio de conexión de la segunda; aún así LtrA sigue conservando las 3 cadenas  $\beta$  que se suponen implicadas en la dimerización de estas RTs (Blocker *et al.*, 2005). La pérdida de estos dominios y las diferencias en la predicción de la estructura secundaria para la parte final del dominio madurasa de RmInt1 podrían apuntar la posibilidad de que nos encontremos ante una IEP que se une al ARN del intrón como un monómero.

Esta misma consecuencia podría explicar las diferencias de comportamiento bioquímico de este intrón respecto a LL.LtrB. En primer lugar, de los datos de que disponemos en relación con la actividad de *splicing* reverso de las RNPs *in vitro* sobre sustratos de ADN (consideremos únicamente el intrón silvestre como pKGEMA4) se

desprende la observación de que la integración completa del intrón es menos frecuente que la integración parcial (figura III.7B, carril 2). Por otro lado, la reacción de *splicing* directo *in vivo* muestra una ausencia de exones ligados (aunque en este trabajo no hemos sido capaces de detectar la acumulación de intrón *lariat* unido al exón 2); este resultado es similar a los datos de *splicing in vitro* en los que la proteína está ausente (Costa *et al.*, 2006a). Si correlacionamos todos los datos podemos deducir que la contribución de la proteína a los procesos en los que se ve implicado el intrón (maduración y movilidad) podría ser suficiente para generar una mínima cantidad de exones ligados o una mínima cantidad de integración completa del intrón. Esto podría estar en relación con que los complejos ribonucleoproteicos formados por este intrón sean de alguna manera más inestables al estar compuestos de ARN:proteína en proporción 1:1. Es probable, por tanto, que este intrón precise de la asistencia de ciertos factores del hospedador cuya ausencia limita su movilidad en los genomas. Esto podría explicar por qué la capacidad de *retroboming* está limitada en bacterias de distinta (*E. coli*; Karberg, 2005; Barrientos-Durán, 2008) o, incluso, de la misma especie (*S. meliloti* 1021 frente a GR4 o RMO17, por ejemplo; Toro *et al.*, 2003; Martínez-Abarca *et al.*, 2004).

El aminoácido prolina en la posición 404 puede suponer un punto de transición estructural y funcional dentro de la porción terminal de la madurasa de RmInt1, puesto que retiene eficiencias de movilidad similares al intrón silvestre independientemente de la orientación de la diana en el plásmido. Por otro lado, la eliminación de los últimos 14 aa de la IEP o mutaciones en algunos de los residuos conservados en la región situada aguas abajo de la posición 404 inhiben completa o parcialmente la capacidad de escisión de RmInt1; además las actividades bioquímicas de las RNPs se ven severamente afectadas a lo que hay que sumar la ausencia de movilidad. Es curioso, que a pesar de todo, existen una serie de mutaciones que retienen la capacidad de movimiento (L406R, V408E y W410F). Así, los datos apuntan a que algunos de los residuos presentes en la cadena  $\beta 2$  se precisan para la estabilización de la conformación activa de la IEP, pero que también podrían estar implicados en el propio mecanismo de movilidad. En concreto, podrían participar en la unión al ADN ya que ninguna de las dos rutas de *retroboming* está diferencialmente afectada. A esto debemos sumarle que la estabilidad de la región parece ser más crítica para los experimentos llevados a cabo *in vitro* con respecto a las actividades *in vivo*, lo que podría sugerir que otros factores no caracterizados podrían ayudar a la funcionalidad del intrón *in vivo*.

Aunque no se ha hecho referencia en ninguno de los apartados de resultados, se trató de comprobar si nuestras preparaciones de RNPs (recordar que se obtienen por fraccionamiento subcelular de lisados de cultivos bacterianos) podrían utilizarse para poner de manifiesto actividad de unión al ADN. Las condiciones se describen en el material y métodos y los resultados se muestran en el anexo (figura AIII.4). Las condiciones se trataron de optimizar con pKGEMA4 pero no obtuvimos ningún resultado. La causa más probable podría ser la heterogeneidad de nuestras preparaciones de RNPs ya que se sabe que los ensayos de retardo en gel funcionan mejor cuando se

utilizan los componentes purificados. De cualquier manera, podremos obtener conclusiones definitivas cuando estos ensayos se lleven a cabo con RNPs reconstituidas *in vitro*.

Finalmente, nuestros datos corroboran observaciones realizadas por otros autores que sugieren la existencia de una conexión funcional y estructural entre los diferentes dominios de las IEPs de los intrones del grupo II (Zimmerly *et al.*, 1999; San Filippo y Lambowitz, 2002; Cui *et al.*, 2004). La alteración de cualquiera de los dominios de la proteína afecta en cierta medida las actividades que implican al resto de los dominios por lo que es posible que todos los dominios de la proteína trabajen cooperativamente. Por ejemplo, trabajos previos (Muñoz-Adelantado *et al.*, 2003) y este trabajo demuestran que una mutación en el centro activo que controla la actividad RT (RT5, mutante YAAH) reduce la capacidad de maduración de RmInt1. Esto se ve reflejado especialmente en los mutantes expuestos en este trabajo que alteran no sólo la capacidad de *splicing* del intrón (madurasa) sino que también disminuyen severamente la actividad RT. El proceso de *retroboming* también se encuentra afectado y dependiendo de la región que se mute podríamos estar alterando el posicionamiento de la proteína o simplemente su unión al ADN diana.

Como decíamos al comienzo de este capítulo, los grandes grupos de intrones descritos hasta la fecha se pueden distinguir por la presencia/ausencia de los dominios D y/o En. Nuestros resultados sugieren que la divergencia entre las IEPs de los intrones del grupo II también afecta a la porción C-terminal del dominio madurasa, ya que en el caso de los intrones IIB3 no puede considerarse como una simple extensión del dominio madurasa. Esta región desaparece en los intrones IIA y IIC aunque se mantiene en otros intrones del grupo IIB (concretamente en los intrones IIB5 caracterizados en el capítulo I). Una de las hipótesis más ampliamente aceptadas en cuanto al origen de estos retroelementos es que una secuencia con capacidad de escisión adquirió una reverso transcriptasa que le permitió su diseminación por los genomas. Esta RT ancestral probablemente carecía de los dominios D y En, lo que sitúa a las proteínas definidas en los intrones de la clase IIB en el origen mismo de los retrointrones. Sin embargo, otros autores consideran que la clase de intrones que apareció en primer lugar fue el grupo IIC (hipótesis del retroelemento ancestral propuesta por Zimmerly y colaboradores; Robart y Zimmerly, 2005); esto puede ser así en cuanto a la particular organización estructural de la ribozima, pero no debemos olvidar que los intrones IIC presentan un dominio D en la secuencia de sus IEPs. Hay un debate evolutivo abierto pero lo que parece indudable es que la adquisición del dominio ADN endonucleasa puede haber influenciado la capacidad de dispersión de esos retroelementos. La presencia de este dominio podría haber permitido la invasión del núcleo de los eucariotas primitivos dando lugar a lo que hoy conocemos como intrones espliceosómicos.

## **Anexo capítulo III**



Tabla AIII.1. Secuencias utilizadas en estudios filogenéticos.

DENOMINACIÓN	ORGANISMO	NºACCESO B d D
<i>Sr.me.I1</i> (RmInt1)	<i>Sinorhizobium meliloti</i>	NP 437164
<i>E.a.I1</i>	<i>Ensifer adhaerens</i>	AAP83798
<i>Sr.med.</i>	<i>Sinorhizobium medicae</i>	YP 001313619
<i>Sr.t.I1</i>	<i>Sinorhizobium terangae</i>	AAU95643
<i>E.c.I2</i>	<i>Escherichia coli</i>	CAA54637
<i>Sb.p</i>	<i>Shewanella putrefaciens</i>	YP 001181807
<i>Az.sp.</i>	<i>Azoarcus</i> sp. EbN1	YP 159836
<i>L.p.</i>	<i>Legionella pneumophila</i>	YP 0012511128
<i>E.c.</i>	<i>Escherichia coli</i> B	ZP 01698243
<i>Pr.ae.I3</i>	<i>Prosthecochloris aestuarii</i>	ZP 00592895
<i>Pr.vi.</i>	<i>Prosthecochloris vibrioformis</i>	YP 001129678
<i>Pe.ph</i>	<i>Pelodyction phaeoclathratiforme</i>	ZP 00589124
<i>Ch.ph</i>	<i>Chlorobium phaeobacteroides</i>	YP 911931
<i>Sy.a.</i>	<i>Syntrophus aciditrophicus</i>	YP460783
<i>M.a.</i>	<i>Methanosarcina acetivorans</i>	NP 619481
UA.I3	uncultured archaeon Gzfos32G12	AAU83697
<i>B. thu.</i>	<i>Bacillus thuringiensis</i>	ZP 00738538
<i>Pa.de.I1</i>	<i>Paracoccus denitrificans</i>	ZP 00628808
<i>Ph.l.</i>	<i>Photobacterium luminescens</i>	NP 928428
<i>Ma.sp.</i>	<i>Magnetococcus</i> sp.	YP 864580
<i>P.ae.</i>	<i>Pseudomonas aeruginosa</i>	ABR13526
<i>P.st.</i>	<i>Pseudomonas stutzeri</i>	YP 001172226
<i>Bu.ph.</i>	<i>Burkholderia phymatum</i>	ZP 01505671
<i>Fr.sp.</i>	<i>Frankia</i> sp.	YP 482811
<i>S.ery.</i>	<i>Saccharopolyspora erythraea</i>	YP 001104541



<i>Pe.a.</i>	<i>Pelobacter acetylenicus</i>	AAQ08377
delta	deltaproteobacterium MLMS-1	ZP 01288325
<i>B.j.</i>	<i>Bradyrhizobium japonicum</i>	NP 768692
<i>S.dy.</i>	<i>Shigella dysenteriae</i>	YP 406035
<i>Al.me.</i>	<i>Alkaliphilus metalliredigens</i>	YP 001321146
<i>B.t.</i>	<i>Bacteroides thetaiotaomicron</i>	NP 811528
<i>Pch.t</i>	<i>Psychroflexus torquis</i>	ZP 01254488
UMB.I1	uncultured marine bacterium 18874275	AAL78688
UMB.I2	uncultured marine bacterium 18874408	AAL78689
UMB.I3	uncultured marine bacterium 18874410	AAL78690

**Tabla AIII.2.** Oligonucleótidos usados para construir los mutantes en el C-terminal por el sistema pALTER-1.

MUTANTE	SECUENCIA	POSICIÓN		CAMBIO	
		IEP	INTRÓN	AA	NT
pKG2.5-A381	5' -CTTGATCAGGCGAG <b>CG</b> TTCAAACGCTTTC-3'	381	1687/89	K → A	AAG → <b>GCG</b>
pKG2.5-E388	5' -CGCTTTCAGTCC <b>GAGA</b> AGACACGCGCC-3'	388	1710	H → E	CAT → <b>GAG</b>
pKG2.5-D389	5' -TTTCAGTCCCAT <b>GAC</b> ACACGCGCCAGC-3'	389	1711/13	K → D	AAG → <b>GAC</b>
pKG2.5-M391	5' -CAGTCCCATAAGACA <b>ATG</b> GCCAGCCTCTTCTTG-3'	391	1717/18/19	R → M	CGC → <b>ATG</b>
pKG2.5-E392	5' -ATAAGACACGCGAG <b>AGC</b> CTCTTCTTG-3'	392	1721/22	A → E	GCC → <b>GAG</b>
pKG2.5-E396	5' -GCCAGCCTCTT <b>CG</b> AGCGAAAGCTGGC-3'	396	1732/33	L → E	TTG → <b>GAG</b>
pKG2.5-ΔC21	5' -TTCTTGCGAAAG <b>TAG</b> TGCGCGGAAAAT-3'	399	1741/42	L → *	CTG → <b>TAG</b>
pKG2.5-E400	5' -CGAAAGCTGG <b>AG</b> CGGGAAAAT-3'	400	1745	A → E	GCG → <b>GAG</b>
pKG2.5-R400	5' -TCTTGCGAAAGCT <b>GCG</b> CCGGGAAAATCCGGG-3'	400	1744/45/46	A → R	GCG → <b>CGC</b>
pKG2.5-T404	5' -GCGGGAAAAT <b>AC</b> GGGGCTGTT-3'	404	1756/57	P → T	CCG → <b>ACG</b>
pKG2.5-ΔC14	5' -GAAAATCCGGGG <b>TAG</b> TTTCGTGCATTGG-3'	406	1762/1763	L → *	CTG → <b>TAG</b>
pKG2.5-R406	5' -AAAATCCGGGG <b>CGC</b> TTTCGTGCATTGG-3'	406	1763/64	L → R	CTG → <b>CGC</b>
pKG2.5-R407	5' -AATCCGGGG <b>CTG</b> CGCGTGCATTGGAA-3'	407	1765/66	F → R	TTC → <b>CGC</b>
pKG2.5-G409	5' -GGGCTGTTCTG <b>GGT</b> TGGAAGCGTT-3'	409	1771/72	H → G	CAT → <b>GGT</b>
pKG2.5-D410	5' -GGGCTGTTCTGTCAT <b>GATA</b> AGCGTTCCGGAACG-3'	410	1774/75/76	W → D	TGG → <b>GAT</b>
pKG2.5-F410	5' -TGTTCTGTCAT <b>TTTA</b> AGGCGTTCGGA-3'	410	1775/76	W → F	TGG → <b>TTT</b>
pKG2.5-P410	5' -CTGTTCTGTCAT <b>CCG</b> AAGGCGTTCGG-3'	410	1774/75	W → P	TGG → <b>CCG</b>

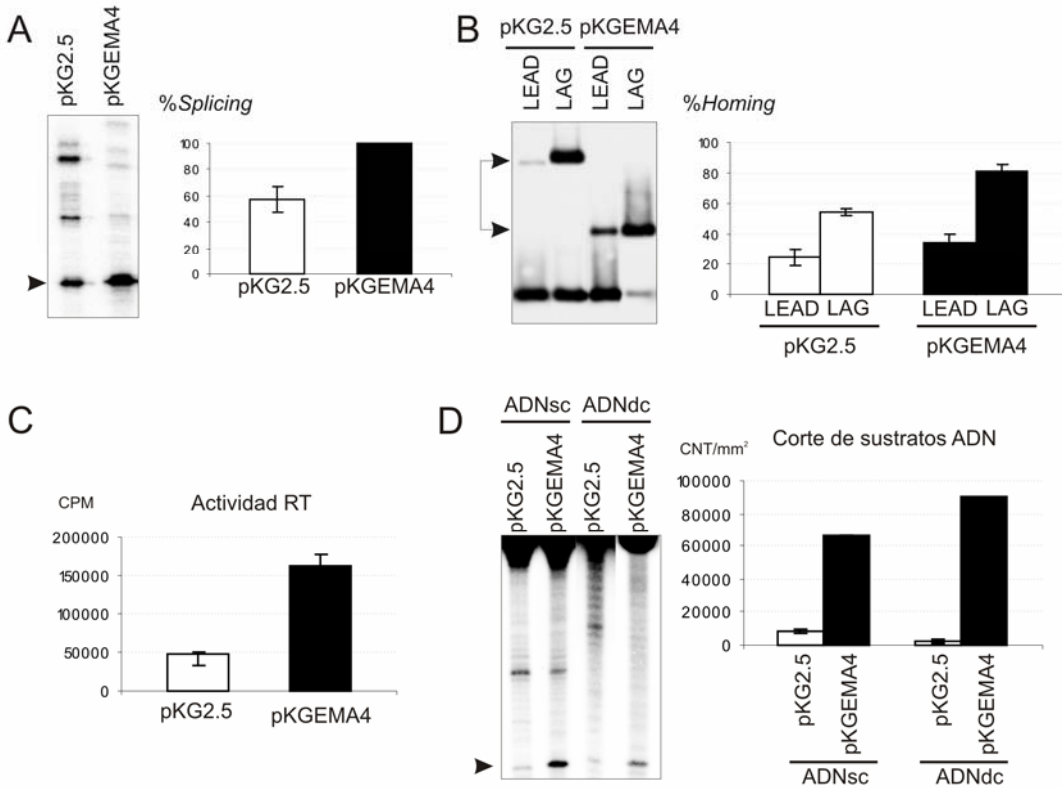
**Tabla AIII.3.** Oligonucleótidos usados para construir el resto de los mutantes en el C-terminal por PCR y para transferir los mutantes a las construcciones pKGEMA4.

NOMBRE	SECUENCIA	APLICACIÓN	REFERENCIA
mut UP	5' -GTCAGCGGTGCTGGAAGTATG-3'	Cebadores exteriores para la construcción de mutantes por PCR <sup>a</sup>	Este trabajo
mut DN	5' -GCGCGCGTAATACGACTCAC-3'		
A400V/UP	5' -GAAAGCTGGTGC GGG AAAATCCGGG-3'	Cebadores internos solapantes para la construcción del mutante pKG2.5-V400	Este trabajo
A400V/DN	5' -ATTTTCCCGC <b>ACC</b> AGCTTTCGCAAGA-3'		
YY-AA/UP	5' -GATGGATTGCC <b>CGCG</b> GGGACGGTACAGTCGTTCCGG-3'	Cebadores internos solapantes para la construcción del mutante pKG2.5-Δ354Δ355	Este trabajo
YY-AA/DN	5' -CTGTACCGTCC <b>CGCG</b> GGGCAATCCATCCCCGAAGG-3'		
SpeI-IEP	5' -GGGGACTAGGTGGAAACAGGATGACTTTCGGA-3'	Amplifica la secuencia de la IEP para clonarla en pKGEMA4 <sup>b</sup>	Muñoz-Adelantado <i>et al.</i> , 2003
IEP-SacXPml	5' -GGGGCACGTGACCTCCTCAGGTAAACGTTCGTTCC-3'		Este trabajo

Las posiciones que se marcan en negrita tanto en esta tabla como en la anterior representan las mutaciones introducidas.

(a) Ver material y métodos, apartado 14.2.

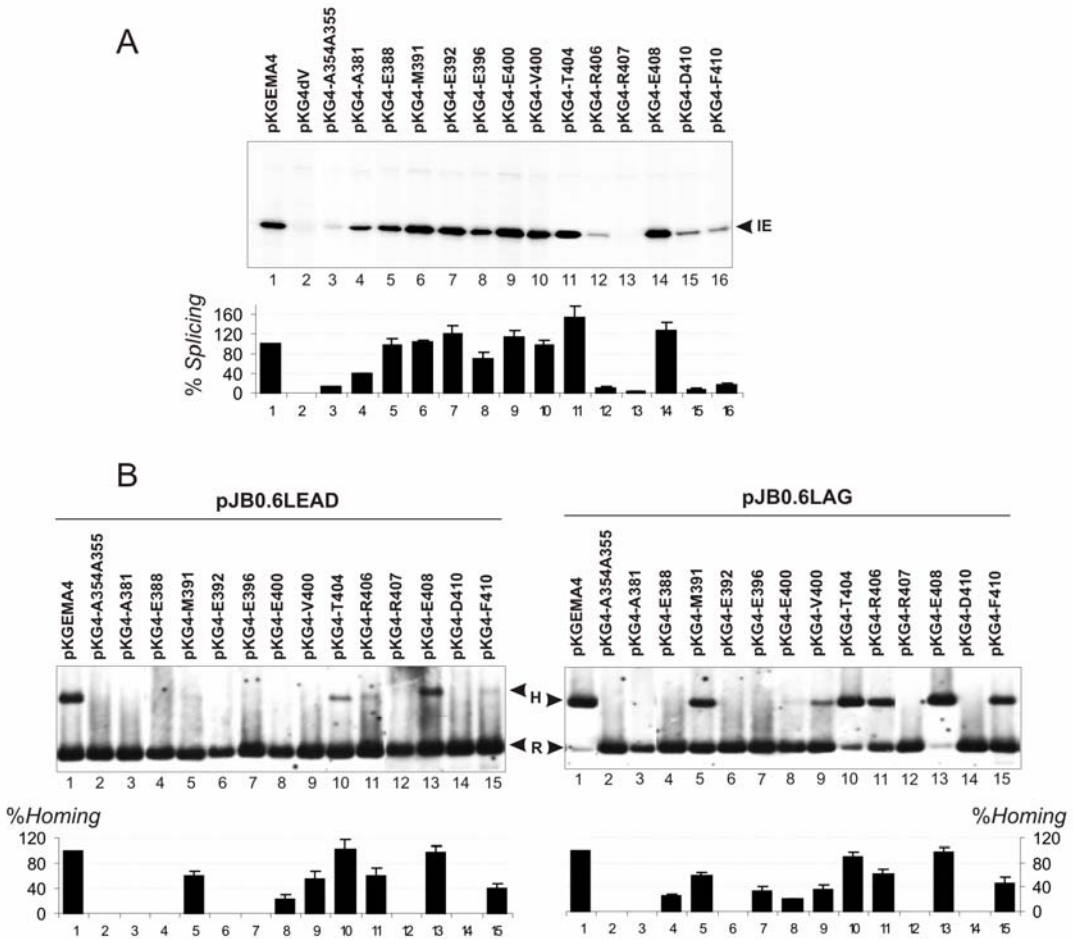
(b) Ambos cebadores se sitúan flanqueando la secuencia de la IEP y contienen enzimas de restricción que nos permitirán clonar el fragmento amplificado. En concreto el cebador IEP-SacXPml contiene la secuencia para la endonucleasa de restricción *SacI* modificada de manera que podamos distinguir por digestión nuestras construcciones mutadas del plásmido silvestre (el plásmido pKGEMA4 contiene dos dianas para la enzima *SacI*).



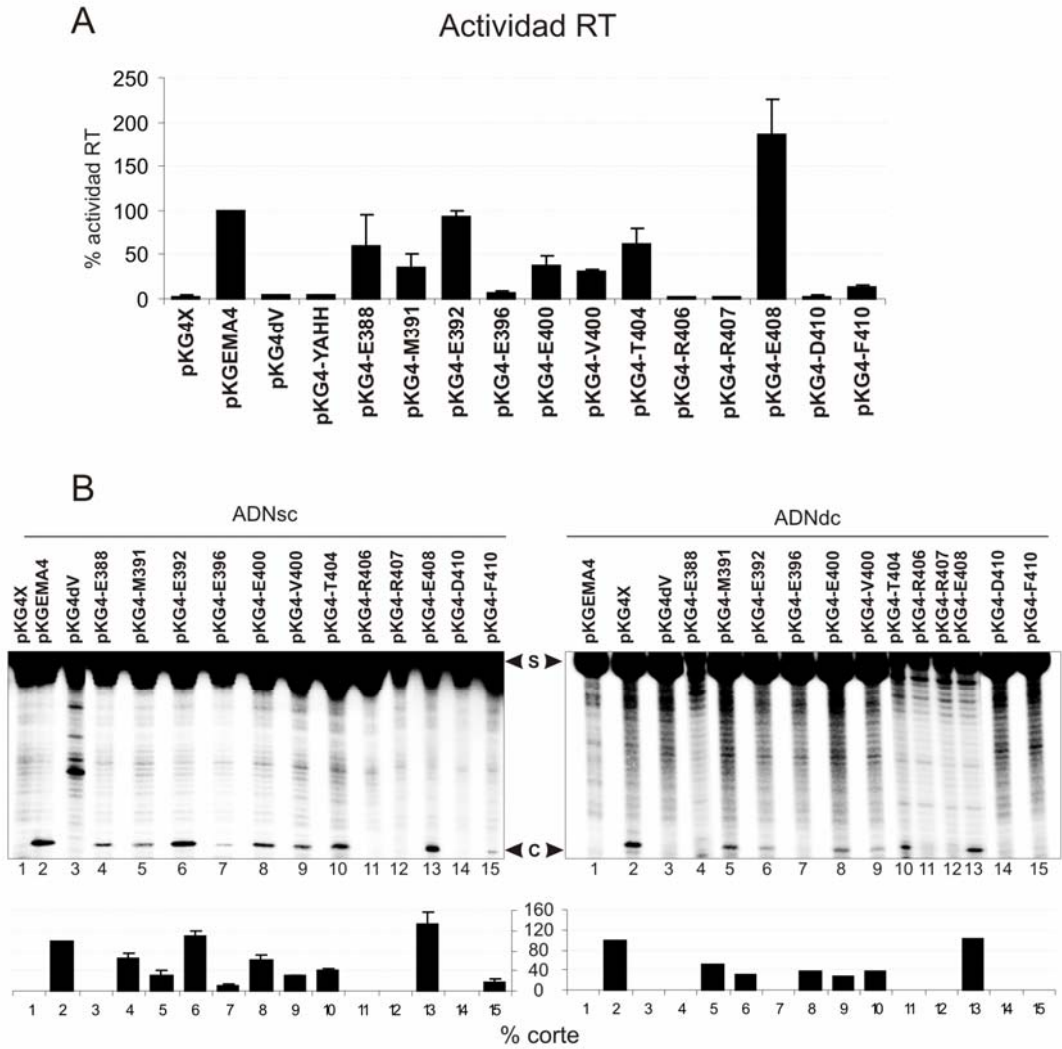
**Figura AIII.1.** Datos comparativos absolutos entre pKG2.5 y pGEMA4. Se ha comparado tanto el comportamiento de las RNPs *in vivo* (*splicing* y *retroboming*, en paneles A y B) como sus actividades bioquímicas *in vitro* (actividad RT y de corte de sustratos ADN, paneles C y D). En paralelo a los geles se muestran una serie de histogramas cada uno correspondiente a la cuantificación de todos los datos disponibles pero que permitan la comparación de los resultados, por ejemplo, ensayos RT o de corte que se hayan desarrollado en paralelo, con el mismo sustrato. **(A)** En el gel se indica con una flecha la banda de ADNc correspondiente al intrón escindido en preparaciones de ARN total procedente de *S. meliloti* RMO17. El porcentaje de *splicing* se ha calculado como la razón entre el producto escindido/todos los productos detectados (derivados de intrón) x 100; en el caso de pKGEMA4, en la mayoría de las ocasiones no se observan bandas de mayor tamaño molecular por lo que la estimación siempre suele corresponder con la banda de producto escindido. **(B)** Comparativa de *homing* entre ambas construcciones y en las dos posibles orientaciones de la diana. Con flechas sobre el gel se indican los productos de *homing* en ambos tipos de construcciones; se diferencian en tamaño como consecuencia de que en el caso de pKG2.5 invade el intrón completo (1884 bp), mientras que en el caso de pKGEMA4 sólo se inserta el ΔORF. El cálculo se corresponde con el cociente entre la intensidad de la banda de *homing* y la suma de ésta más la banda de receptor no invadido (en la parte inferior del gel) x 100. **(C)** Diagrama de barras en el que se muestran en cuentas por minuto correspondientes a la incorporación de [ $\alpha$ -<sup>32</sup>P]dTTP como consecuencia de la incubación de RNPs extraídas de *S. meliloti* RMO17 con un sustrato externo consistente en poli(rA)/oligo(dT)<sub>18</sub>. Estos productos de alto peso molecular son retenidos en papel Watman DE81 y cuantificados en contador de centelleo. **(D)** Comparativa del corte en sustratos de ADN de cadena simple (ADSc) y doble (ADNdc) de las RNPs producidas a partir de ambos plásmidos. Con una flecha se indica la banda de corte que se corresponde con el producto cuantificado con el programa informático Quantity One® (BIO-RAD). Este es el dato representado en el histograma.

**Tabla AIII.4.** Actividades medidas *in vivo* e *in vitro* para las construcciones derivadas de pKGEMA4 que presentan mutaciones en la IEP de RmInt1.

MUTANTES	<i>SPLICING</i> ARN	<i>HOMING</i> <i>LEAD</i>	<i>HOMING</i> <i>LAG</i>	Actividad RT	Corte en ssDNA70mer	Corte de la cadena superior en dsDNA70bp	<i>Splicing</i> reverso en dsDNA70bp
pKGEMA4	100	100	100	100	100	100	100
pKGEMA4X	0	0	0	3 ± 1	0	0	0
pK4YAHH	78 ± 18	0	0	3 ± 0	105 ± 11	60	16 ± 8
pKG4dV	0	0	0	4 ± 1	0	0	0
pKG4-A354A355	12 ± 1	0	0	0,8 ± 0	0	0	0
pKG4-A381	38 ± 3	0	0	46 ± 7	28	0	11 ± 7
pKG4-E388	97 ± 12	0	0	60 ± 34	65 ± 10	0	0
pKG4-M391	103 ± 3	61 ± 7	59 ± 5	36 ± 15	30 ± 9	53	17 ± 3
pKG4-E392	118 ± 18	0	0	93 ± 7	110 ± 11	30	0
pKG4-E396	71 ± 12	0	0	6 ± 2	11 ± 4	0	0
pKG4-E400	114 ± 13	22 ± 7	20 ± 2	38 ± 10	60 ± 13	38	0
pKG4-V400	98 ± 10	56 ± 11	37 ± 6	32 ± 1	29 ± 3	28	0
pKG4-T404	153 ± 24	102 ± 16	90 ± 6	61 ± 18	41 ± 4	38	46 ± 11
pKG4-R406	11 ± 3	59 ± 12	61 ± 8	2 ± 0	0	0	0
pKG4-R407	2 ± 0	0	0	2 ± 0	0	0	0
pKG4-E408	127 ± 15	96 ± 11	97 ± 8	185 ± 40	131 ± 26	101	133 ± 16
pKG4-D410	8 ± 3	0	0	3 ± 0	0	0	0
pKG4-F410	16 ± 5	39 ± 8	47 ± 8	12 ± 4	19 ± 4	0	8



**Figura AIII.2.** Actividades *in vivo* medidas en los distintos mutantes en la región C-terminal de la proteína codificada por RmInt1 con construcciones pKGEMA4. En ambos casos los datos son relativos al WT considerado pKGEMA4. **(A)** Evaluación de los niveles de maduración de los diferentes mutantes. Se muestra un gel representativo acompañado de un diagrama de barras donde se representa la intensidad relativa de la banda de intrón escindido (IE). **(B)** Cuantificación de la eficiencia de invasión de dianas situadas en la misma (pJB0.6LAG) u orientación opuesta (pJB0.6LEAD) a la dirección de la horquilla de replicación. Los gels se corresponden con el análisis del conjunto de plásmidos presentes en la bacteria mediante digestión *SalI*, electroforesis, posterior transferencia a membrana de nylon e hibridación con una sonda específica del exón 2. Esta sonda hibrida tanto en la diana invadida (H) como en la diana sin invadir (R). El porcentaje de *homing* se calcula (H/H + R) y luego se refiere al valor obtenido para pKGEMA4

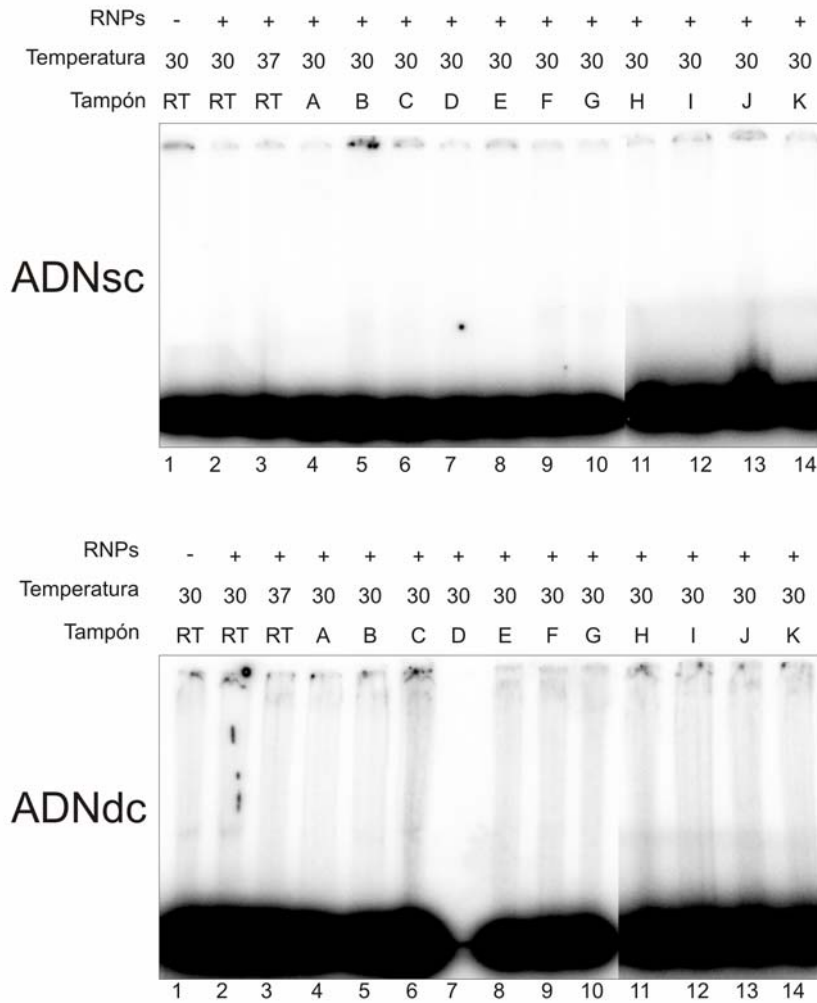


**Figura AIII.3.** Actividades bioquímicas cuantificadas *in vitro* para las fracciones enriquecidas en RNPs obtenidas para los diferentes mutantes indicados que expresan el intrón a partir de construcciones derivadas de pKGEMA4. En esta ocasión, con el fin de poder comparar los múltiples experimentos realizados en cada caso, la evaluación de las actividades también se refiere a los niveles silvestres que se cuantificaron en cada ensayo. En las gráficas se representa el valor de la media de al menos dos experimentos, acompañada de las bandas de error que se corresponden con la desviación estándar. **(A)** Histograma que representa la capacidad de polimerización de las RNPs usando un sustrato poli(rA)/oligo(dT)<sub>18</sub>. La actividad se mide en cuentas por minuto y se corresponde con la incorporación de [ $\alpha$ -<sup>32</sup>P]dTTP cuantificada en contador de centelleo. **(B)** Evaluación de los niveles de corte de las RNPs en sustratos de ADN de cadena simple (ADNsc) y doble (ADNdc). S representa a los sustratos, diferentes en cada panel pero marcados en ambos casos en el extremo 5' de la cadena que se corta; C, se corresponde con la banda de corte. En el caso de cadena simple, se pudieron evaluar más de un experimento mientras que para sustratos de doble cadena encontramos problemas para la cuantificación de la banda de corte por lo que sólo se muestra un valor. No ponemos en duda la veracidad del resultado porque en paralelo se han podido evaluar los niveles de *splicing* reverso y nuestra experiencia nos dice que la eficiencia de corte depende en gran medida de las preparaciones de sustrato.

**Tabla AIII.5.** Datos de movilidad basados en selección de colonias resistentes a tetraciclina.

Cruce	%invasion	Cruce	%invasion
pKGEA4 <sub>T7</sub> /pACEs	<10 <sup>-4</sup>	pKGEA4 <sub>T7</sub> /pACEs	<10 <sup>-4</sup>
pKGEA4 <sub>T7</sub> /pACE0.6s	5,7±1,8	pKGEA4 <sub>T7</sub> /pACE0.6as	54,3±9,6
pKGEA4X <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKGEA4X <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4YAHH <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4YAHH <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4A354A355 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4A354A355 <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4A381 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4A381 <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4E388 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4E388 <sub>T7</sub> /pACE0.6as	0,2±0,1
pKG4D389 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4D389 <sub>T7</sub> /pACE0.6as	9,6
pKG4M391 <sub>T7</sub> /pACE0.6s	2,7±0,9	pKG4M391 <sub>T7</sub> /pACE0.6as	66,3±33,1
pKG4E392 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4E392 <sub>T7</sub> /pACE0.6as	49,3±9,8
pKG4E396 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4E396 <sub>T7</sub> /pACE0.6as	0,2±0,1
pKG4E400 <sub>T7</sub> /pACE0.6s	0,1±0,0	pKG4E400 <sub>T7</sub> /pACE0.6as	8,2±2,4
pKG4R400 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4R400 <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4V400 <sub>T7</sub> /pACE0.6s	0,5±0,4	pKG4V400 <sub>T7</sub> /pACE0.6as	37,2±2,7
pKG4T404 <sub>T7</sub> /pACE0.6s	3,1±2,4	pKG4T404 <sub>T7</sub> /pACE0.6as	67,2±18,2
pKG4R406 <sub>T7</sub> /pACE0.6s	0,7	pKG4R406 <sub>T7</sub> /pACE0.6as	62,3
pKG4R407 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4R407 <sub>T7</sub> /pACE0.6as	0,1±0,0
pKG4E408 <sub>T7</sub> /pACE0.6s	2,5	pKG4E408 <sub>T7</sub> /pACE0.6as	73,4±43,1
pKG4G409 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4G409 <sub>T7</sub> /pACE0.6as	1,2±0,3
pKG4D410 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4D410 <sub>T7</sub> /pACE0.6as	<10 <sup>-4</sup>
pKG4F410 <sub>T7</sub> /pACE0.6s	1,1±0,1	pKG4F410 <sub>T7</sub> /pACE0.6as	39,5±18,6
pKG4P410 <sub>T7</sub> /pACE0.6s	<10 <sup>-4</sup>	pKG4P410 <sub>T7</sub> /pACE0.6as	11,7

El porcentaje de invasión se calcula según la razón ( $A_{p^R Tc^R} / Tc^R$ ) x 100. Los datos se expresan como la media de al menos dos ensayos independientes ± S.D.



**Figura AIII.4.** Comprobación de la capacidad de unión de RNPs extraídas de construcciones pKGEMA4 a sustratos de ADN de cadena doble y simple por retardo en gel en distintas condiciones (ver material y métodos). Se comprobaron hasta 12 tampones diferentes y para alguno de ellos dos temperaturas. El tercer carril se corresponde con las condiciones que se ensayan para la detección del corte de sustratos ADN. Es evidente que esta técnica no es lo suficientemente sensible para revelarnos capacidad de unión.





*CONCLUSIONES*



1. El plásmido críptico pRmeGR4b de *Sinorhizobium meliloti* GR4 se revela como un sumidero de elementos **móviles** probablemente adquiridos por transferencia horizontal. El porcentaje de elementos móviles encontrados tras la secuenciación parcial es superior a la presente en el genoma de la cepa de referencia 1021.
2. El intrón del grupo II *S.m.*I4, encontrado durante la secuenciación parcial del plásmido críptico pRmeGR4b, pertenece a una nueva clase de intrones bacterianos denominada clase bacteriana E o intrones del grupo IIB5. La predicción de la estructura secundaria de su ribozima muestra rasgos diferenciales que definen esta nueva clase.
3. El intrón del grupo II de *Sinorhizobium meliloti* RmInt1 madura *in vivo* dando lugar a tres posibles productos de escisión: intrón en lazo, circular y lineal. De los tres productos, la forma en lazo parece ser mayoritaria y es probablemente la que se encuentra formando parte de los complejos de ribonucleoproteína activos.
4. El mecanismo de escisión propuesto para las moléculas circulares de RmInt1 implica la liberación en primer lugar del exón 2. Este extremo 3' hidroxilo libre ataca entonces el sitio de maduración 5' liberando el ARN circular.
5. La proteína codificada por RmInt1 está implicada en la escisión precisa de este intrón manteniendo a su vez el balance entre formas de escisión circular y en lazo.
6. La adenosina desapareada del dominio VI de la ribozima es esencial para que este intrón madure a formas en lazo.
7. La región de coordinación en el dominio I es indispensable para la conformación del sitio activo de la ribozima y su alteración no sólo afecta a la escisión sino también a la movilidad de RmInt1, pudiendo estar implicada en la coordinación de metales en el sitio catalítico de la ribozima.
8. La ligación de exones en el proceso de maduración de RmInt1 se encuentra muy disminuida en relación a la formación de intrón escindido. Este hecho podría tener efectos negativos sobre la movilidad del elemento genético en el que se inserta: la secuencia de inserción *ISRm2011-2*.
9. Las proteínas codificadas por intrones de la subclase bacteriana IIB3 muestran un elevado grado de conservación de su región C-terminal lo que permite inferir un motivo característico: LX<sub>3</sub>AX<sub>3</sub>PXLF(V/A)HW así como una característica conformación diferencial integrada por una alfa hélice ( $\alpha$ 4), y una lámina beta ( $\beta$ 2).

**10.** Esta región C-terminal, aunque presenta características funcionales particulares, forma parte del dominio madurasa. Los residuos presentes en la hélice  $\alpha 4$  podrían ser requeridos durante el mecanismo de movilidad para el posicionamiento de la reverso transcriptasa una vez que el ARN del intrón se ha integrado en su cadena diana. Por otro lado, la región  $\beta 2$  tendría un papel más amplio, contribuyendo a la estabilización estructural de la proteína y permitiendo una conformación activa de los complejos de ribonucleoproteína tanto *in vitro* como *in vivo*.

## *BIBLIOGRAFÍA*





## BIBLIOGRAFÍA

Abelson, J., Trotta, C.R., y Li, H. (1998). tRNA splicing. *J. Biol. Chem.* **273**; 12685–12688.

Abramovitz, D.L., Friedman, R.A., y Pyle, A.M. (1996). Catalytic role of 29-hydroxyl groups within a group II intron active site. *Science* **271**; 1410–1413.

Adamidi, C., Fedorova, O., y Pyle, A.M. (2003). A group II intron inserted into a bacterial heat-shock operon shows autocatalytic activity and unusual thermostability. *Biochemistry* **42**; 3409–3418.

Agrawal, S., Mayrand, S.H., Zamecnik, P.C. y Pederson, T. (1990) Site-specific excision from RNA by RNase H and mixed-phosphate-backbone oligodeoxynucleotides. *Proc. Natl. Acad. Sci. USA*, **87** (7); 1401-1405.

Aizawa, Y., Xiang, Q., Lambowitz, A.M., y Pyle, A.M. (2003). The pathway of DNA recognition and RNA integration by a group II intron retrotransposon. *Mol. Cell* **11**; 795–805.

Arnberg, A.C., Van Ommen, G.J., Grivell, L.A., Van Bruggen, E.F., y Borst, P. (1980). Some yeast mitochondrial RNAs are circular. *Cell* **19**; 313–319.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. y Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**(3); 403-410.

Augustin, S., Mueller, M.W., y Schweyen, R.J. (1990). Reverse self-splicing of group II intron RNAs *in vitro*. *Nature* **343**; 383–386.

Bachl, J. y Schmelzer, C. (1990). Effect of deletions at structural domains of group II intron bI1 on selfsplicing *in vitro*. *J. Mol. Biol.* **212**; 113–125.

Barloy-Hubler, F., Capela, D., Barnett, M.J., Kalman, S., Federspiel, N.A., Long, S.R. y Galibert, F. (2000a) High-resolution physical map of the *Sinorhizobium meliloti* 1021 pSyma megaplasmid. *J. Bacteriol.* **182**; 1185-1189.

Barloy-Hubler, F., Capela, D., Batut, J. y Galibert, F. (2000b) High resolution physical map of the pSymb megaplasmid and comparison of the three replicons of *Sinorhizobium meliloti*. *Curr. Microbiol.* **41**; 109-113.

Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J. *et al.* (2001) Nucleotide sequence and



predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. *Proc. Natl. Acad. Sci. USA* **98**; 9883-9888.

**Barrientos-Durán, A.M.** (2008) Bases moleculares para la aplicación biotecnológica del intrón del grupo II RmInt1 de *Sinorhizobium meliloti*. Tesis doctoral. Universidad de Granada.

**Beauregard, A., Chalamcharla, V.R., Piazza, C.L., Belfort, M. y Coros, C.J.** (2006) Bipolar localization of the group II intron Ll.LtrB is maintained in *Escherichia coli* deficient in nucleoid condensation, chromosome partitioning and DNA replication. *Mol Microbiol*, **62(3)**; 709-722.

**Begel, O., Boulay, J., Albert, B., Dufour, E., y Sainsard-Chanet, A.** (1999). Mitochondrial group II introns, cytochrome c oxidase, and senescence in *Podospora anserina*. *Mol. Cell. Biol.* **19**; 4093–4100.

**Belfort, M. y Roberts, R.J.** (1997) Homing endonucleases: keeping the house in order. *Nucleic Acids Res.*, **25 (17)**; 3379-3388.

**Belfort, M., Derbyshire, V., Parker, M.M., Cousineau, B., y Lambowitz, A.M.** (2002). Mobile introns: pathways and proteins. In: *Mobile DNA II*. pp. 761–783. Craig, N.L., Craigie, R., Gellert, M., and Lambowitz, A.M., Eds., ASM Press Publishers, Washington, DC.

**Belhocine, K., Plante, I. y Cousineau, B.** (2004) Conjugation mediates transfer of the Ll.LtrB group II intron between different bacterial species. *Mol. Microbiol.*, **51**; 1459-1469.

**Berget, M. S., Moore, C., y Sharp, P. A.** (1977) Spliced segments at the 5' terminus of adenovirus 2 late mRNA *Proc. Natl. Acad. Sci. USA*, **74**; 3171 - 3175.

**Beringer, J.E.** (1974) R factor transfer in *Rhizobium leguminosarum*. *J. Gen. Microbiol.* **84**: 188-198.

**Birnboim, H.C. y Doly, J.** (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucl. Acids Res.* **7**: 1513-1524.

**Blocker, F.J.H., Mohr, G., Conlan, L.H., Qi, L., Belfort, M. y Lambowitz, A.M.** (2005) Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA*, **11**; 14–28.

**Bonen, L.** (1993) *Trans*-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* **7**; 40–46.

**Bonen, L. y Vogel, J.** (2001). The ins and outs of group II introns. *Trends Genet.* **17**; 322–331.



---

**Bonen, L.** (2008) *Cis-* and *trans-*splicing of group II introns in plant mitochondria. *Mitochondrion*, **8**(1); 26-34.

**Boorstein, W.R. y Craig, E.A.** (1989) Primer extension analysis of RNA. *Methods Enzymol.*, **180**; 347-369.

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

**Boudvillain, M., de Lencastre, A., y Pyle, A.M.** (2000). A tertiary interaction that links active-site domains to the 5' splice site of a group II intron. *Nature* **406**; 315–318.

**Boulanger, S.C., Belcher, S.M., Schmidt, U., Dib-Hajj, S.D., Schmidt, T., y Perlman, P.S.** (1995). Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. *Mol. Cell. Biol.* **15**; 4479–4488.

**Boulanger, S.C., Faix, P.H., Yang, H., Zhuo, J., Franzen, J.S., Peebles, C.L., y Perlman, P.S.** (1996). Length changes in the joining segment between domains 5 and 6 of a group II intron inhibit self-splicing and alter 3' splice site selection. *Mol. Cell. Biol.* **16**; 5896–5904

**Brown, T., Mackey, K. y Du, T.** (2004) Analysis of RNA by Northern and Slot Blot hybridization. Current Protocols in Molecular Biology 4.9.1.-4.9.19. John Wiley & Sons, Inc.

**Buratowski, S. y Chodosh, L.A.** (1996) Mobility shift DNA-binding assay using gel electrophoresis. Current Protocols in Molecular Biology 12.2.1.-12.2.11. John Wiley & Sons, Inc.

**Burge, C.B., Tuschl, T., y Sharp, P.A.** (1999). Splicing of precursors to mRNAs by the spliceosomes. In: *The RNA World, 2nd edition*. pp. 525–560. Gesteland, R.F., Cech, T.R., and Atkins, J.F., Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

**Burkardt, B., Schillik, D. y Puhler, A.** (1987) Physical characterization of *Rhizobium meliloti* megaplasmids. *Plasmid*, **17**(1); 13-25.

**Burt, A.** (2003) Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. Lond. B*, **270**; 921-928.

**Cabanes, D., Boistard, P. y Batut, J.** (2000) Identification of *Sinorhizobium meliloti* genes regulated during symbiosis. *J. Bacteriol.*, **182** (13); 3632-3637.

**Cadwell, R. C. y Joyce, G. F.** (1992). Randomization of genes by PCR mutagenesis. *PCR. Methods Appl.*, **2**; 28–33.

**Capela, D., Barloy-Hubler, F., Gatiús, M-T., Gouzy, J. y Galibert, F.** (1999) A high-density physical map of *Sinorhizobium meliloti* 1021 chromosome derived from bacterial artificial chromosome library. *Proc. Natl. Acad. Sci. USA* **96**; 9357-9362.

**Capela, D., Barloy-Hubler, F., Gouzy, J., Bothe, G., Ampe, F., Batut, J., Boistard, P., Becker, A., Boutry, M., Cadieu, E. et al.** (2001) Analysis of the chromosome sequence of the legume symbiont *Sinorhizobium meliloti* strain 1021. *Proc. Natl. Acad. Sci. USA* **98**; 9877-9882.

**Carignani, G., Groudinsky, O., Frezza, D., Schiavon, E., Bergantino, E., y Slonimski, P.P.** (1983). An mRNA maturase is encoded by the first intron of the mitochondrial gene for the subunit I of cytochrome oxidase in *S. cerevisiae*. *Cell* **35**; 733–742.

**Casadesús, J. y Olivares, J.** (1979) Rough and fine linkage mapping of the *Rhizobium meliloti* chromosome. *Mol. Gen. Genet.* **174**; 203-209.

**Cavalier-Smith T** (1991) Intron phylogeny: a new hypothesis. *Trends Genet* **7**; 145–148.

**Cech, T.K.** (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*, **44**; 207-210.

**Cech, T.R., Moras, D., Nagai, K. y Williamson, J.R.** (2006) The RNP world. *RNAWorld*. 3rd edn (Getsland, R.F., Ceh, T.R. y Atkins, J.F., eds), pp. 309-326. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Chanfreau, G. y Jacquier, A.** (1993). Interaction of intronic boundaries is required for the second splicing step efficiency of a group II intron. *EMBO J.* **12**; 5173–5180.

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

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

**Chen, Y., McClane, B. A., Fisher, D.J., Rood, J. I., y Gupta, P.** (2005) Construction of an alpha toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron. *Appl Environ Microbiol* **71**; 7542–7547.

**Chevalier, B.S. y Stoddard, B.L.** (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.*, **29** (18); 3757-3774.



**Chin, K. y Pyle, A.M.** (1995). Branch-point attack in group II introns is a highly reversible transesterification, providing a potential proofreading mechanism for 5'-splice site selection. *RNA* **1**; 391–406.

**Chomczynski, P. y Sacchi, N.** (1987) Single-step methods of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**; 156-159.

**Chow, L., Gelinas, R.E., Broker, T.R. y Roberts, R. J.** (1977) An amazing sequence arrangement at the 5' end of adenovirus 2 mRNA. *Cell*, **12**; 1-8.

**Christopher, D.A. y Hallick, R.B.** (1989). *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic Acids Res.* **17**; 7591–7608.

**Chu, V.T., Liu, Q., Podar, M., Perlman, P.S., y Pyle, A.M.** (1998). More than one way to splice an RNA: branching without a bulge and splicing without branching in group II introns. *RNA* **4**; 1186–1202.

**Chu, V.T., Adamidi, C., Liu, Q., Perlman, P.S., y Pyle, A.M.** (2001). Control of branch-site choice by a group II intron. *EMBO J.* **20**; 6866–6876.

**Conklin, J.F., Goldman, A. y Lopez, A.J.** (2005) Stabilization and analysis of intron lariats in vivo. *Methods*, **37**; 368-375.

**Copertino, D.W. y Hallick, R.B.** (1993). Group II and group III introns of twintrons: potential relationship with nuclear pre-mRNA introns. *Trends Biochem. Sci.* **18**; 467–471.

**Copertino, D.W., Hall, E.T., Van Hook, F.W., Jenkins, K.P. y Hallick, R.B.** (1994) A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic Acids Res.*, **22**; 1029-1036.

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

**Costa, M. y Michel, F.** (1995). Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J.* **14**; 1276–1285.

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

**Costa, M. y Michel, F.** (1999). Tight binding of the 5' exon to domain I of a group II self-splicing intron requires completion of the intron active site. *EMBO J.* **18**; 1025–1037.

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

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

**Costa, M., Christian, E.L., y Michel, F.** (1998). Differential chemical probing of a group II self-splicing intron identifies bases involved in tertiary interactions and supports and alternative secondary structure model of domain V. *RNA* **4**; 1055–1068.

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

**Costa, M., Michel, F., Molina-Sanchez, M.D., Martinez-Abarca, F., y Toro, N.** (2006a). An alternative intron-exon pairing scheme implied by unexpected *in vitro* activities of group II intron RmInt1 from *Sinorhizobium meliloti*. *Biochimie* **88**; 711–717.

**Costa, M., Michel, F., y Toro, N.** (2006b). Potential for alternative intron-exon pairings in group II intron RmInt1 from *Sinorhizobium meliloti* and its relatives. *RNA* **12**; 338–341.

**Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J.E., Yang, J., Mills, D., Manias, D., Dunny, G.M., Lambowitz, A.M., y Belfort, M.** (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**; 451–462.

**Cousineau, B., Lawrence, S., Smith, D., y Belfort, M.** (2000). Retrotransposition of a bacterial group II intron. *Nature* **404**; 1018–1021.

**Cousineau, B., Lawrence, S., Smith, D., y Belfort, M.** (2001). Retrotransposition of a bacterial group II intron (Erratum). *Nature*, **414**; 84.

**Crick, F.** (1970) *Central Dogma of Molecular Biology*, *Nature* **227**; 561-563

**Cui, X., Matsuura, M., Wang, Q., Ma, H. y Lambowitz, A.M.** (2004) A group II intron-encoded maturase functions preferentially in *cis* and requires both the reverse transcriptase and X domains to promote RNA splicing. *J Mol Biol*, **340**; 211–231.



- Cui, X.** (2005) RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells. PhD dissertation, the University of Texas at Austin.
- Cui, X. y Davis, G.** (2007) Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes. *Frontiers in Bioscience*, **12**; 4972-4985.
- Curcio, M.J. y Belfort, M.** (1996). Retrohoming: cDNA mediated mobility of group II introns requires a catalytic RNA. *Cell* **84**; 9–12.
- D'Souza LM, y Zhong J.** (2002). Mutations in the *Lactococcus lactis* L1.LtrB group II intron that retain mobility in vivo. *BMC Mol. Biol.* **3**;17
- Dai, L. y Zimmerly, S.** (2002). Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res.* **30**; 1091–1102.
- Dai, L. y Zimmerly, S.** (2003). ORF-less and reversetranscriptase-encoding group II introns in archaeobacteria, with a pattern of homing into related group II intron ORFs. *RNA* **9**; 14–19.
- Dai, L., Toor,N., Olson,R., Keeping,A. y Zimmerly, S.** (2003) Database for mobile group II introns *Nucleic Acids Res.*, **31**; 424 - 426.
- Daniels, D., Michels, W.J., y Pyle, A.M.** (1996). Two competing pathways for self-splicing by group II introns; a quantitative analysis of *in-vitro* reaction rates and products. *J. Mol. Biol.* **256**; 31–49.
- Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. y Scazzochio, C.** (1982). Making ends meet: a model for RNA splicing in fungal mitochondria. *Nature* **300**; 719–724.
- de Lencastre, A., Hamill, S., y Pyle, A.M.** (2005). A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.* **12**; 626–627.
- de Lencastre, A. y Pyle, A.M.** (2008) Three essential and conserved regions of the group II intron are proximal to the 5'-splice site. *RNA*, **14**; 11-24.
- del Val, C., Rivas, E., Torres-Quesada, O., Toro, N. y Jimenez-Zurdo, J.I.** (2007) Identification of differentially expressed small non-coding RNAs in the legume endosymbiont *Sinorhizobium meliloti* by comparative genomics. *Mol. Microbiol.*, **66(5)**; 1080-1091.

**Deppenmeir, U., Johann, A., Hartsch, T., Merkl, R., Schimitz, R.A. et al.** (2002) The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between Bacteria and Archaea. *J. Mol. Micro. Biotechnol.*, **4**; 453-461.

**Dib-Hajj, S.D., Boulanger, S.C., Hebbar, S.K., Peebles, C.L., Franzen, J.S., y Perlman, P.S.** (1993). Domain 5 interacts with domain 6 and influences the second transesterification reaction of group II intron selfsplicing. *Nucleic Acids Res.* **21**; 1797–1804.

**Dickson, L., Huang, H.R., Liu, L., Matsuura, M., Lambowitz, A.M., y Perlman, P.S.** (2001). Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc. Natl. Acad. Sci. USA* **98**; 13207–13212.

**Ding, J., Hughes, S. H. y Arnold, E. A.** (1997). Protein-nucleic acid interactions and DNA conformation in a complex of human immunodeficiency virus type 1 reverse transcriptase with a double-stranded DNA template-primer. *Biopolymers*, **44**; 125–138.

**Ditta, G., Stanfield, S., Corbin, D. y Helinski, D.R.** (1980) Broad host range DNA cloning system for Gram-negative bacteria. Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **77**: 7347-7351.

**Ditta, G., Schmidhauser, T., Yakobson, E., Lu, P., Liang, X-W., Finlay, D.R., Guiney, D. y Helinski, D.R.** (1985) Plasmid related to the broad host vector, pRK290, useful for gene cloning and for monitoring gene expression. *Plasmid*, **13**; 149-153.

**Doetsch, N.A., Thompson, M.D., y Hallick, R.B.** (1998). A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? *Mol. Biol. Evol.* **15**; 76–86.

**Doetsch, N.A., Thompson, M.D., Favreau, M.R., y Hallick, R.B.** (2001). Comparison of psbK operon organization and group III intron content in chloroplast genomes of 12 Euglenoid species. *Mol. Gen. Genet.* **264**; 682–690.

**Eickbush, T. H.** (1994) Origins and evolutionary relationships of retroelements. The Evolutionary Biology of Viruses (Morse SS, ed), pp. 121–157. Raven Press, New York.

**Eickbush, T.H.** (1999). Mobile introns: retrohoming by complete reverse splicing. *Curr. Biol.* **9**; R11–14.

**Eickbush, T.H.** (2000) Molecular biology. Introns gain ground. *Nature*, **404**; 940-1, 943.

**Eldho, N.V. y Dayie, K.T.** (2007) Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: implications for catalysis. *J. Mol. Biol.*, **365**; 930-944.



Ems, S.C., Morden, C.W., Dixon, C.K., Wolfe, K.H., dePamphilis, C.W., y Palmer, J.D. (1995). Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Mol. Biol.* **29**; 721–733.

Erat, M.C., Zerbe, O., Fox, T., y Sigel, R.K. (2007). Solution structure of domain 6 from a self-splicing group II intron ribozyme: a Mg<sup>2+</sup> binding site is located close to the stacked branch adenosine. *Chem. BioChem* **8**; 306–314.

Eskes, R., Yang, J., Lambowitz, A.M., y Perlman, P.S. (1997). Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell* **88**; 865–874.

Eskes, R., Liu, L., Ma, H., Chao, M.Y., Dickson, L., Lambowitz, A.M., y Perlman, P.S. (2000). Multiple homing pathways used by yeast mitochondrial group II introns. *Mol. Cell. Biol.* **20**; 8432–8446.

Fassbender, S., Brühl, K.H., Ciriacy, M., y Kück, U. (1994). Reverse transcriptase activity of an intron encoded polypeptide. *EMBO J.* **13**; 2075–2083.

Fabrizio, P., y Abelson, J. (1990) Two domains of yeast U6 small nuclear RNA required for both steps of nuclear precursor messenger RNA splicing. *Science*, **250**; 404–409.

Ferat, J. L. y Michel, F. (1993) Group II self-splicing introns in bacteria. *Nature*, **364**; 358–361.

Ferat, J. L., Le Gouar, M. y Michel, F. (1994) Multiple group II selfsplicing introns in mobile DNA from *Escherichia coli*. *C R Acad Sci III* **317**;141–148.

Ferat, J. L., Le Gouar, M. y Michel, F. (2003) A group II intron has invaded the genus *Azotobacter* and inserted within the termination codon of the essential *groEL* gene. *Mol Microbiol.*, **49**; 1407–1423.

Fedorova, O., Mitros, T., y Pyle, A.M. (2003). Domains 2 and 3 interact to form critical elements of the group II intron active site. *J. Mol. Biol.* **330**; 197–209.

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

Fedorova, O. y Zingler, N. (2007) Group II introns: structure, folding and splicing mechanism. *Biol. Chem.*, **388**; 665–678

Fedorova, O., Waldsich, C., y Pyle, A.M. (2007). Group II intron folding under near-physiological conditions: collapsing to the near-native State. *J. Mol. Biol.* **366**; 1099–1114.



**Feselstein, J.** (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**;783-791.

**Feselstein, J.** (1995) PHYLIP (Phylogenie Inference Package) version 3.55c. Distributed by the author. Department of genetics, University of Washington, Seattle. <http://evolution.genetics.washington.edu/phylip/getme.html>.

**Fernández-López, M., Muñoz-Adelantado, E., Gillis, M., Willems, A. y Toro, N.** (2005) Dispersal and evolution of the *Sinorhizobium meliloti* group II RmInt1 intron in bacteria that interact with plants. *Mol Biol Evol* **22**; 1518–1528

**Finan, T.M., Weidner, S., Wong, K., Buhrmester, J., Chain, P., Vorhölter, F.J., Hernández-Lucas, I., Becker, A., Cowie, A., Gouzy, J. et al.** (2001) The complete sequence of the 1,683-kb pSymB megaplasmid from the N<sub>2</sub>-fixing endosymbiont *Sinorhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **98**; 9889-9894.

**Franzen, J.S., Zhang, M., y Peebles, C.L.** (1993). Kinetic analysis of the 5' splice junction hydrolysis of a group II intron promoted by domain 5. *Nucleic Acids Res.* **21**; 627–634.

**Frazier, C.L., San Filippo, J., Lambowitz, A.M., y Mills, D.A.** (2003). Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl. Environ. Microbiol.* **69**; 1121–1128.

**Friedmann, H.C.** (2004) From 'butyribacterium' to 'E. coli': an assay on unity in biochemistry. *Perspect Biol Med* **47**;47–66.

**Frost, L.S., Leplae, R., Summers, A.O. y Toussaint, A.** (2005) Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.*, **3 (9)**; 722-732.

**Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., et al.** (2002) The Genome of *Methanosarcina acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**; 532-542.

**Galibert, F., Finan, T.M., Long, S.R., Pühler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P. et al.** (2001) The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**; 668-672.

**García-Rodríguez, F.M. y Toro, N.** (2000) *Sinorhizobium meliloti nfe* (nodulation formation efficiency) genes exhibit temporal and spatial expression patterns similar to those of genes involved in symbiotic nitrogen fixation. *Mol. Plant-Microbe Interact.* **13**; 583-591.



- García-Rodríguez, F.M., Zekrí, S. y Toro, N.** (2000) Characterization of the *Sinorhizobium meliloti* genes encoding a functional dihydropicolinate synthase (*dapA*) and dihydropicolinate reductase (*dapB*). *Arch. Microbiol.* **173**; 438-444.
- Gaur, R.K., McLaughlin, L.W. y Green, M.R.** (1997) Functional group substitutions of the branchpoint adenosine in a nuclear pre-mRNA and a group II intron. *RNA*, **3**; 861-869.
- Goldschmidt-Clermont, M., Girard-Bascou, J., Choquet, Y., y Rochaix, J.D.** (1990). *Trans*-splicing mutants of *Chlamydomonas reinhardtii*. *Mol. Gen. Genet.*, **223**; 417–425.
- Goldschmidt-Clermont, M., Choquet, Y., Girard-Bascou, J., Michel, F., Schirmer-Rahire, M., y Rochaix, J.D.** (1991). A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. *Cell*, **65**; 135–143.
- Gorbalenya, A.E.** (1994). Self-splicing group I and group II introns encode homologous (putative) DNA endonucleases of a new family. *Protein Sci.* **3**; 1117–1120.
- Gordon, P.M., Sontheimer, E.J., y Piccirilli, J.A.** (2000a). Kinetic characterization of the second step of group II intron splicing: role of metal ions and the cleavage site 2'-OH in catalysis. *Biochemistry* **39**; 12939–12952.
- Gordon, P.M., Sontheimer, E.J., y Piccirilli, J.A.** (2000b). Metal ion catalysis during the exon-ligation step of nuclear pre-mRNA splicing: extending the parallels between the spliceosome and group II introns. *RNA* **6**; 199–205.
- Gordon, P.M. y Piccirilli, J.A.** (2001). Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat. Struct. Biol.* **8**; 893–898.
- Gordon, P.M., Fong, R. y Piccirilli, J.A.** (2007). A second divalent metal ion in the group II intron reaction center. *Chem. & Biol.*, **14**; 607-612.
- Göttfert, M., Röthlisberger, S., Kundig, Ch., Beck, Ch., Marty, R. y Hennecke, H.** (2001) Potential Symbiosis-Specific genes uncovered by sequencing a 410-kilobase DNA region of the *Bradyrhizobium japonicum* chromosome. *J. Bacteriology*, **183**; 1405- 1412.
- Granlund M, Michel F y Norgren M** (2001) Mutually exclusive distribution of IS1548 and GBSi1, an active group II intron identified in human isolates of group B Streptococci. *J Bacteriol* **183**; 2560–2569.
- Gregan, J., Kolisek, M., y Schweyen, R.J.** (2001). Mitochondrial Mg<sup>2+</sup> homeostasis is critical for group II intron splicing *in vivo*. *Genes Dev.* **15**; 2229–2237.

**Gregory TR** (2001). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews*, **76**; 65-101.

**Griffin, E.A., Jr., Qin, Z., Michels, W.J., Jr., y Pyle, A.M.** (1995). Group II intron ribozymes that cleave DNA and RNA linkages with similar efficiency, and lack contacts with substrate 2'-hydroxyl groups. *Chem. Biol.* **2**; 761–770.

**Gumbs, O.H., Padgett, R.A., y Dayie, K.T.** (2006). Fluorescence and solution NMR study of the active site of a 160-kDa group II intron ribozyme. *RNA* **12**; 1693–1707.

**Guo, H., Zimmerly, S., Perlman, P.S., y Lambowitz, A.M.** (1997). Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J.* **16**; 6835–6848.

**Guo, H., Karberg, M., Long, M., Jones, J. P. 3rd, Sullenger, B. y Lambowitz, A. M.** (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* **289**; 452–57

**Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A., y Stutz, E.** (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.*, **21**; 3537 - 3544.

**Hamill, S. y Pyle, A.M.** (2006). The receptor for branch-site docking within a group II intron active site. *Mol. Cell* **23**; 831–840.

**Harris-Kerr, C.L., Zhang, M., y Peebles, C.L.** (1993). The phylogenetically predicted base-pairing interaction between  $\alpha$  and  $\alpha'$  is required for group II splicing *in vitro*. *Proc. Natl. Acad. Sci. USA* **90**; 10658–10662.

**Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E. R., Karol, K. G., McCourt, R. M., y Zimmerly, S.** (2006) Origin and Evolution of the Chloroplast *trnK* (*matK*) Intron: A Model for Evolution of Group II Intron RNA Structures. *Mol. Biol. Evol.*, **23**; 380 - 391.

**Hebbar, S.K., Belcher, S.M., y Perlman, P.S.** (1992). A maturase-encoding Group IIA intron of yeast mitochondria selfsplices *in vitro*. *Nucleic Acids Res.* **20**; 1747–1754.

**Hess, W.R., Hoch, B., Zeltz, P., Hübschmann, T., Kössel, H., y Börner, T.** (1994). Inefficient rpl2 splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* **6**; 1455–1465.

**Heuer, H., Engelen, B. y Smalla, K.** (1997) A Modern Soil Microbiology (van Elsas, J. D., Wellington, F. M. H., y Trevors, J. T., eds.), pp. 353–373, Dekker, New York.



- Hsu, M.-Y., Inouye, M. e Inouye, S.** (1990) Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: a potential transposable element encoding both reverse transcriptase and Dam methylase functions. *Proc. Natl. Acad. Sci. U.S.A.*, **87 (23)**; 9454-9458.
- Hetzer, M., Wurzer, G., Schweyen, R.J., y Mueller, M.W.** (1997a) *Trans*-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386**; 417–420.
- Hetzer, M., Schweyen, R.J., y Mueller, M.W.** (1997b) DNA polymerization by a group II intron *in vitro*. *Nucleic Acids Res.*, **25**; 1825-1829.
- Higgins, D.G., Thompson, J.D. y Gibson, T.J.** (1996). Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**; 383-402.
- Higuchi, R., Krummel, B. y Saiki, R.** (1988) A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.*, **16 (15)**; 7351-67.
- Hiller, R., Hetzer, M Schweyen, R.J. y Mueller, M.W.** (2000) Transposition and exon shuffling by group II intron RNA molecules in pieces. *J Mol Biol*, **297(2)**; 301-8.
- Holländer, V. y Kück, U.** (1999a) Group II intron splicing in chloroplasts: identification of mutations determining intron stability and fate of exon RNA. *Nucleic Acids Res.*, **27**; 2345–2353.
- Holländer, V. y Kück, U.** (1999b) Group II intron splicing in *Escherichia coli*: phenotypes of *cis*-acting mutations resemble splicing defects observed in organelle RNA processing. *Nucleic Acids Res.* **27**; 2339–2344.
- Huang, H., Chopra, R., Verdin, G. F. y Harrison, S. C.** (1998) Structure of a covalently trapped catalytic complex of HIV-1 RT: implications for drug resistance. *Science* **282**;1669–1675.
- Huang, H.R., Chao, M.Y., Armstrong, B., Wank, Y, Lambowitz, A.M. y Perlman, P.S.** (2003) The DIVa maturase binding site in the yeast group II intron aI2 is essential for intron homing but not for *in vivo* splicing. *Mol. Cell. Biol.*, **23**; 8809-8819.
- Huang, H-R., Rowe, C.E., Mohr, S., Jiang, Y., Lambowitz, A.M. y Perlman, P.S.** (2005) The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc. Natl. Acad. Sci. U.S.A.*, **102**; 163 - 168.
- Hübschmann, T., Hess, W.R., y Börner, T.** (1996). Impaired splicing of the rps12 transcript in ribosome deficient plastids. *Plant Mol. Biol.* **30**; 109–123.

- Ichiyanagi, K., Beauregard, A., Lawrence, S., Smith, D., Cousineau, B., y Belfort, M.** (2002). Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol. Microbiol.* **46**; 1259–1272.
- Ichiyanagi K, Beauregard A y Belfort M** (2003) A bacterial group II intron favors retrotransposition into plasmid targets. *Proc. Natl Acad Sci USA* **100**; 15742–15747.
- Inouye, M. e Inouye, S.** (1991a) msDNA y bacterial reverse transcriptase. *Annu. Rev. Microbiol.*, **45**; 163-186.
- Inouye, M. e Inouye, S.** (1991b) Retroelements in bacteria. *TIBS*, **16**; 18-21.
- Inouye, M. e Inouye, S.** (1993) The retron: a bacterial retroelement required for the synthesis of msDNA. *Curr. Op. Gen. and Dev.*, **3**; 713-718.
- Inouye, S. e Inouye, M.** (1996) Structure, Function and Evolution of bacterial reverse transcriptase. *Virus Genes*, **11**; 81-94.
- Ish-Horowicz, D. y Burke, J.F.** (1981) Rapid and efficient cosmid vector cloning. *Nucleic Acids Res.* **9**: 2989-2998.
- Jacquier, A.** (1990). Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem. Sci.* **15**; 351–354.
- Jacquier, A.** (1996). Group II introns: elaborate ribozymes. *Biochimie* **78**; 474–487.
- Jacquier, A. y Rosbash, M.** (1986). Efficient *trans*-splicing of a yeast mitochondrial RNA group II intron implicates a strong 5' exon-intron interaction. *Science* **234**; 1099–1104.
- Jacquier, A. y Michel, F.** (1987). Multiple exon-binding sites in class II self-splicing introns. *Cell* **50**; 17–29.
- Jacquier, A.** (1990). Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem Sci*, **15(9)**; 351-4.
- Jacquier, A. y Michel, F.** (1990). Base-pairing interactions involving the 5' and 3'-terminal nucleotides of group II self-splicing introns. *J. Mol. Biol.* **213**; 437–447.
- Jacquier, A. y Jacquesson-Breuleux, N.** (1991). Splice site selection and role of the lariat in a group II intron. *J. Mol. Biol.* **219**; 415–428



---

**Jarrell, K.A., Peebles, C.L., Dietrich, R.C., Romiti, S.L., y Perlman, P.S.** (1988). Group II intron self-splicing: alternative reaction conditions yield novel products. *J. Biol. Chem.* **263**; 3432–3439.

**Jeffreys, A. J. y Flavell, R.A.** (1977). The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell*, **12(4)**; 1097-108.

**Jenkins, B.D., Kulhanek, D.J., y Barkan, A.** (1997). Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* **9**; 283–296.

**Jenkins, B.D. y Barkan, A.** (2001). Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J.* **20**; 872–879.

**Jestin, J.-L., Deme, E., y Jacquier, A.** (1997). Identification of structural elements critical for inter-domain interactions in a group II self-splicing intron. *EMBO J.* **16**; 2945–2954.

**Jiménez-Zurdo, J. I., García-Rodríguez, F. M., Barrientos-Durán, A., Toro, N.** (2003). DNA target site requirements for homing in vivo of a bacterial group II intron encoding a protein lacking the DNA endonuclease domain. *J. Mol. Biol.* **326**;413–23

**Jones III, J.P., Kerlin, M.N., Coon; R.G., Perutka, J., Lambowitz, A.M. y Sullenger, B.A.** (2005). Retargeting mobile group II introns to repair mutant genes. *Mol. Therapy*, **11 (5)**; 687-694.

**Jurica, M.S. y Stoddard, B.L.** (1999). Homing endonucleases: structure, function and evolution. *Cel. Mol. Life Sci.*, **55**; 1304-1326.

**Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., y Lambowitz, A.M.** (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat. Biotechnol.* **19**; 1162–1167.

**Karberg, M.S.** (2005). Group II intron mobility and its applications in biotechnology and gene therapy. PhD. Dissertation, University of Texas at Austin.

**Kennell, J.C., Moran, J.V., Perlman, P.S., Butow, R.A., y Lambowitz, A.M.** (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell* **73**; 133–146.

**Kingston, E. Chomcsynski, P. y Sacchi, N.** (1996) Guanidine methods for total RNA preparation. Current Protocols in Molecular Biology 4.2.1.-4.2.9. John Wiley & Sons, Inc.

**Klein, J.R. y Dunny, G.M.** (2002). Bacterial group II introns and their association with mobile genetic elements. *Frontiers in Bioscience*, **7**; 1843-1856.

**Knoop, V., Brennicke, A.** (1993). Group II introns in plant mitochondria - Trans-splicing, RNA editing, evolution, and promiscuity. In: Plant Mitochondria. Brennicke, A., Kück, U. (eds.) Weinheim: VCH Chemie, pp. 221-232

**Knoop, V., Kloska, S., y Brennicke, A.** (1994). On the identification of group II introns in nucleotide sequence data. *J Mol Biol*, **242(4)**; 389-96.

**Knoop, V. y Brennicke, A.** (1994). Promiscuous mitochondrial group II intron sequences in plant nuclear genomes. *J Mol Evol*, **39(2)**; 144-50.

**Knoop, V., Altwasser, M. y Brennicke, A.** (1997). A tripartite group II intron in mitochondria of an angiosperm plant. *Mol Gen Genet*, **255(3)**; 269-76.

**Koch, J.L., Boulanger, S.C., Dib-Hajj, S.D., Hebbar, S.K., y Perlman, P.S.** (1992). Group II introns deleted for multiple substructures retain self-splicing activity. *Mol. Cell. Biol.* **12**; 1950–1958.

**Kohlstaedt, L.A., Wang, J., Friedman, J.M., Rice, P.A., y Steitz, T.A.** (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**; 1783–1790.

**Kholodii, G.Y., Mindlin, S.Z., Gorlenko, Z.M., Bass, I.A., Kalyaeva, E.S. y Nikiforov, V.** (2000). Host-dependent transposition of Tn5041. *Russ. J. Genet.* **36**; 365-373.

**Konforti, B.B., Liu, Q., y Pyle, A.M.** (1998a). A map of the binding site for catalytic domain 5 in the core of a group II intron ribozyme. *EMBO J.* **17**; 7105–7117.

**Konforti, B.B., Abramovitz, D.L., Duarte, C.M., Karpeisky, A., Beigelman, L., y Pyle, A.M.** (1998b). Ribozyme catalysis from the major groove of group II intron domain 5. *Mol. Cell* **1**; 433–441.

**Kono, M., Satoh, H., Okabe, Y., Abe, Y., Nakayama, K. y Okada, M.** (1991). Nucleotide sequence of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from the green alga *Bryopsis maxima*. *Plant Mol Biol*, **17(3)**; 505-8.

**Koonin, E. V.** (2006). The origins of introns and their role in eukaryogenesis: a compromise solution to introns-early vs. introns-late debate? *Biol Direct* **1**; 39.

**Kück, U., Choquet, Y., Schneider, M., Dron, M. y Bennoun, P.** (1987). Structural and transcription analysis of two homologous genes for the P700 chlorophyll a-



---

apoproteins in *Chlamydomonas reinhardtii*: evidence for in vivo *trans*-splicing. *EMBO J*, **6(8)**; 2185-2195.

**Kück, U.** (1989). The intron of a plastid gene from a green alga contains an open reading frame for a reverse transcriptase-like enzyme. *Mol Gen Genet*, **218(2)**; 257-65

**Kulaeva, O.I., Koonin, E.V., Wootton, J.C., Levine, A.S. y Woodgate, R.** (1998). Unusual insertion element polymorphisms in the promoter and terminator regions of the *mutAB*-like genes of R471a and R446b. *Mutation Research*, **397**; 247-262.

**Kwakman, J.H., Konings, D., Pel, H.J., y Grivell, L.A.** (1989). Structure-function relationships in a selfsplicing group II intron: a large part of domain II of the mitochondrial intron aI5 is not essential for self-splicing. *Nucleic Acids Res.* **17**; 4205-4216.

**Lambowitz AM.** (1989). Infectious introns. *Cell* **56**; 323-26

**Lambowitz, A.M. y Perlman, P.S.** (1990). Involvement of aminoacyl-tRNA synthetases and other proteins in group I and group II intron splicing. *Trends Biochem. Sci.* **15**; 440-444.

**Lambowitz, A.M. y Belfort, M.** (1993). Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**; 587-622

**Lambowitz, A.M., Caprara, M.G., Zimmerly, S., Perlman, P.S.** (1999). Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In *TheRNAWorld*, ed. RF Gesteland, TR Cech, JF Atkins, pp. 451-85. Cold Spring Harbor, NY: Cold Spring Harbor Lab. Press. 2nd ed.

**Lambowitz, A.M. y Zimmerly, S.** (2004). Mobile group II introns. *Ann Rev Genet* **38**; 1-35.

**Landegren, U., Kaiser, R., Sanders, J. y Hood, L.** (1988). A ligase-mediated gene detection technique. *Science*, **241**; 1077 - 1080.

**Lampson, B.C., Inouye, M. e Inouye, S.** (2005). Retrons, msDNA, and the bacterial genome. *Cytogenet. Genome Res.*, **110(1-4)**; 491-499.

**Lazowska, J., Meunier, B., y Macadre, C.** (1994). Homing of a group II intron in yeast mitochondrial DNA is accompanied by unidirectional co-conversion of upstreamlocated markers. *EMBO J.* **13**; 4963-72

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



**Lewin, B.** (2007) Genes IX. Ed. Jones & Barlett Publishers.

**Lewin, A.S. y Hauswirth, W.W.** (2001). Ribozyme gene therapy: applications for molecular medicine. *TRENDS Mol. Med.*, **7** (5); 221-228.

**Lim, D.** (1991). Structure of two retrons of *Escherichia coli* and their common chromosomal insertion site. *Mol. Microbiol.*, **5** (8); 1863-1872.

**Lim, D.** (1995). Analysis of a retron EC86 and EC67 insertion site in *E. coli*. *Plasmid*, **34**; 58-61.

**Lin, X., Kaul, S., Rounsley, S., Shea, T.P., Benito, M.I. et al.** (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature*, **402**; 761-768.

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

**Liu, Q., Green, J.B., Khodadadi, A., Haerberli, P., Beigelman, L. y Pyle, A.M.** (1997). Branch-site selection in a group II intron mediated by active recognition of the adenine amino group and steric exclusion of non-adenine functionalities. *J. Mol. Biol.* **267**; 163-171.

**Long, M.B., Jones III, J.P., Sullenger, B.A. y Byun, J.** (2003). Ribozyme-mediated revision of RNA and DNA. *J. Clin. Invest.*, **112** (3); 312-318.

**Lorsch, J.R., Bartel, D.P., y Szostak, J.W.** (1995). Reverse transcriptase reads through a 29-59 linkage and a 29-thiophosphate in a template. *Nucl. Acids Res.* **23**; 2811-2814.

**Lowe, D.M., Parmar, V., Kemp, S.D. y Larder, B.A.** (1991). Mutational analysis of two conserved sequence motifs in HIV-1 reverse transcriptase. *FEBS Letters* **282**; 231-234.

**Luan, D.D., Korman, M.H., Jakubczak, J.L y Eickbush, T.H.** (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**; 595-605.

**Lykke-Andersen, J., Aagaard, C., Semionenkov, M., y Garrett, R.A.** (1997). Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci.* **22**; 326-331.

**Madhani, H.D., BordonnB,Ft., y Guthrie,C.** (1990). Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev.*, **4**; 2264-2277.

**Madhani, H.D. y Guthrie, C.** (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell*, **71**; 803-817.



**Manhart, J.R. y VonderHaar, R.A.** (1991). Intron revealed by nucleotide sequence of large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Codium fragile* (Chlorophyta): Phylogenetic analysis. *J. Phycol.* **27**; 613-617.

**Marienfeld, J., Unseld, M. y Brennicke, A.** (1999). The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends Plant Sci.*, **4**; 495-502.

**Marqués, S., Ramos, J.L. y Timmis, K.N.** (1993) Analysis of the mRNA structure of the *Pseudomonas putida* TOL meta fission pathway operon around the transcription initiation point, the xylTE and the xylFJ regions. *Biochem. Biophys. Acta* **1.216**: 227-236.

**Martínez-Abarca, F., Zekri, S. y Toro, N.** (1998). Characterization and splicing *in vivo* of a *Sinorhizobium meliloti* group II intron associated with particular insertion sequences of the IS630- Tc1/IS3 retroposon superfamily. *Mol Microbiol* **28**; 1295–1306.

**Martínez-Abarca, F., García-Rodríguez, F.M., Muñoz, E., y Toro, N.** (1999). Biochemical demonstration of reverse transcriptase activity and splicing efficiency of a bacterial group II intron from *Sinorhizobium meliloti*. *Nucleic Acids Symp. Ser.* **41**; 117–119.

**Martínez-Abarca, F. y Toro, N.** (2000a). Group II introns in the bacterial world. *Mol. Microbiol.* **38**; 917–926.

**Martínez-Abarca, F. y Toro, N.** (2000b). RecA-independent ectopic transposition *in vivo* of a bacterial group II intron. *Nucleic Acids Res.* **28**; 4397-4402.

**Martínez-Abarca, F., García-Rodríguez, F.M. y Toro, N.** (2000). Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol Microbiol* **35**; 1405–1412.

**Martínez-Abarca, F., Barrientos-Durán, A., Fernández-López, M. y Toro, N.** (2004). The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res* **32**; 2880–2888.

**Mathews, D.H., Sabina, J., Zuker, M. y Turner, D.H.** (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**; 911-940.

**Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., Cavanagh, S., Dunny, G.M., Belfort, M., y Lambowitz, A.M.** (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical

demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11**; 2910–2924.

**Matsuura, M., Noah, J.W., y Lambowitz, A.M.** (2001). Mechanism of maturase-promoted group II intron splicing. *EMBO J.* **20**; 7259–7270.

**McPheeters, D.S. y Abelson, J.** (1992). Mutational analysis of the yeast U2 snRNA suggests a structural similarity to the catalytic core of group II introns. *Cell*, **71**; 819–831.

**Meng, Q., Wang, Y. y Liu, X.Q.** (2005). An intron-encoded protein assists RNA splicing of multiple similar introns of different bacterial genes. *J Biol Chem* **280**; 35085–35088.

**Mercado-Blanco, J.** (1993). Estudio de algunas características de los plásmidos no simbióticos de *Rhizobium meliloti* GR4. Análisis genético-molecular de la producción de melanina ligada al pRmeGR4b y de la estabilidad y replicación del pRmeGR4a. Tesis Doctoral. Universidad de Granada.

**Mercado-Blanco, J., García, F., Fernández-López, M. y Olivares, J.** (1993). Melanin production by *Rhizobium meliloti* GR4 is linked to nonsymbiotic plasmid pRmeGR4b: cloning, sequencing y expression of tyrosinase gene *mepA*. *Journal of Bacteriology*, **175**; 5403–5410.

**Mercado-Blanco, J. y Olivares, J.** (1994). The large nonsymbiotic plasmid pRmeGR4a de *Rhizobium meliloti* GR4 encodes a protein involved in replication that has homology with the RepC protein of *Agrobacterium* plasmids. *Plasmid*, **32**; 1–5.

**Meunier, B., Tian, G.-L., Macadre, C., Slonimski, P.P. y Lazowska, J.** (1990). Group II introns transpose in yeast mitochondria. In *Structure Function and Biogenesis of Energy Transfer Systems*, ed. E Quagliariello, S Papa, F Palmieri, C Saccone, pp. 169–74. Amsterdam: Elsevier

**Michel, F., Jacquier, A., y Dujon, B.** (1982). Comparison of fungal mitochondrial introns reveals extensive homologies in RNA secondary structure. *Biochimie* **64**; 867–861.

**Michel, F. y Lang, B.F.** (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature* **316**; 641–43

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

**Michel, F. y Ferat, J.L.** (1995). Structure and activities of group II introns. *Annu Rev Biochem* **64**; 435–461.



---

**Michel, F., Costa, M., Doucet, A.J. y Ferat, J.-L.** (2007). Specialized lineages of bacterial group II introns. *Biochimie*, **89**; 542-553.

**Michels, W.J. y Pyle, A.M.** (1995). Conversion of a group II intron into a new multiple-turnover ribozyme that selectively cleaves oligonucleotides: elucidation of reaction mechanism and structure/function relationships. *Biochemistry* **34**; 2965–2977.

**Mikheeva, S., Murray, H.L., Zhou, H., Turczyk, B.M., y Jarrell, K.A.** (2000). Deletion of a conserved dinucleotide inhibits the second step of group II intron splicing. *RNA* **6**; 1509–1515.

**Mills, D.A., McKay, L.L., y Dunny, G.M.** (1996). Splicing of a group II intron involved in the conjugative transfer of pRS01 in Lactococci. *J. Bacteriol.* **178**; 3531–3538.

**Mills, D.A., Manias, D.A., McKay, L.L., y Dunny, G.M.** (1997). Homing of a group II intron from *Lactococcus lactis* subsp. *lactis* ML3. *J. Bacteriol.* **179**; 6107–6111.

**Mills, D.A.** (2001). Mutagenesis in the post genomics era: tools for generating insertional mutations in the lactic acid bacteria. *Curr. Opin. Biotech.* **12**; 503-509.

**Mohr, G., Perlman, P.S., y Lambowitz, A.M.** (1993). Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res.* **21**; 4991–4997.

**Mohr, G., Smith, D., Belfort, M. y Lambowitz, A.M.** (2000). Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* **14**; 559– 73

**Mohr, S., Stryker, J. y Lambowitz, A.M.** (2002). A DEAD-box protein functions as an ATP-dependent RNA chaperone in group I intron splicing. *Cell*, **109**; 769-779.

**Mohr, G. y Lambowitz, A.M.** (2003). Putative proteins related to group II reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res.* **31**; 647–652.

**Molina-Sánchez, M.D., Martínez-Abarca, F. y Toro, N.** (2006). Excision of the *Sinorhizobium meliloti* group II intron RmInt1 as circles in vivo. *J Biol Chem* **281**; 28737–28744.

**Moore, M.J., Query, C.C. y Sharp, P.A.** (1993). Splicing of precursors to mRNA by the spliceosome. In *The RNA World*, Gesteland, R. and Atkins, J. (eds) pp. 303-357. New York: Cold Spring Harbor Laboratory Press.

**Moore, M.J. y Query, C.C.** (2000) Joining of RNAs by splinted ligation. *Methods in enzymology*, **317**; 109-123.

**Moore, P.B. y Steitz, T.A.** (2006). The roles of RNA in the synthesis of protein. RNAWorld. 3rd edn (Getsland RF, Ceh TR & Atkins JF, eds), pp. 257–285. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Moran, J.V., Mecklenburg, K.L., Sass, P., Belcher, S.M., Mahnke, D., Lewin, A., y Perlman, P.** (1994). Splicing defective mutants of the *COXI* gene of yeast mitochondrial DNA: initial definition of the maturase domain of the group II intron aI2. *Nucleic Acids Res.* **22**; 2057–2064.

**Moran, J.V., Zimmerly, S., Eskes, R., Kennell, J.C., Lambowitz, A.M., Butow, R.A., y Perlman, P.S.** (1995).. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol. Cell. Biol.* **15**; 2828–2838.

**Mörl, M. y Schmelzer, C.** (1990a). Integration of group II intron bI1 into a foreign RNA by reversal of the self-splicing reaction *in vitro*. *Cell* **60**; 629–636.

**Mörl, M. y Schmelzer, C.** (1990b). Group II intron RNA-catalyzed recombination of RNA *in vitro*. *Nucleic Acids Res.* **18**; 6545–6551.

**Mörl, M., Niemer, I., y Schmelzer, C.** (1992). New reactions catalyzed by a group II intron ribozyme with RNA and DNA substrates. *Cell* **70**; 803–810.

**Morozova, T., Seo, W., y Zimmerly, S.** (2002). Noncognate template usage and alternative priming by a group II intron-encoded reverse transcriptase. *J. Mol. Biol.* **315**; 951–963.

**Mueller, M.W., Stocker, P., Hetzer, M., y Schweyen, R.J.** (1991). Fate of the junction phosphate in alternating forward and reverse self-splicing reactions of group II intron RNA. *J. Mol. Biol.* **222**; 145–154.

**Mueller, M.W., Hetzer, M., y Schweyen, R.J.** (1993a). Group II intron RNA catalysis of progressive nucleotide insertion: a model for RNA editing. *Science* **261**; 1035–1038.

**Mueller, M.W., Allmaier, M., Eskes, R., y Schweyen, R.J.** (1993b). Transposition of group II intron aI1 in yeast and invasion of mitochondrial genes at new locations. *Nature* **366**; 174–176.

**Muñoz, E., Villadas, P.J., y Toro, N.** (2001). Ectopic transposition of a group II intron in natural bacterial populations. *Mol. Microbiol.* **41**; 645–652.

**Muñoz-Adelantado, E., San Filippo, J., Martínez-Abarca, F., García-Rodríguez, F.M., Lambowitz, A.M. y Toro, N.** (2003). Mobility of the *Sinorhizobium meliloti* group



II intron RmInt1 occurs by reverse splicing into DNA, but requires an unknown reverse transcriptase priming mechanism. *J Mol Biol* **327**; 931–943.

**Muñoz Adelantado, E.** (2003). El Intrón del Grupo II de *Sinorhizobium meliloti*, RmInt1: Estudios Ecológicos y Caracterización Molecular de las Actividades Bioquímicas Implicadas en su Movilidad. Tesis Doctoral. Universidad de Granada.

**Murray, H.L., Mikheeva, S., Coljee, V.W., Turczyk, B.M., Donahue, W.F., Bar-Shalom, A. y Jarrel, K.A** (2001). Excision of group II introns as circles. *Mol Cell* **8**; 201–211.

**Nakamura, H., Oda, Y., Iwai, S., Inoue, H., Ohtsuka, E., Kanaya, S. et al.** (1991) How does RNase H recognize a DNA-RNA hybrid? *Proc. Natl. Acad. Sci. USA*, **88**; 11535-11539.

**Newman, A.J. y Norman, C.** (1991). Mutations in yeast U5 snRNA alter the specificity of 5' splice-site cleavage. *Cell*, **65**; 115-123.

**Newman, A.J. y Norman, C.** (1992). U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell*, **68**; 743-754.

**Ng, B., Nayak, S., Gibbs, M.D., Lee, J. y Bergquist, P.L.** (2007). Reverse transcriptases: intron-encoded proteins found in thermophilic bacteria. *Gene*, **393(1-2)**; 137-44.

**Nielsen, T.W.** (1998). RNA-RNA interactions in nuclear pre-mRNA splicing. In *RNA Structure and Function*, Simons, R.W. and Grunberg-Manago, M. (eds) pp.279-307. New York: Cold Spring Harbor Laboratory Press.

**Nisa-Martínez, R., Jiménez-Zurdo, J.I., Martínez-Abarca, F., Muñoz-Adelantado, E. y Toro, N.** (2007). Dispersion of the RmInt1 group II intron in the *Sinorhizobium meliloti* genome upon acquisition by conjugative transfer. *Nucleic Acids Res* **35**; 214–222.

**Noah, J. y Lambowitz, A.** (2003). Effects of maturase binding and Mg<sup>2+</sup> concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry* **42**; 12466–12480.

**Noah, J.W., Park, S., Whitt, J.T., Perutka, J., Frey, W. y Lambowitz, A.M.** (2006). Atomic force microscopy reveals DNA bending during group II intron ribonucleoprotein particle integration into double-stranded DNA. *Biochemistry* **45**; 12424–12435.

**Ochman, H., Lawrence, J.G. y Groisman, E.A.** (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**; 299-304.

**Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., et al.** (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223** (1); 1-7

**Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., et al.** (1986). Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature*, **322** (6079); 572-74.

**Osiewacz, H.D., y Esser, K.** (1984). The mitochondrial plasmid of *Podospora anserina*: A mobile intron of a mitochondrial gene. *Curr. Genet.* **8**; 299–305.

**Ostersetzer, O., Cooke, A.M., Watkins, K.P., y Barkan, A.** (2005). CRS1, a chloroplast group II intron splicing factor, promotes intron folding through specific interactions with two intron domains. *Plant Cell* **17**; 241–255.

**Ostheimer, G.J., Williams-Carrie, R., Belcher, S., Osborne, E., Gierke, J. y Barkan, A.** (2003) Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. *EMBO J.* **22**; 3919–3929.

**Ostheimer, G.J., Hadjivasiliou, H., Kloer, D.P., Barkan, A. y Matthews, B.W.** (2005) Structural analysis of the group II intron splicing factor CRS2 yields insights into its protein and RNA interaction surfaces. *J. Mol. Biol.* **345**; 51-68.

**Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. y Sharp, P.A.** (1984) Lariat RNAs as intermediates and products in the splicing of messenger RNA precursors. *Science*, **225**; 89-903.

**Padgett, R.A., Podar, M., Boulanger, S.C., y Perlman, P.S.** (1994). The stereochemical course of group II intron self-splicing. *Science*, **266**; 1685–1688.

**Palmer, J.D. y Logsdon, J.M. Jr.** (1991) The recent origins of introns. *Curr. Opin. Gen. Develop.*, **1**; 470-477.

**Parker, R., Siliciano, P.G. y Guthrie, C.** (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, **49**; 229-239.

**Peebles, C.L., Perlman, P.S., Mecklenburg, K.L., Petrillo, M.L., Tabor, J.H., Jarrell, K.A., y Cheng, H.L.** (1986). A self-splicing RNA excises an intron lariat. *Cell*, **44**; 213–223.

**Peebles, C.L., Benatan, E.J., Jarrell, K.A., y Perlman, P.S.** (1987). Group II intron self-splicing: development of alternative reaction conditions and identification of a predicted intermediate. *Cold Spring Harb. Symp. Quant. Biol.* **52**; 223–232.



- Peebles, C.L., Belcher, S.M., Zhang, M., Dietrich, R.C., y Perlman, P.S.** (1993). Mutation of the conserved first nucleotide of a group II intron from yeast mitochondrial DNA reduces the rate but allows accurate splicing. *J. Biol. Chem.* **268**; 11929–11938.
- Peebles, C.L., Zhang, M., Perlman, P.S., y Franzen, J.S.** (1995). Catalytically critical nucleotide in domain 5 of a group II intron. *Proc. Natl. Acad. Sci. USA* **92**; 4422–4426.
- Perron, K., Goldschmidt-Clermont, M. y Rochaix, J-D.** (2004) A multiprotein complex involved in chloroplast group II intron splicing. *RNA*, **10**; 704-711.
- Perutka, J., Wang, W., Goerlitz, D., y Lambowitz, A.M.** (2004). Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J. Mol. Biol.* **336**; 421–39
- Pearson, W. R. y Lipman, D. J.** (1988) Improved Tools for Biological Sequence Comparison. *Proc. Natl. Acad. Sci. USA*, **85**; 2444 - 2448.
- Plante, I. y Cousineau, B.** (2006) Restriction for gene insertion within the *Lactococcus lactis* Ll.LtrB group II intron. *RNA*, **12**; 1980–1992.
- Podar, M., Dib-Hajj, S., y Perlman, P.S.** (1995a). A UV-induced  $Mg^{2+}$ -dependent cross-link traps an active form of domain 3 of a self-splicing group II intron. *RNA*, **1**; 828–840.
- Podar, M., Perlman, P.S., y Padgett, R.A.** (1995b) Stereochemical selectivity of group II intron splicing, reverse splicing, and hydrolysis reactions. *Mol. Cell. Biol.* **15**; 4466–4478.
- Podar, M., Zhuo, J., Zhang, M., Franzen, J.S., Perlman, P.S., y Peebles, C.L.** (1998a). Domain 5 binds near a highly conserved dinucleotide in the joiner linking domains 2 and 3 of a group II intron. *RNA*, **4**; 151–166.
- Podar, M., Chu, V.T., Pyle, A.M., y Perlman, P.S.** (1998b). Group II intron splicing *in vivo* by first-step hydrolysis. *Nature*, **391**; 915–918.
- Podar, M., Perlman, P.S., y Padgett, R.A.** (1998c). The two steps of group II intron self-splicing are mechanistically distinguishable. *RNA*, **4**; 890-900.
- Podar, M. y Perlman, P.S.** (1999) Photocrosslinking of 4-thio uracil-containing RNAs supports a side-by-side arrangement of domains 5 and 6 of a group II intron. *RNA*, **5**; 318-329.



**Podar, M., Mullineaux, L., Huang, H.R., Perlman, P.S. y Sogin, M.L.** (2002) Bacterial Group II Introns in a Deep-Sea Hydrothermal Vent Environment. *Appl. Environ. Microbiol.*, **68** (12); 6392-6398.

**Puerta-Fernández, E., Romero-López, C., Barroso-delJesus, A. y Berzal-Herranz, A.** (2003) Ribozymes: recent advances in the development of RNA tools. *FEMS Microbiol. Rev.* **27**; 75-97.

**Pyle, A.M. y Green, J.B.** (1994). Building a kinetic framework for group II intron ribozyme activity: quantitation of interdomain binding and reaction rate. *Biochemistry* **33**; 2716–2725.

**Pyle, A.M., Chu, V.T., Jankowsky, E. y Boudvillain, M.** (2000) Using DNazymes to cut, process, and map RNA molecules for structural studies or modification. *Methods Enzymol.*, **317**; 140-146.

**Pyle, A.M. y Lambowitz, A.M.** (2006) Group II introns: ribozymes that splice RNA and invade DNA. RNAWorld. 3rd edn (Getsland RF, Ceh TR & Atkins JF, eds), pp. 469–505. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Pyle, A.M., Fedorova, O., y Waldsich, C.** (2007). Folding of group II introns: a model system for large, multidomain RNAs? *Trends Biochem. Sci.* **32**; 138–145.

**Qin, P.Z. y Pyle, A.M.** (1997). Stopped-flow fluorescence spectroscopy of a group II intron ribozyme reveals that domain 1 is an independent folding unit with a requirement for specific Mg<sup>2+</sup> ions in the tertiary structure. *Biochemistry* **36**; 4718–4730.

**Qin, P.Z. y Pyle, A.M.** (1998). The architectural organization and mechanistic function of group II intron structural elements. *Curr. Opin. Struct. Biol.* **8**; 301–308.

**Qiu, Y.L. y Palmer, J.D.** (2004) Many independent origins of *trans*-splicing of a plant mitochondrial group II intron. *J Mol Evol*, **59**(1); 80-9.

**Query, C.C. Moore, M.J. y Sharp, P.A.** (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes and Dev.*, **8**; 587-597.

**Query, C.C., Strobel, S.A. y Sharp, P.A.** (1996) Three recognition events at the branch-site adenine. *EMBO J.* **15**; 1392-1402.

**Rambo, R.P. y Doudna, J.A.** (2004) Assembly of an active group II intron-maturase complex by protein dimerization. *Biochemistry*, **43**; 6486-6497.

**Rawsthorne, H., Turner, K.N. y Mills, D.A.** (2006) Multicopy integration of heterologous genes, using the Lactococcal group II intron targeted to bacterial insertion sequences. *Appl. Env. Microbiol.* **72** (9); 6088–6093.



- Reed, R.** (2000). Mechanisms of fidelity in pre-mRNA splicing. *Curr. Opin. Cell Biol.* **12**; 340–345.
- Rest, J.S. y Mindell, D.P.** (2003) Retroids in archaea: phylogeny and lateral origins. *Mol Biol Evol* **20**; 1134–1142.
- Rice, S.A., Bieber, J., Chun, J.-Y., Stacey, G. y Lampson, B.C.** (1993) Diversity of retron elements in a population of Rhizobia and other Gram-negative bacteria. *J. Bacteriol.* **175(13)**; 4250-4254.
- Rice, S.A. y Lampson, B.C.** (1996) Bacterial reverse transcriptase and ms DNA. *Virus Genes*, **11**; 95-104.
- Robart, A.R. y Zimmerly, S.** (2005) Group II intron retroelements: function and diversity. *Cytogenet Genome Res* **110**; 589–597.
- Robertsen, B.K., Aiman, P., Darvill, A.G., McNeil, M. y Albersheim, P.** (1981) The structure of acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* y *Rhizobium trifolii*. *Plant Physiol.* **67**: 389-400.
- Ruskin, B., Krainer, A.R., Maniatis, T. y Green, M.R.** (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing *in vitro*. *Cell*, **38**; 317-331.
- Sainsard-Chanet, A., Begel, O., y Belcour, L.** (1994). DNA double-strand break in vivo at the 3' extremity of exons located upstream of group II introns. Senescence and circular DNA introns in *Podospora* mitochondria. *J. Mol. Biol.* **242**; 630–643.
- Saldanha, R., Mohr, G., Belfort, M., y Lambowitz, A.M.** (1993). Group I and group II introns. *FASEB J.* **7**; 15–24.
- Saldanha, R., Chen, B., Wank, H., Matsuura, M., Edwards, J., y Lambowitz, A.M.** (1999). RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*, **38**; 9069–9083.
- Sambrook, J., Fritsch, E.F. y Maniatis, T.** (1989) Molecular cloning. A laboratory Manual. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
- San Filippo, J. y Lambowitz, A.M.** (2002). Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J. Mol. Biol.* **324**; 933–951.

**Sanjuan, J. y Olivares, J.** (1989) Implication of *nifA* in regulation of genes located on a *Rhizobium meliloti* cryptic plasmid that affect nodulation efficiency. *J. Bacteriol.* **171**; 4154-4161.

**Santoro, S. W. y Joyce, G.F.** (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. USA*, **94**; 4262-4266.

**Santoro, S. W. y Joyce, G.F.** (1998) Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry*, **37**; 13330-13342.

**Sashital, D.G., Cornilescu, G., McManus, C.J., Brow, D.A. y Butcher, S.E.** (2004) U2-U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat Struct Mol Biol*, **11(12)**; 1237-42.

**Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E. y Tabata, S.** (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. *DNA Res.*, **6**; 283-290.

**Schäfer, B., Gan, L. y Perlman, P.S.** (2003) Reverse transcriptase and reverse splicing activities encoded by the mobile group II intron COB1 of fission yeast mitochondrial DNA. *J. Mol. Biol.*, **329**; 191-206.

**Schlatterer, J.C., Crayton, S.H., y Greenbaum, N.L.** (2006). Conformation of the Group II intron branch site in solution. *J. Am. Chem. Soc.* **128**; 3866–3867.

**Schmelzer, C., Schmidt, C., y Schweyen, R.J.** (1982). Identification of splicing signals in introns of yeast mitochondrial split genes: mutational alterations in intron bI1 and secondary structures in related introns. *Nucleic Acids Res.* **10** ; 6797–6808.

**Schmelzer, C., Schmidt, C., May, K., y Schweyen, R.J.** (1983). Determination of functional domains in intron bI1 of yeast mitochondrial RNA by studies of mitochondrial mutations and a nuclear suppressor. *EMBO J.* **2**; 2047–2052.

**Schmelzer, C. y Schweyen, R.J.** (1986). Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell*, **46**; 557–565.

**Schmelzer, C. y Mueller, M.W.** (1987) Self-splicing of group II introns *in vitro*: lariat formation and 3' splice site selection in mutant RNAs. *Cell*, **51**; 753-762.

**Schmidt, U., Kosack, M., y Stahl, U.** (1987). Lariat RNA of a group II intron in a filamentous fungus. *Curr. Genet.* **12**; 291–295.

**Schmidt, U., Riederer, B., Morl, M., Schmelzer, C., y Stahl, U.** (1990). Self-splicing of the mobile group II intron of the filamentous fungus *Podospora anserina* (COI I1) *in vitro*. *EMBO J.* **9**; 2289–2298.



- Schmidt, U., Sägebarth, R., Schmelzer, C. y Stahl, U.** (1993) Self-splicing of a *Podospora anserina* group II intron *in vitro*. Effects of 3'-terminal intron alterations on cleavage at the 5' and 3' splice site. *J. Mol. Biol.*, **231**; 559-568.
- Schmidt, W.M., Schweyen, R.J., Wolf, K., y Mueller, M.W.** (1994). Transposable group II introns in fission and budding yeast. Site-specific genomic instabilities and formation of group II IVS pDNAs. *J. Mol. Biol.* **243**; 157-166.
- Schmidt, U., Podar, M., Stahl, U., y Perlman, P.S.** (1996). Mutations of the two-nucleotide bulge of D5 of a group II intron block splicing *in vitro* and *in vivo*: phenotypes and suppressor mutations. *RNA*, **2**; 1161-1172.
- Selbitschka, W., Arnol, W., Jording, D., Kosier, B., Toro, N. y Pühler, A.** (1995) The insertion sequence element *IS<sub>Rm2011-2</sub>* belongs to the *IS630-Tc1* family of IS elements and is abundant in *Rhizobium meliloti*. *Gene*, **163**; 59-64.
- Sellem, C.H., Lecellier, G., y Belcour, L.** (1993) Transposition of a group II intron. *Nature*, **366**; 176-178.
- Séraphin, B., Simon, M., Boulet, A., y Faye, G.** (1989). Mitochondrial splicing requires a protein from a novel helicase family. *Nature*, **337**; 84-87.
- Sharp, P.A.** (1985) On the origin of RNA splicing and introns. *Cell*, **42**; 397-400.
- Sharp, P.A.** (1991) "Five easy pieces". *Science*, **254**; 663.
- Shearman, C., Godon, J.J., y Gasson, M.** (1996). Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*. *Mol. Microbiol.* **21**; 45-53.
- K Shinozaki, M Ohme, M Tanaka, T Wakasugi, N Hayashida, et al.** (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J*, **5(9)** ; 2043-2049.
- Shub, D.A., Goodrich-Blair, H., y Eddy, S.R.** (1994). Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.* **19**; 402-404.
- Shukla, G.C. y Padgett, R.A.** (2002) A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell*, **9(5)**; 1145-50.
- Sigel, R., Vaidya, A., y Pyle, A.** (2000). Metal ion binding sites in a group II intron core. *Nat. Struct. Biol.* **7**; 1111-1116.

**Sigel, R.K., Sashital, D.G., Abramovitz, D.L., Palmer, A.G., Butcher, S.E., y Pyle, A.M.** (2004). Solution structure of domain 5 of a group II intron ribozyme reveals a new RNA motif. *Nat. Struct. Mol. Biol.* **11**; 187–192.

**Sigel, R.K.O.** (2005) Group II intron ribozymes and metal ions- A delicate relationship. *Eur. J. Inorg. Chem.*, **2005**; 2281-2292.

**Simpson, L.** (2006) Uridine insertion/deletion RNA editing as a paradigm for site-specific modifications of RNA molecules. RNAWorld. 3rd edn (Getsland RF, Ceh TR & Atkins JF, eds), pp. 401-417. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Singh, N.N. y Lambowitz, A.M.** (2001). Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J. Mol. Biol.* **309**; 361–386.

**Singh, R.N., Saldanha, R.J., D'Souza, L.M., y Lambowitz, A.M.** (2002). Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J. Mol. Biol.* **318**; 287–303.

**Skelly, P.J., Hardy, C.M., y Clark-Walker, G.D.** (1991). A mobile group II intron of a naturally occurring rearranged mitochondrial genome in *Kluyveromyces lactis*. *Curr. Genet.* **20**; 115–120.

**Smith D, Zhong J, Matsuura M, Lambowitz AM y Belfort M** (2005) Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev.* **19**; 2477–2487.

**Sobral, B., Honeycutt, R.J., Atherly, A.G. y McClelland, M.** (1991) Electrophoretic separation of the three *Rhizobium meliloti* replicons. *J. Bacteriol.* **173**; 5173-5180.

**Sontheimer, E.J. y Steitz, J.A.** (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**; 1989-1996.

**Sontheimer, E.J., Sun, S. y Piccirilli, J.A.** (1997) Metal ion catalysis during splicing of premessenger RNA. *Nature*, **388**; 801-805.

**Sontheimer, E.J., Gordon, P.M., y Piccirilli, J.A.** (1999). Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**; 1729–1741.

**Soto, M.J., Zorzano, A., Olivares, J. y Toro, N.** (1992a) Nucleotide sequence of *Rhizobium meliloti* GR4 insertion sequence *ISRm3* linked to the nodulation competitiveness locus *nfe*. *Plant Mol. Biol.* **20N** 307-309.



- Soto, M.J., Zorzano, A., Olivares, J. y Toro, N.** (1992b) Sequence of *ISRM4* from *Rhizobium meliloti* strain GR4. *Gene* **120**; 125-126.
- Soto, M.J., Zorzano, A., Mercado-Blanco, J., Lepek, V., Olivares, J. y Toro, N.** (1993) Nucleotide sequence and characterization of *Rhizobium meliloti* nodulation competitiveness genes *nfe*. *J. Mol. Biol.* **229**; 570-576.
- Stabell, F.B., Tourasse, N.J., Ravnum, S., y Kolstø, A.B.** (2007). Group II intron in *Bacillus cereus* has an unusual 39 extension and splices 56 nucleotides downstream of the predicted site. *Nucleic Acids Res.*, **35**; 1612 - 1623.
- Steitz, T.A. y Steitz, J.A.** (1993). A general two-metalion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**; 6498–6502.
- Studier, F.W.** (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol Biol.* **219**: 37-44.
- Su, L., Qin, P., Michels, W., y Pyle, A.** (2001). Guiding ribozyme cleavage through motif recognition: the mechanism of cleavage site selection by a group II intron ribozyme. *J. Mol. Biol.* **306**; 665–668.
- Su, L.J., Brenowitz, M., y Pyle, A.M.** (2003). An alternative route for the folding of large RNAs: apparent two-state folding by a group II intron ribozyme. *J. Mol. Biol.* **334**; 639–652.
- Su, L.J., Waldsich, C., y Pyle, A.M.** (2005). An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res.* **33**; 6674–6687.
- Swisher, J., Duarte, C.M., Su, L.J., y Pyle, A.M.** (2001). Visualizing the solvent-inaccessible core of a group II intron ribozyme. *EMBO J.* **20**; 2051–2061.
- Swisher, J.F., Su, L.J., Brenowitz, M., Anderson, V.E., y Pyle, A.M.** (2002). Productive folding to the native state by a group II intron ribozyme. *J. Mol. Biol.* **315**; 297–310.
- Tabak, H.F., Van der Horst, G., Smiyh, J., Winter, A.J., Mul, Y. y Groot Koerkamp, M.J.A.** (1988). Discrimination between RNA circles, interlocked RNA circles and lariats using two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **16 (14)**; 6597-6605
- Tanner, N.K.** (1999). Ribozymes: the characteristics and properties of catalytic RNAs. *FEMS Microbiol. Rev.* **23**; 257–275.

- Thomas, C.A.** (1971). The genetic organization of chromosomes. *Annual Review of Genetics* **5**; 237-256.
- Thompson, J.D., Higgins, D.G. y Gibson, T.J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**; 4673-4680.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. y Higgins, D.G.** (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**; 4876-4882.
- Till, B., Schmitz-Linneweber, C., Williams-Carrier, R., y Barkan, A.** (2001). CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA* **7**; 1227-1238.
- Toor, N., Hausner, G. y Zimmerly, S.** (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**; 1142-1152.
- Toor, N. y Zimmerly, S.** (2002) Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns. *RNA*, **8**; 1373-1377.
- Toor, N., Robart, A.R., Christianson, J. y Zimmerly, S.** (2006) Selfsplicing of a group IIC intron: 50 exon recognition and alternative 50 splicing events implicate the stem-loop motif of a transcriptional terminator. *Nucleic Acids Res* **34**; 6561-6573.
- Toro, N.** (1985) Estudio genético de las características simbióticas de *Rhizobium meliloti*. Tesis Doctoral. Universidad de Granada.
- Toro, N. y Olivares, J.** (1986) Characterization of a large plasmid of *Rhizobium meliloti* involved in enhancing nodulation. *Mol. Gen. Genet.* **202**; 331-335.
- Toro, N., Molina-Sánchez, M.D. y Fernández-López, M.** (2002) Identification and characterization of bacterial class E group II introns. *Gene* **299**; 245-250.
- Toro, N., Martínez-Abarca, F., Fernandez-Lopez, M. y Muñoz-Adelantado, E.** (2003) Diversity of group II introns in the genome of *Sinorhizobium meliloti* strain 1021: splicing and mobility of RmInt1. *Mol. Genet. Genomics*, **268(5)**; 628-36.
- Toro, N.** (2003) Bacteria and archaea group II introns: additional mobile elements in the environment. *Environ Microbiol* **5**; 143-151.
- Toro, N., Jiménez-Zurdo, J.I. y García-Rodríguez, F.M.** (2007). Bacterial group II introns: not just splicing. *FEMS Microbiol Rev* **31**; 342-358



**Tourasse, N.J., Stabell, F.B., Reiter, L., y Kolstø, A.B.** (2005). Unusual group II introns in bacteria of the *Bacillus cereus* group. *J. Bacteriol.* **187**; 5437–5451.

**Tycowski, K.T., Kolev, N.G., Conrad, N.K., Fok, V. y Steitz, J.A.** (2006). The ever-growing world of small nuclear ribonucleoproteins. RNAWorld. 3rd edn (Getsland RF, Ceh TR & Atkins JF, eds), pp. 327-368. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Valadkhan, S.** (2007). The spliceosome: a ribozyme at heart? *Biol. Chem.*, **388**; 693–697

**Vallès, Y., Halanych, K.M. y Boore, J.L.** (2008) Group II introns break new boundaries: presence in a bilaterian's genome. *PLoS ONE*, **3(1)**; e1488.

**Van der Veen, R., Arnberg, A.C., van der Horst, G., Bonen, L., Tabak, H.F., y Grivell, L.A.** (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing *in vitro*. *Cell*, **44**; 225–234.

**Van der Veen, R., Kwakman, J.H., y Grivell, L.A.** (1987a). Mutations at the lariat acceptor site allow self-splicing of a group II intron without lariat formation. *EMBO J.* **6**; 3827–3831.

**Van der Veen, R., Arnberg, A.C., y Grivell, L.A.** (1987b). Self-splicing of a group II intron in yeast mitochondria: dependence on 5' exon sequences. *EMBO J.* **6**; 1079–1084.

**Venkatesan, M.M., Goldberg, M.B., Rose, D.J., Grotbeck, E.J., Burland, V. y Blattner, F.R.** (2001) Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. *Infection and Immunity*, **69**; 3271-3285.

**Villadas, P.J., Velázquez, E., Martínez-Molina, E. y Toro, N.** (1995) Identification of nodule-dominant *Rhizobium meliloti* strains carrying pRmeGR4b-type plasmid within indigenous soil populations by PCR using primers derived from specific DNA sequences. *FEMS Microbiol. Ecol.* **17**;161-168.

**Vogel, J., Hübschmann, T., Börner, T., y Hess, W.R.** (1997). Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J. Mol. Biol.* **270**; 179–187.

**Vogel, J., Börner, T., y Hess, W.R.** (1999). Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Res.* **27**; 3866–3874.

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



**Waldsich, C. y Pyle, A.M.** (2007). A folding control element for tertiary collapse of a group II intron ribozyme. *Nat. Struct. Mol. Biol.* **14**; 37–44.

**Wang, H. y Lambowitz, A.M.** (1993) The Mauriceville plasmid reverse transcriptase can initiate cDNA synthesis de novo and may be related to reverse transcriptase and DNA polymerase progenitor. *Cell*, **75**; 1071–1081.

**Wank, H., San Filippo, J., Singh, R.N., Matsuura, M. y Lambowitz, A.M.** (1999) A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol. Cell*, **4**; 239-250.

**Watanabe, K. y Lambowitz, A.M.** (2004). High-affinity binding site for a group II intron-encoded reverse transcriptase/maturase within a stem-loop structure in the intron RNA. *RNA*, **10**; 1433–1443.

**Will, C.L. y Lührmann, R.** (2006) Spliceosome structure and function. *RNAWorld*. 3rd edn (Getsland RF, Ceh TR & Atkins JF, eds), pp. 369-400. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

**Wolfe, K.H., Morden, C.W. y Palmer, J.D.** (1992) Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci USA.*, **89**; 10648-10652.

**Xiang, Q., Qin, P.Z., Michels, W.J., Freeland, K., y Pyle, A.M.** (1998). Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry* **37**; 3839–3849.

**Xiong, Y. y Eickbush, T.H.** (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**; 3353–3362.

**Yang, J., Zimmerly, S., Perlman, P.S., y Lambowitz, A.M.** (1996). Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature*, **381**; 332–335.

**Yang, J., Mohr, G., Perlman, P.S., y Lambowitz, A.M.** (1998). Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of aI1 and reverse splicing into DNA transposition sites *in vitro*. *J. Mol. Biol.* **282**; 505–523.

**Yao, J., Zhong, J. y Lambowitz, A.M.** (2005). Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. *Nucleic Acids Res.*, **33**; 3351-62.

**Yao, J., Zhong, J., Fang, Y., Geisinger, E., Novick, R.P. y Lambowitz, A.M.** (2006). Use of targetrons to disrupt essential and non-essential genes in *Staphylococcus aureus* reveals temperature-sensitivity of group II intron splicing. *RNA*, **12**; 1271-81.



- Yao, J. y Lambowitz, A.M.** (2007). Gene targeting in Gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. *Appl Environ Microbiol*, **73**; 2735-43.
- Yeo, C.C., Tham, J.M., Yap, M.W. y Poh, C.L.** (1997). Group II intron from *Pseudomonas alcaligenes* NCIB 9867 (P25X): entrapment in plasmid RP4 and sequence analysis. *Microbiology*, **143**; 2833-2840.
- Zekrí, S. y Toro, N.** (1996) Identification and nucleotide sequence of *Rhizobium meliloti* insertion sequence *ISRM6*, a small transposable element that belongs to the *IS3* family. *Gene* **175**; 43-48.
- Zekrí, S., Soto, M.J. y Toro, N.** (1998) *ISRM4-1* y *ISRM9*, two novel insertion sequences from *Sinorhizobium meliloti*. *Gene* **207**; 93-96.
- Zhang, L. y Doudna, J.A.** (2002). Structural insights into group II intron catalysis and branch-site selection. *Science*, **295**; 2084–2088.
- Zhao, J. y Lambowitz, A.M.** (2005) A bacterial group II intron encoded reverse transcriptase localizes to cellular poles. *Proc Natl Acad Sci USA* **102**; 16133–16140.
- Zhong, J. y Lambowitz, A.M.** (2003) Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J.* **22**; 4555–65
- Zhong J, Karberg M, Lambowitz AM.** (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res.* **31**; 1656–64
- Zimmerly, S., Guo, H., Eskes, R., Yang, J., Perlman, P.S., y Lambowitz, A.M.** (1995a). A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* **83**; 529–538.
- Zimmerly, S., Guo, H., Perlman, P.S., y Lambowitz, A.M.** (1995b). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**; 545–554.
- Zimmerly, S., Moran, J.V., Perlman, P.S., y Lambowitz, A.M.** (1999). Group II intron reverse transcriptase in yeast mitochondria. Stabilization and regulation of reverse transcriptase activity by the intron RNA. *J. Mol. Biol.* **289**; 473–490.
- Zimmerly, S., Hausner, G., y Wu, X.** (2001). Phylogenetic relationships among group II intron ORFs *Nucleic Acids Res.* **29**; 1238–1250.

**Zuker, M.** (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31** (13); 3406-3415.

# *ABREVIATURAS Y ANGLICISMOS*



- A:** Adenina
- A<sub>260nm</sub>:** Absorbancia a 600 nm de longitud de onda
- ADN:** Ácido Desoxirribonucleico
- Ap:** Amplicilina
- APS:** Persulfato amónico
- ARN:** Ácido Ribonucleico
- C:** Citosina
- Ci:** Curio
- cm:** centímetro
- c.p.m.:** cuentas por minuto
- dNTP:** Deoxinucleósido trifosfato
- ddNTP:** Di-deoxinucleósido trifosfato
- DMSO:** Dimetil sulfóxido
- D.O.<sub>260nm</sub>:** Densidad óptica a 260 nm
- DTT:** 1,4 Ditiotreitol
- EBS1, EBS2, EBS3:** Secuencias de unión al exón, del inglés Exon Binding Sites
- EDTA:** Ácido etilendiaminotetraacético
- et al.:** y colaboradores
- G:** Guanina
- Homing:** Proceso de inserción del intrón del grupo II en una copia del alelo de su elemento hospedador original
- IBS1, IBS2, IBS3:** Secuencias de unión al intrón, del inglés Intron Binding Sites
- IEP:** Proteína codificada por el intrón, del inglés Intron Encoded Protein
- IS:** Secuencia de inserción, del inglés, Insertion Sequence
- Km:** Kanamicina
- Kb:** Kilobases
- Kit:** Equipo
- LB:** medio Luria-Bertani
- LTR:** terminaciones largas repetidas del inglés (Long Terminal Repeats)
- M:** Molar
- Mb:** Megabase
- min:** Minuto
- ml:** Mililitro
- mM:** Milimolar
- mol/l:** Moles/litro
- mm:** Milímetro
- mmol/l:** Milimol/litro

**μl:** Microlitro

**μg:** Microgramo

**ng:** Nanogramo

**nm:** Nanometro

**nts:** nucleótidos

**ORF:** Marco abierto de lectura, del inglés Open Reading Frame

**pb:** Pares de bases

**PCR:** Reacción en cadena de la polimerasa

**pmol:** picomol

**p/v:** peso/volumen

**RFLP:** Polimorfismo para la Longitud de los Fragmentos de Restricción

**RNP, RNPs:** Partículas de ribonucleoproteína

**r.p.m.:** Revoluciones por minuto

**RT:** Reverso transcriptasa

***Splicing:*** Maduración del intrón del grupo II mediante un proceso por corte y empalme

**SDS:** Sodio dodecil sulfato

**Tc:** Tetraciclina

**TEMED:** N,N,N,N.tetrametiletilenodiamina

**Tn:** transposón

**TPRT:** Transcripción Reversa a partir de Cebador, del inglés Target DNA-Primed

Reverse Transcription

**TY:** medio Triptona-Yeast

**X-Gluc:** 5-bromo-4-cloro-3-indolil-β-D-glucuronida

**U:** Uracilo

**V:** Voltio

**v/v:** volumen/volumen

# *AGRADECIMIENTOS*







