**[Frontiers in Bioscience 7, d1843-1856, August 1, 2002]**

# BACTERIAL GROUP II INTRONS AND THEIR ASSOCIATION WITH MOBILE GENETIC ELEMENTS

**Joanna R. Klein and Gary M. Dunny**

*Department of Microbiology, University of Minnesota Medical School, 1460 Mayo Bldg/Box 196, 420 Delaware St. SE, Minneapolis, MN, 55455-0312*

## TABLE OF CONTENTS



## 1. ABSTRACT

Group II introns are an abundant class of self-splicing RNAs, found primarily in the organelles of plants and lower eukaryotes and in bacteria. The first bacterial group II intron identified to be functional for splicing *in vivo* was the Ll.ltrB intron of *Lactococcus lactis*. It has served as an excellent model for the study of group II intron structure and function. Taking advantage of the tools of bacterial genetics and biochemical methodologies, details of Ll.ltrB splicing and homing reactions have been elucidated and are similar to those of fungal group II introns. This review provides a summary of these results. Of particular interest is the potential use of Ll.ltrB as an agent for targeted gene disruption. In addition, the development of a genetic system to analyze Ll.ltrB splicing promises to provide new insight into group II intron structure and function. Identification and analysis of group II introns in other bacterial species is a continuing process, and a discussion of published reports on these introns is provided here. Limited functional data is available for most of these introns, but sequence analysis points out several common themes, most notably that bacterial group II introns are almost always carried on mobile genetic elements.

## 2. INTRODUCTION

Introns are segments of intervening DNA that disrupt coding sequences called exons. Introns are excised precisely from messenger RNA, a process called splicing, fusing together the mRNA of the disrupted exons so that a functional product is made. Although the presence of introns in bacteria has been known for some time, introns are still largely thought to be the hallmark of eukaryotic cells. However, the discovery of introns in bacteria has

provided great insight into intron structure, function and evolution. Two major classes of introns have been identified in bacteria: group I and group II. These introns differ rather substantially, and this review will focus only on group II introns.

A group II intron can be described as a self-splicing, mobile genetic element. The information required for excision from the sequence it disrupts is carried in the highly conserved secondary structure of the intron RNA. An intron-encoded protein (IEP) facilitates the splicing reaction. In addition to self-splicing activity, group II introns are mobile. Once spliced from an mRNA transcript, they are able to insert into intronless alleles with high efficiency, a process called homing. At lower efficiency, group II introns are capable of inserting into ectopic locations in a process called transposition.

Group II introns are found both in gram-positive and gram-negative bacterial species, organelles of lower eukaryotes and plants, and in the progenitors of these organelles, proteobacteria and cyanobacteria. In fact, much of our knowledge about the structure and function of group II introns was learned by studying organellar introns. Since the discovery of the Ll.ltrB intron in the gram-positive bacteria *Lactococcus lactis* in 1996, a significant body of work has accumulated about its specific mechanisms of splicing and mobility. Identification and characterization of group II introns in other bacterial species is also lending insight into group II intron function and distribution. Of significant interest is the observation that group II introns in bacteria are often found within mobile genetic elements such as transposons and conjugative plasmids. The focus of this review is two-fold. First, it is intended to provide a comprehensive summary of the studies performed on Ll.ltrB. Second, it will discuss other bacterial group II introns identified to date with special emphasis on the association between group II introns and mobile genetic elements.

## 3. *L. LACTIS* GROUP II INTRON Ll.ltrB

The Ll.ltrB group II intron in *Lactococcus lactis* was the first group II intron identified in bacteria shown to function *in vivo*. The intron was discovered in *L. lactis* independently by two different groups. Shearman and colleagues discovered an ORF with significant homology to ORFs within mitochondrial group II introns while sequencing the sex-factor DNA cloned from the chromosome of *L. lactis* 712 (1). Mills and colleagues discovered the same sequence in a region of the *L. lactis* ML3 conjugative plasmid, pRS01, which had been defined genetically to be required for conjugal transfer (2). The ORF in this sequence was called *ltrA*. Similar to other intron-encoded proteins, *ltrA* had seven domains characteristic of reverse transcriptases (RT1 to RT7), a maturase domain (X), and a zinc finger-like domain (Zn) characteristic of endonucleases (figure 1a). Furthermore, the *ltrBE1* and *ltrBE2* ORFs flanking *ltrA* both revealed homology to conjugative relaxase enzymes. Together, this data indicated that the conjugative relaxase of pRS01 was interrupted by a group II intron and that splicing of the intron would join the two exons *ltrBE1* and *ltrBE2* to produce a functional relaxase. This intron was named Ll.ltrB and subsequent analysis indicated that it was functional for splicing and homing *in vivo*.

### 3.1. Secondary Structure

Although group II introns are fairly divergent in their primary sequence, they fold into a complex and highly conserved secondary structure. Computer modeling of the Ll.ltrB intron sequence showed that it conformed to this characteristic group II intron structure (2) (figure 1b). The conserved structure consists of a central core with 6 radiating stem loop domains (I-VI). Based on work done with a number of group II introns (for a review see ref. 3), it is generally believed that domains I, V and VI have crucial roles in the splicing reaction. Domain V is the most highly conserved domain. It is usually 34 nt in length and folds into a hairpin with an 8 bp lower stem separated from a 6 bp upper stem by a 2 nt bulge. A 3 bp stretch at the base is nearly always conserved. Domain VI contains a non-base paired adenine residue, the "bulging A". This residue is involved in the splicing reaction, as described below. Domain I has several sequences involved in base pairing interactions. Exon binding site (EBS) 1 and EBS2 are short RNA sequences found in the loop structures of domain I. They pair with sequences in the 5' exon, called Intron Binding Site (IBS) 1 and IBS2. This interaction is important for both splicing and mobility. Additional complementary sequences within the intron are believed to be involved in the proper tertiary folding of the intron structure (2, 3). Intron domains II and III are also implicated in maintaining the proper folded structure. Domain IV consists of the IEP gene and does not have a known role in the folding or catalytic activity. This domain is looped out from the functional ribozyme molecule.

### 3.2. Splicing

The mechanism of splicing is very similar between group II introns and nuclear spliceosomal introns. However, unlike spliceosomal introns, which require an array of snRNA and protein molecules, group II intron splicing is catalyzed by the intron RNA, which folds into a catalytically active structure. The intron-encoded protein facilitates the splicing reaction, a function encoded by the maturase domain of the protein. It is thought that when the intron RNA and intron-encoded protein interact to form a ribonucleoprotein (RNP) complex, the IEP is involved in maintaining the folded structure of the RNA required for splicing. The basic splicing reaction involves two consecutive transesterification reactions, diagramed schematically in figure 2a. The first reaction involves nucleophilic attack of the 5' splice site (5' intron/exon boundary) by the 2' hydroxyl group of an adenosine residue ("bulging A") located near the 3' end of the intron. This frees exon 1 and results in a 2'-5' phosphodiester branch in the intron. In the second reaction, the 3' hydroxyl of exon 1 attacks the 3' intron/exon junction, ligating the two exons via a 3'-5' phosphodiester bond. This releases the intron as a branched lariat structure. An alternate splicing pathway may be used in which the first step occurs by hydrolysis, ultimately releasing the intron as a linear molecule rather than a lariat.

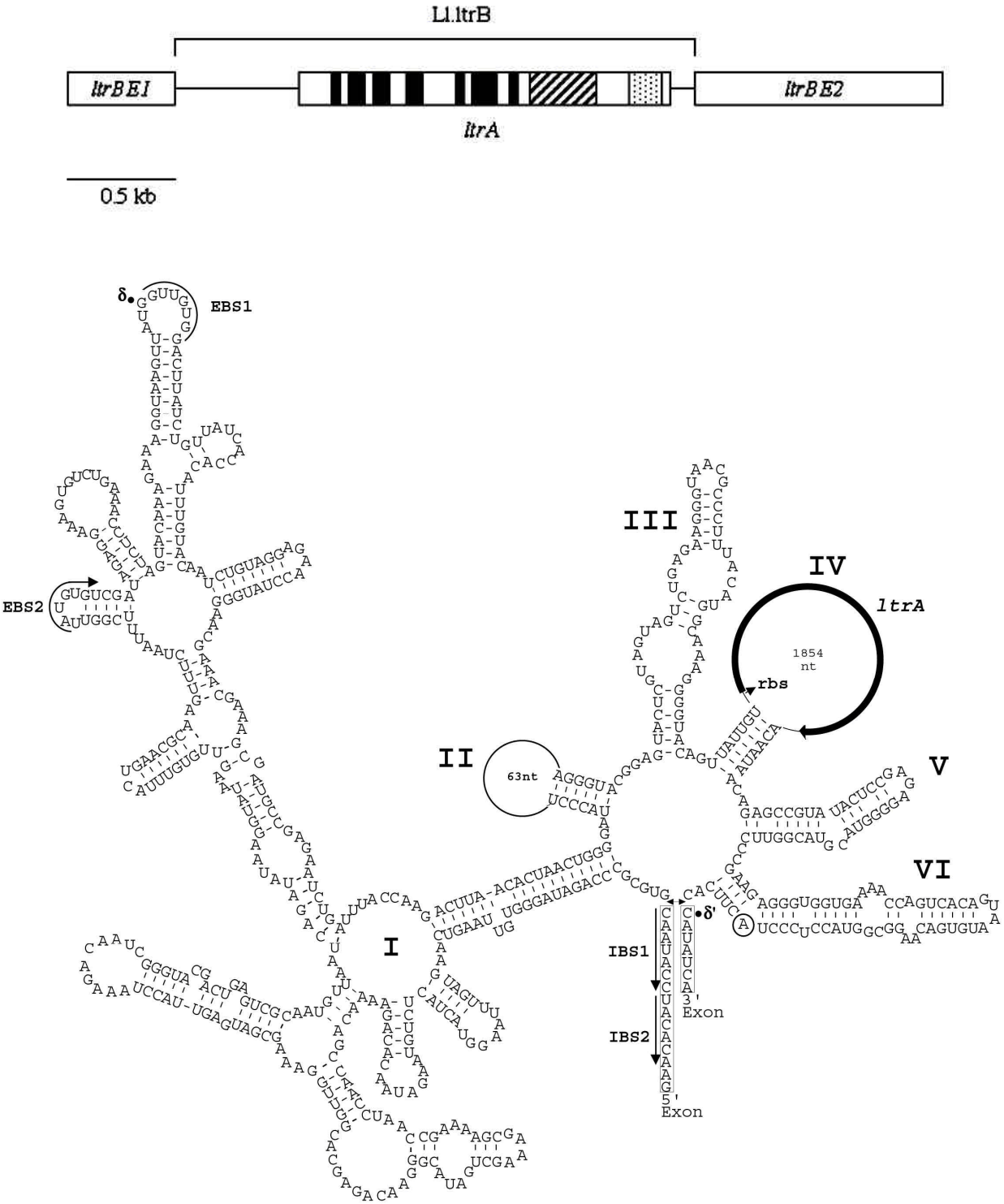


**Figure 1.** The lactococcal Ll.ltrB intron. a. Schematic diagram of Ll.ltrB DNA. The *ltrBE1* and *ltrBE2* exons flank the Ll.ltrB intron. The conserved domains of LtrA are drawn as boxes and include seven RT domains (grey), the maturase domain (striped) and the Zn domain (stippled). The internal *ltrA* promoter is indicated by an arrow. b. Secondary structure of Ll.ltrB RNA. The six major domains are indicated by roman numerals. Regions involved in base pairing to exon sequences (EBS1, EBS2, delta) are indicated. The bulging A in domain VI is circled. Exon sequences are boxed and splice sites are indicated by a double arrow.

Direct evidence that Ll.ltrB was spliced *in vivo* was obtained by doing RT-PCR analysis with primers specific to sequences within the exons *ltrBE1* and *ltrBE2* (2). An RT-PCR product was observed of a size indicating the Ll.ltrB intron had been spliced out. Sequence analysis of the RT-PCR product demonstrated that the two exons were fused together in frame. With these findings, it was clear that Ll.ltrB was a functional group II intron.

To study the biochemical reactions involved in splicing, it was desirable to purify intron-containing RNP particles. This task was made possible by expressing the Ll.ltrB intron in *E. coli* (4). RNP particles containing intron RNA and LtrA were purified and found to have splicing activity in *E. coli*. Furthermore, splicing did not occur when the IEP was deleted. However, *in vitro* transcribed RNA from a construct lacking the IEP was able to self-splice under non-physiological high salt conditions, but not under low salt conditions. When LtrA was added back to the low salt reaction, splicing was induced. These experiments were the first biochemical evidence for the maturase activity of a group II intron-encoded protein.

**3.3. Mobility**

A remarkable characteristic of group II introns is their ability to act as mobile genetic elements. The major process of mobility occurs when an intron that is present in one copy of a certain gene inserts into to an intronless allele of the same gene, a process called homing. Movement of the intron into an ectopic location by a process called transposition also occurs, but much less efficiently. Early after the discovery of the Ll.ltrB intron, studies were directed at determining whether this newly identified intron was mobile (5). Mills *et al.* demonstrated that Ll.ltrB was a mobile element and that the *ltrB* splice site region was a target site for intron insertion. In this study, a plasmid containing an intronless *ltrB* gene was introduced into two lactococcal strains, one that had a functional Ll.ltrB intron and one that was disrupted for splicing function. By examining plasmid pools, a plasmid of larger size than the original plasmid was observed only from the splicing proficient strain. The larger plasmid resulted from acquisition of intron sequences, suggesting that the Ll.ltrB intron is mobile. It was found that homing was precise and reconstituted the original intron/exon junctions. Further experiments showed that homing did not occur when the splice site region of *ltrB* was deleted, suggesting that this specific sequence is recognized.

Once it was recognized that Ll.ltrB was mobile, it became an important model for studying mobility of group II introns. Considerable work has been done on mobility of the yeast aI1 and aI2 introns (6). However, work on these yeast mitochondrial introns was somewhat limited due to several technical problems. Therefore, when the Ll.ltrB intron was discovered as a functional bacterial group II intron, it provided a genetic system in which the complex homing pathway could be dissected. A number of detailed studies of mobility of Ll.ltrB have been performed, the details of which will be discussed in subsequent paragraphs. The basic model of intron homing follows and is diagramed in figure 2b. After splicing from RNA, the excised intron lariat and IEP remain complexed in an RNP particle. The RNP recognizes a specific double stranded DNA target site. RNA catalyzed cleavage of the sense strand of the target site results in the intron RNA inserting into the target site in a process called reverse splicing. Subsequently, the antisense strand is cleaved by the endonuclease activity of the IEP. The 3' end of the cleaved antisense strand acts as a primer for reverse transcription (catalyzed by the IEP) of the reverse spliced intron RNA. Using host repair mechanisms, the cDNA copy is then incorporated into the DNA, resulting in a complete copy of the intron in a new location.

Expression of Ll.ltrB in *E. coli* has greatly facilitated the mechanistic study of homing. As mentioned above, RNP particles isolated from *E. coli* had maturase activity important for Ll.ltrB splicing. Based on the presence of conserved domains, LtrA was predicted to possess reverse transcriptase and endonuclease activities important for homing. In fact, RNP particles isolated from *E. coli* did have both RT and endonuclease activity as determined by *in vitro* assays (4). *In vivo* assays demonstrated that Ll.ltrB could home into an intronless allele in *E. coli* (7). This species was then utilized for numerous experiments examining the molecular events involved in homing. It was found that homologous recombination mediated by RecA protein is not involved in Ll.ltrB homing in *E. coli* or *L. lactis* (5, 7). Furthermore, homing uses an RNA intermediate in a process called retrohoming, depicted in figure 2b. Compelling genetic evidence for retrohoming was obtained by examining the mobility of an engineered "Twintron" derivative of Ll.ltrB, where a small, self-splicing group I intron was inserted into a region of domain IV not essential for proper folding of the Ll.ltrB RNA. Homing products obtained using the Twintron had lost the internal group I intron in most cases. This indicated that an Ll.ltrB RNA intermediate from which the group I intron had self-spliced was used in the homing reaction. Further analysis of the homing products showed that the exon sequence flanking the donor intron was not carried over into the homing product, arguing against any form of recombination between the donor and recipient DNA flanking the intron insertion sites.

Two endonuclease activities are important for Ll.ltrB homing into a target DNA site. First, reverse splicing into the sense strand results from cleavage at the exon 1/exon 2 junction. The twintron experiments outlined above implicated a process of complete reverse splicing in which the intron RNA is reverse spliced into the target DNA so that its 5' and 3' ends are joined to the respective exons. This cleavage step is dependent on the RNA component of the RNP particle (4, 7). After reverse splicing, the antisense strand is cleaved at a location 9 nt to the 3' side of the exon junction. This cleavage reaction is dependent on the IEP zinc finger domain since RNP particles consisting of a protein with a deletion of this domain did not catalyze antisense strand cleavage (4).

In order for reverse splicing to occur, the appropriate target site for homing must be recognized. Several studies have examined the specificities of target

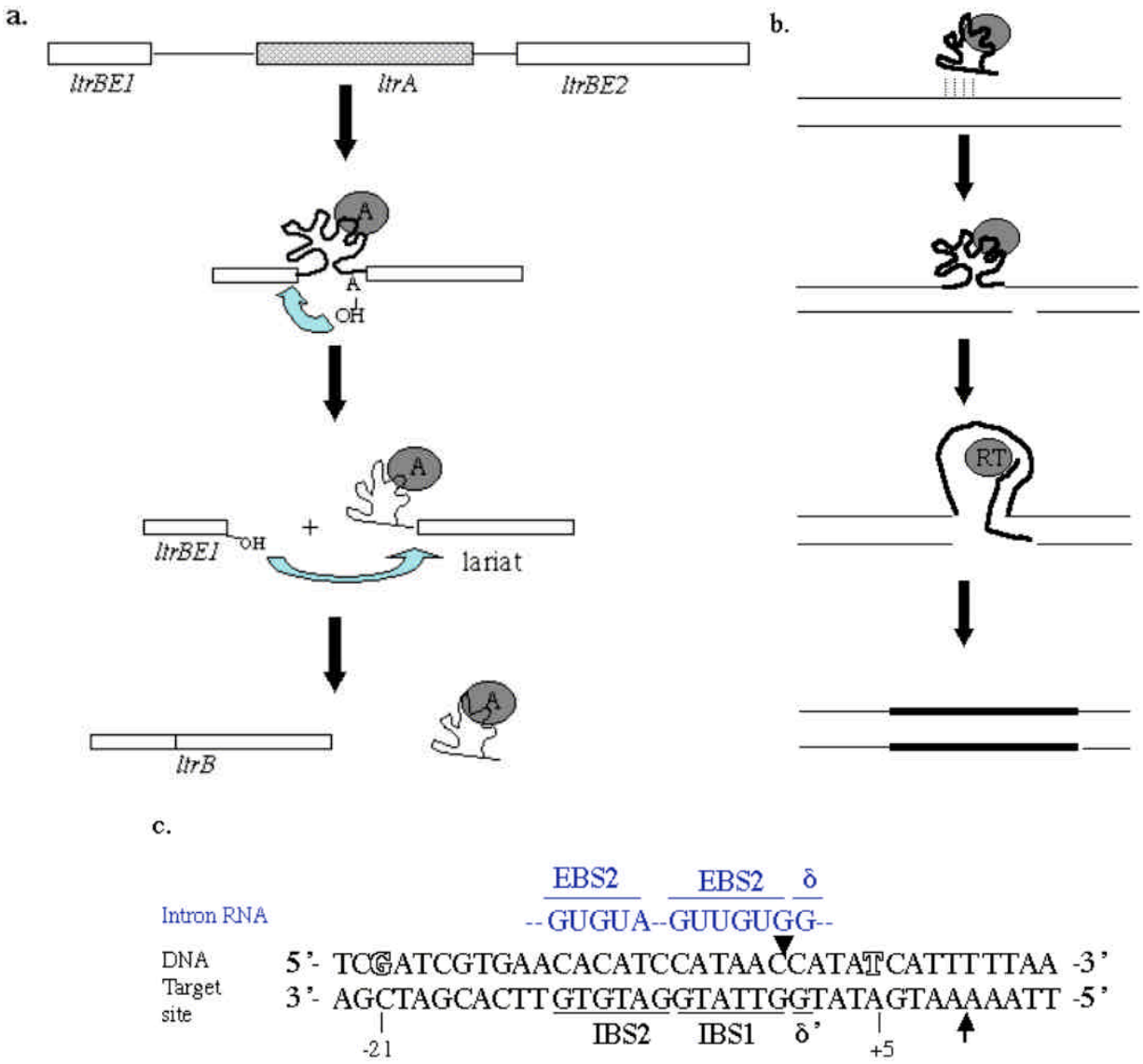


**Figure 2.** Schematic diagrams of the Ll.ltrB splicing and homing pathways. a. Splicing pathway. A transcript extending from *ltrBE1* through *ltrBE2* is depicted at the top of the diagram. LtrA is translated and interacts with the folded RNA structure to form the catalytically active ribozyme. Subsequently there is nucleophilic attack of the 5' splice junction by the hydroxyl group of the bulging A residue. The 5' exon is released and a 2'-5' phosphodiester bond within the intron forms a 'lariat' structure. A second transesterification reaction follows in which the hydroxyl group of the 5' exon attacks the 3' splice junction. The two exons are joined and the intron lariat/LtrA RNP particle is released. b. Homing pathway. The Ll.ltrB intron containing RNP particle released after splicing recognizes a double stranded DNA target site. RNA reverse splices into the exon junction of the sense strand and the antisense strand is cleaved by the endonuclease activity of LtrA. Target DNA primed reverse transcription of intron RNA is catalyzed by LtrA. Host repair mechanisms allow for the cDNA copy to be incorporated into the DNA. c. Features important for target site recognition. Shown are complementary EBS/IBS and delta/delta prime interactions between the intron RNA and DNA target site. The exon junction is indicated by an arrowhead and the site of antisense strand cleavage is indicated by an arrow. G-21 and T+5 residues critical for recognition by LtrA are outlined.

site recognition, ultimately defining the target site as a region extending from –25 to +9 nt from the intron insertion site (figure 2c). This site is recognized by both RNA and protein components of the RNP. A 14-16 nt stretch of the DNA target site is composed of short sequence elements designated IBS1, IBS2 and delta-prime. These sequences can be recognized by complementary sequences in the intron RNA (EBS1, EBS2 and delta). When certain nucleotides in these regions of the target DNA were mutated, reduced levels of reverse splicing and homing endonuclease activity were observed (4, 8). Compensatory changes in the complementary binding sites

in the intron RNA often increased function, indicating that the interaction between specific intron RNA nucleotides and DNA nucleotides in this region is direct. Regions flanking the IBS and delta-prime sites are important for the protein component of the RNP to recognize and cleave the target site. Mutagenesis experiments defined several positions in these flanking regions to be required for reverse splicing and antisense strand cleavage events. In particular, the G-21 nucleotide in the 5' exon was required for reverse splicing and bottom strand cleavage while the T+5 nucleotide in the 3' exon was required for bottom strand cleavage. Subsequently, both mutations were found to inhibit intron mobility *in vivo* (8, 9).

A subsequent study by Singh and colleagues provided a detailed view of the interaction between Ll.ltrB RNP particles and the DNA target site using various biochemical methods including DNase footprinting and modification interference mapping (10). Their results were in good agreement with the mutagenesis experiments described above. The region protected by the RNP was very similar to target site as defined by mutagenesis. In addition, interaction with G-21 and T+5 in the distal exon regions were found to be critical for RNP interaction with the target site. Experiments in this paper led to a model in which the LtrA protein first recognizes the DNA target site by major groove interactions with a small number of nucleotides in the 5' exon region including G-21. These interactions are followed by DNA unwinding and base pairing between intron RNA and the target site to allow for reverse splicing. Bottom strand cleavage proceeds through recognition of nucleotides in the 3' exon including T+5.

Based on the rather simple rules for target site selection outlined above, the idea was conceived that Ll.ltrB could be targeted to insert into alternative DNA sequences with some modification of the intron sequence. This was shown to be possible by employing two experimental approaches to target specific genes (8, 9). One method scanned the target gene for matches to the positions recognized by LtrA (G-21 and T+5). The Ll.ltrB intron was then modified to allow base pairing with the surrounding sequence. In a second method, a pool of Ll.ltrB introns mutagenized in the EBS region was screened for the ability to home into a defined target gene. Both of these methods were successful in identifying modified introns that could home into various target genes. In all homing products, the intron was found in the desired target site, indicating the intron could be reengineered with remarkable specificity. An added feature of Ll.ltrB insertion into target genes is that two outcomes are possible, depending on the DNA strand into which Ll.ltrB is inserted. If the Ll.ltrB is directed to insert into the antisense strand, it cannot subsequently splice and results in gene disruption. Alternatively, insertion into the sense strand provides for the possibility of forward splicing to occur so that gene expression is not significantly affected. Furthermore, conditional gene expression could be achieved by controlling splicing through modulation of LtrA synthesis with an inducible promoter. The ability to target Ll.ltrB to insert into specific sequences presents several practical applications. First, it provides a system for directed mutagenesis in organisms in which homologous recombination systems are lacking or not amenable, such as plants, flies, worms, mice and human cells. Second, if it is found to proceed efficiently in human cells it could have direct therapeutic applications by modifying gene expression.

While the most efficient mechanism of group II intron mobility is retrohoming, group II introns can also insert into a different location within the genome in a process called retrotransposition. Retrotransposition of Ll.ltrB into the *L. lactis* chromosome was examined in a report by Cousineau *et al.* (11). Using a temperature sensitive plasmid system to select for retrotransposition events, eight ectopic insertion sites (insertions of Ll.ltrB into regions of the chromosome other than the *ltrB* gene) were identified. A fairly well conserved IBS1 was located upstream of all of the insertion sites, suggesting that there may have been some recognition of IBS1 by EBS1 leading to reverse splicing into locations that resemble the native exon. However, DNA target sites recognized by the IEP in the retrohoming pathway (G-21and T+5) were not conserved in these sites of retrotransposition. Since the authors observed that a portion of the retrotransposition events were recombination dependent and IEP endonuclease independent, they proposed a model in which the intron reverse splices into mRNA containing the target site, followed by reverse transcription and homologous recombination between the resulting cDNA and the intronless allele of the target gene. A more recent paper by Dickson *et al.* on the retrotransposition mechanism of the yeast aI2 intron, to which the Ll.ltrB intron is quite similar, has thrown this model into question (12). First, citing a personal communication, the authors mention that the RecA dependence of Ll.ltrB retrotransposition is in question. Second, they present evidence that suggests DNA is the major target for transposition of aI2. They measured ectopic transposition into a site in either of two contexts: a) present in the wild type sense orientation or b) inverted so as to be in the antisense orientation. Their reasoning was that if RNA is the target of retrotransposition, then the inverted site should not support retrotransposition because no transcript contains the correct sequence. They found that retrotransposition occurred when the target site was in either orientation, supporting DNA as the target molecule for transposition. Taken together, the data from these two studies suggests the possibility that both mechanisms might occur in retrotransposition. Further study will be required to define the predominant pathway for Ll.ltrB transposition in bacterial hosts.

### 3.4. RNP particles

RNP particles comprised of intron RNA and LtrA protein carry out the Ll.ltrB splicing and mobility reactions. Several studies have provided insight into the nature of this RNP particle. A study by Saldanha *et al.* looked at Ll.ltrB intron splicing and mobility catalyzed by purified RNA and LtrA components (13). They found that LtrA protein stability was affected by substrate binding and in the absence of intron RNA, the protein was degraded. LtrA was a specific splicing factor, binding only to Ll.ltrB intron

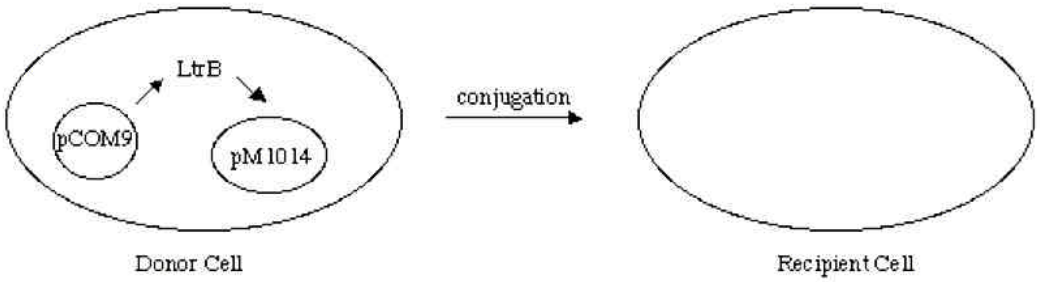


**Figure 3.** Conjugation-based assay for splicing. Splicing of Ll.ltrB from pCOM9 results in LtrB expression, which acts in trans to nick the oriT of pM1014, allowing for plasmid mobilization.

but not other group II intron RNA. In addition, LtrA functioned as an endonuclease only when the RNP particle was composed of intron lariat RNA; unspliced RNA was not suitable. They also determined that LtrA bound to the intron in a stoichiometry of 2:1, possibly binding as a dimer. Wank and colleagues localized the LtrA-binding site in the Ll.ltrB intron (14). LtrA was found to bind to a portion of domain IV that included the LtrA ribosome binding site. Furthermore, deletions in domain IV that prevented LtrA binding, prevented protein dependent splicing in assays performed *in vitro*. These deletions did not affect the catalytic activity of the intron, since splicing was observed in all mutants.

### 3.5. Conjugation based genetic assay to analyze Ll.ltrB splicing

In addition to the various biochemical assays that have been used to analyze the structure and function of the Ll.ltrB intron, a genetic system has recently been developed to study Ll.ltrB splicing. To understand this system, it is necessary to give some background on conjugation in *Lactococcus*. The initial experiments that led to the discovery of Ll.ltrB were designed to identify the genes on pRS01 involved in conjugation. The ability to transfer genetic material by conjugation is widespread in Lactococcus. Conjugative plasmid transfer requires direct contact between a donor cell containing the plasmid and a recipient cell that lacks the plasmid. The plasmid is subsequently nicked in a region called the origin of transfer by a complex called the relaxosome and a single stranded copy of the plasmid is transferred to the recipient cell. Second strand synthesis then occurs resulting in a complete copy of the plasmid in both the donor and recipient cells. To identify the genes required for conjugation of pRS01 in *L. lactis*, Mills and colleagues generated a series of insertion mutations by integration of a second plasmid, pTRK28, into pRS01 (15). Mutant plasmids were then screened for a conjugation defect and the location of insertions was identified in mating deficient cointegrants. In this way, several regions of pRS01 were identified as being required for conjugation. Sequencing one of these regions revealed several open reading frames, one of which, *ltrB*, had significant homology to relaxase enzymes in other conjugation systems (2). Relaxase enzymes are part of the relaxosome complex involved in plasmid nicking, and therefore are required for the conjugation process. In the case of pRS01, the relaxase *ltrB* was disrupted by the Ll.ltrB intron. As discussed above (section 3.2), further investigation revealed that the Ll.ltrB intron was readily spliced out, allowing production of LtrB protein and subsequent conjugation. Of the many insertions into pRS01 that resulted in conjugation defects, several were within or near *ltrA*, and presumably prevented *ltrA* expression or disrupted the conserved RNA secondary structure required for splicing. Furthermore, it was found that the conjugation defects of these mutations could be complemented in trans by a smaller plasmid carrying the intact intron region. Presumably, the ability of the cell to undergo conjugation was dependent on expression of LtrB, which in turn was dependent upon intron splicing. Thus it was hypothesized that a conjugation-based assay could be used to analyze group II intron splicing.

To test whether a conjugation assay could be used to study Ll.ltrB, a two-plasmid system was developed (figure 3)(16). Plasmid pM1014 is a cointegrant of plasmid pRS01 and pTRK28. Integration of pTRK28 occurred in the 5' portion of LtrA, disrupting its expression so that splicing did not occur and consequently LtrB was not produced. Therefore, pM1014 had very low frequency of transfer ($<10^{-9}$ transconjugants per donor). However, pM1014 had all of the other genes required for conjugation, and it was possible to complement the conjugation defect with a second plasmid, pCOM9. The shuttle vector pCOM9 has the entire Ll.ltrB intron and flanking *ltrBE1* and *ltrBE2* exon sequences. When pM1014 and pCOM9 were present in the same cell, the conjugation frequency increased to $4.9 \times 10^{-3}$, indicating that excision of Ll.ltrB in pCOM9 allowed production of LtrB that could function in trans to nick pM1014 and lead to plasmid transfer. In fact, a pCOM9 derivative that lacked the intron and simply had the two *ltrB* exons fused together further increased conjugation frequency to $1.3 \times 10^{-2}$. These data confirmed that expression of LtrB was required for conjugation, and not intron sequences or LtrA. Further experiments demonstrated that conjugation efficiency correlates to intron splicing *in vitro*.

The two-plasmid system comprised of pM1014 and pCOM9 has several advantages. First, splicing is measured not by a biochemical assay, but by a simple screen relying on the genetics of the organism. The

conjugation assay allows many intron constructs to be analyzed in a short period of time, and is a quantitative and extremely sensitive measure of splicing. Second, since the pCOM9 plasmid is comparatively small (13 kb vs. 50 kb for pRS01) it is relatively easy to manipulate the intron sequence in this context. Mutations can be easily introduced into the intron and the effect on splicing function can be determined by assaying for plasmid transfer. For example, residues critical for eukaryotic group II intron function have been mutated in pCOM9 and found to affect conjugation (personal observation). It is also possible to introduce mutations in the intron by random processes (personal observation). Taken together, these features of the two-plasmid system suggest that it is a convenient system to randomly generate mutations in the intron and subsequently screen for affects on splicing. Therefore, it is possible to take a genetic approach towards understanding group II intron function. Experiments up until this time have traditionally looked at the effect of mutating well conserved sequences or structures on intron function. By adding a random mutagenesis approach to the study of group II introns, it is likely that entirely unexpected residues and interactions will be found to be involved in the splicing reaction.

### 3.6. Regulation of Ll.ltrB

Although quite a bit is known about the mechanism of splicing and mobility of group II introns, relatively little is known regarding the regulation of the expression of intron function. Furthermore, the effect of introns on host gene expression is unclear. For Ll.ltrB, these questions become especially interesting considering its residence on the conjugative plasmid pRS01. Since the intron disrupts a gene that is required for mobility of the plasmid, intron splicing is intimately linked to the mobility of the plasmid and one must question to what extent intron function is related to pRS01 mobility.

A simple model for group II intron splicing would predict that a single transcript is made that includes two exons and an intron. The intron-encoded protein is translated, binds to the unspliced RNA forming an RNP particle and initiates the splicing reaction. The intron is subsequently spliced from the transcript, joining the two exons. However, recent results suggest that for the Ll.ltrB intron the picture is much more complex, involving multiple transcripts. This is due to a surprising result in which a promoter was identified within the Ll.ltrB intron, immediately upstream of the *ltrA* coding region (16) (figure 1a). This is the first report of a promoter localized within a bacterial group II intron sequence. This promoter is responsible for the majority of the *ltrA* mRNA and LtrA protein made in the cell since a mutation in the –35 or –10 region of the promoter resulted in about a 90% reduction of the *ltrA* containing message and nearly abolished LtrA protein production. Furthermore, the *ltrA* promoter is required for efficient *in vivo* splicing. New data suggests that an additional promoter likely exists upstream of *ltrBE1*. (Klein and Dunny, unpublished results). The length of this transcript is unknown, but is predicted to extend through *ltrBE2*. If this is the case, it is in relatively minor abundance, accounting for 10% of the *ltrA* containing transcripts. Based on these findings, there are possibly two *ltrA* containing transcripts, a longer one initiating upstream of *ltrBE1* and extending through *ltrBE2*, and a shorter more abundant transcript initiating upstream of ltrA. LtrA could presumably be translated from either of these messages.

A possible explanation for the strong *ltrA* promoter comes from studies outlined in section 3.4 demonstrating the nature of the LtrA containing RNP particle. LtrA binds to intron RNA in a stoichiometry of 2:1 to form RNP particles. LtrA appears to bind intron RNA in a region that possibly overlaps with the LtrA ribosome binding site. This binding would putatively repress further translation of LtrA, resulting in LtrA autoregulating its own synthesis. Therefore, it can be imagined that a second abundant transcript would be necessary to produce enough LtrA to result in productive splicing. Clearly, more examination of the transcription and translation of the Ll.ltrB region needs to be done before an accurate picture of intron regulation is in place.

### 3.7. Distribution of Ll.ltrB in *L. lactis*

When Ll.ltrB was first discovered, it was found in two different contexts, in the chromosomally located sex factor of *L. lactis* 712 and in the conjugative plasmid pRS01 of *L. lactis* ML3 (1, 2). In the sex factor, the intron disrupts the *mobA* gene, a conjugation associated mobilization protein. In pRSO1, the intron disrupts the *ltrB* gene, a conjugative relaxase. These two proteins are 99% identical. Recently, an intron nearly identical to the sex factor and pRS01 introns was found in the *Lactococcal* plasmid pAH90 (17). This intron is functional for splicing *in vivo*. The intron disrupts the putative relaxase *mobD*, whose protein sequence is 33% and 29% identical to the MobA and LtrB proteins respectively. Although the *mobD* sequence differs from the other host genes, the DNA target site is quite conserved, having two changes in the IBS pairing region and conserved G-21 and T+5 residues. The presence of the intron in these different contexts suggests that retrohoming and retrotransposition occur in nature.

## 4. DISTRIBUTION OF GROUP II INTRONS IN BACTERIA

Ll.ltrB is the most well studied bacterial group II intron, however, it seems to be an atypical example. When Ll.ltrB was first identified, it was recognized to have greater similarity to the group II intron-encoded ORFs of fungal mitochondrial introns than to the IEPs of the few bacterial group II introns that had been identified at that time (2). Recent phylogenetic analysis comparing the reverse transcriptase and maturase domains of the IEPs of both bacterial and eukaryotic group II introns placed Ll.ltrB on a branch with mitochondrial group II introns, while other bacterial introns group elsewhere on the tree (18, 19). A distinct difference among bacterial IEPs, is the presence or absence of the Zn domain. While the Zn domain is required for Ll.ltrB homing, only about half of the bacterial group II introns have this domain. Together, these observations indicate that the studies on Ll.ltrB that have provided insight into the mechanisms of intron function,

may not apply to all bacterial group II introns. Therefore, it is important to investigate the structure and function of group II introns in other bacterial species. A discussion of the work done on other bacterial group II introns is provided in the following sections.

Group II introns were first identified in bacteria by Ferat and Michel in 1993 (20). Up to this time, group II introns had been identified exclusively in mitochondria and chloroplast genomes. Since the proteobacteria and cyanobacteria are progenitors of mitochondria and chloroplast, the authors sought to determine whether group II introns were present in these bacteria as well. PCR using primers specific to the highly conserved intron domain V and the reverse transcriptase domain of the intron-encoded ORF, led to the identification of group II introns in the cyanobacteria *Calothrix* (Cal.x1) and the proteobacteria *Azotobacter vinelandii* (AvUWR.x1). Further experiments showed that Cal.x1 was a function intron and underwent splicing *in vitro*. Since that time, numerous introns in both gram-negative and gram-positive bacterial species have been identified.

### 4.1 Group II introns in gram-negative bacterial species

#### 4.1.1. *E. coli*

Following the discovery of group II introns in cyanobacteria and proteobacteria, Ferat *et al.* expanded the search for group II introns into *E. coli* (21). Using the ECOR collection of *E. coli* strains and a similar PCR methodology as above, four introns were identified. These introns, designated *IntA,B,C* and *D* were all found to be inserted in host genes that were known or putative mobile genetic elements. *IntA* and *IntB* were both inserted into different sites of a putative insertion sequence-like region called an H-repeat. *IntC* interrupted an ORF related to polypeptides carried by the mobile T-DNA of *Agrobacterium tumefaciens.* *IntD* was found in an insertion sequence similar to IS*3411*. Also in 1994, Knoop and Brennicke reported the discovery of an *E. coli* intron located in a host gene designated orf104, the same gene that *IntC* is inserted in (22). This intron was identified by searching the database with an intron domain V consensus sequence (23). Recently, a group II intron has been identified on a 69 kb plasmid in an enteropathogenic *E. coli* strain (24). The intron, called EPEC.IntA is located within IS*911* on the plasmid. It is 99.7% similar to an intron found in *Shigella flexneri*. The ability of these introns to undergo splicing *in vivo* has not been reported, so it is unclear whether they are functional.

#### 4.1.2. *Sinorhizobium meliloti*

Next to Ll.ltrB, the *Sinorhizobium meliloti* RmInt1 is the most well characterized bacterial group II intron. The first report of this intron was in 1998 by Martinez-Abarca and colleagues (25). In the process of sequencing plasmid pRmeGR4b, this group discovered a novel insertion sequence called IS*Rm2011-2*. This insertion sequence itself was interrupted by an 1884 bp sequence that was 63% identical to the group II intron IntB from *E. coli*. The intron-encoded IEP has all 7 RT domains and the X domain, but it lacks a Zn domain. The surrounding RNA can fold into the secondary structure typical of group II introns. When cloned under control of the T7 promoter and expressed in *E. coli*, the intron underwent splicing as determined by RT-PCR analysis, demonstrating that it is functional *in vivo*. Furthermore, when the IEP was mutated, reduced splicing was observed, indicating that the IEP has maturase function.

Work with other group II introns has indicated that mobility by the process of homing requires the IEP endonuclease activity to cleave the antisense strand of the target DNA (4). Since the RmInt1 IEP has no endonuclease domain, experiments were designed to examine the mobility of the intron (26). An intronless allele of the host gene was cloned into a medium copy number, broad host range vector and moved by conjugation from *E. coli* into two strains of *S. meliloti*: GR4, which has 9 copies of the intron, and RMS16, which does not carry the intron. The plasmid pool was examined and a larger plasmid characteristic of a homing product was seen in GR4 but not RMS16. This suggested that the intron homed from the GR4 chromosome into the target site in the plasmid. The homing product was not observed if the insertion target site was deleted from the plasmid. In addition, homing was observed in a *recA* mutant strain indicating recombination mechanisms were not involved, similar to results seen with the lactococcal Ll.ltrB intron. As described above, work done with the Ll.ltrB intron indicated that homing of group II introns occurs by a process of complete reverse splicing into the target DNA site followed by cleavage of the antisense strand by the endonuclease domain of the IEP. Since the RmInt1 IEP does not have the complete endonuclease domain, it is unclear how homing occurs. Perhaps an additional host protein with endonuclease activity is required. Alternatively, a different mechanism of homing that doesn't require the endonuclease activity may be used.

The distribution of the RmInt1 intron is widespread in natural isolates, and it is often abundant in individual strains (25). For example, in examining the distribution of RmInt1 in 85 field strains, only 2 strains were found to lack RmInt1 and the strains had between 1 and 11 copies of the intron. The strain GR4, in which RmInt1 was identified, has nine copies of the intron. While most copies of RmInt1 are inserted into the ISRm2001-2 elements, there are several examples of the intron being inserted into similar IS elements (IS*Rm10-1*, IS*Rm10-2)* or a divergent site, a putative oxide reductase gene (*oxi1)*.

The finding of RmInt1 in divergent sites, indicates that ectopic transposition may have occurred in nature. A subsequent study found that within GR4-type isolates, 4.1% had an *oxi1* locus disrupted by the RmInt1 intron (27). This provides evidence that ectopic transposition does occur at an appreciable level in nature and could lead to the spread of the intron in bacterial populations. When tested in the lab, RmInt1 was able to transpose into these ectopic target sites *in vivo*, albeit at a reduced frequency from homing into the wild type target site (28). The additional insertion sites of RmInt1 all contain sequences immediately upstream of the insertion site that are nearly identical to the IBS1, IBS2 and delta-

prime sequences found in the wild type IS*Rm2011-2*. As discussed with the Ll.ltrB intron, these sequences are involved in base pairing interactions with the intron RNA in target site selection. When present at these divergent sites, RmInt1 is able to forward splice, further supporting a productive interaction between EBS and IBS sequences.

To date, Ll.ltrB and RmInt1 are the only intons that have been experimentally demonstrated to be mobile *in vitro*. The ability of both introns to transpose to novel ectopic sites suggests a mechanism by which introns can spread within the genome and throughout the bacterial population. However, several differences in the structure and function of the group II introns Ll.ltrB and RmInt1 have been observed. This fact underscores the need to study multiple bacterial group II introns in detail in order to fully understand their function, structure and evolution.

#### 4.1.3. *Pseudomonas sp.*

In 1997, a report by Yeo *et al.* described a group II intron from *Pseudomonas alcaligenes* P25X (29). The group II intron, called Xln6, was recognized based on homology of a newly identified ORF to maturase proteins of eukaryotic group II introns. Intron Xln6 has the 7 characteristic RT domains and a maturase domain X, but it lacks the Zn endonuclease domain. Xln6 exhibits several additional features of group II introns including EBS1 and IBS1 sequences, conserved splice site consensus sequences, and an appropriately placed bulging A residue. The intron RNA is also able to fold into the typical secondary structure of group II introns. Xln6 was found on the conjugative plasmid RP4, but did not disrupt an ORF. The plasmid RP4 had acquired the Xln6 intron from the chromosome in an entrapment experiment done to identify mobile genetic elements, therefore, the ability of the Xln6 intron to insert into plasmid RP4 was an indication of its implied mobility, either its own homing or transposition activity or its presence on a mobile genetic element.

As a follow up to this study, Yeo *et al.* identified three additional group II introns in other *Pseudomonas* species (30). The Xln3 intron from *P. alcaligenes* P25X was identified by entrapment on plasmid pRA2 and is 87% identical to Xln6. An intron identified in plasmid pRA500 of *P. putida* strain P35X is 92% identical to Xln6. A group II intron 95% identical to the Xln6 intron was found in *P. putida* strain KT2440. Like the Xln6 intron, these three introns also lack a Zn domain. These introns do not interrupt any ORFs and were all located downstream of hairpin loop structures that resemble rho-independent transcriptional terminators.

#### 4.1.4. *Shigella flexneri*

The *she* pathogenicity island of *Shigella flexneri* is a 25 kb region harboring several virulence genes as well as a putative group II intron called Sf.IntA (31). This intron is capable of folding into the conserved secondary structure characteristic of group II introns and the IEP (SfiA) has conserved RT and X domains, but lacks the zinc finger domain (19). Intron Sf.IntA disrupts an IS*629*-like ORF on the pathogenicity island, another example of a group II intron being associated with a mobile genetic element.

#### 4.1.5. *Shingomonas aromaticivorans*

In the process of sequencing the pNL1 conjugative plasmid from *Shingomonas aromaticivorans*, two ORFs with homology to maturase proteins were identified, MatRa and MatRb (32). Although not discussed in the original publication, recent compilations of data on group II introns indicate that these ORFs have RT, X and Zn domains as well as a conserved group II intron secondary structure (18, 19). MatRa is flanked by ORFs with homology to the TraC replication primase, a protein required for replication of the F plasmid in *E. coli*. Conjugative plasmid transfer of the F plasmid involves transfer of a single strand of the F plasmid into a recipient cell followed by second strand synthesis; therefore the replication protein TraC is important for conjugative plasmid transfer. From these sketchy details, it can be speculated that this is an example of a group II intron disrupting a gene required for plasmid conjugation in *S. aromaticivorans*, much like the Ll.ltrB intron of *L. lactis*.

### 4.2 Group II introns in gram-positive bacterial species

In addition to the Ll.ltrB intron of *Lactococcus lactis*, which has already been discussed in great detail, a number of group II introns have been identified in gram-positive bacteria.

#### 4.2.1. *Clostridium difficile*

In 1996, Mullany reported the identification of a group II intron in the Tn*5397* conjugative transposon of *Clostridium difficile* (33). The intron ORF is homologous to group II intron-encoded proteins and has RT, X and Zn domains. The secondary structure of the surrounding RNA conforms to that of other group II introns. A subsequent study found that the intron was spliced *in vivo* in *C. difficile* as well as when expressed in *B. subtilis* (34). A striking feature of this intron and the transposon it disrupts is the higher G+C content compared to the rest of the *C. difficile* chromosome. This suggests that the transposon originated in another bacteria, and transferred into *C. difficile* rather recently in evolutionary history.

#### 4.2.2. *Streptococcus sp.*

Group II introns have been identified in two species of Streptococcus. *Streptococcus pneumoniae* encodes a protein homologous to the maturase protein of group II introns (35). This protein has RT and X domains but lacks a Zn domain (18, 19). The surrounding region can fold into the characteristic group II intron secondary structure. The intron is found in an intergenic region in the capsular polysaccharide biosynthetic locus. A putative transposase protein is located near by, consistent with the trend of group II introns to associate with mobile genetic elements.

Several isolates of Group B Strep (*S. agalactiae*) were sequenced in the region of a previously identified insertion sequence element (36). In one highly virulent isolate, a group II intron 90% identical to the *S. pneumoniae* intron was identified and designated GBSi1. GBSi1 has an RT domain and a maturase domain X, but lacks the Zn domain associated with endonuclease activity. The secondary structure of this intron reveals the characteristic 6 helical domains, however domain V is

slightly divergent, being only 30 nt in length (compared to 34 nt in other introns) and having one nucleotide change in a highly conserved location. The intron does not disrupt an ORF and the EBS2 and IBS2 binding pairs are missing. The intron insertion site is preceded by an inverted repeat followed by a poly(T) tail, which resembles a rho-independent transcriptional terminator. It was suggested that this sequence may add specificity to intron-target site recognition and compensate for the lack of EBS2 and IBS2. Despite the differences in domain V and EBS/IBS pairing, an *in vitro* self-splicing assay indicated that this intron is capable of splicing. Splicing occurred via a hydrolysis reaction and no lariat intermediate was detected, however, this could be an artifact of the assay or the fact that the IEP was not included in the assay. In contrast to other bacterial group II introns, GBSi1 is not associated with a mobile genetic element.

#### 4.2.3. *Bacillus sp.*

Several group II introns have been identified in various species of *Bacillus*. A putative intron was identified in *Bacillus megaterium* in the process of analyzing a mercury resistance operon (37). This 2717 bp region encodes an ORF with reverse transcriptase, maturase and endonuclease domains. It has consensus splice sites, a bulging A residue and IBS/EBS binding pairs. The intron is found in a class II transposon, but does not disrupt an exon gene. Recently, several copies of a group II intron 47.6 % identical to the *C. difficle* group II intron have been found on the chromosome of *Bacillus halodurans* strain C-125 (38). The intron is called Bh.Int, but little information about its structure is available. Two ORFs with homology to reverse transcriptases of group II introns are present on the large pXO1 plasmid of *Bacillus anthracis* (39). ORF07 has RT, X and Zn domains, while ORF23 has only RT and X domains (19). Both ORFs are located within sequences that are capable of folding into a conserved intron structure (19). In neither of these instances is it mentioned whether an ORF is disrupted.

### 4.3. Common themes

Of the bacterial group II introns described above, only a handful have been studied in much detail. Five of the introns have been shown to undergo splicing *in vivo* or *in vitro* and include Ll.ltrB, RmInt1, GBSiI, Cal.xI and the *C. difficle* intron. Only Ll.ltrB and RmInt1 have been demonstrated to be mobile. Despite the lack of functional data, sequence analysis of the introns and flanking regions has provided additional insight into the introns. First, nearly all of the introns are located on mobile genetic elements including insertion sequences, conjugative plasmids and transposons and sex factors. Second, some of the introns described above do not disrupt host ORFs, but are found downstream of a rho-independent terminator sequence. This stem-loop secondary structure may aid in target site selection in intron mobility events, especially if EBS/IBS pairings are weak or non-existent. Insertion downstream of terminators also suggests that there may be some selection against disrupting host ORFs. Third, all of the bacterial group II introns so far identified encode a protein. In fact, the presence of proteins with reverse transcriptase and maturase domains is precisely what allowed these sequences to be noticed and classified as group II introns. Since it is difficult to identify group II introns in genome databases based on their conserved intron RNA structure, it is unknown whether there are group II introns in bacteria that do not encode IEPs. In addition, some bacterial IEPs have a Zn domain associated with endonuclease activity, while others lack this domain. Since endonuclease activity is necessary for homing in fungal and lactococcal systems, introns without this domain may employ different mechanisms for mobility. Fourth, slight differences in the highly conserved domain V of the intron RNA secondary structure are observed among introns (18). The effect of these differences on splicing or mobility is unknown.

## 5. EVOLUTIONARY PERSPECTIVE ON GROUP II INTRON DISTRIBUTION

As well as being found in an increasing number of bacterial species, group II introns are abundant in eukaryotes. Roughly 120 unique group II introns are found in the chloroplasts and mitochondria of plants, fungi and protists (40). An obvious variation among these introns is in the IEPs. About half of lower eukaryotes have an IEP, while IEPs are almost completely absent from group II introns in higher plants. There has been much speculation about the origin of group II introns and the evolutionary steps that lead to the present day variation of group II intron structure. Recent phylogenetic analyses comparing the IEP of group II introns from many organisms support theories in which group II introns originated in bacteria. For example, comparing domain X and the reverse transcriptase domains of 71 group II intron ORFs, Zimmerly *et al.* proposed a phylogenetic tree that was divided into two clusters designated the mitochondrial and chloroplast-like lineages (19). Bacterial ORFs were positioned at the base of each branch, supporting bacteria as a common ancestor. Also comparing the conserved RT domains, Martinez and colleagues determined the phylogenitic relationships of bacterial group II intron-encoded proteins compared to organellar counterparts (18). They found that bacteria grouped into three branches. Of note is the fact that one branch exclusively contains members that both lack an endonuclease domain in the IEP and have a distinct intron RNA domain V. This suggests that there was coevolution of intron RNA structure and IEPs in these species. In other words, rather than being the result of invasion of a ribozyme by a mobile retroelement, the core RNA and intron ORFs of group II introns likely evolved together as part of the same molecule. A study by Toor and colleagues further analyzed coevolution of group II intron RNA structure and IEP reverse transcriptase domains (40). From their analysis, they proposed the "retroelement ancestor hypothesis" for the basis of intron evolution. In this model of coevolution, mobile group II introns arose in bacteria with both IEP and RNA structural features present. The RNA structure further evolved into mitochondrial and chloroplast-like lineages, eventually resulting in ORF loss in these lineages. So there was a progressive ORF loss from bacteria to plants, but they all evolved from retroelements similar to the present day bacterial group II introns.

Aside from the ancestral relationship between bacterial and eukaryotic group II introns, bacterial group II introns are related to two other eukaryotic genetic elements. First, phylogenetic analysis places the reverse transcriptase of group II intron-encoded proteins in the same family as non-LTR retrotransposable elements (41). Second, group II introns are similar to nuclear spliceosomal introns in the mechanism of the RNA-splicing reaction. In fact, it is often suggested that group II introns are the progenitors of spliceosomal introns. Perhaps once group II introns were present in the organelles of eukaryotes, they then spread to the nuclear genome where they evolved into the present day spliceosomal intron. Several differences between group II introns and spliceosomal introns are apparent. Group II introns are self-splicing and rely on a highly conserved and catalytic intron RNA secondary structure, and often times on an IEP, whereas nuclear introns require a complex of RNAs and protein called the spliceosome. Also, spliceosomal introns are non-mobile while group II introns are capable of homing or transposition.

In addition to vertical inheritance of group II introns as one species evolved from another, horizontal transfer of bacterial group II introns between species has been suggested to occur. This is supported by two observations. First, the G+C content of group II introns is often different from the average G+C content of the rest of the genome (18). This suggests that group II introns were acquired relatively recently in evolutionary history. Second, group II introns are often located on mobile genetic elements, providing a vehicle for horizontal spread among species.

The processes of retrohoming and retrotransposition are additional mechanisms that can lead to the spread of group II introns. In homing, allelic loci acquire an intron with high efficiency. This probably led to the presence of the Ll.ltrB intron in more than one location in *lactococcal* strains and to the proliferation of the RmInt1 intron in *S. meliloti*. Retrotransposition provides a mechanism whereby group II introns can spread throughout the chromosome by inserting into ectopic sites. Both the Ll.ltrB and RmInt1 introns appear to have moved by this process as well. Based on these examples, an obvious question arises as to why group II introns are not more abundant in bacteria if they can retrotranspose.

## 6. SUMMARY

Group II introns are found in a wide variety of bacteria and it is likely that more examples will be identified as bacterial genomes continue to be sequenced and annotated. Two processes have likely influenced the distribution of bacterial group II introns. First, the inherent mobility of group II introns may have allowed for proliferation within individual isolates. Second, their frequent localization on mobile genetic elements likely provided a means for horizontal gene transfer between species. Much work has been done on the lactococcal Ll.ltrB intron and it is evident that the mechanisms of intron splicing and mobility are quite similar to those of fungal group II introns. However, differences in the structure and context of bacterial group II introns emphasize the need to continue studying intron splicing and mobility mechanisms of a variety of bacterial introns.

## 7. REFERENCES

1.Shearman, C., J. J. Godon and M. Gasson: Splicing of a group II intron in a functional transfer gene of Lactococcus lactis. *Mol Microbiol* 21, 45-53 (1996)

2.Mills, D. A., L. L. McKay and G. M. Dunny: Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol* 178, 3531-3538 (1996)

3.Qin, P. Z. and A. M. Pyle: The architectural organization and mechanistic function of group II intron structural elements. [Review] [58 refs]. *Curr Opin Struct Biol* 8, 301-308 (1998)

4.Matsuura, M., R. Saldanha, H. Ma, H. Wank, J. Yang, G. Mohr, S. Cavanagh, G. M. Dunny, M. Belfort and A. M. Lambowitz: 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 (1997)

5.Mills, D. A., D. A. Manias, L. L. McKay and G. M. Dunny: Homing of a group II intron from Lactococcus lactis subsp. lactis ML3. *J Bacteriol* 179, 6107-6111 (1997)

6.Lambowitz, A. M., M. G. Caprara, S. Zimmerly and P. S. Perlman: Group I and group II ribozymes as RNPs: clues to the past and guides to the future. In: The RNA world. Eds: Atkins, J. F., Cold Spring Harbor Laboratory Press, New York, 451-485 (1999)

7.Zimmerly, S., H. Guo, R. Eskes, J. Yang, P. S. Perlman and A. M. Lambowitz: A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. *Cell* 83, 529-538 (1995)

8.Mohr, G., D. Smith, M. Belfort and A. M. Lambowitz: Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev* 14, 559-573 (2000)

9.Guo, H., M. Karberg, M. Long, J. R. Jones, B. Sullenger and A. M. Lambowitz: Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science* 289, 452-457 (2000)

10.Singh, N. N. and A. M. Lambowitz: 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 (2001)

11.Cousineau, B., S. Lawrence, D. Smith and M. Belfort: Retrotransposition of a bacterial group II intron. *Nature* 404, 1018-1021 (2000)

12.Dickson, L., H. Huang, L. Liu, M. Matsuura, A. M. Lambowitz and P. S. Perlman: Retrotransposiiton of a yeast

group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc. Natl. Acad. Sci. USA* 98, 13207-13212 (2001)

13.Saldanha, R., B. Chen, H. Wank, M. Matsuura, J. Edwards and A. M. Lambowitz: RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry* 38, 9069-9083 (1999)

14.Wank, H., J. SanFilippo, R. N. Singh, M. Matsuura and A. M. Lambowitz: A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell* 4, 239-250 (1999)

15.Mills, D. A., C. K. Choi, G. M. Dunny and L. L. McKay: Genetic analysis of regions of the Lactococcus lactis subsp. lactis plasmid pRS01 involved in conjugative transfer. *Appl Environ Microbiol* 60, 4413-4420 (1994)

16.Zhou, L., D. A. Manias and G. M. Dunny: Regulation of intron function: efficient splicing *in vivo* of a bacterial group II intron requires a functional promoter within the intron. *Mol Microbiol* 37, 639-651 (2000)

17.O'Sullivan, D., R. P. Ross, D. P. Twomey, G. F. Fitzgerald, C. Hill and A. Coffey: Naturally occurring lactococcal plasmid pAH90 links bacteriophage resistance and mobility functions to a food-grade selectable marker. *Appl Environ Microbiol* 67, 929-937 (2001)

18.Martinez, A. F. and N. Toro: Group II introns in the bacterial world. *Mol Microbiol* 38, 917-926 (2000)

19.Zimmerly, S., G. Hausner and X. c. Wu: Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Res* 29, 1238-1250 (2001)

20.Ferat, J. L. and F. Michel: Group II self-splicing introns in bacteria. *Nature* 364, 358-361 (1993)

21.Ferat, J. L., G. M. Le and F. Michel: Multiple group II self-splicing introns in mobile DNA from Escherichia coli. *C. R. Acad. Sci., Ser. III* 317, 141-148 (1994)

22.Knoop, V. and A. Brennicke: Evidence for a group II intron in Escherichia coli inserted into a highly conserved reading frame associated with mobile DNA sequences. *Nucleic Acids Res* 22, 1167-1171 (1994)

23.Knoop, V., S. Kloska and A. Brennicke: On the identification of group II introns in nucleotide sequence data. *J Mol Biol* 242, 389-396 (1994)

24.Tobe, T., T. Hayashi, C. Han, G. K. Schoolnik, E. Ohtsubo and C. Sasakawa: Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. *Infect Immun* 67, 5455-5462 (1999)

25.Martinez, A. F., S. Zekri and N. Toro: 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 (1998)

26.Martinez, A. F., R. F. Garcia and N. Toro: Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Mol Microbiol* 35, 1405-1412 (2000)

27.Munoz, E., P. J. Villadas and N. Toro: Ectopic transposition of a group II intron in natural bacterial populations. *Mol Microbiol* 41, 645-652 (2001)

28.Martinez, A. F. and N. Toro: RecA-independent ectopic transposition *in vivo* of a bacterial group II intron. *Nucleic Acids Res* 28, 4397-4402 (2000)

29.Yeo, C. C., J. M. Tham, M. W. Yap and C. L. Poh: Group II intron from Pseudomonas alcaligenes NCIB 9867 (P25X): entrapment in plasmid RP4 and sequence analysis. *Microbiology* (1997)

30.Yeo, C. C., S. Yiin, B. H. Tan and C. L. Poh: Isolation and characterization of group II introns from Pheudomonas alcaligenes and Pseudomonas putida. *Plasmid* 45, 233-239 (2001)

31.Rajakumar, K., C. Sasakawa and B. Adler: Use of a novel approach, termed island probing, identifies the Shigella flexneri she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect Immun* 65, 4606-4614 (1997)

32.Romine, M. F., L. C. Stillwell, K. K. Wong, S. J. Thurston, E. C. Sisk, C. Sensen, T. Gaasterland, J. K. Fredrickson and J. D. Saffer: Complete sequence of a 184-kilobase catabolic plasmid from Sphingomonas aromaticivorans F199. *J Bacteriol* 181, 1585-1602 (1999)

33.Mullany, P., M. Pallen, M. Wilks, J. R. Stephen and S. Tabaqchali: A group II intron in a conjugative transposon from the gram-positive bacterium, Clostridium difficile. *Gene* 174, 145-150 (1996)

34.Roberts, A. P., V. Braun, E. von, C. Streiber and P. Mullany: Demonstration that the group II intron from the Clostridial Conjugative transposon Tn5397 undergoes splicing *In vivo*. *J Bacteriol* 183, 1296-1299 (2001)

35.Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, J. C. Paton and B. G. Spratt: Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae. *Mol Microbiol* 27, 73-83 (1998)

36.Granlund, M., F. Michel and M. Norgren: 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 (2001)

37.Huang, C. C., M. Narita, T. Yamagata, Y. Itoh and G. Endo: Structure analysis of a class II transposon encoding

the mercury resistance of the Gram-positive Bacterium bacillus megaterium MB1, a strain isolated from minamata bay, Japan. *Gene* 234, 361-369 (1999)

38.Takami, H., C. G. Han, Y. Takaki and E. Ohtsubo: Identification and distribution of new insertion sequences in the genome of alkaliphilic Bacillus halodurans C-125. *J Bacteriol* 183, 4345-4356 (2001)

39.Okinaka, R. T., K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, G. Lamke, S. Kumano, J. Mahillon, D. Manter, Y. Martinez, D. Ricke, R. Svensson and P. J. Jackson: Sequence and organization of pXO1, the large Bacillus anthracis plasmid harboring the anthrax toxin genes. *J Bacteriol* 181, 6509-6515 (1999)

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

41.Curcio, M. J. and M. Belfort: Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* 84, 9-12 (1996)

**Abbreviations:** IEP, intron encoded protein; RT, reverse transcriptase; X, maturase; Zn, zinc finger-like domain; EBS, exon binding site; IBS, intron binding site; RNP, ribonucleoprotein

**Key Words:** Microbiology, Bacteria, Group II Intron, Splicing, *Lactococcus lactis*, Review

**Send correspondence to:** Dr. Gary Dunny, Department of Microbiology, University of Minnesota Medical School, 1460 Mayo Bldg, 420 Delaware St. SE, Minneapolis, MN, 55455. Tel: 612-625-9930, Fax: 612-626-0623, E-mail: gary-d@biosci.cbs.umn.edu