The Dissertation Committee for Travis Brandon White Certifies that this is the approved version of the following dissertation:

Group II Intron Retrohoming and Gene Targeting Reactions in *Drosophila melanogaster*

Committee:

Alan M. Lambowitz, Supervisor

Jim Bull

Paul Macdonald

Tanya Paull

Scott Stevens

Group II Intron Retrohoming and Gene Targeting Reactions in *Drosophila melanogaster*

by

Travis Brandon White, B.S.Bio.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2011

Dedication

To my partner and my parents for all of their love and encouragement.

Acknowledgements

I owe many thanks to the people who have made this dissertation possible. Foremost, I would like to express my gratitude to my adviser, Dr. Alan M. Lambowitz, for the opportunity to perform research in his laboratory. His ability to incorporate several areas of study into a compelling scientific narrative has been a great influence.

I would also like to thank my graduate committee, Dr. James J. Bull, Dr. Paul M. Macdonald, Dr. Tanya T. Paull and Dr. Scott W. Stevens, for their guidance and thoughtful input on my research.

My thanks go to many members of the Lambowitz Lab who have been integral in my education of scientific methods and growth as a scientist.

I would like to thank Dr. Paul M. Macdonald and the Macdonald Lab for providing resources and expertise for many of my *Drosophila* experiments.

For his immense support both in and out of the lab, my extreme gratitude is given to my partner, Dr. Jacob T. Whitt. Without his presence this work would not have been possible.

Finally, I would like to thank my family for instilling a belief of limitless potential from an early age. My mother, Carol White, taught me that education is one of the most important journeys of our lives, and it never ends. My father, Dallas White, showed me to strive for the most out of life, but always appreciate what you have. My grandparents, Anna White, Albin and Joyce Petter, set great examples of a successful life I hope to achieve. My brother, Tyler White, and sister, Toni Sevener, always make me feel proud of my hard work, as I am of theirs. My gratitude is owed to my entire family, in-laws, and friends for all of their love and encouragement.

Group II Intron Retrohoming and Gene Targeting Reactions in *Drosophila melanogaster*

Travis Brandon White, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Alan M. Lambowitz

Mobile group II introns are retroelements that insert site-specifically into double-stranded DNA sites by a process called retrohoming. Retrohoming activity rests in a ribonucleoprotein (RNP) complex that contains an intron-encoded protein (IEP) and the excised intron RNA. The intron RNA uses its ribozyme activity to reverse splice into the top strand of the DNA target site, while the IEP cleaves the bottom DNA strand and reverse transcribes the inserted intron. My dissertation focuses on the *Lactococcus lactis* Ll.LtrB group II intron and its IEP, denoted LtrA. First, I investigated the ability of microinjected Ll.LtrB RNPs to retrohome into plasmid target sites in *Drosophila melanogaster* precellular blastoderm stage embryos. I found that injection of extra Mg^{2+} into the embryo was crucial for efficient retrohoming. Next, I compared retrohoming of linear and lariat forms of the intron RNP. Unlike lariat RNPs, retrohoming products of linear intron RNPs displayed heterogeneity at the 5'-intron insertion junction, including 5'-exon resection, intron truncation, and/or repair at regions of microhomology. To investigate whether these junctions result from cDNA ligation by non-homologous end-joining (NHEJ), I analyzed retrohoming of linear and lariat intron RNPs in *D. melanogaster* embryos with null mutations in the NHEJ genes *lig4* and *ku70*, as well as the DNA repair polymerase *polQ*. I found that null mutations in each gene decreased

retrohoming of linear compared to lariat intron RNPs. To determine whether novel activities of the LtrA protein contributed to the linear intron retrohoming 5' junctions, I assayed the polymerase, non-templated nucleotide addition and template-switching activities of LtrA on oligonucleotide substrates mimicking the 5'-intron insertion junction *in vitro*. Although LtrA efficiently template switched to 5'-exon DNA substrates, the junctions produced differed from those observed *in vivo*, indicating that template switching is not a significant alternative to NHEJ *in vivo*. Finally, I designed and constructed retargeted Ll.LtrB RNPs to site-specifically insert into endogenous chromosomal DNA sites in *D. melanogaster*. I obtained intron integration efficiencies into chromosomal targets up to 0.4% in embryos and 0.021% in adult flies. These studies expand the utility of group II intron RNPs as gene targeting tools in model eukaryotic organisms.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Group II intron characteristics

Group II introns are mobile retroelements found principally within eubacteria and the organelles of certain eukaryotes (Lambowitz & Zimmerly, 2004). These introns are catalytic RNA molecules capable not only of splicing from a precursor transcript, but also of integrating into double-stranded DNA (dsDNA) with the aid of an intron-encoded protein (IEP) in a process called retrohoming (Lambowitz & Zimmerly, 2004, 2010). Although group II introns have not been identified in the nuclear genomes of eukaryotes, retrotransposons and spliceosomal introns, hypothesized evolutionary descendents of group II introns, form a substantial portion of many eukaryotic genomes (Lambowitz & Zimmerly, 2010).

1.1.1 Overview of group II intron structure

Group II introns have a conserved secondary structure comprised of six domains (I-VI) radiating from a central wheel, and rely on tertiary interactions between domains to form the catalytically active ribozyme (Michel & Ferat, 1995). Differences in intron-exon interactions are found in the subgroup IIA, IIB and IIC introns (Lambowitz & Zimmerly, 2010). For group IIA introns, forward splicing from the precursor RNA depends upon base-pairing interactions between exon-binding sites 1 and 2 (EBS1 and 2) located in domain I of the intron RNA, and intron-binding sites 1 and 2 (IBS1 and 2) within the 5' exon of the precursor RNA, in addition to the δ-δ' tertiary interaction between the intron and the first 1-3 nucleotides (nt) of the 3' exon (Figure 1.1) (Michel & Ferat, 1995, Costa et al., 2000). These interactions bring both the 5' and 3' exons in proximity to the intron's active site located in domain V and an unpaired A residue located in domain VI (Michel

& Ferat, 1995, Qin & Pyle, 1998). Integral to group II intron ribozyme activity is the presence of coordinated Mg^{2+} ions at the active site (Qin & Pyle, 1998). The IEP open reading frame (ORF) is located within DIV, a domain dispensable for catalysis, whereas domains II and III contribute to overall stability of the molecule and reaction efficiency (Fedorova et al., 2003, Lambowitz & Zimmerly, 2010).

1.1.2 Group II intron-encoded proteins

For mobile group II introns, the IEP is required for splicing. The IEP binds specifically to the intron RNA to stabilize the catalytically active structure and thereby promotes the splicing of the intron from the precursor RNA ("maturase" activity) (Matsuura et al., 1997, Lambowitz & Zimmerly, 2004). Following splicing, the IEP remains bound to the excised intron RNA in a ribonucleoprotein particle (RNP) capable of promoting intron insertion into dsDNA sites (Lambowitz & Zimmerly, 2010). The IEPs are expressed from sequences located within DIV of the intron and have been shown to bind specifically to these same regions, thereby regulating their own translation (Wank et al., 1999, Singh & Lambowitz, 2001). IEPs are multifunctional proteins which posses reverse transcriptase, maturase, DNA binding, and DNA endonuclease activities. Mobile group II intron IEPs are composed of RT, X, D, and En domains (Figure 1.2). The N-terminal RT domain contains the conserved sequence blocks RT-1 to -7 corresponding to the fingers and palm regions of retroviral and other retroelement RTs, as well as an upstream conserved sequence block, RT-0, characteristic of the RTs of non-LTR retrotransposons (Malik et al., 1999, Zimmerly et al., 2001, Lambowitz & Zimmerly, 2004). The X domain functions together with the RT domain to specifically bind the intron RNA in its active structure and thus promote splicing (Wank et al., 1999, Cui et al., 2004). The D domain is required for dsDNA binding and intron integration in

dsDNA sites (Zimmerly et al., 1995a, Guo et al., 1997, San Filippo & Lambowitz, 2002). The C-terminal En domain is responsible for bottom-strand DNA cleavage during retrohoming. The 3'-DNA end generated by this cleavage is then used as a primer for reverse transcription of the integrated intron RNA (San Filippo & Lambowitz, 2002, Lambowitz & Zimmerly, 2004).

1.1.3 Group II intron splicing mechanism

Splicing of a group II intron from a precursor RNA is facilitated by base pairing between the intron EBS sequences and the 5'-exon IBS sequences, as well as the δ-δ' interaction between the intron and 3' exon (Michel & Ferat, 1995). These interactions bring the 5'- and 3'-splice sites in proximity to the intron's active site. Splicing occurs via two successive transesterification reactions. First, the 2' OH of a bulged A residue in DVI acts as the nucleophile for attack of the phosphate at the 5' end of the intron (Michel & Ferat, 1995). This reaction yields a branched intron lariat still attached to the 3' exon. During the second step of splicing, the free 3' OH of the 5' exon then attacks the intron-3'-exon junction (Michel & Ferat, 1995), which results in the formation of ligated exons and freed lariat intron RNA (Figure 1.3 A). For mobile group II introns, these splicing reactions are assisted by the IEP, which binds specifically to the intron RNA to stabilize the catalytically active RNA structure (Matsuura et al., 2001, Lambowitz & Zimmerly, 2004).

Group II intron RNAs can also splice without lariat formation by an alternative pathway termed "hydrolytic splicing" (Figure 1.3 B). In this pathway 5'-splice site cleavage occurs by attack of the phosphate at the 5' end of the intron by a water molecule (hydrolysis) rather than the 2' OH of a bulged A residue in DVI. The second step of splicing proceeds as with lariat intron splicing but now yields ligated exons and freed

linear intron RNA (Figure 1.3) (Michel & Ferat, 1995). Hydrolytic splicing was first observed during group II intron self-splicing reactions under non-physiological salt conditions (Schmelzer & Schweyen, 1986). It was then shown to occur *in vivo* for a mutant yeast mitochondrial intron with a branch-point A residue deletion (Podar et al., 1998) and for a subclass of plant chloroplast DNA introns that naturally lack the branch-point A residue (Vogel & Borner, 2002). Linear group II intron RNAs can also be generated from excised intron lariat RNAs by debranching, which is believed to accelerate RNA turnover (Green, 1986).

The splicing mechanism of group II introns is remarkably similar to that of eukaryotic nuclear introns. However, all steps for group II intron splicing are catalyzed by the intron RNA itself, whereas eukaryotic nuclear introns require the participation of several small nuclear RNAs and many proteins. Several studies have described group II intron splicing reactions *in trans* using RNA domains from separate molecules (Goldschmidt-Clermont et al., 1991, Qiu & Palmer, 2004, Glanz & Kuck, 2009). These studies along with the similarity in splicing mechanisms led to the hypothesis that group II introns are the evolutionary predecessor to spliceosomal introns by separation of structural features in the group II intron that were pressured to achieve splicing *in trans* and evolved to become the general snRNP factors of the spliceosome (Martin & Koonin, 2006, Rodriguez-Trelles et al., 2006, Lambowitz & Zimmerly, 2010).

1.2 GROUP II INTRON RETROHOMING REACTION

The IEP remains bound to the excised intron lariat RNA after protein-assisted splicing of mobile group II introns forming an RNP that promotes intron mobility (Zimmerly et al., 1995b, Yang et al., 1998, Saldanha et al., 1999, Matsuura et al., 2001). Group II intron RNPs retrohome to the ligated-exon junction ("homing site") in intronless

alleles at high frequency and retrotranspose to ectopic sites that resemble the normal homing site at low frequency (Dickson et al., 2001, Ichiyanagi et al., 2002). These processes enabled the dispersal of mobile group II introns to diverse bacteria and may have been used for the invasion and proliferation of group II introns in the nuclear genomes of early eukaryotes, where they evolved into spliceosomal introns.

Group II intron RNPs initiate intron mobility by recognizing a relatively large DNA target site spanning 20 to 25 bp upstream of the intron insertion site in the 5' exon and 10 bp downstream in the 3' exon (Yang et al., 1996, Singh & Lambowitz, 2001, Lambowitz & Zimmerly, 2004, Perutka et al., 2004). Intron RNA base-paring of the EBS1, EBS2 and δ sequences with the DNA target site's IBS1, IBS2 and δ' sequences contribute to DNA target site recognition and promote reverse-splicing of the intron RNA into the top DNA strand (Singh & Lambowitz, 2001, Aizawa et al., 2003). IEP contacts with distal portions of the 5'- and 3'-exon DNA sequences also contribute to recognition of the DNA target site (Zimmerly et al., 1995a, Mohr et al., 2000). The intron lariat RNA splices directly into the top DNA strand by a reversal of the group II intron splicing reaction (Figure 1.4 A). In the first step of reverse spicing, the 3' end of the intron RNA attacks the phosphodiester bond at the ligated exon homing site on the top DNA strand, which results in the ligation of the 3' end of the intron RNA to the 5' end of the 3'-exon DNA (Zimmerly et al., 1995a, Yang et al., 1996, Aizawa et al., 2003). In the second step, the free 3' OH group at the 3' end of the 5'-exon DNA attacks the branch-point phosphodiester bond in the intron, resulting in the ligation of the 5' end of the intron RNA to the 3' end of the 5'-exon DNA. The intron RNA cleaves precisely at the ligated exon junction in the top DNA strand, whereas the IEP cleaves at +9 or +10 in the 3' exon, depending on the group II intron, in the bottom DNA strand (Zimmerly et al., 1995a, Yang et al., 1996, Guo et al., 1997, Matsuura et al., 1997, Eskes et al., 2000).

After reverse splicing, the inserted intron RNA is reverse transcribed by the IEP in a process called target-primed reverse transcription (TPRT), using either a DNA strand cleaved by the IEP or a nascent strand at a DNA replication fork to prime cDNA synthesis (Zimmerly et al., 1995b, Matsuura et al., 1997, Yang et al., 1998, Lambowitz & Zimmerly, 2010). The latter pathway is used for the mobility of group II introns with endonuclease-deficient IEPs (Zhong & Lambowitz, 2003, Lambowitz & Zimmerly, 2004). Finally, the resulting intron cDNA is integrated into the recipient DNA by host cell DNA recombination or repair mechanisms (Eskes et al., 1997, Smith et al., 2005).

Linear group II intron RNAs cannot carry out both steps of reverse splicing, as the energy to drive the second step of the reaction originates from the 2'-5' phosphodiester bond present in the branched lariat structure (Figure 1.4 B) (Mörl et al., 1992). Linear group II introns can, however, carry out the first step of reverse splicing into either RNA or DNA sites, thereby ligating their 3' end to the 5' end of the 3' exon (Mörl et al., 1992, Mastroianni et al., 2008, Roitzsch & Pyle, 2009). This partial reverse splicing reaction could potentially be used for intron mobility. If partial reverse splicing occurs into an RNA site, then the recombined RNA would need to be reverse transcribed and the resulting cDNA integrated into the host genome by recombination, while if partial reverse splicing occurs into a DNA site, the attached intron RNA could be reverse transcribed and the intron cDNA integrated by DNA repair. The latter pathway is more likely for group II introns that encode RTs because association with the IEP in RNPs strongly biases the intron RNA to reverse splice into DNA, especially dsDNA, rather than RNA sites (Zimmerly et al., 1995a).

1.3 THE LL.LTRB GROUP II INTRON

Although the retrohoming mechanism was first elucidated in studies with yeast mitochondrial group II introns (Zimmerly et al., 1995b, Yang et al., 1996), it has been studied in the greatest detail for the *Lactococcus lactis* Ll.LtrB group IIA intron and its cognate IEP, denoted LtrA (Matsuura et al., 1997, Cousineau et al., 1998, Saldanha et al., 1999, Singh & Lambowitz, 2001). This intron was discovered in the relaxase gene, *ltrB* in the *L. lactis* conjugative element pRS01, and splicing of this intron from the *ltrB* precursor RNA is essential for translation of the functional relaxase involved in conjugation (Mills et al., 1996, Shearman et al., 1996). Unlike many other group II intron RNPs, this complex expresses well and is functional in *E. coli*, which has enabled its use for numerous studies of the retrohoming process and allowed preparation of purified RNPs for gene targeting experiments.

1.3.1 Retrohoming of the Ll.LtrB group II intron

During retrohoming of the Ll.LtrB group II intron, the spliced intron RNA, in linear or lariat form, remains bound to LtrA, forming an RNP that is able to scan dsDNA for the correct target site and facilitate local DNA unwinding. Contacts between the intron RNA EBS and δ sequences with top DNA strand IBS and δ' sequences promotes insertion of the intron RNA into the DNA target site in a reversal of the forward splicing reaction (Matsuura et al., 1997, Singh & Lambowitz, 2001, Aizawa et al., 2003). Intron RNA-DNA target site interactions, as well as contacts between key residues in the 5'- and 3'-exon DNA with LtrA, allows endonucleolytic cleavage of the bottom DNA strand between the +9 and +10 positions relative to the intron RNA insertion site. Subsequently, LtrA carries out TPRT in which the RT uses the cleaved bottom-strand DNA as a primer for DNA synthesis beginning in the 3' exon and extending into the inserted intron RNA to create a cDNA copy of the intron on the bottom DNA strand (Figure 1.4). A nascent

DNA strand at a replication fork can also serve as a primer for LtrA-mediated reverse transcription of the intron (Zhong & Lambowitz, 2003). Following reverse transcription, host enzymes repair the RNA-DNA hybrid at the DNA target site resulting in ligation of the cDNA to the DNA 5' exon and replacement of the intron RNA strand with DNA. These processes have been studied extensively in *E. coli* and have been shown to depend upon DNA exonuclease, ligase and RNase H enzymes, as well as the DNA replication polymerase Pol III and possibly the repair DNA polymerases Pol II, IV, and V (Smith et al., 2005).

1.4 ALTERATION OF TARGET-SITE SPECIFICITY FOR GROUP II INTRON RNPS

Group II intron target site interactions are determined primarily through EBS-IBS base-paring, with relatively few IEP sequence contacts contributing to DNA target site recognition (Guo et al., 1997, Yang et al., 1998). Concomitant changes to the EBS and IBS residues, as well as the δ and δ' residues in the precursor RNA enable efficient forward splicing of the intron, as well as specific retrohoming into DNA target sites containing the same IBS and δ' sequences (Eskes et al., 1997, Guo et al., 2000, Mohr et al., 2000, Karberg et al., 2001). Although alteration of the RNA/DNA target site base pairing to change target specificity was first demonstrated for the yeast aI1 intron (Eskes et al., 1997), the Ll.LtrB intron has been the subject of extensive engineering to allow retrohoming into desired DNA sequences.

1.4.1 Development of the Ll.LtrB group II intron into a gene targeting vector

The initial work needed for retargeting the Ll.LtrB group II intron included identification of nucleotides important for target site recognition by the IEP (Mohr et al., 2000). These studies were achieved through biochemical analysis of the reverse splicing and bottom-strand cleavage steps using DNA target sites with single nucleotide

substitutions in portions of the 5'- and 3'-exon DNA residing outside of the EBS-IBS and δ-δ' interactions (Mohr et al., 2000). Two nucleotide residues, G-21 and T+5, were identified as critical for the reverse splicing and bottom-strand cleavage steps of retrohoming, respectively. Additionally, it was found that the GC content present in the wild-type target site was important for reverse splicing, as DNA unwinding by the IEP is necessary for subsequent RNA base pairing with the DNA target site. Single nucleotide substitution experiments were also used to test the efficiency of reverse splicing for mutant RNA/DNA target site interactions in the EBS-IBS and δ-δ' regions. Tolerable deviations from wild-type EBS-IBS and δ-δ' interactions, nucleotide residues important for LtrA activity, and GC content were all considered in choosing target sites in the *E. coli thyA* gene present on a recipient plasmid (Mohr et al., 2000). These studies showed the Ll.LtrB group II intron could be efficiently retargeted to DNA sites of interest.

An efficient plasmid assay was devised for the investigation and quantification of retargeted Ll.LtrB group II intron mobility frequencies in *E. coli* (Guo et al., 2000, Karberg et al., 2001). Studies of retargeting of the Ll.LtrB intron in bacteria into plasmid sequences through randomization of EBS and δ sequences enabled selection of targeted events (Guo et al., 2000, Karberg et al., 2001). Large-scale analysis of retargeted intron retrohoming events allowed development of a computer algorithm for predicting ranking orders of efficient DNA target sites (Perutka et al., 2004). This algorithm has been used to obtain gene knock-outs and knock-ins in diverse Gram-negative and Gram-positive bacteria using retargeted Ll.LtrB introns ("targetrons") expressed from transformed vectors (Frazier et al., 2003, Chen et al., 2005, Yao et al., 2006, Heap et al., 2007, Shao et al., 2007, Yao & Lambowitz, 2007, Rodriguez et al., 2008, Zoraghi et al., 2010, Carter et al., 2011). This system of group II intron gene targeting has allowed highly efficient

chromosomal gene knockouts in bacterial systems at frequencies approaching 100%, six orders of magnitude higher than the incidence of ectopic integration.

Reconstitution of purified targetron RNPs *in vitro* has allowed investigation of both plasmid and chromosomal gene targeting in eukaryotes through introduction of RNPs by microinjection (Mastroianni et al., 2008, Zhuang et al., 2009). Such RNPs retained activity in human cell assays in which plasmid-borne target sites and targetron RNPs were introduced separately by liposome-mediated transfection (Guo et al., 2000). In these initial experiments, however, group II intron integration into plasmid target sites in human cells was much less efficient than in bacteria, requiring nested PCR for detection.

1.5 EUKARYOTIC GENE TARGETING TECHNOLOGY

It would be highly desirable to develop group II intron-based gene targeting methods for higher organisms analogous to those that currently exist for prokaryotes. With the exception of mouse embryonic stem cells and a few other cell types, mammalian cells and those of other higher eukaryotes lack efficient homologous recombination systems for gene targeting (Capecchi, 1989, Bunz, 2002). Although RNAi has been a major advance in genetic manipulation in higher organisms, it yields "knock downs" rather than "knock outs", requires continuous expression or repeated introduction, has off-target effects, and does not permit the introduction of new genes (Hannon, 2002, Jackson et al., 2003, Dorsett & Tuschl, 2004). Currently available gene therapy vectors, either do not integrate stably (adenovirus) or insert at multiple, semi-random DNA sites with possible deleterious effects (retroviruses, rep-deficient adeno-associated virus (AAV))(Verma & Somia, 1997, Anderson, 1998, Marshall, 2002, Baum et al., 2004, Donsante et al., 2007).

To overcome such deficiencies, alternative gene targeting methods are being developed, employing different methods to increase the frequency of homologous recombination. One of the best developed of such methods for human cells, homologous recombination with recombinant AAV, generally yields 0.1-15% of integrations at the desired target site (Russell & Hirata, 1998, Rago et al., 2007). Other methods with potentially higher specificity involve the use of site-specific DNA endonucleases, such as meganucleases or artificial Zn-finger nucleases, to make a targeted double-strand break that promotes homologous recombination (Jasin, 1996, Porteus & Carroll, 2005, Pâques & Duchateau, 2007, Cathomen & Keith Joung, 2008). Efficient gene targeting in mammalian cells has been achieved by using Zn-finger nucleases in this way to stimulate homologous recombination with a co-transformed donor DNA (Urnov et al., 2005, Lombardo et al., 2007, Moehle et al., 2007). This method has been reported to give recombination frequencies up to 29% in mammalian cells and is efficient enough to target both alleles in diploids. Zn-finger nucleases have also been used to obtain gene disruptions in mammalian cells by mutagenic nonhomologous end-joining (NHEJ) (Santiago et al., 2008) and are being applied for gene targeting in plants and eukaryotic model organisms (Bibikova et al., 2002, Lloyd et al., 2005, Beumer et al., 2006, Doyon et al., 2008, Meng et al., 2008). Ultimately, the utility of this and other methods will depend on the ease of designing and optimizing protein endonucleases to recognize different DNA target sequences, their specificity and toxicity, and their cost and availability compared to other gene targeting methods.

1.5.1 Development of targetrons for eukaryotic gene targeting

Mobile group II introns can potentially be employed in the same way as Zn-finger nucleases and meganucleases to introduce a targeted double-strand break that stimulates

homologous recombination or yields mutations by inaccurate NHEJ (Karberg et al., 2001). In the case of group II introns, the double-strand break results from the initial reverse splicing and second-strand cleavage reactions. Reverse splicing of the lariat form of the intron occurs in two steps, the first of which is attachment of intron lariat RNA to the 5' end of the 3' exon leaving a strand break, while second-strand cleavage is a conventional DNA endonuclease reaction catalyzed by the IEP (Lambowitz & Zimmerly, 2004). Use of the linear form of group II introns allows the DNA top-strand break to persist, since the second step of reverse splicing does not occur (Mastroianni et al., 2008). Both linear and lariat intron RNPs have been shown to stimulate homologous recombination between a donor DNA and plasmids bearing the wild-type target site in *Xenopus laevis* oocyte nuclei (Mastroianni et al., 2008). Compared to protein endonucleases, group II introns offer the potential advantage that the double-strand break can be targeted to different locations simply by modifying the base-pairing sequences in the intron RNA, thereby obviating the need for protein engineering to change target specificity (Karberg et al., 2001).

Mobile group II introns, like retroviruses to which they are related, can also be used as integrating vectors, enabling the insertion of cargo genes at desired sites, with the advantage of high, controllable DNA target specificity (Mohr et al., 2000, Frazier et al., 2003, Jones et al., 2005). This gene delivery is made possible through replacement of the LtrA ORF in domain IV of the intron with genetic information of interest. Such an approach was used to successfully integrate a β-globin repair construct into a plasmid-borne mutant β-globin gene which mimics that found in human patients suffering from β-thalassemia (Jones et al., 2005). Site-specific integration of the targetron delivered a eukaryotic splice acceptor sequence and downstream repaired exons to an intron of the mutant β-globin gene. Subsequent transfection of targetron-integrated plasmids into

mammalian cells allowed expression of wild-type β-globin transcript and protein, highlighting the potential for targetron-mediated gene delivery in human cells.

Another example of the utilization of cargo sequences in DIV of targetrons is the retrotransposition activated selectable marker (RAM) (Zhong et al., 2003). These RAM-targetrons enable selection of mobility events, as has been used for targeting chromosomal genes in *E. coli* and members of the *Clostridium* genus (Zhong et al., 2003, Yao et al., 2005, Heap et al., 2007). For use in prokaryotes, the RAM consists of a selectable antibiotic resistance marker with its promoter (the trimethoprim resistance gene, DHFR, for *E. coli*, erythromycin resistance gene, *ermB*, for members of the *Clostridium* genus) inserted within DIV of the targetron in the antisense orientation. The selectable marker is disrupted by the efficient self-splicing *td* group I intron. During retrotransposition of the targetron via an RNA intermediate, the *td* group I intron is spliced, forming the intact genetic marker. Completion of retrohoming enables expression of the integrated marker and subsequent selection of these cells.

Identification of targetron integration at genomic locations in *Xenopus laevis* sperm nuclei has been greatly improved with the development of targetrons with RAM-GFP. In this RAM construct, a GFP eukaryotic expression cassette is placed within DIV of the targetron (Zhuang, 2009). Gene targeting of decondensed chromosomes within *Xenopus laevis* sperm nuclei was investigated with RAM-GFP RNPs harboring randomized EBS1, EBS2 and δ sequences. Fertilization of oocytes through microinjection of these modified sperm nuclei allowed rapid identification of oocytes containing intron integration into the genome by GFP expression in embryos and tadpoles (Zhuang, 2009). Analysis of genomic DNA from targeted tadpoles by thermal asymmetric interlaced (TAIL) PCR (Liu et al., 1995) identified the genomic location of integration and EBS-IBS and δ-δ' interactions at eukaryotic chromosomal loci (Zhuang,

2009). This process relied wholly on the use of the linear group II intron, since splicing of an *in vitro* prepared randomized EBS-IBS library is inefficient and produces few lariat introns that would generate functional RNPs.

1.6 OVERVIEW OF DISSERTATION RESEARCH

This dissertation research focuses on retrohoming of Ll.LtrB group II intron RNPs to both plasmid and chromosomal sites in the model eukaryotic organism *Drosophila melanogaster*. The dissertation is divided into three related subjects, which are described in the next three chapters. In the next chapter (Chapter 2), I describe the retrohoming of wild-type Ll.LtrB intron RNPs to plasmid-borne DNA target sequences microinjected into precellular blastoderm stage *D. melanogaster* embryos. First, I investigated the retrohoming efficiency of lariat intron RNPs in this eukaryotic environment. In these experiments I employed a previously described bacterial genetic assay of retrohoming efficiency where integration of the Ll.LtrB group II intron RNP into the wild-type DNA target site present on an Amp^R plasmid activated a downstream *tet*^{R} gene. Transformation of recovered plasmids into *E. coli* enabled quantification of retrohoming efficiency by determining the ratio of (Amp^R + Tet^R)/Amp^R colonies. I found addition of Mg^{2+} was crucial for retrohoming in the embryo. Retrohoming efficiencies of up to 9.5% were obtained using a microinjection protocol with an initial injection of 17 mM dNTPs and 250 mM $MgCl_2$, followed by microinjection of lariat intron RNPs in a low salt buffer. Furthermore, retrohoming was dependent upon injection into the embryo, indicating host factors were required for resolution of the retrohoming product.

Next, I compared retrohoming of linear and lariat forms of the intron RNP. Linear forms of group II introns are generated as a product of the hydrolytic splicing reaction or by run-off transcription of PCR products, as was done in my experiments. The linear

intron RNP, while able to site-specifically insert into a dsDNA target and undergo TPRT as with the lariat intron RNP, cannot fully reverse splice into the top DNA strand. This reaction effectively produces a DSB with the intron RNA-cDNA hybrid at one side of the DNA break. I found retrohoming efficiencies using linear intron RNPs were greatly reduced compared with that of lariat intron RNPs. Sequence analysis of retrohoming products showed both linear and lariat intron RNPs produced the accurate 3'-intron integration junction. When I analyzed the 5'-intron integration junction of lariat intron RNP retrohoming products, I found that lariat intron RNPs undergo complete intron integration with high fidelity. In contrast, retrohoming of linear intron RNPs in the embryo produces heterogeneous 5'-intron integration junctions. These heterogeneous 5' junctions have 5'-exon DNA resections, intron cDNA truncations, insertion of extra nucleotide residues, and/or indication of DNA repair at sites of microhomology. These characteristics are hallmarks of the eukaryotic process of non-homologous end-joining (NHEJ) DNA repair.

Next, I examined the retrohoming efficiencies of linear and lariat intron RNPs in wild-type embryos and embryos harboring homozygous null mutations in genes encoding the NHEJ factors DNA ligase IV (Lig4) and Ku70. I also investigated retrohoming in embryos lacking DNA polymerase θ (PolQ), the *Drosophila* ortholog of the vertebrate DNA polymerase θ, which is involved in DNA repair. Retrohoming of the linear intron, but not lariat, was specifically reduced in a *lig4*$^{-}$ background compared with wild-type embryos. Rescue experiments in which the wild-type *lig4* gene was introduced in the *lig4*$^{-}$ background by P-element transformation regained wild-type levels of linear intron retrohoming. I found both types of intron RNPs show a decrease in retrohoming efficiency in the *ku70*$^{-}$ and *polQ*$^{-}$ background relative to wild-type embryos, although the linear intron was affected to a significantly greater extent in the *polQ*$^{-}$ embryos. These

findings provide evidence that the linear intron retrohoming product is subject to repair by Lig4-dependent and, to a lesser extent, Lig4-independent NHEJ in the *D. melanogaster* embryo.

Although the characteristics of linear intron RNP retrohoming 5' junctions were most likely produced by NHEJ in the embryo, I also investigated whether novel activities of the IEP, LtrA, may have contributed to their formation. Biochemical assays of LtrA activities on oligonucleotide substrates are discussed in Chapter 3. I assayed LtrA non-templated nucleotide addition and template-switching activities *in vitro* using oligonucleotide mimics of the 5' junction produced upon linear intron RNP retrohoming. I found that LtrA was able to add non-templated nucleotides to the 3' end of cDNAs and template switch to both DNA and RNA oligonucleotides bearing the 5'-exon sequence. I analyzed the sequences of these cDNA products and found that the composition and length of the non-templated nucleotide residues added prior to template switching to 5'-exon DNA substrates differed from insertions produced at the 5' junctions from linear intron RNP retrohoming in embryos. Furthermore, template switching to internal sites within the DNA oligonucleotide was not observed in these assays. Thus, I conclude that LtrA template switching is not likely to have generated 5'-exon deletions of sequences nor the majority of inserted sequences found from linear intron RNP retrohoming in embryos.

Next, I investigated LtrA activities at the retrohoming 3'-integration junction. During retrohoming of the Ll.LtrB group II intron, LtrA cleaves the bottom DNA strand between residues +9 and +10 relative to the intron insertion site on the top DNA strand. LtrA then uses the generated 3' end to serve as a primer for reverse transcription of the inserted intron RNA. Thus, LtrA must copy 9-nt of the 3'-exon DNA prior to switching to the intron RNA template. It was unclear how well LtrA could use DNA templates and

previous studies have suggested that host DNA polymerases may have a role this process (Smith et al. 2005). However, these activities had not been investigated systematically for LtrA and my finding that LtrA could efficiently template switch to 5'-exon DNA substrates *in vitro* reopened this issue. I assayed the DNA polymerase and template-switching activities of LtrA on oligonucleotide substrates that mimic the 3'-intron integration junction. I found that while LtrA efficiently copied 3'-exon RNA substrates, polymerase activity opposite 3'-exon DNA was inefficient. To characterize LtrA activity at DNA to RNA template transitions encountered at the 3' junction, I also tested substrates that append 20-nt of either DNA or RNA of Ll.LtrB intron 3'-end sequence to the 5' end of the 3'-exon DNA, the latter mimicking the 3' junction formed *in vivo*. I found LtrA exhibited similar activities on these two substrates in time-course assays, indicating initiation of cDNA synthesis opposite the 3'-exon DNA substrate was the limiting step in cDNA synthesis opposite these two substrates. I conclude that initiation of LtrA polymerase activity opposite 3'-exon DNA during retrohoming is an inefficient step and the participation of a host polymerase at the Ll.LtrB intron retrohoming 3' junction is possible.

Finally, the finding that Ll.LtrB RNPs function efficiently in the environment of the *D. melanogaster* precellular blastoderm stage embryo led me to investigate chromosomal gene targeting reactions using retargeted Ll.LtrB RNPs ("targetrons"). In Chapter 4, I describe targetron-mediated gene targeting experiments for the two *D. melanogaster* genes *yellow* (*y*) and *white* (*w*). Using a computer algorithm for efficient target site predictions (Perutka et al., 2004), I designed and constructed two *y* gene-directed targetrons, Y18a and Y3776s, and one *w* gene-directed targetron, W3711a, where the number indicates the nucleotide position of the targetron-insertion site relative to the first nucleotide of the ATG start codon and "a" or "s" indicate the antisense and

sense strands, respectively. For gene targeting experiments, these targetron RNPs were introduced into embryos by microinjection. By nested PCR analysis of genomic DNA extracted from injected embryos, I detected the 3' and 5' junctions from site-specific integration of Y18a into the *y* gene. In similar experiments using Y3776s RNPs, I detected the 3'-integration junction by direct PCR. As with plasmid targeting experiments, detection of site-specific targetron integration was dependent upon an initial microinjection of a buffer containing Mg^{2+} prior to injection of targetron RNPs. I investigated the gene targeting abilities of Y3776s RNPs in a solution containing polyamines and found these were capable of site-specific integration as detected by nested PCR of the 3'-integration junction. This procedure allowed for gene targeting reactions to be achieved using a single injection protocol, which increases embryo survival compared with a double injection protocol. I developed TaqMan quantitative PCR (qPCR) assays for detection of Y18a and W3711a 3'-integration junctions. In Y18a gene targeting experiments, I found integration efficiencies of up to 0.12% of endogenous *y* targets in wild-type embryos. Integration efficiencies up to 0.021% were seen in adult flies, indicating Y18a integration in embryos are stable genomic alterations capable of persisting through adulthood. I found W3711a integration efficiencies of up to 0.4% in embryos. This work describes generation of the first viable, fertile eukaryotic organism to have an integrated retargeted group II intron at a single-copy gene in its genome and highlights the potential of group II introns for gene therapy applications.

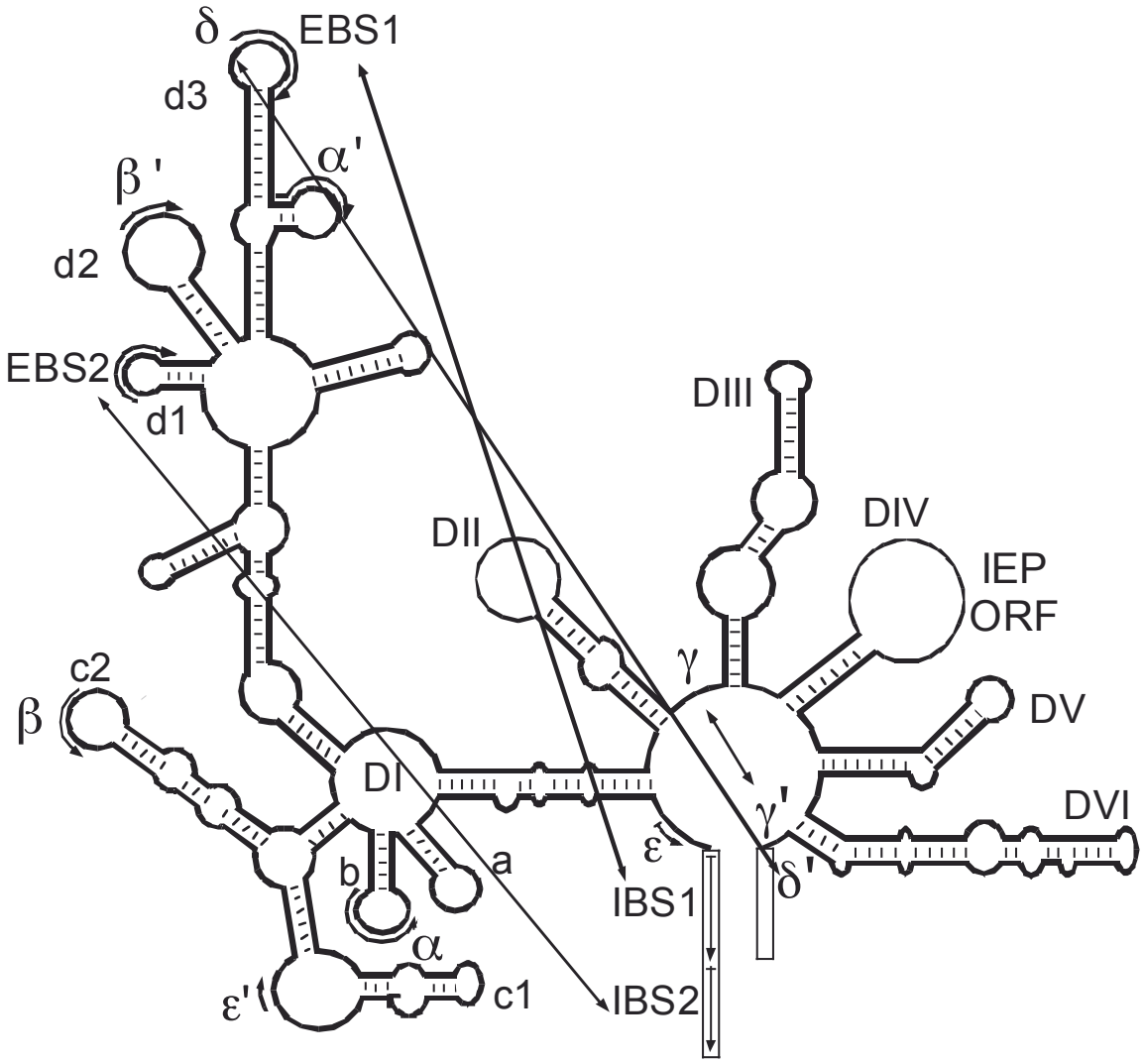


Figure 1.1: Group IIA intron secondary structure

The conserved secondary structure of a group IIA intron consists of six double-helical domains D1-DVI. The IEP ORF is located in the "loop" region of DIV. The EBS1, EBS2, and δ sequences located in the stem-loops Id1 and Id3 of DI base pair to sequences IBS1, IBS2, and δ' sequences in the 5' and 3' exons, shown with arrows, for RNA splicing and reverse splicing. Greek letters indicate sequence elements involved in long-range tertiary interactions. The figure is adapted from (Lambowitz & Zimmerly, 2004).

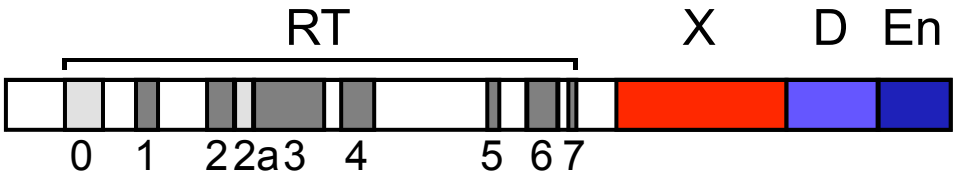


Figure 1.2: Group II intron-encoded protein (IEP) domains

IEP domains conserved among mobile group II intron IEPs are shown for the *Lactococcus lactis* Ll.LtrB group II intron IEP, LtrA. In the RT domain, conserved sequence blocks 1-7, which are found in all RTs, are shown in dark gray, and conserved sequence block RT-0 and 2a, which are characteristic of group II intron and other non-LTR-retroelement RTs are shown in light gray. The X, maturase, domain is shown in red; the DNA binding domain (D) is shown in light blue; and the DNA endonuclease domain (En) is shown in dark blue. The figure is adapted from (Lambowitz & Zimmerly, 2004).

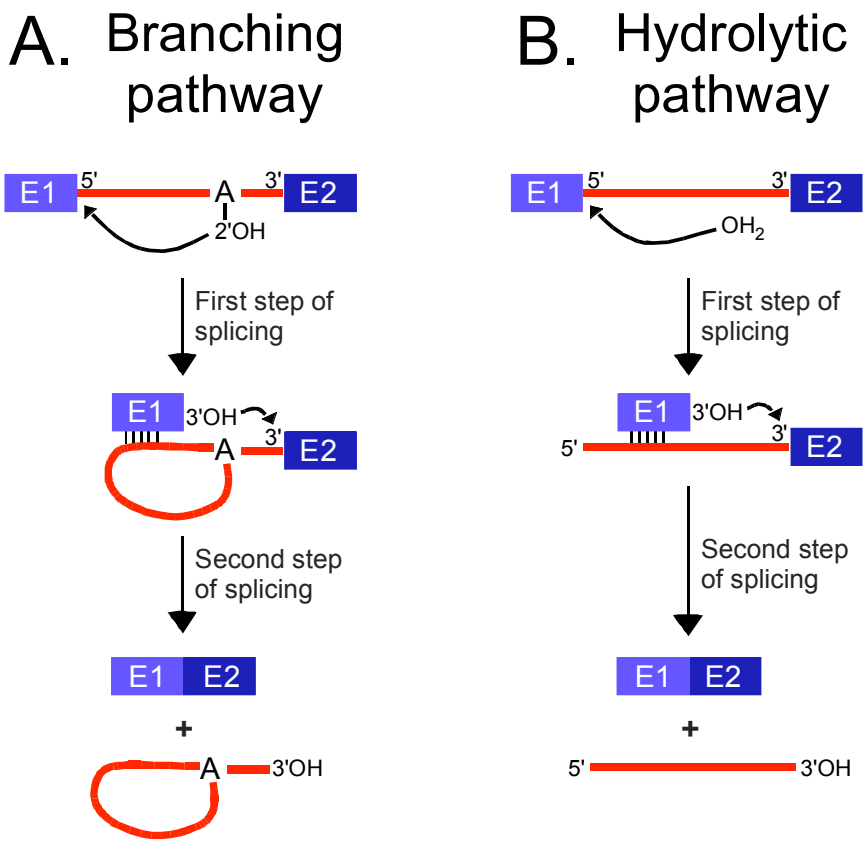


Figure 1.3: Group II intron splicing pathways

Group II introns splice by two successive transesterification reactions. (A) In the first step of the branching pathway of splicing, nucleophilic attack by the 2' OH of a nucleotide residue within DVI (typically an unpaired A) at the 5' exon-intron junction results in formation of a branched intron lariat-3' exon intermediate. In the second step, the free 3' OH of the 5' exon attacks to intron lariat-3' exon junction to produce ligated exons and freed lariat intron. (B) In the first step of the hydrolytic pathway of splicing, nucleophilic attack by a water molecule at the 5' exon-intron junction results in formation of a linear intron-3' exon intermediate. In the second step, the free 3' OH of the 5' exon attacks to linear intron-3' exon junction to produce ligated exons and freed linear intron.

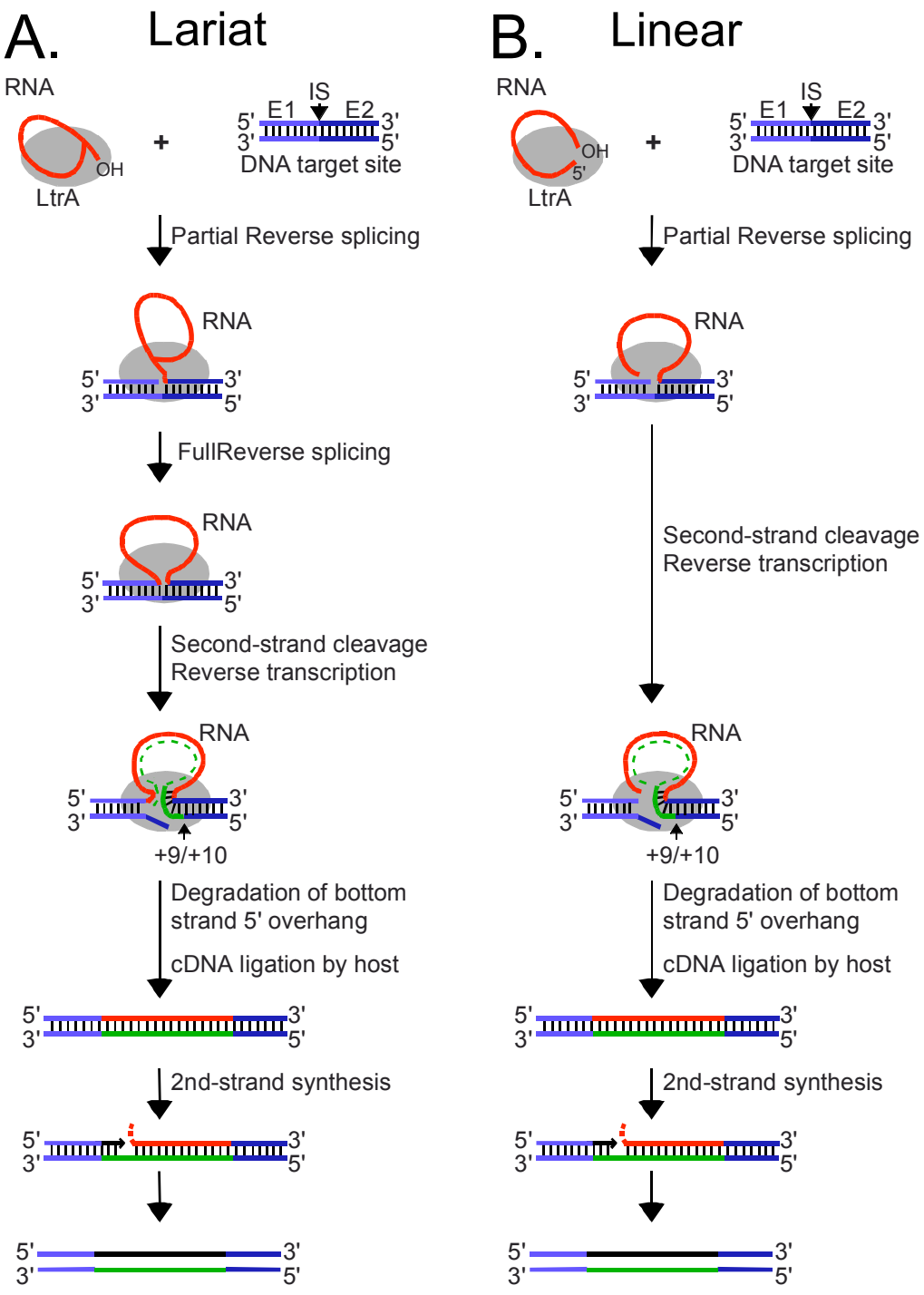


Figure 1.4: Group II intron retrohoming mechanism

Mobile group II introns insert site-specifically into double-stranded DNA target sites in a process called retrohoming. (A) In lariat intron retrohoming, excised intron lariat RNA (shown in red) in a ribonucleoprotein (RNP) complex with the IEP (shown in gray) first uses its ribozyme activity to partially reverse splice into the intron-insertion site (IS) between the ligated 5'-exon (E1; shown in light blue) and 3'-exon (E2; shown in dark blue) sequences in the top DNA strand. The free 3' OH of the 5' exon then attacks the branch-point phosphodiester bond to ligate the 5' end of the intron to the 3' end of the 5' exon, resulting a fully reverse-spliced intron in the top DNA strand. The IEP uses its DNA endonuclease activity to cleave the bottom DNA strand at 3'-exon position +9 or +10, depending on the group II intron. Next, the IEP uses the 3' end of the cleaved bottom DNA strand as a primer for reverse transcription of the inserted intron RNA and intron cDNA synthesis (shown in green) continues through the intron into the 5' exon. Resection of 5' overhang on the 3' exon and replacement of the intron RNA strand with DNA (shown in black) are done by host enzymes. (B) In linear intron retrohoming, linear intron RNA (shown in red) in an RNP complex with the IEP (shown in gray) uses its ribozyme activity to partially reverse splice into the same intron-insertion site (IS) in the top DNA strand. The IEP then uses its DNA endonuclease activity to cleave the bottom DNA strand at 3'-exon position +9 or +10 depending on the intron, and the IEP uses the 3' end on the bottom DNA strand as a primer for reverse transcription of the attached reverse-spliced linear intron RNA, yielding an intron cDNA whose free end is ligated to the 5' exon by a host cell DNA repair mechanism. Resection of the 5' overhang of the 3' and replacement of the intron RNA strand with DNA (shown in black) are done by host enzymes, as for lariat intron retrohoming.

Chapter 2: Analysis of linear and lariat Ll.LtrB group II intron RNP retrohoming into plasmid target sites in *Drosophila melanogaster* embryos

2.1 INTRODUCTION

In order to extend research of group II intron gene targeting reactions in eukaryotic cells, I investigated the ability of microinjected Ll.LtrB RNPs to retrohome in *D. melanogaster* precellular blastoderm stage embryos. By assaying retrohoming efficiencies into injected target plasmids, I was able to find Mg^{2+} concentration and temperature optima for linear and lariat RNP retrohoming. I compared the retrohoming products of linear and lariat RNPs and describe the characteristics of the 5'-linear intron insertion junction. The sequence characteristics surrounding these 5' junctions suggested the process of NHEJ participated in resolution of the DSB resulting from linear intron retrohoming. I investigated the retrohoming of both linear and lariat intron RNPs in embryos that have null mutations in the NHEJ genes *lig4* and *ku70*, as well as DNA polymerase θ (*polQ*). Specific reduction of linear intron retrohoming in the *lig4*$^-$ mutant embryos further indicates involvement of the NHEJ pathway for repair of the linear intron RNP retrohoming product. This defect was rescued by expression of wild-type Lig4 protein from a P-element transgene in the *lig4*$^-$ background. I found both linear and lariat intron retrohoming was reduced in embryos lacking Ku70 or PolQ, but the linear intron to a greater extent. Sequence analysis of 5' junctions from linear intron RNP retrohoming experiments in *polQ*$^-$ embryos show a large decrease in addition of extra nucleotides compared to wild-type embryos.

2.2 RETROHOMING IN WILD-TYPE EMBRYOS

2.2.1 Effect of Mg^{2+} on lariat RNP retrohoming

There was reason to believe that group II introns might be particularly sensitive to low Mg^{2+} concentrations in eukaryotic cells. In the yeast *Saccharomyces cerevisiae*, mutations in mitochondrial Mg^{2+}-transport proteins, which result in decreased intramitochondrial Mg^{2+} concentrations, strongly inhibited the splicing of all four mtDNA group II introns, while having much less effect on the splicing of group I introns (Wiesenberger et al., 1992, Gregan et al., 2001). *In vitro* at 100 mM KCl, the Ll.LtrB intron has a Mg^{2+} optimum of 5 mM for RNA splicing promoted by the IEP and 10 mM for reverse splicing of RNPs into DNA target sites (Saldanha et al., 1999). Thus, it was possible that group II intron RNPs cannot function optimally in eukaryotic nuclei where the free Mg^{2+} concentration is estimated to be 1-2 mM (Romani & Maguire, 2002). To investigate the functionality of group II intron RNPs in the eukaryotic environment, I assayed their retrohoming into plasmid targets in *D. melanogaster* precellular blastoderm stage embryos after microinjection of components, with an initial focus on the Mg^{2+}-dependence of the intron integration reaction.

To test whether the group II intron-integration reaction could occur in *D. melanogaster* embryos, I used a plasmid assay patterned after one developed to assay group II intron mobility in *E. coli* (Figure 2.1) (Guo et al., 2000, Karberg et al., 2001). In this assay, an Ll.LtrB intron with a phage T7 promoter inserted near its 3' end integrates into a target site (the ligated E1-E2 sequence of the *ltrB* gene) cloned upstream of a promoterless tet^R gene in an Amp^R target plasmid. Insertion of the intron containing the T7 promoter into the target site activates the tet^R gene, enabling intron-integration efficiencies to be measured by the ratio of $(Tet^R + Amp^R)/Amp^R$ resistant colonies in transformed *E. coli*. For plasmid retrohoming assays, I used a streamlined 0.9-kb

derivative of the Ll.LtrB intron from which ORF sequences encoding the IEP have been deleted. This Ll.LtrB-ΔORF intron was shown previously to have much higher integration efficiency and nuclease-resistance than does the full-length intron (Guo et al., 2000).

Standard microinjection procedures (Ashburner et al., 2005) were used for retrohoming assays in *D. melanogaster* embryos. The target plasmid (pBRR3-1trB) containing the *ltrB* target site was injected into the posterior of 100 embryos on a single slide covered in halocarbon oil first, followed within 5 min by Ll.LtrB RNPs reconstituted with self-spliced Ll.LtrB-ΔORF lariat intron RNA and purified recombinant LtrA protein (Figure 2.2 and Materials and Methods sections 2.5.1-2.5.4). In all the *D. melanogaster* experiments, the DNA target plasmid and RNPs were injected using different needles to avoid prior mixing. After incubation for 1 h at 30°C in a humidified chamber, the embryos for each condition were pooled and nucleic acids were extracted and electroporated into *E. coli* HMS174(DE3), which expresses phage T7 RNA polymerase. Serial dilutions of transformed cells were plated in triplicate on LB medium containing ampicillin with or without tetracycline, and averages were used to determine mobility frequency by the ratio of $(Amp^R + Tet^R)/Amp^R$ colonies.

As shown in Table 2.1, when the target plasmid DNA in solution with 17 mM dNTPs and intron RNPs were injected without further additions, no Tet^R colonies were obtained. In light of the previous studies indicating that group II introns are particularly sensitive to low Mg^{2+} concentrations, I repeated the experiment, now injecting the target plasmid and dNTPs in a solution containing 100, 250, and 500 mM $MgCl_2$. Under these conditions, the RNP integration efficiency increased dramatically, up to 9.5% with target plasmid and dNTPs in solution with 250 mM $MgCl_2$. Because the injection volumes were ~0.3 nl and the reported embryo volume is 7.3 nl (Tuschl et al., 1999), the calculated

increase in intracellular Mg^{2+} resulting from this injection would be ~10 mM, consistent with optimal Mg^{2+} concentrations for Ll.LtrB intron RNP integration in the eukaryotic intracellular environment of *Xenopus laevis* oocyte nuclei and *Danio rerio* embryos (Mastroianni et al., 2008). These findings suggest that a similar requirement for additional Mg^{2+} for maximal group II intron RNP integration may be a common feature of eukaryotic organisms. Precise group II intron integration at the target site was confirmed by PCR and sequencing of both the 5'- and 3'-integration junctions in five Tet^R + Amp^R colonies. The hatch rate after a single injection of 100 mM $MgCl_2$, a Mg^{2+} concentration that supports chromosomal gene targeting (see Chapter 4), was 69%. However, a double injection of 100 mM $MgCl_2$ and RNP buffer decreased the hatch rate to 31%.

As a control, I carried out experiments mimicking the *D. melanogaster* microinjection assays above in the absence of embryos . Here, I microinjected target plasmid and 17 mM dNTPs in solution with 100 mM $MgCl_2$ into halocarbon oil on a slide followed by microinjection of intron RNPs into these droplets of target plasmid solution. Incubation of the slide and nucleic acid extraction was done as with embryos and recovered plasmid was used to transform *E. coli* HMS174(DE3). Using these conditions, no Tet^R + Amp^R colonies are observed, indicating that factors in the embryo are essential for retrohoming at plasmid sites to form a product that yields Tet^R + Amp^R colonies upon transformation.

2.2.2 Retrohoming of RNPs containing linear intron RNA

Group II intron RNAs can also splice without lariat formation by an alternative pathway, termed "hydrolytic splicing" (Michel & Ferat, 1995). In this pathway, the first transesterification, 5'-splice site cleavage, occurs by hydrolysis rather than branching,

while the second transesterification again yields ligated exons but now accompanied by the release of linear intron RNA. For both linear and lariat intron RNA, DNA cleavage occurs by a transesterification reaction with the attachment of the intron RNA to the DNA target strand and not by hydrolysis, a mechanism used for group II intron ribozyme cleavage of RNA substrates (Michel & Ferat, 1995, Fedorova et al., 2002). However, linear group II intron RNAs cannot carry out both steps of reverse splicing and were thus thought to be immobile. They can, however, carry out the first step of reverse splicing into either RNA or DNA sites, thereby ligating their 3' end to the 5' end of the 3' exon (Mörl et al., 1992).

One advantage of using the linear intron is ease of preparation, since it can be obtained directly by *in vitro* transcription without the self-splicing step needed for the preparation of lariat RNA. Biochemical assays showed that Ll.LtrB RNPs containing linear intron RNA could efficiently carry out the first step of reverse splicing into a DNA target site, after which the IEP uses its En domain to cleave the bottom strand and reverse transcribes the attached linear intron RNA (Mastroianni et al., 2008). Linear intron RNPs microinjected into *Xenopus laevis* oocyte nuclei were shown to use these reactions *in vivo* to introduce a site-specific double-strand break that stimulates gene targeting by homologous recombination with a co-injected DNA (Mastroianni et al., 2008). In addition to demonstrating the potential utility of group II introns for gene targeting by double-strand-break-stimulated homologous recombination, these experiments raised the possibility that linear group II intron RNAs might be able to retrohome *in vivo* by a mechanism involving partial reverse splicing into the DNA target site followed by reverse transcription of the attached RNA, provided the resulting intron cDNA could be integrated into the target DNA by DNA repair (Mastroianni et al., 2008).

I used the plasmid-based retrohoming microinjection assay to compare the retrohoming efficiencies of Ll.LtrB RNPs containing lariat or linear intron RNA in *D. melanogaster* embryos. Two forms of the linear intron RNA, one whose 5' end corresponds precisely to that of the Ll.LtrB intron and the other with two extra 5' G residues (GG-Lin), enabling more efficient transcription by phage T3 RNA polymerase (Bailey et al., 1983), were tested. These extra 5' residues are not expected to affect the ability of the GG-Lin RNA to carry out the first step of reverse splicing (Mörl & Schmelzer, 1990), nor to serve as a 5' exon for RNA splicing to produce lariat RNA. The latter was confirmed by *in vitro* splicing assays (Jacquier & Rosbash, 1986, Zhuang et al., 2009).

Because group II intron integration into double-stranded DNA is more efficient at higher temperatures (Zimmerly et al., 1995b, Singh & Lambowitz, 2001), the injected *D. melanogaster* embryos were incubated at 37°C for a short time (30 min) prior to analysis of the retrohoming products, as done previously for chromosomal gene targeting (Mastroianni et al., 2008). Initial experiments in which retrohoming products were recovered from 80 injected embryos by transforming extracted nucleic acids into *E. coli* showed that the retrohoming efficiency measured by the ratio of $(Tet^R + Amp^R)/Amp^R$ colonies was much lower for GG-Lin RNPs than for lariat RNPs (0.002 and 5%, respectively; 3.4 mg/ml RNPs injected in both cases).

To confirm retrohoming of the linear intron RNA, I carried out PCR for the 5'- and 3'-intron insertion junctions of randomly selected $Tet^R + Amp^R$ colonies. For lariat RNPs, sequencing of the PCR products for 24 colonies all showed the correct 5'- and 3'-junction sequences. For GG-Lin RNPs, analysis by colony PCR showed the correct 3'-junction sequence for all 24 colonies, as expected for accurate completion of the first step of reverse splicing into the DNA target site, subsequent endonucleolytic cleavage and

reverse transcription through DV of the intron, which includes the forward primer binding site for the 3'-junction PCR.

By contrast, colony PCR of the 5' junction gave products of the size expected for insertion of the full-length intron, along with smaller products (Fig 2.3 A). Sequencing these products revealed the majority of events analyzed for GG-Lin RNPs gave insertion of full-length intron with an accurate 5'-junction sequence (Figure 2.3 C). The remainder had 5'-junctions with some combination of: (i) resection of 5'-exon sequences; (ii) insertion of 5'-truncated introns; and/or (iii) insertion of extra nucleotide residues at the integration junction (Figure 2.3 C). These anomalies presumably reflect errors during the DNA repair process used for ligation of the intron cDNA to the 5' exon. Some of the junction sequences, particularly those with extensive 5'-intron truncations, showed microhomologies of 1-4 nucleotide residues between the 5' end of the intron and the ligated 5'-exon sequences (nucleotides in parentheses in Figure 2.3 C). The characteristics of the 5' junctions for retrohoming of GG-Lin RNPs are similar to those of ligation junctions formed during DNA repair by NHEJ (Hagmann et al., 1998, Ducau et al., 2000, Adams et al., 2003, Lieber et al., 2003, McVey et al., 2004, Weterings & van Gent, 2004). Similar results were described for retrohoming of the linear intron RNP in *X. laevis* oocyte nuclei (Zhuang et al., 2009).

To more fully analyze the spectrum of retrohoming products for linear RNPs and confirm that cDNA ligation occurred in the eukaryotic host, I carried out PCR of 5'- and 3'-integration junctions in the nucleic acids extracted from *D. melanogaster* embryos without transformation into *E. coli* (Figure 2.3 B and D). As expected from the colony PCR results, the 5'-junction PCR for GG-Lin RNPs showed some products of the same size as those for lariat RNPs, along with smaller products (Figure 2.3 B, left panel). Sequencing of cloned PCR products confirmed that the larger products (band a) had

insertions of full-length or almost full-length introns, while smaller PCR products (band b) had insertions with progressively longer 5'-intron truncations. Some of the 5'-junction sequences also showed variably sized 5'-exon resections, insertion of extra nucleotide residues, and/or repair at microhomologies, similar to the junction sequences analyzed by colony PCR. Some insertions were apparently templated from nearby 5'-exon or intron sequence (underlined nucleotides in Fig 2.3 D). These types of insertions have been described previously as a characteristic of NHEJ-mediated DNA DSB repair in *D. melanogaster* (Adams et al., 2003, McVey et al., 2004).

Although the transcription of GG-Lin RNA was confirmed by single-nucleotide omission experiments to initiate preferentially at the upstream G-residues (Zhuang et al., 2009), only a few PCR products retained one or both of the extra 5' G-residues at the ligation junction, possibly reflecting their preferential removal or exclusion at some step during retrohoming. One possibility is that because these extra nucleotide residues lie outside the folded RNA structure in RNPs, they are prone to removal by RNase trimming.

By contrast to the 5' junction, the 3'-junction PCR for retrohoming products of GG-Lin RNPs gave a single strong band of the same size as that for lariat RNPs (Figure 2.3 B, right panel), and sequencing showed the expected junction corresponding to the 3' end of the intron inserted precisely at the target site by partial reverse splicing.

2.2.3 Temperature dependence of linear and lariat RNP retrohoming

The linear intron retrohoming assays described above were done by incubation of the microinjected *D. melanogaster* embryos at 37°C for 30 min. While these conditions are optimal for RNP activity, flies, which are typically reared at 25°C, are known to undergo a heat-shock response at these temperatures (Tissières et al., 1974). To compare

the results obtained above at a temperature that facilitates normal fly development, retrohoming assays were done at 25°C. Analysis of intron integration products by 5'- and 3'-junction PCR on extracted nucleic acids show similar results for lariat and GG-Lin as observed for retrohoming assays done at 37°C. With the exception of a smaller proportion of completely reverse-transcribed introns, sequence analysis of randomly selected Tet^R + Amp^R colonies as well as cloned 5'-junction PCR products from experiments with GG-Lin RNPs indicated no major differences from the 5' junctions formed during retrohoming at 37°C (Figures 2.4 and 2.5).

Since the overall sequence composition of 5'-intron insertion junctions with GG-Lin RNPs was found to be similar for reactions done at 37 and 25°C, I was tested what effect temperature has on the retrohoming efficiency of the linear intron RNP. For these experiments the linear intron RNPs without the two G residues at the 5' end were used. Table 2.2 summarizes the results of two experiments in which retrohoming assays with linear and lariat RNPs were tested in embryos that were incubated for 1 h at 37, 30 or 25°C. As expected, the retrohoming efficiency of lariat RNPs was reduced roughly 3-fold in embryos incubated at 25°C compared with 30 or 37°C. Surprisingly, temperature had an inverse effect on the retrohoming efficiency of linear intron RNPs, whose retrohoming efficiency was highest at 25°C in these assays. On the average, the retrohoming efficiencies of linear intron RNPs decreased more than seven-fold at 30°C compared with 25°C and an additional six-fold at 37°C compared with 30°C.

Together, the above findings suggest that retrohoming of linear intron RNA in *D. melanogaster* embryos occurs by a mechanism depicted in Figure 2.6 B in comparison with that of lariat intron RNA (Figure 2.6 A). First, the linear intron RNA carries out the first step of reverse splicing, precisely ligating its 3' end to the 5' end of the 3' exon. As for retrohoming of lariat RNA, the IEP then cleaves the bottom strand and synthesizes an

intron cDNA, but the intron cDNA now has a free end, which must be linked to the 5'-exon DNA. Sequencing of 5'-integration junctions showed that this step occurs by an error-prone process with the characteristics of non-homologous end joining (NHEJ).

2.3 RETROHOMING IN NHEJ MUTANT EMBRYOS

To preserve genetic information and the viability of eukaryotic cells, mechanisms have evolved to repair chromosomes that have undergone DNA damage. When the cell encounters a DSB, repair can proceed via NHEJ or homologous recombination (HR) (Sonoda et al., 2006, Wyman & Kanaar, 2006). Preferential use of one or the other pathway is dependent on the organism, cell type and stage in the cell cycle. In *D. melanogaster*, both types of DNA repair are active and have been described in the early embryo (Gorski et al., 2003). The process of NHEJ involves recruitment of the Ku70-Ku80 heterodimer to the site of DNA DSB where it binds DNA ends and protects them from further degradation. Subsequently, the catalytic subunit of the DNA-dependent protein kinase complex (DNA-PKcs) and DNA ligase IV (Lig4) are recruited to the repair site (Sonoda et al., 2006, Wyman & Kanaar, 2006).

The requirement of Lig4 for NHEJ in *D. melanogaster* has been investigated by analyzing DSB repair in *lig4* null mutants at the embryonic, larval and adult stages. There is a definite role for Lig4 protein that is maternally deposited in the embryo in repair of ionizing radiation (IR) induced DSBs for embryos 0-24 h of age (Gorski et al., 2003). Because Lig4 is required only for the NHEJ process and not HR, this Lig4-dependent defect is attributed specifically to loss of NHEJ repair of DSBs caused by IR. However, *lig4*$^{-}$ third-instar larvae (approximately 82 h old) are not compromised in their ability to repair DSBs caused by IR (Gorski et al., 2003). Intriguingly, an assay for NHEJ repair of a DSB caused by excision of a P-element transposon in the premeiotic germline

cells of adult males displayed no reduction of NHEJ events with the loss of functional Lig4 (McVey et al., 2004). These data clearly point to an alternative DNA end-joining (alt-EJ) pathway that is Lig4 independent and may be attributed to the fact that P-element excision generates 3'-non-complementary overhangs that are not efficient substrates for Lig4-dependent, canonical NHEJ (c-NHEJ). Studies of DSB repair induced by the I-*Sce*I meganuclease at chromosomal locations in *D. melanogaster* germline cells show up to a 4-fold decrease in NHEJ in Lig4 mutant flies (Johnson-Schlitz et al., 2007).

The gene encoding DNA polymerase θ (PolQ) in *Drosophila* is *mus308* (mutagen sensitive 308) and was identified in a genetic screen for mutants exhibiting sensitivity to DNA interstrand crosslinking agents, but normal resistance to alkylating agents (Boyd et al., 1990). This repair polymerase has been implicated in end-joining repair of DSBs produced by the P-element excision at chromosomal locations in *D. melanogaster* germline cells (Chan et al., 2010). Mutant flies lacking PolQ have been shown to have a decrease in the incidence and size of insertions found at DSB repair junctions generated by the I-*Sce*I meganuclease at chromosomal locations in *D. melanogaster* germline cells (Yu & McVey, 2010).

Studies of the retrohoming of linear Ll.LtrB intron RNPs in *D. melanogaster* embryos, as well as in *Xenopus laevis* oocyte nuclei, have led to a model of retrohoming outlined in Figure 2.6 which includes partial reverse splicing of the intron RNA into the DNA target site, reverse transcription of the attached intron RNA, and ligation of the free end of the intron cDNA to the upstream exon by a mechanism akin to NHEJ (Zhuang et al., 2009).

To investigate the involvement of NHEJ activities in linear group II intron retrohoming, I assayed retrohoming in *D. melanogaster* embryos with null mutations in the genes encoding the c-NHEJ factors Lig4 and Ku70 (Materials and Methods section

2.5.6). In these experiments, retrohoming efficiencies of linear and lariat Ll.LtrB RNPs were determined by using the plasmid-based bacterial genetic assay described above. Table 2.3 and Figure 2.7 summarize the results of nine experiments, with the retrohoming efficiency for each strain in each experiment determined by pooling 80 injected embryos prior to nucleic acid extraction and transformation of *E. coli*. The retrohoming efficiency of lariat RNPs was undiminished in *lig4* deletion mutant embryos, whereas the retrohoming efficiency of linear RNPs was strongly decreased (18% wild type), but restored completely by ectopic expression of Lig4 in a P-element transformant in the *lig4*$^-$ background (*lig4*$^-$; *P{lig4}* embryos). In the *ku70* deletion mutant embryos, the retrohoming efficiencies of both lariat and linear RNPs were moderately decreased, although the decrease appeared to be somewhat greater for linear than for lariat RNPs (46% and 67% wild type), respectively. These findings show that Lig4 is required specifically for linear RNP retrohoming and raise the possibility that Ku70 plays some role in the retrohoming of both lariat and linear RNPs (see Discussion section 2.4).

In the *polQ* null mutant embryos, the retrohoming efficiencies of both lariat and linear RNPs were drastically decreased, and the decrease was greater for linear than for lariat RNPs (<3% and 27% wild type). No Tet^R + Amp^R colonies were recovered from experiments with the linear intron RNP in *polQ*$^-$ embryos, so the values given are the upper limit of retrohoming efficiencies. However, 5'-intron integration junctions could be detected in *polQ*$^-$ embryos by the more sensitive PCR assay and are described below.

As described previously in experiments with GG-lin RNPs, analysis of 5'-linear intron integration sequences from retrohoming assays in wild-type as well as *lig4*$^-$, *ku70*$^-$, and *lig4*$^-$; *P{lig4}* embryos revealed the presence of small insertions, deletions, repair at regions of microhomology, as well as intron truncations (Figures 2.8 and 2.9). These features were also observed in 5'-retrohoming junctions obtained from raandomly

selected Tet^R + Amp^R colonies (Figure 2.10) These characteristics are similar to those of junctions formed during double-strand break repair by NHEJ in *D. melanogaster* (Hagmann et al., 1998, Ducau et al., 2000, Adams et al., 2003, McVey et al., 2004). Notably, I could discern no significant differences in 5'-junction characteristics between wild-type embryos and any of the mutant or transformant stocks. Both Lig4 and Ku70 have been shown to be required for efficient repair by NHEJ, comprising components of the canonical NHEJ (c-NHEJ) pathway in *D. melanogaster*, although alternative end-joining (alt-EJ) pathways independent of either factor have been described (Ma et al., 2003, Wang et al., 2003). Overall NHEJ is decreased by 76% and 79% in *lig4*$^-$ and *ku70*$^-$ flies, respectively, in a genetic assay of I*Sce*I-induced DSB repair (Johnson-Schlitz et al., 2007). However, no difference between wild-type and *lig4*$^-$ or *ku70*$^-$ flies has been found during analysis of repair junctions (McVey et al., 2004, Yu & McVey, 2010). Thus, both the similar junction characteristics and the degree of inhibition in the *lig4*$^-$ and *ku70*$^-$ embryos are consistent with the use of both *D. melanogaster* NHEJ pathways for linear intron RNA retrohoming. A minority of 5' junctions for *lig4*$^-$, *ku70*$^-$, and *lig4*$^-$; *P{lig4}* embryos tested show templated insertion of 5'-exon and intron sequences (underlined nucleotides in Figures 2.8 B-C, 2.9 A-B, and 2.10 B), as found for NHEJ in many model systems (Hagmann et al., 1998, Ducau et al., 2000, Jager et al., 2000, Adams et al., 2003, McVey et al., 2004, Preston et al., 2006). The extent of intron cDNA truncation is more dramatic in *lig4*$^-$; *P{lig4}* embryos (an average of 491 nt truncation) as compared with wild-type embryos (average 203 nt), *lig4*$^-$ (average 304 nt), *ku70*$^-$ (average 174 nt), and *polQ*$^-$ (average 331 nt) mutant embryos, but the proportion of full-length and shorter 5'-junction products was variable for each stock in different experiments (Figures 2.3-2.5 and 2.8-2.11). Together, these results suggest that the retrohoming of linear Ll.LtrB RNPs occurs by both the Lig4-dependent and Lig4-independent NHEJ pathways, which

have been observed previously in *D. melanogaster* embryos (McVey et al., 2004, Yu & McVey, 2010).

In contrast with *lig4*$^{-}$ and *ku70*$^{-}$ embryos, analysis of 5'-linear intron integration sequences from retrohoming assays in wild-type and *polQ*$^{-}$ embryos from the same experiments (Table 2.3 experiments 8 and 9) showed a striking decrease in addition of extra nucleotides in *polQ* mutants compared with wild type (2/46 for PolQ mutant compared to 8/10 for wild type) (Figure 2.11). These results indicate PolQ is likely the repair polymerase responsible for the majority of inserted nucleotides found at the 5'-linear intron retrohoming junction in embryos.

2.4 DISCUSSION

2.4.1 Lariat intron retrohoming in embryos

Here, by using group II intron RNP microinjection assays, I show that group II intron-based gene targeting reactions can occur efficiently in *D. melanogaster* precellular blastoderm stage embryos, but are dependent upon the injection of additional Mg^{2+}. The finding that Ll.LtrB intron RNP activity is limited by Mg^{2+} concentration in eukaryotes differs from the situation in bacteria, the natural hosts of group II introns, where free Mg^{2+} concentrations are apparently sufficient to support efficient mobility of Ll.LtrB and other group II introns. Group II introns seem to be particularly sensitive to low Mg^{2+} concentrations. Studies of yeast mtDNA group II introns showed that mutations in Mg^{2+} transporters that result in decreased intramitochondrial Mg^{2+} concentrations strongly inhibit the splicing of all four group II introns, which belong to different structural subclasses, while having relatively little effect on the splicing of group I introns (Wiesenberger et al., 1992, Gregan et al., 2001). The group II introns affected by low Mg^{2+} concentrations in yeast mitochondria use different protein co-factors to promote

RNA splicing (Gregan et al., 2001), suggesting that the high Mg^{2+} requirement for group II intron function is a property of the intron RNA.

Although group II intron RNPs can by themselves initiate site-specific DNA integration, host enzymes are required to complete the process (see Chapter 1). An important finding here is that *D. melanogaster* embryos appear to have the enzymatic machinery necessary to complete these reactions. The late steps that require host enzymes likely include degradation of the intron RNA template by the endogenous RNase H, followed by second-strand DNA synthesis and DNA ligation, which require DNA polymerase and DNA ligase activities. Previous studies of group II intron retrohoming showed that cDNA integration could occur by different DNA recombination or repair mechanisms in different organisms (Cousineau et al., 1998, Eskes et al., 2000, Smith et al., 2005). These findings suggest that the availability of host enzymes to complete group II intron gene targeting reactions is not a limitation in *D. melanogaster*.

2.4.2 Linear intron retrohoming in embryos

In addition to their practical applications, my findings for RNPs containing linear intron RNA also have implications for group II intron mobility mechanisms. Some group II introns splice *in vivo* by using hydrolysis rather than branch-point formation for cleavage at the 5'-splice site, leading to the production of linear intron RNA (Podar et al., 1998, Granlund et al., 2001, Vogel & Borner, 2002). These findings suggest that such group II introns could be mobile by a mechanism involving partial reverse splicing into the target DNA, leading to attachment of linear intron RNA to the 5' end of the 3' exon, followed by second-strand cleavage and target DNA-primed reverse transcription of the linear intron RNA, provided that cellular enzymes could complete ligation of the cDNA into the target site.

My results indicate that RNPs containing linear group II intron RNA can retrohome *in vivo* by the mechanism proposed in Figure 2.6. The pathway begins with the linear intron RNA catalyzing the first step of reverse splicing into a DNA target site, resulting in the ligation of the 3' end of the intron to the 5' end of the 3'-exon DNA. The associated IEP then uses its En domain to cleave the opposite strand and synthesizes a cDNA copy of the linear intron RNA. These initial reactions were demonstrated previously for GG-Lin RNPs *in vitro* (Mastroianni et al., 2008), and their occurrence *in vivo* is indicated here by the expected 3'-junction sequences for retrohoming of Lin and GG-Lin RNPs.

A key step found here to complete retrohoming *in vivo* is the ligation of the free end of the linear intron cDNA to the upstream exon DNA. This step occasionally leads to the precise insertion of the full-length intron, just as for retrohoming of lariat RNA, but more frequently occurs with loss of 5'-exon or 5'-intron sequences and insertion of additional nucleotide residues at the junction. These characteristics suggest that the ligation step occurs by a mechanism akin to NHEJ. An alternative possibility that the LtrA protein, upon completion of cDNA synthesis of the integrated linear intron, generates the 5' junctions described here through non-templated nucleotide addition and/or template switching to the 5' exon is a topic of investigation in Chapter 3.

As I observed no duplication of target site sequences resulting from the initial staggered double-strand break made by the RNPs, I infer that the single-stranded 5' overhang attached to the 5'-exon DNA bottom strand is resected prior to cDNA ligation. Like retrohoming of lariat RNA, the retrohoming of linear intron RNA is presumably completed by the degradation or displacement of the intron RNA template strand followed by second-strand DNA synthesis and the sealing of nicks by host DNA polymerases and ligases (Smith et al., 2005).

In *D. melanogaster* embryos, the insertion of 5'-truncated introns most likely results either from degradation of the linear intron RNA or resection of unattached cDNA ends, rather than abortive cDNA synthesis, as such 5' truncations are not observed for retrohoming of lariat RNA. The latter is identical to the linear intron RNA, except that its 5' end is protected by the 2'-5' linkage and then ligated directly to the 5' exon by the second-step of reverse splicing. As a result, DNA synthesis can extend continuously into the upstream exon without leaving an unattached cDNA end, as is the case for linear intron RNA. Additionally, most of the truncated introns (67%) have 5' ends at positions consistent with termination of cDNAs synthesis within single-stranded RNA regions, which may be prone to RNase cleavage (Figure 2.12).

Linear and lariat retrohoming assays done in parallel indicate that linear intron RNPs are most efficient at retrohoming in *D. melanogaster* embryos at 25°C and efficiency decreases dramatically with increasing temperatures, in contrast to the trend seen for the retrohoming efficiency of lariat intron RNPs. The free 5' end of the reverse-spliced linear intron may allow exonucleolytic degradation within the *D. melanogaster* embryo, generating the 5'-intron truncations observed in the retrohoming products. Furthermore, this degradation may accelerate at temperatures above 25°C, due to dissociation of LtrA at the intron integration site or heat-induced stress response in the embryo, resulting in an overall decrease in retrohoming efficiency at 30°C or 37°C. It is also possible that the linear intron RNP is more susceptible than the lariat intron RNP to exonucleolytic degradation at elevated temperatures prior to contact with the target plasmid within the embryo.

Previous biochemical assays showed that linear and lariat Ll.LtrB RNAs can carry out the first step of reverse splicing into DNA target sites with similar efficiency *in vitro* (Mastroianni et al., 2008), and this step also appears to be efficient for linear intron RNA

in vivo, as judged by PCR analysis of 3'-integration junctions (Figure 2.3 B). Thus, the lower retrohoming efficiency of linear intron RNA is likely due mainly to its greater nuclease sensitivity and/or low NHEJ activity in the *D. melanogaster* embryo. If so, the retrohoming efficiency of linear intron RNA may be increased by RNA modifications that confer nuclease resistance and could be higher in other hosts with higher NHEJ activity. I also note that the retrohoming efficiency of the linear intron RNA is likely underestimated by this genetic assay, which requires integration of sufficiently long cDNA to introduce the T7 promoter upstream of the promoterless tet^R gene (Figure 2.1).

NHEJ is ubiquitous in eukaryotes and a related NHEJ pathway with a Ku homolog and ATP-dependent DNA ligase exists in many prokaryotes, albeit not in *E. coli* K12 (Bowater & Doherty, 2006). Thus, the mechanism elucidated here could be used in both prokaryotes and eukaryotes for the retrohoming of linear group II intron RNAs generated by hydrolytic splicing or by the action of host debranching enzymes on lariat intron RNAs. Hydrolytic splicing is a side pathway for group II introns that splice via branching but is the exclusive splicing pathway for group II introns that cannot branch (see Introduction). It is possible that ancestral group II introns spliced exclusively via the hydrolytic pathway and were mobile by the type of mechanism elucidated here, while intron branching was a later evolutionary adaptation that ultimately predominated because it increased the efficiency and fidelity of retromobility.

2.4.3 Linear and lariat intron retrohoming in NHEJ mutant embryos

Here I report genetic evidence that group II linear intron retrohoming products are subject, in part, to repair via the canonical NHEJ (c-NHEJ) pathway. Specific reduction of linear intron retrohoming efficiency in the Lig4 null embryos indicates Lig4 is involved in resolution of the retrohoming product for the linear intron but not the lariat.

Presumably, Lig4 is responsible for ligation of the reverse transcribed intron following NHEJ-mediated repair at the majority of 5' junctions formed during linear intron integration. Furthermore, the data shows that the pathway involved in ligation of the reverse transcribed intron differs for lariat and linear forms of introns, since loss of Lig4 has no effect in lariat intron retrohoming efficiency.

The remaining retrohoming observed in Lig4 null mutant embryos is likely the result of Lig4-independent NHEJ event or so called "backup" or "alternative" end-joining (alt-EJ) (Ma et al., 2003, Wang et al., 2003, Yu & McVey, 2010). My experiments show a Ku70-dependent reduction in retrohoming efficiency of not only the linear, but also the lariat form of the intron. Although linear intron retrohoming is impaired to a greater extent than the lariat intron, it is possible that the Ku70/Ku80 heterodimer plays a general role in resolution of the DNA aberration created by reverse-splicing of the intron RNP at the target site. The Ku complex has known functions in maintenance of chromosome ends in *D. melanogaster*, and loss of Ku70 impairs the telomere-elongation mechanism governed by the HeT-A and TART non-LTR retrotransposons (Melnikova et al., 2005). Further, the Ku70 protein has been shown to interact with the RNA component of yeast telomerase, TLC1, *in vitro* (Peterson et al., 2001, Stellwagen et al., 2003). Thus, Ku70 may have specific functions in recognition of an attached RNA at a DNA site and serve to protect it from degradation or to recruit repair factors. Such an activity would be expected to affect both lariat and linear intron retrohoming.

Interestingly, retrohoming by the linear intron is not impaired to the same extent in the Ku70 mutant background as with the Lig4 mutant background. Characterizations of *lig4*$^{-}$ and *ku70*$^{-}$ DT40 cells showed that Lig4 acts exclusively in Ku70-dependent NHEJ (Adachi et al., 2001). In retrohoming of linear group II intron RNA, however, the group II intron RNP could itself contribute to such tethering, as it binds to both 5'- and

3'-exon sequences for DNA target site recognition (Singh & Lambowitz, 2001, Perutka et al., 2004), and the RNP contacts with both exons appear to be maintained at least through the initiation of cDNA synthesis (Noah et al., 2006). A bridging role for the RNPs could explain why microhomologies between the ligated ends are infrequent for long intron cDNAs, which extend up to the 5' end of the intron, but more frequent for truncated cDNAs, whose free ends are farther from the eventual site of ligation.

I found retrohoming of both linear and lariat RNPs was decreased in *polQ*$^-$ embryos compared with wild-type embryos. While it is unclear how PolQ affects the retrohoming of lariat RNPs, it is evident from sequence analysis of linear intron 5'-retrohoming junctions that PolQ contributes to the insertion of extra nucleotide residues. Whereas over half of the 5'-junction sequences in wild-type embryos have extra nucleotide residues, only 2 of 46 junctions analyzed from *polQ*$^-$ embryos have insertions. These inserted sequences show no apparent difference from those found from wild-type embryos. These results indicate PolQ may contribute to efficient end-joining repair by adding templated or non-templated nucleotides at the DSB formed upon linear intron retrohoming. These inserted sequences could then base-pair with the resected 5' exon, stabilizing the two ends of the DNA molecule at the DSB to be ligated efficiently. In most of the 5' junctions analyzed from *polQ*$^-$ embryos, deletion of 5'-exon sequences of at least 10 bp and repair at microhomologies of 4 nt or more is seen, suggesting annealing at microhomologies between the intron cDNA and the 5' exon is a requirement for efficient repair in the absence of insertion at the DSB.

Finally, my results suggest that RNPs containing linear group II intron RNAs could be used for site-specific insertion in gene targeting. Targeted integration of linear Ll.LtrB intron RNA into the *D. melanogaster yellow* gene (Mastroianni et al., 2008), likely occurred by the mechanism elucidated here (see Chapter 4). The lower

retrohoming efficiency of linear intron RNA might be offset both by its greater ease of preparation, which does not require the self-splicing and purification steps needed for lariat RNA, and by the facility of incorporating selectable or screenable genetic markers, which can substantially decrease the splicing and reverse splicing efficiency of lariat RNA. *In vitro* preparation of randomly targeted RNPs via introduction of random EBS sequences in the RNA template requires use of the linear intron, since efficient splicing of such a library is not possible due to the low probability of complementary IBS sequence presence in the donor RNA. The use of such randomly targeted linear intron RNPs with a GFP marker cassette was recently used for eukaryotic transgenesis in decondensed *X. laevis* sperm nuclei and introduced its potential application as a general forward genetic tool for screening of disrupted protein coding sequences within eukaryotic cells.

2.5 Materials and Methods

2.5.1 Recombinant plasmids

pACD2 is the intron-donor plasmid used for *in vitro* transcription of wild-type Ll.LtrB-ΔORF precursor RNA, which is then self-spliced and reconstituted with purified LtrA protein to form RNPs for the plasmid targeting reactions investigating Mg^{2+}-dependence described in section 2.1.1 (Guo et al., 2000, Karberg et al., 2001). This plasmid contains the 0.9-kb Ll.LtrB-ΔORF intron and flanking exons cloned downstream of a T7lac promoter in a pACYC184-based vector with a cam^R gene, and has the LtrA ORF cloned downstream of the 3' exon. The Ll.LtrB-ΔORF intron contains an additional phage T7 promoter inserted in intron domain IV (DIV) for use in plasmid-based DNA-integration assays. Ll.LtrB-ΔORF intron RNAs used in all other plasmid targeting experiments were transcribed from DNA templates generated by PCR of plasmid

pACD5C, a derivative of pACD4C (Perutka et al., 2004) with a T7 promoter sequence inserted at the SalI site in DIV.

pBRR3-ltrB, the target plasmid for intron-integration assays, contains the Ll.LtrB homing site (ligated exon 1 and 2 of the *ltrB* gene from positions -178 upstream to +91 downstream of the intron-insertion site) cloned upstream of a promoterless tet^R gene in an Amp^R pBR322-based vector (Guo et al., 2000, Mohr et al., 2000, Karberg et al., 2001).

pIMP-1P, used for expression of the LtrA protein, contains the LtrA ORF cloned downstream of a tac promoter and Φ10 Shine-Dalgarno sequence in the protein-expression vector pCYB2 (New England Biolabs, Ipswich, MA) (Saldanha et al., 1999). LtrA is expressed from this plasmid as a fusion protein with a C-terminal tag containing an intein-linked chitin-binding domain, enabling LtrA purification via a chitin-affinity column, followed by intein-cleavage.

2.5.2 Preparation of Ll.LtrB lariat and linear RNAs

For plasmid targeting reactions investigating Mg^{2+}-dependence described in section 2.1.1, Ll.LtrB-ΔORF precursor RNA was transcribed with phage T7 RNA polymerase (Megascript T7 Kit; Ambion, Austin, TX) from pACD2 which had been linearized with NheI. The resulting precursor RNAs containing the Ll.LtrB-ΔORF intron and flanking exon sequences were self-spliced in 1.25 M NH_4Cl, 50 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5 for 3 h at 37˚C, then ethanol-precipitated and dissolved in distilled water.

For all other plasmid targeting experiments, Ll.LtrB-ΔORF intron RNAs were transcribed from DNA templates generated by PCR of plasmid pACD5C, a derivative of pACD4C (Perutka et al., 2004) with a T7 promoter sequence inserted at the SalI site in DIV. DNA oligonucleotides were obtained from Integrated DNA Technologies (IDT;

Coralville, IA). For the lariat precursor RNA, the PCR primers were pACD-T3 (5'-GGAGTCTAGAAATTAACCCTCACTAAAGGGAATTGTGAGCG) and NheIR (5'-CTAGCAGCACGCCATAGTGACTGGCG), and for linear intron RNA, the PCR primers were T3LIS-1G (5'-AATTAACCCTCACTAAAGTGCGCCCAGATAGGGTGTTAAGTCAAG) and HPLC-purified LtrB940a (5'-GTGAAGTAGGGAGGTACCGCCTTGTTC) (Mastroianni et al., 2008). In both cases, the upstream primer appends a T3 promoter sequence (underlined). The DNA template for GG-Lin RNA were made by PCR of pACD5C with T3LIS-3G (5'-AATTAACCCTCACTAAAGGGTGCGCCCAGATAGGGTGTTAAGTCAAG) and HPLC-purified LtrB940a. In these cases, the upstream primer appends a T3 promoter sequence (underlined). The PCR products were purified by using the Wizard SV Gel and PCR Clean-up System (Promega), extracted with phenol-chloroform-isoamyl alcohol (phenol-CIA; 25:24:1 by volume), ethanol precipitated, and dissolved in nuclease-free water. *In vitro* transcription was done with phage T3 RNA polymerase (Megascript T3 Kit; Ambion, Austin, TX). For preparation of lariat Ll.LtrB RNA, the precursor RNAs were self-spliced as described above.

2.5.3 Preparation of LtrA protein and RNPs

The LtrA protein used for RNP reconstitution was expressed in *E. coli* BL21(DE3) from the intein-based expression vector pImp-1P and purified via a chitin-affinity column and intein cleavage, as described (Saldanha et al., 1999), except that the column buffer contained 50 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, and 0.1% NP-40.

To reconstitute Ll.LtrB-ΔORF RNPs, the self-spliced RNA (100 nM) was re-natured by heating to 50˚C in 10 ml of 450 mM NaCl, 5 mM $MgCl_2$, 40 mM Tris-HCl,

pH 7.5 and slow cooling to 30˚C prior to addition of 200 nM of purified LtrA protein and further incubation for 30 min at 30ºC. The resulting RNPs were pelleted by ultracentrifugation in Beckman 50.2 Ti rotor at 145,000 x g for 16 h at 4˚C and resuspended in 20 μl of 10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0 for the plasmid targeting reactions investigating Mg^{2+}-dependence described in section 2.1.1. RNP resuspension buffer used in all other plasmid targeting experiments was the same except that it contained 5 mM $MgCl_2$ instead of 10 mM $MgCl_2$. The RNP preparations typically contain 60-70% intron lariat RNA, with the remainder being precursor RNA plus smaller amounts of linear intron and ligated exons.

2.5.4 Retrohoming assays

D. melanogaster stocks were grown in standard fly media at 23˚C and embryos were collected on apple juice agar plates with yeast paste (Ashburner et al., 2005) for collection times not exceeding 40 min. The embryos were then bleach dechorionated, rinsed in egg wash [0.7% (w/v) NaCl, 0.04% (v/v) Triton X-100], adhered to a 25 x75 mm slide, desiccated in a dish that contains Drierite and covered in halocarbon oil (700 Series; Halocarbon Products, River Edge, NJ). Microinjection was done using filament needles with a bore size of ~3 μm attached to an air-filled 50-ml syringe. A micromanipulator (MO-150; Narishige) was used to manipulate the injection needles. The injection volumes were ~ 0.3 nl. Target plasmid (pBRR3-ltrB) at 1.4 mg/ml in solution with indicated $MgCl_2$,where indicated, and 17 mM dNTPs, followed within 5 min by approximately 300 pl of Ll.LtrB RNPs (3.4 mg/ml based on A_{260} in plasmid targeting reactions investigating Mg^{2+}-dependence described in section 2.1.1 and 2.6 mg/ml based on A_{260} in all other plasmid targeting reactions) containing lariat or linear intron RNA with a phage T7 promoter inserted in intron domain IV. The target plasmid

contains the Ll.LtrB intron target site (ligated exon 1 and 2 sequences of the *ltrB* gene; E1 and E2) cloned upstream of a promoterless *tet*^{R} gene in a pBR322-based vector carrying an Amp^{R} marker. Site-specific integration of the intron into the target site introduces the T7 promoter upstream of the promoterless *tet*^{R} gene, thereby activating that gene. One hundred injected embryos in plasmid targeting reactions investigating Mg^{2+}-dependence described in section 2.1.1 or eighty injected embryos in all other plasmid targeting reactions were collected for each condition. Embryos were incubated for 30 min or 1 h, where indicated, at 25, 30 or 37˚C, where indicated. Embryos were recovered by pipetting off the slide with a micropipette tip. The pooled embryos were incubated in lysis buffer [20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 400 mM NaCl, 1% SDS (w/v), 400 μg/ml proteinase K (Molecular Biology Grade; Sigma-Aldrich)] for 1 h at 55˚C, and then extracted with phenol-CIA. Nucleic acids were precipitated with ethanol and dissolved in 12 μl of water. Four μl of the nucleic acid preparation was electroporated into electrocompetent *E. coli* HMS174(DE3) F^{-}, *hsdR, recA, rif*^{r} (Novagen, EMD Chemicals, Gibbstown, NJ), which expresses T7 RNA polymerase. Cells were plated at different dilutions on LB containing ampicillin (50 μg/ml) plus tetracycline (25 μg/ml) or the same concentration of ampicillin alone. Colonies were counted after overnight incubation at 37˚C, and the integration efficiency was calculated as the ratio of (Amp^{R}+Tet^{R})/Amp^{R}.

2.5.5 PCR and sequencing of 5' and 3' junctions from retrohoming assays

Amp^{R}+Tet^{R} colonies and nucleic acids isolated from *D. melanogaster* embryos after retrohoming assays were used as a template to amplify the 5'-Ll.LtrB integration junction using primers P1 (5′-CTGATCGATAGCTGAAACGC) and P2 (5'-CGGCTCTGTTATTGTTCGTTCG) for plasmid targeting experiments done at 37˚C and

25˚C (Figures 2.3 and 2.4, respectively) or P1 (see above) and LtrB933a (5′-AGGGAGGTACCGCCTTGTTCACATTAC) for plasmid targeting experiments done at 30˚C (Figure 2.7). Amp^R+Tet^R colonies and nucleic acids isolated from *D. melanogaster* embryos after retrohoming assays were used as a template to amplify the 3'-Ll.LtrB integration junction using primers P3 (5'-CAGTGAATTTTTACGAACGAACAATAAC) and P4 (5'-AATGGACGATATCCCGCA). The colony PCR products were sequenced by using the primers Rseq (5'-CCATGCGAGAGTAGGGAAC) and P3 (see above) for the 5'- and 3'-integration junctions, respectively. 5'-junction PCR products from extracted nucleic acids from embryos were purified using a QIAquick Gel Extraction Kit (Qiagen) for plasmid targeting experiments done at 37˚C and 25˚C (Figures 2.3 and 2.4, respectively) or a MinElute PCR purification Kit (Qiagen) for plasmid targeting experiments done at 30˚C (Figure 2.7) and cloned into a TOPO TA cloning vector (pCR2.1-TOPO; Invitrogen) according to the manufacturer's protocol. Random colonies were picked and the cloned PCR products were amplified by colony PCR using Phusion High Fidelity PCR Master Mix with HF buffer with primers M13 F(-20) (5'- GTAAAACGACGGCCAGT) and M13 R(-26) (5'-CAGGAAACAGCTATGAC), then sequenced using the M13 Rev(-24) (5'-GGAAACAGCTATGACCATG) primer. When necessary, sequencing of the PCR product was also done with M13 F(-20). Apparently templated nucleotides present in insertions were underlined in figures if 4 or more nucleotides were found to have homology to proximal inserted sequences, intron cDNA or 5'-exon sequences within 15 nucleotides of the DSB interrupted by at most three mismatches.

2.5.6 *D. melanogaster* stocks

The *lig4*^{169} mutant fly stock was obtained from M. McVey (Tufts University, Medford, MA). This allele was generated by imprecise P-element excision that deleted 2,368 bp of the 2,942-bp gene, including the start codon and most of the coding region for the ATPase and adenylation domains (McVey et al., 2004).

The *ku70*$^{7B2.2}$ and *ku70*^{Ex8} mutants were obtained from William Engels (University of Wisconsin, Madison, WI). Both alleles were also generated by imprecise P-element excision. The *ku70*$^{7B2.2}$ allele deletes the 400 C-terminal amino acid residues of the 535 amino acid protein, including most of the known DNA and Ku80-interaction domains (Johnson-Schlitz et al., 2007). The *ku70*^{Ex8} allele deletes at least 1 kb including all of exon 1 and the start codon (Johnson-Schlitz et al., 2007). The *lig4*^{169} and *ku70*$^{7B2.2}$ mutant flies were genotyped by PCR and sequencing across the deletion junction. Microinjection assays were done with homozygous *lig4*^{169} and heterozygous *ku70*$^{7B2.2}$/*ku70*^{Ex8} female flies were used for microinjection assays. The latter were obtained by crossing yw; *ku70*$^{7B2.2}$ virgin females to yw; *ku70*^{Ex8}/TM6B males and the resulting virgin female yw; *ku70*$^{7B2.2}$/*ku70*^{Ex8} and male yw; *ku70*$^{7B2.2}$/*ku70*^{Ex8} progeny were crossed to provide embryos used in microinjection assays. The *ku70*$^{7B2.2}$/*ku70*^{Ex8} heterozygotes were used previously to study DNA DSB repair pathways (Johnson-Schlitz et al., 2007, Yu & McVey, 2010).

The *mus308*^{D2} stock was obtained from the Bloomington Stock Center (Bloomington, IA). This allele harbors an unknown mutation residing outside of the coding region which manifests in undetectable levels of PolQ protein expression (Chan et al., 2010).

Transgenic flies harboring the *lig4* rescue fragment were generated as follows: a 6-kb DNA fragment containing the *lig4* gene was amplified from wild-type (*w*^{1118}) flies using the Expand High Fidelity PCR System (Roche Applied Science, Indianapolis, IN)

with primers Lig4 F1 BamHI (5'-AAGAGGATCCAGTAGCTGTAGAAGCAGCCAAC) and Lig4 R1 XhoI 5'-AAGACTCGAGCAGCAGTTCCTCCGACATGAAG). This PCR product was inserted between BamHI and XhoI sites of the P-element transformation vector pCaSpeR4 (Thummel, 1992). Flies of the genotype yw; Ki delta-2-3 were transformed with the P-element *lig4* rescue construct (*P{lig4}*) by GenetiVision (Houston, TX) and transformants containing *P{lig4}* on chromosome 2 were isolated by standard fly methods. (Injected male flies were crossed to yw virgin females to score for P{lig4} transformation by w^+ progeny. w^+ male progeny were crossed to w; Sco/CyO; MKRS/TM6B virgin females and resulting w^+ male progeny of the genotype w; CyO; TM6B were crossed to yw virgin females. The progeny of this cross were used to ascertain the chromosome on which the P-element resides based on segregation of w^+ and CyO or TM6B. To combine the $lig4^{169}$ allele and the *P{lig4}* genomic rescue fragment on the second chromosome in the same fly stock, FM7; CyO males were crossed to $lig4^{169}$ homozygous virgin females to generate $lig4^{169}$/FM7; CyO female progeny. w; *P{lig4}*/CyO virgin females were crossed to FM7; CyO males to generate FM7; *P{lig4}*/CyO male progeny. $lig4^{169}$/FM7; CyO virgin females were crossed to FM7; *P{lig4}*/CyO males to generate $lig4^{169}$/FM7; *P{lig4}*/CyO virgin female progeny and $lig4^{169}$; *P{lig4}*/CyO male progeny which were crossed to each other. The resulting $lig4^{169}$; *P{lig4}* virgin female and male progeny were crossed to each other to generate the fly stock used in *P{lig4}* rescue experiments.

Condition	$Amp^R + Tet^R$	Amp^R	Retrohoming efficiency (%)
RNP	0	2.1×10^5	0
RNP + TP + 0.1 M $MgCl_2$ (- embryos)	0	2.9×10^5	0
RNP + TP + 0.1 M $MgCl_2$	1.3×10^2	9.3×10^4	0.14
RNP + TP + 0.25 $MgCl_2$	4.0×10^4	4.2×10^5	9.5
RNP + TP + 0.5 M $MgCl_2$	8.4×10^3	3.2×10^5	2.7

Table 2.1: Plasmid assays for site-specific group II intron integration in *D. melanogaster* embryos

Site-specific integration of group II intron RNPs into a plasmid target site in *D. melanogaster* embryos was assayed as described in Figs 2.1, 2.2 and Materials and Methods section 2.5.4. Target plasmid (TP) at 1.4 mg/ml with 17 mM of each dNTP, with or without the indicated concentrations of $MgCl_2$, and Ll.LtrB lariat RNPs were injected separately into halocarbon oil (- embryos) or the posterior of 100 precellular blastoderm stage embryos, which were then incubated for 1 h at 30ºC. After the incubation, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency.

Experiment	Intron	Temperature (°C)	Retrohoming efficiency (%)
1	Lariat	25	1.0
	Lariat	30	2.8
	Lariat	37	4.3
	Linear	25	0.01400
	Linear	30	0.00310
	Linear	37	0.00014
2	Lariat	25	0.95
	Lariat	30	4.2
	Lariat	37	3.7
	Linear	25	0.00310
	Linear	30	0.00014
	Linear	37	0.00004

Table 2.2: Temperature dependence on linear and lariat group II intron targeting in *D. melanogaster* embryos

Retrohoming assays using lariat and linear RNPs were done in *D. melanogaster* precellular blastoderm embryos, as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 1 h at the temperature indicated, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency.

Table 2.3

Experiment	Stock	Intron	Retrohoming efficiency (%)	Retrohoming efficiency relative to WT (%)
1	Or-R (WT)	Linear	0.11	
	lig4^{-}	Linear	0.047	43
	Or-R (WT)	Lariat	3.2	
	lig4^{-}	Lariat	3.7	116
2	Or-R (WT)	Linear	0.042	
	lig4^{-}	Linear	0.005	12
	Or-R (WT)	Lariat	2.4	
	lig4^{-}	Lariat	3.3	138
3	w^{1118} (WT)	Linear	0.081	
	ku70^{-}	Linear	0.052	64
	lig4^{-}	Linear	0.021	25
	w^{1118} (WT)	Lariat	2.3	
	ku70^{-}	Lariat	1.8	78
	lig4^{-}	Lariat	2.0	86
4	Or-R (WT)	Linear	0.14	
	ku70^{-}	Linear	0.048	34
	lig4^{-}	Linear	0.012	9
	Or-R (WT)	Lariat	3.5	
	ku70^{-}	Lariat	3.5	100
	lig4^{-}	Lariat	2.5	71
5	w^{1118} (WT)	Linear	0.023	
	ku70^{-}	Linear	0.018	78
	lig4^{-}	Linear	0.0002	9
	lig4^{-}; *P{lig4}*	Linear	0.034	148
	w^{1118} (WT)	Lariat	2.8	
	ku70^{-}	Lariat	1.4	50
	lig4^{-}	Lariat	3.7	132
	lig4^{-}; *P{lig4}*	Lariat	3.4	121

Table 2.3 continued

Experiment	Stock	Intron	Retrohoming efficiency (%)	Retrohoming efficiency relative to WT (%)
6	w^{1118} (WT)	Linear	0.071	
	ku70^{-}	Linear	0.0027	4
	lig4^{-}	Linear	0.0051	7
	lig4^{-}; *P{lig4}*	Linear	0.06	86
	w^{1118} (WT)	Lariat	2.5	
	ku70^{-}	Lariat	0.9	36
	lig4^{-}	Lariat	2.8	112
	lig4^{-}; *P{lig4}*	Lariat	1.3	52
7	w^{1118} (WT)	Linear	0.06	
	ku70^{-}	Linear	0.031	52
	lig4^{-}	Linear	0.012	20
	lig4^{-}; *P{lig4}*	Linear	0.051	86
	w^{1118} (WT)	Lariat	3.0	
	ku70^{-}	Lariat	2.1	70
	lig4^{-};	Lariat	2.8	93
	lig4^{-}; *P{lig4}*	Lariat	4.0	133
8	Or-R (WT)	Linear	0.031	
	polQ^{-}	Linear	<0.0007	<2
	Or-R (WT)	Lariat	0.9	
	polQ^{-}	Lariat	0.32	36
9	Or-R (WT)	Linear	0.019	
	polQ^{-}	Linear	<0.0007	<4
	Or-R (WT)	Lariat	3.1	
	polQ^{-}	Lariat	0.54	17

Table 2.3: Linear and lariat group II intron retrohoming efficiencies in wild-type, *lig4*$^{-}$, *ku70*$^{-}$, *polQ*$^{-}$, and *lig4*$^{-}$; *P{lig4}* *D. melanogaster* embryos

Retrohoming assays using lariat and linear RNPs were done in *D. melanogaster* precellular blastoderm embryos, as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 1 h at 30°C, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency. WT, wild type.

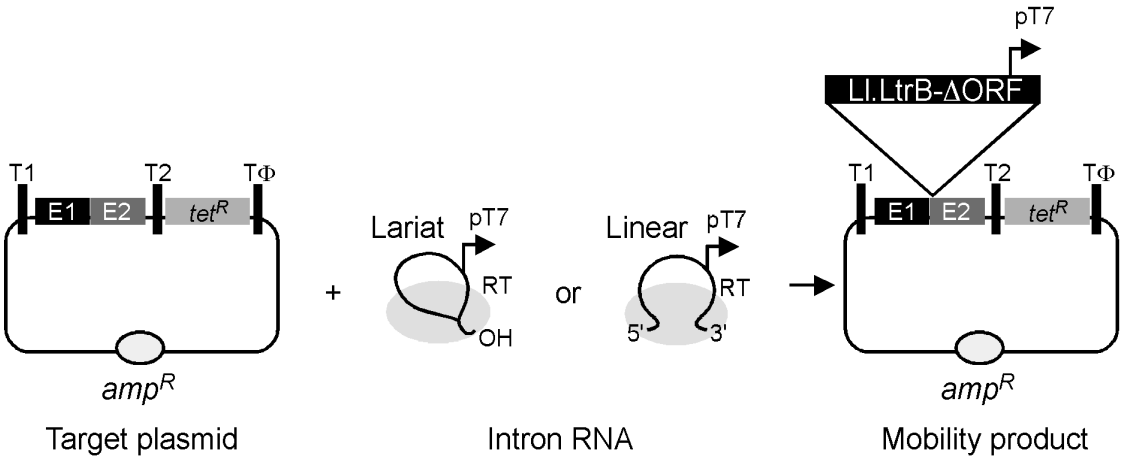


Figure 2.1: Retrohoming assay of lariat and linear group II intron RNPs in *D. melanogaster* embryos

D. melanogaster precellular blastoderm stage embryos are microinjected first with the Amp^R target plasmid pBRR3-ltrB, which contains an Ll.LtrB target site (ligated E1-E2 sequence of the *ltrB* gene) cloned upstream of a promoterless tet^R gene, followed within 5 min by Ll.LtrB RNPs containing linear or lariat intron RNA with a phage T7 promoter sequence inserted near its 3' end. The embryos are incubated at 30°C for 1 h during which time the intron integrates into the target site, positioning the T7 promoter upstream of the promoterless tet^R gene. Nucleic acids are then extracted and transformed into *E. coli* HMS174(DE3), and retrohoming efficiencies are calculated as the ratio of $(Tet^R + Amp^R)/Amp^R$ colonies. T1 and T2 are *E. coli* rrnB transcription terminators, and Tϕ is a phage T7 transcription terminator.

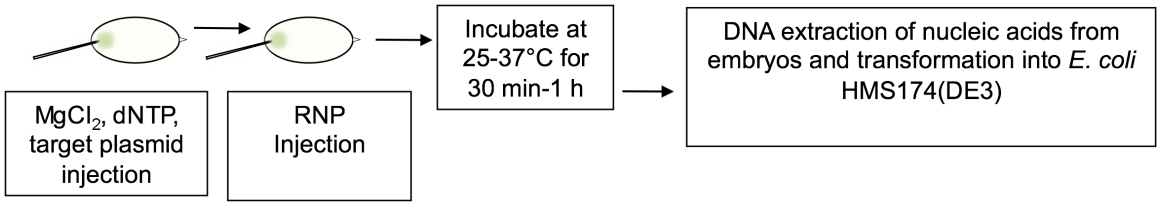
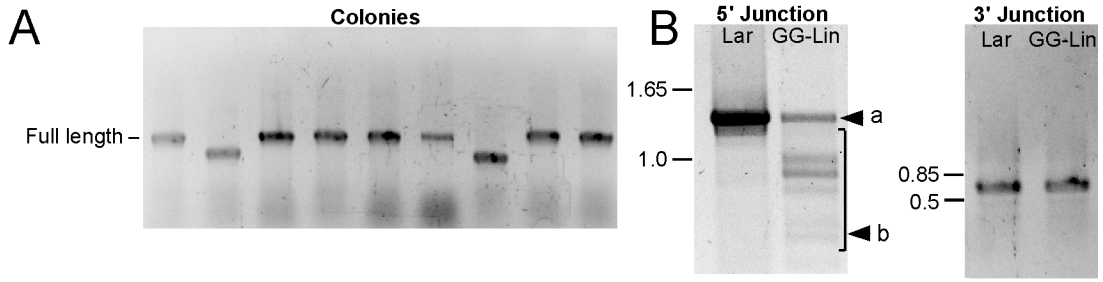


Figure 2.2: Microinjection procedure for plasmid retrohoming assays in *D. melanogaster* embryos

D. melanogaster embryos are collected as described in Materials and Methods section 2.5.4 and microinjected with approximately 0.3 nl target plasmid pBRR3-ltrB at 1.4 mg/ml in solution with or without indicated $MgCl_2$,where indicated, and 17 mM dNTPs, followed within 5 min by approximately 0.3 nl of Ll.LtrB RNPs. Embryos are incubated for 30 min or 1 h at 25, 30 or 37˚C, as indicated for individual experiments. Embryos are recovered by pipetting off the slide with a micropipette tip. Total nucleic acids are extracted (Materials and Methods section 2.5.4) and used to transform *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency.



C

5' Exon	Insertion	Intron	Freq
GAACACATCCATAAC		+1 GTGCGCCCAGATAGG	10
GAACACATCCA----	gactggctaaccca-tcc	+1 GTGCGCCCAGATAGG	1
GAACACATCCAT(A)--		+267 AGGTAAGTTATGGTT	1
GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	1
GAACACATCCAT---	gataagcagaaat-ggagttgtta-gagctgttcggt-tcggcaaaaaaaac tgatatcatat-caaaatcc	+341 ACGAAAGCGATGCCG	1

D

	5' Exon	Insertion	Intron	Freq
Band a	GAACACATCCATAAC		+1 GTGCGCCCAGATAGG	21
	GAACACATCCATAAC	agg	+1 GTGCGCCCAGATAGG	1
	GAACACATCCA----		+1 GTGCGCCCAGATAGG	1
	GAACACATCCATAAC	caatttggtctg	+1 GTGCGCCCAGATAGG	1
	GAACACATCCATAAC	accg	+1 GTGCGCCCAGATAGG	1
	GAACACATCCATA--	tatacgga	+2 -TGCGCCCAGATAGG	1
	AACCCACGTCGATCG-15	gggt	+2 -TGCGCCCAGATAGG	1
	GAACACATCCATAA(C)		+4 ---CGCCCAGATAGG	1
	GAACACATCCA----	tccatta	+4 ---CGCCCAGATAGG	1
	GAACACATCC-----		+39 TACTACTCTGTAAGA	1
	GAACACATCCAT(A)		+323 AGAACCTATGGGAAC	1
	GAACAC---------	ggtaaagttggcg-cagcattcaaggg	+346 AGCGATGCCGAGAAT	1
	GAACACATCCATAAC	gccgatgccga	+349 GATGCCGAGAATCTG	1
	GAACACATCCATAAC	gtatctcatgt-gatgcgacaccac-catacgcacct	+427 AAAAAGGCTATAGCA	1
Band b	GAACACATCCAT(AA)		+265 AAAGGTAAGTTATGG	1
	GAACACATCCATAAC	taat	+265 AAAGGTAAGTTATGG	1
	GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	1
	GAACACATCCAT---	g	+341 ACGAAAGCGATGCCG	2
	GAACACATCCAT(AA)		+344 AAAGCGATGCCGAGA	1
	GAACACATCC(AT)		+350 ATGCCGAGAATCTGA	1
	GAACACATCCATAAC		+357 GAATCTGAATTTACC	1
	GAACACATCCA(TAAC)		+384 TAACTGGGGATACCC	1
	GAACACATCCATA(AC)		+508 ACATGGCAAAGGGGT	1
	GAACACATCC(A)		+583 ACAATGGCAATTTTA	1

Figure 2.3. Retrohoming of Ll.LtrB GG-Lin RNPs in *D. melanogaster* embryos.

Retrohoming assays using lariat and GG-Lin RNPs were done in *D. melanogaster* precellular blastoderm embryos, as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 30 min at 37ºC, nucleic acids were extracted and either transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency and colony PCR or used directly for PCR analysis of integration products (*A*) *E. coli* colony PCR. Tet^R + Amp^R colonies resulting from retrohoming of GG-Lin RNPs were analyzed by colony PCR using the primers P1, located 388-bp upstream of the intron insertion site in pBRR3-ltrB, and reverse primer, P2, located 850-nt downstream of the 5' end of the intron bearing the antisense sequence (Materials and Methods section 2.5.5). The PCR products were analyzed in a 1% agarose gel, which was stained with ethidium bromide. Full-length indicated to the left is the size of the PCR products resulting from amplification of the 5'-intron insertion junction without 5'-exon resection, intron cDNA truncation or insertion, as observed with retrohoming of lariat intron RNPs. (B) PCR of the 5' junction using primers P1 and P2, and PCR of the 3' junction using primers P3 and P4 (Materials and Methods section 2.5.5) amplified from extracted DNA without transformation into *E. coli* from lariat (Lar) and GG-Lin retrohoming assays. (C) Colony PCR sequences of 5'-integration junctions resulting from retrohoming of GG-Lin RNPs. The sequences were determined by sequencing the colony PCR products using the primer Rseq. (D) Sequences of 5'-integration junctions resulting from retrohoming of GG-Lin RNPs. PCR products corresponding to regions "a" and "b" of the gel in (B, left panel) were cloned into TOPO-TA pCR2.1 vector (Invitrogen) and sequenced (Materials and Methods section 2.5.5). 5'-exon sequences and intron sequences are in upper case letters, and extra nucleotide residues inserted at the junction are in lower case letters. The position numbers at the beginning of the intron sequences are indicated. Microhomologies at the ligation junction in the 5' exon are shown in parentheses. Inserted sequence that has homology with nearby 5'-exon or intron sequences is underlined. The frequency of occurrence of a sequence (Freq) is indicated to the right.

5' Exon	Insertion	Intron	Freq
E. coli colony PCR			
GAACACATCCATAA-	gaaaacc	+1 GTGCGCCCAGATAGG	1
GAACACATCCATAA-	agtaggacgctcaccgctct	+1 GTGCGCCCAGATAGG	1
GAACACATCCATA--		+1 GTGCGCCCAGATAGG	1
GAACACATCCA----	agtaggacgctcaccgctct	+1 GTGCGCCCAGATAGG	1
GAACACA--------		+1 GTGCGCCCAGATAGG	1
CACGTCGATC(GTG)-15		+1 GTGCGCCCAGATAGG	1
GAACACATCCATAAC		+2 -TGCGCCCAGATAGG	1
GAACACATCCATAAC	atccattcctc	+2 -TGCGCCCAGATAGG	1
GAACACATCCATAAC	gccttacgactggacgcactcaca	+2 -TGCGCCCAGATAGG	1
GAACA(C)		+4 ---CGCCCAGATAGG	1
GAACA(CA)		+8 --------CAGATAGG	1
GAACACATCCATA(A)		+10 ---------GATAGG	1
GAACACATCCATAA-	atccattcctc	+27 AGTAGTTTAAGGTAC	1
GAACACATCCATAA-	tcca	+35 AAGGTACTACTCTGT	1
GAACACATCCATA--	tctcgttatcttaacgaga	+49 TAAGATAACACAGAA	1
GAACACATCCATAA-	gtggatgtataccttacc	+84 AGCGAAAGCTGATAC	1
GAACACATC(CA)		+109 CACGGTTGGAAAGCG	1
GAACACATCCATA--	tgtgttacctaa	+127 AGTTACCTAAAGACA	1
GAACACATCCATA--		+174 ATATAAGGTATAAGT	1
GAACACA--------	cccataacgggatatcttaagtc	+185 AAGTTGTGTTTACTG	1
GAACACATCC(AT)		+214 ATTTCGGTTATGTGT	1
GAACACATCCATAAC	atccca	+242 TGTCTGAAACCTCTA	1
GAACACA--------	atccata	+256 AGTACAAAGAAAGGT	1
GAACACATCCATAA-	tgtcccctgagtaggac	+265 AAAGGTAAGTTATGG	1
GAACACATCCATA(A)		+265 AAAGGTAAGTTATGG	1
GAACACATCCATAAC		+269 GTAAGTTATGGTTGT	1
GAACACATCCATAAC	atcc	+271 AAGTTATGGTTGTGG	1
GAACACATCCATAAC	tggacctgcgttatgaccacataac	+273 GTTATGGTTGTGGAC	1
GAACACAT-------	aagtcc	+276 ATGGTTGTGGACTTA	1
GAACACATCCA(T)		+283 TGGACTTATCTGTTA	1
GAACACATCCATAAC	ggattattaatcg	+298 TCACCACATTTGTAC	1
GAACACATCCATAA-		+306 TTTGTACAATCTGTA	1
GAACACATCCATAAC	atcc	+330 ATGGGAACGAAACGA	1
GAACA----------	tccatatggatatggatcataac	+331 TGGGAACGAAACGAA	1
GAACACATCCATAA-	t	+340 AACGAAAGCGATGCC	1
GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	3
GAACACATCCAT---	cacgaaagcgatgcgatgcacacatgtag	+342 CGAAAGCGATGCCGA	1
GAACACATCCATAAC	agaacatccaggca	+343 GAAAGCGATGCCGAG	1
GAACACATCCAT---	gcgtctgggtgtt	+346 AGCGATGCCGAGAAT	1
GAACACATCCATAAC	t	+348 CGATGCCGAGAATCT	1
GAACACATCCATAA-	gtaatgtaccccgccatatatgtac-tataagtttatggcatatc	+357 GAATCTGAATTTACC	1
GAACACATCCATAAC	gatgttatgatcttgttgtatgctttcatt-[1]gtcatcgtcacgtgattcataaacacaagt-gaatgtcgactaatacgacatttta	+358 AATCTGAATTTACCA	1
GAACACATCCATAA-	tcttatattac	+421 AGGAGGAAAAAGGCT	1
GAACACA--------	c	+428 AAAAGGCTATAGCAC	1
GAACACATCC(A)		+578 AACCAACAATGGCAA	1
GAACACATCCAT---	gatttc	+590 CAATTTTAGAAAGAA	1

1. Possible template for insertion: intron sequence nucleotides 732 to 800

Figure 2.4: Colony PCR sequences of 5'-integration junctions resulting from retrohoming of GG-Lin RNPs in *D. melanogaster* embryos at 25°C

Retrohoming assays using GG-Lin RNPs were done in *D. melanogaster* precellular blastoderm embryos, as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 1 h at 25°C, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency and colony PCR. The sequences of 5′ junctions were obtained from colony PCR products in randomly selected Tet^R + Amp^R colonies as described in Figure 2.3.

Intron, exon and insertion sequences are depicted as in Figure 2.3. Inserted sequences that have homology with nearby 5'-exon or intron sequences are underlined. The frequency of occurrence of a sequence (Freq) is indicated to the right.

	5' Exon	Insertion	Intron	Freq
Band a	GAACACATCCAT---	tt	+1 GTGCGCCCAGATAGG	1
	GAACACATCCA----	gttatgaacatccataact	+3 --GCGCCCAGATAGG	1
	GAACACATCCATAA-	gcggctaacgttatttacgcaccaa	+4 ---CGCCCAGATAGG	1
	GAACACATCCATA--	gcacc	+4 ---CGCCCAGATAGG	1
	GAACA(C)		+4 ---CGCCCAGATAGG	1
	GAAC----------	ccaa	+4 ---CGCCCAGATAGG	1
	G----------	ta	+5 ----GCCCAGATAGG	1
	AACCCACGTCGATCG-17	gtaatatgttatct	+44 CTCTGTAAGATAACA	1
	GAACACATCCAT---		+191 TGTTTACTGAACGCA	1
	GAACACATCCATAAC	c	+265 AAAGGTAAGTTATGG	1
	GAACACATCCAT(AA)		+266 AAGGTAAGTTATGGT	1
	GAACACATCCATAA-	aacac	+340 AACGAAAGCGATGCC	1
	GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	1
	GAAC----------	gaccgttagcgaataatagcat	+340 AACGAAAGCGATGCC	1
	GAACACATCCA----	cacatcc	+341 ACGAAAGCGATGCCG	1
	GAACACATCCAT---	gtgt	+343 GAAAGCGATGCCGAG	1
	GAACACATCCA----	gtgt	+345 AAGCGATGCCGAGAA	1
Band b	GAACACATCCA(T)		+2 -TGCGCCCAGATAGG	1
	GAACACATCCATAtC	gtgt	+159 CGCAATGTTAATCAG	1
	GAACACATCC(A)		+266 AAGGTAAGTTATGGT	1
	GAACACATCCATA(AC)		+300 ACCACATTTGTACAA	1
	GAACACATC------	tggtt	+302 CACATTTGTACAATC	1
	GAACACATCCATAAC	tgtaact	+303 ACATTTGTACAATCT	1
	GAACACATCCATA--	gcgccctttgg	+340 AACGAAAGCGATGCC	1
	GAACACATCCATAAC	at	+341 ACGAAAGCGATGCCG	1
	GAACACATCC-----	gtc	+341 ACGAAAGCGATGCCG	1
	GAACACATCCATA(A)		+344 AAAGCGATGCCGAGA	1
	GAACACATCCAT(AA)		+344 AAAGCGATGCCGAGA	2
	GAACACATCC(A)		+345 AAGCGATGCCGAGAA	1
	GAACACATCCATAA(C)		+348 CGATGCCGAGAATCT	1
	GAACACATCCATAAC	t	+357 GAATCTGAATTTACC	1
	GAACACATCCATAAC	gc	+358 AATCTGAATTTACCA	1
	GAACACATCCATAA-	ggaaaata	+424 AGGAAAAAGGCTATA	1
	GAACACATCCATAAC	gttt	+425 GGAAAAAGGCTATAG	1
	GAACACATCCATAAC	gt	+426 GAAAAAGGCTATAGC	1
	GAACACATCCAT(AA)		+428 AAAAGGCTATAGCAC	1
	GAACACATC------	acacccgattgtttgt	+429 AAAGGCTATAGCACT	1
	GAACACATCCAT---	taaaccttaacatcgc	+430 AAGGCTATAGCACTA	1
	GAACACATCCAT(AA)		+452 AAAATCTTGCAAGGG	1
	GAACACATCCATAA-		+462 AAGGGTACGGAGTAC	1
	GAACACATCCATAAC	gatgtta	+503 CCTTTACATGGCAAA	1
	GAACACATCCATAA-	gaatggatgtacggaccatgaccatc-cataacaatgtgg	+578 AACCAACAATGGCAA	1
	GAACACATCCATAA-	tg	+580 CCAACAATGGCAATT	1

Figure 2.5: Sequences of 5'-integration junctions resulting from retrohoming of GG-Lin RNPs in *D. melanogaster* embryos at 25°C

Retrohoming assays using GG-Lin RNPs were done in *D. melanogaster* precellular blastoderm embryos, as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 1 h at 25°C, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency and colony PCR. The sequences of 5′ junctions were obtained from PCR products amplified from extracted DNA without transformation into *E. coli* as described in Figure 2.3. Intron, exon and insertion sequences are depicted as in Figure 2.4. Mutant nucleotide residues in 5'-exon or intron sequences are in lower case letters. The frequency of occurrence of a sequence (Freq) is indicated to the right.

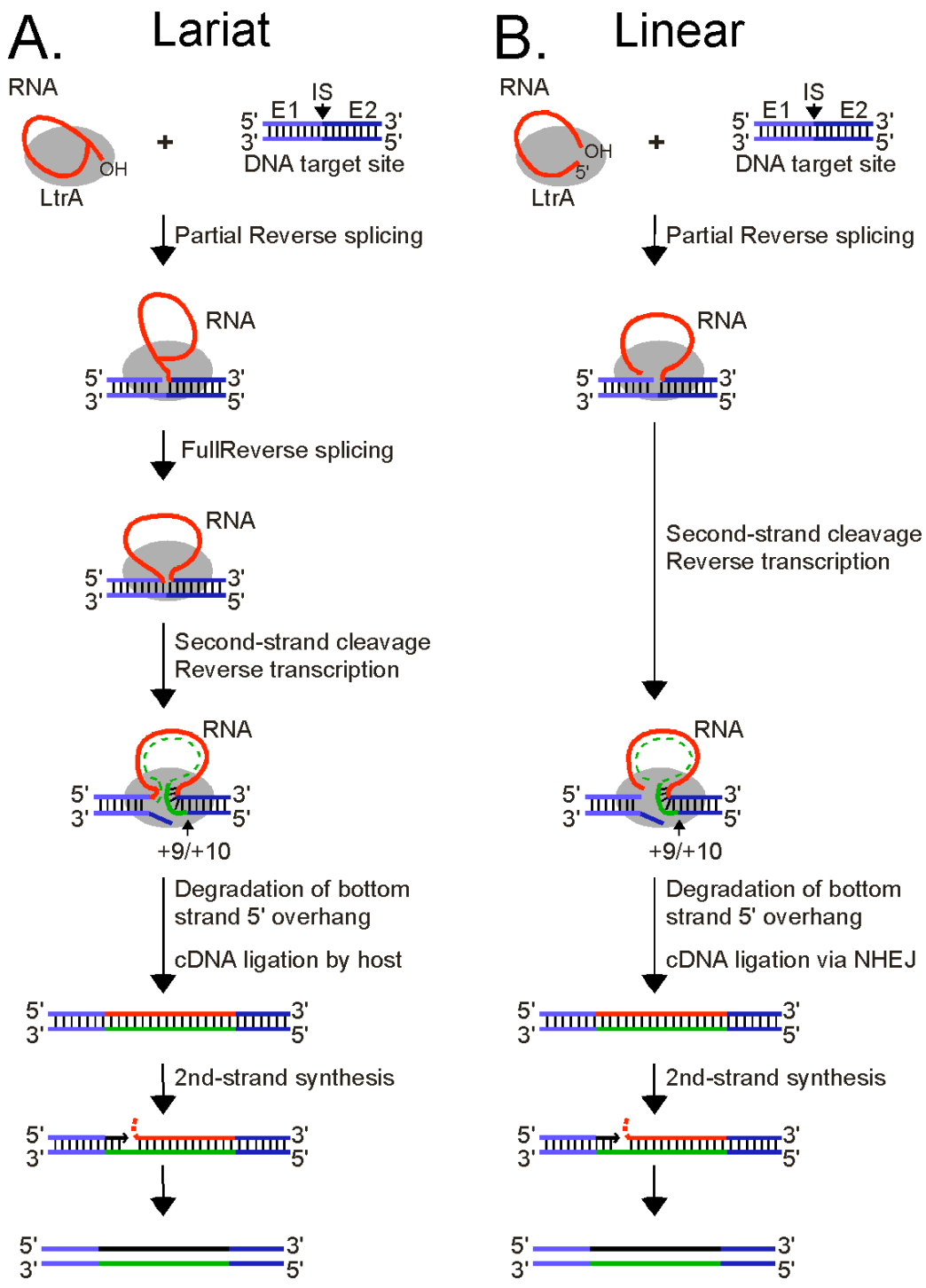


Figure 2.6: Model for retrohoming of lariat and linear group II intron RNAs

(A) Lariat RNPs recognize the DNA target site (ligated *ltrB* E1-E2 sequence) and carry out both steps of reverse splicing, resulting in insertion of linear intron RNA between E1 and E2. The IEP (LtrA) uses its En domain to cleave the bottom strand between positions +9 and +10 of the 3' exon, and then uses the 3' end of the cleaved DNA strand as a primer for reverse transcription of the inserted intron RNA. Retrohoming is thought to be completed by a process that includes resection of the 5' overhang on the bottom strand, extension of cDNA synthesis into exon 1, degradation or displacement of the intron RNA template strand, second-strand DNA synthesis, and sealing of nicks using host enzymes (Smith et al., 2005). (B) Linear RNPs recognize the DNA target site and carry out the first step of reverse splicing, resulting in ligation of the 3' end of the intron RNA to the 5' end of the E2. The IEP (LtrA) then uses its En domain to cleave the bottom strand between positions +9 and +10 of E2 and reverse transcribes the attached linear intron RNA. The intron cDNA is linked to the 5'-exon DNA by an error-prone process that sometimes leads to precise insertion of the intron RNA, but often results in loss of E1 sequences due to excessive resection, insertion of 5'-truncated introns due to incomplete cDNA synthesis or degradation, and/or insertion of extra nucleotide residues at the ligation junction. As for lariat RNA, retrohoming of the linear intron RNA is presumably completed by degradation or displacement of the intron RNA template strand, second-strand DNA synthesis, and sealing of nicks using host enzymes. CS, bottom-strand cleavage site; IS, intron insertion site.

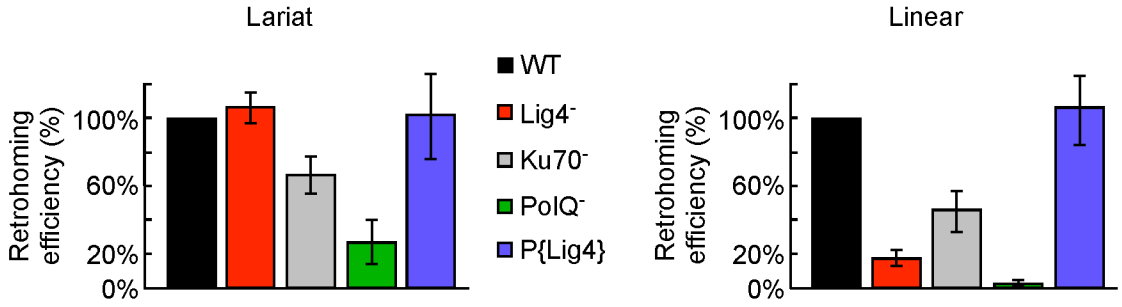
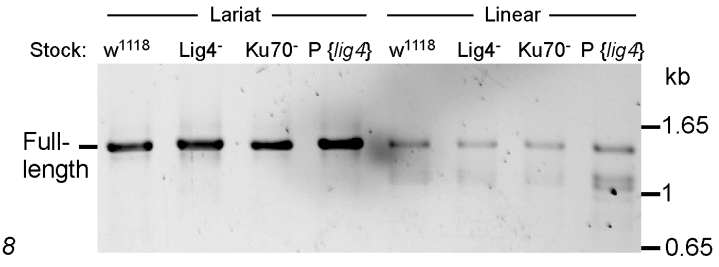


Figure 2.7: Retrohoming of linear and lariat group II intron RNAs in wild-type and mutant *D. melanogaster* embryos

Retrohoming efficiencies of lariat and linear Ll.LtrB RNPs in *D. melanogaster* wild-type, *lig4*$^{-}$, *ku70*$^{-}$, *polQ*$^{-}$, and *lig4*$^{-}$; *P{lig4}* embryos were assayed as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. The bar graphs show the average retrohoming efficiency of retrohoming assays described in Table 2.3 as a percentage of wild-type embryos assayed in parallel with the error bars indicating the standard error.

A



B *w*^{1118}

5' Exon	Insertion	Intron	Freq
GAACACATCCATAA-	aaaaatc	+1 GTGCGCCCAGATAGG	1
GAACACATCCAT---		+1 GTGCGCCCAGATAGG	1
GAACACATCCATAA-		+2 -TGCGCCCAGATAGG	1
GAACACATCCA(T)		+2 -TGCGCCCAGATAGG	1
-AACACA(T)		+2 -TGCGCCCAGATAGG	1
GAACACATCCATAAC	atggaatcac	+3 --GCGCCCAGATAGG	1
GAACACATCCATAA(C)		+4 ---CGCCCAGATAGG	1
GAACACATCCATAA-	tac	+4 ---CGCCCAGATAGG	1
GAACACATCCATAA(C)		+7 ------CCAGATAGG	1
GAACACATCCATAAC	tgatgcgttcagggaatcgaatgcc	+46 CTGTAAGATAACACA	1
GAACACATCCATA--	ca	+146 GGTACGACTGAGTCG	1
GAACACATCCATAAC	gaata	+183 ATAAGTTGTGTTTAC	1
GAACACATCCATA--	caca	+198 TGAACGCAAGTTTCT	1
GAACACATCCATAAC	gtgtcatattc	+239 AAGTGTCTGAAACCT	1
GAACACATCCATAAC	ttctaccttgtctgcctatggattag	+330 ATGGGAACGAAACGA	1
GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	3
GAACACATCCA----	ctgagcagacaccat	+340 AACGAAAGCGATGCC	1
GAACACATC------		+341 ACGAAAGCGATGCCG	1
GAACACATCCATAAC	gaattt	+342 CGAAAGCGATGCCGA	1
GAACACATCCAT---	tcc	+342 CGAAAGCGATGCCGA	1
GAACACATCCAT(AA)		+344 AAAGCGATGCCGAGA	1
GAACACATCC(A)		+345 AAGCGATGCCGAGAA	1
GAACACATCCATAAC	aaaa	+427 AAAAAGGCTATAGCA	1
GAACAC(ATC)		+653 ATCGTTATCTTTTAC	1
GAACACATCCATAAC	ctgcgcctatcgaccgc	+808 AGGGTCGACAGTGAA	1

C *lig4*$^{-}$

5' Exon	Insertion	Intron	Freq
GAACACATCCATAAC		+1 GTGCGCCCAGATAGG	1
GAACACATCCATAA-	g	+1 GTGCGCCCAGATAGG	1
GAACACATCCA(T)		+2 -TGCGCCCAGATAGG	2
GAACACATCCATAA-	tgccataag	+3 --GCGCCCAGATAGG	1
GAACACATCCATAA-	taatgtgtcaattaccgctttata	+3 --GCGCCCAGATAGG	1
GAACACATCCATAA(C)		+4 ---CGCCCAGtTAGG	1
GAACACATCCAT---	ggggatgt	+5 ----GCCCAGATAGG	1
GAACACATCCATAAC	aa	+76 AACCGAAAAGCGAAA	1
GAACACATCC(ATAA)		+176 ATAAGGTATAAGTTG	1
GAACACATCCATAAC	c	+244 TCTGAAACCTCTAGT	1
GAACACAT-------	gt	+268 GGTAAGTTATGGTTG	1
GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	1
GAACACATCCATAAC	caagg	+343 GAAAGCGATGCCGAG	1
GAACAC---------	ga	+344 AAAGCGATGCCGAGA	1
GAACACATCCATAAC	gtgt	+427 AAAAAGGCTATAGCA	1
GAACACATCCA----	aaagcgcccttt	+427 AAAAAGGCTATAGCA	1
GTGGGGTCTCTATTC-341	tattc	+427 AAAAAGGCTATAGCA	1
ACGTCGATCGTG(AA)-15		+428 AAAAGGCTATAGCAC	1
GAACACATCCAT(AA)		+429 AAAGGCTATAGCACT	1
GAACACATCCATAAC	taact	+429 AAAGGCTATAGCACT	1
^C			
GAACACATC------		+433 GCTATAGCACTAGAG	1
CAACCCACGTCGATC-18	cat	+473 GTACTCGTAGTAGTC	1
GAACACATCCATAA-		+479 GTAGTAGTCTGAGAA	1
GAACACATCCATAAC	t	+558 GGAGGAAAACCTCAA	1
GAACACATCCA(T)		+732 TCTTGTTGTATGCTT	1
GAACACATCCATAA-	gtggataggttatgccaccgcataa-cagccactc	+883 AGAGGGTGGTGCAAA	1

Figure 2.8: 5' junctions resulting from retrohoming of Ll.LtrB linear intron RNPs in *D. melanogaster* wild-type and *lig4*$^{-}$ embryos

Retrohoming assays were done using lariat and linear RNPs as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. (A) Direct PCR analysis of 5'-integration junctions for retrohoming assay samples listed in Table 2.3, experiment 6 using forward primer, P1, located 388-bp upstream of the target site in pBRR3-ltrB, and reverse primer, LtrB933a, located 933-nt downstream of the 5' end of the intron bearing the antisense sequence (Materials and Methods section 2.5.5). (B) Sequences of 5'-junction products from PCR shown in panel A, cloned into the TOPO-TA pCR2.1 vector (Invitrogen) (Materials and Methods section 2.5.5) in wild-type embryos or (C) *lig4*$^{-}$ embryos. Intron, exon and insertion sequences are depicted as in Figure 2.5. The frequency of occurrence of a sequence (Freq) is indicated to the right.

A *ku70*$^{-}$

5' Exon	Insertion	Intron	Freq
GAACACATCCATAAC		+1 GTGCGCCCAGATAGG	4
GAACACATCCATAAC		+1 GgGCGCCCAGATAGG	1
GAACACATCCATAA-	tccataag	+1 GTGCGCCCAGATAGG	1
GAACACATCCATAAC	cgggggatgtgacgtactagcaaac	+1 GTGCGCCCAGATAGG	1
GAACACATCCATAAC	ca	+2 -TGCGCCCAGATAGG	1
GAACACATCCAT---	cc	+2 -TGCGCCCAGATAGG	1
GAACACATCCA(T)		+2 -TGCGCCCAGATAGG	1
GAACACATCCATAAC	ataacatccaaa	+2 -TGCGCCCAGATAGG	1
GAACACATCCATAAC	tgtggatattgcttatagatagat-taactatatataa	+4 ---CGCCCAGATAGG	1
GAACACATCCATAA(C)		+4 ---CGCCCAGATAGG	2
GAACAC---------	gat	+4 ---CGCCCAGATAGG	1
GAACACATCCATA(A)		+9 --------AGATAGG	1
GAACACATCCAT(AA)		+63 AAACAGCCAACCTAA	1
GAACAC(ATC)		+169 ATCAGATATAAGGTA	1
GAACACATCCAT(AAC)		+249 AACCTCTAGTACAAA	1
GAACACATCCATAAC	ctttact	+264 GAAAGGTAAGTTATG	1
GAACACATCCATAAC	c	+265 AAAGGTAAGTTATGG	1
GAACACATCCAT(AA)		+266 AAGGTAAGTTATGGT	1
GAACACATCC-----	tcacgat	+266 AAGGTAAGTTATGGT	1
GGCTCAGTCGAAAGA-282	tctgagttgcatgcgaa	+274 TTATGGTTGTGGACT	1
GAACACATCCAT(AAC)		+340 AACGAAAGCGATGCC	1
GAACACATCCAT---	gttcgcctcggtaaacatggattacat	+340 AACGAAAGCGATGCC	1
GAACACATCCATAA-	ttgcgatattgcgaaatttccgtat	+342 CGAAAGCGATGCCGA	1
GAACACATCCATA--	caacgaaaccg	+342 CGAAAGCGATGCCGA	1
GAACACATCCAT(AA)		+344 AAAGCGATGCCGAGA	1
TGCCAGGCATCAAAT-306	caacctactactgggcaacctactac	+578 AACCAACAATGGCAA	1
GAACACATCCAT(AAC)		+582 AACAATGGCAATTTT	1

B *lig4*$^{-}$; *P{lig4}*

5' Exon	Insertion	Intron	Freq
GAACACATCCATAAC	cacagtttaggtcatccggcatcg	+8 --------CAGATAGG	1
GAACACATCCATAA-	ta	+246 TGAAACCTCTAGTAC	1
ACGTCGATCGTG(AAC)-12		+249 AACCTCTAGTACAAA	1
GAACACATCCAT---	ttcgtc	+265 AAAGGTAAGTTATGG	1
GAACACATCC(A)		+272 AGTTATGGTTGTGGA	1
GAACACATCCATAA-	tgt	+273 GTTATGGTTGTGGAC	1
GAACACATCCATA--	caaaaag	+342 CGAAAGCGATGCCGA	1
GAACACATCCATAAC		+343 GAAAGCGATGCCGAG	1
GAACACATCCATAAC	cagttc	+344 AAAGCGATGCCGAGA	1
GAACACATCCA----	g	+346 AGCGATGCCGAGAAT	1
GAACACATCCAT---	ggatg	+349 GATGCCGAGAATCTG	1
GAACACATCCATAAC	cggattccagat	+357 GAATCTGAATTTACC	1
GAACACATCCAT(AA)		+427 AAAAAGGCTATAGCA	1
GAACACATCCAT(AA)		+429 AAAGGCTATAGCACT	1
GAACACATCCA----	c	+445 GAGCTTGAAAATCTT	1
GAACACATCC(AT)		+586 ATGGCAATTTTAGAA	1
GAACACATCCATAAC	gatgt	+634 GAAGTTTTTACAAGA	1
GAACACATCCATAAC		+647 GACTTTATCGTTATC	1
GAACACATCCATAAC	tctagaggcataagaaaggtacaa	+716 CGAAACAACGTAAAA	1
GAACACATCCATAA-	gagta	+726 TAAAACTCTTGTTGT	1
GAACACATCCATAAC	gataacatacaacatagc	+733 CTTGTTGTATGCTTT	1
GAACACATCCATAA-	gctgagggatgcac	+882 AAGAGGGTGGTGCAA	1
GAACACATCCATAA-	t	+888 GTGGTGCAAACCAGT	1
GAACACATCCATAAC	tgagcacaccaagaa	+893 GCAAACCAGTCACAG	1
GAACACATCCATAAC	atccacattttcacgctttcgtt	+896 AACCAGTCACAGTAA	1

Figure 2.9: 5' junctions resulting from retrohoming of Ll.LtrB linear intron RNPs in *D. melanogaster ku70*$^{-}$ and *lig4*$^{-}$; *P{lig4}* embryos

Retrohoming assays were done using lariat and linear RNPs as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. (A) Sequences of 5'-junction products from PCR shown in Figure 2.8 panel A, cloned into the TOPO-TA pCR2.1 vector (Invitrogen) (Materials and Methods section 2.5.5) in *ku70*$^{-}$ embryos or (B) *lig4*$^{-}$; *P{lig4}* embryos. Intron, exon and insertion sequences are depicted as in Figure 2.5. The frequency of occurrence of a sequence (Freq) is indicated to the right.

A ***E. coli* colony PCR- w^{1118}**

5' Exon	Insertion		Intron	Freq
GAACACATCC-----		+1	GTGCGCCCAGATAGG	1
GAACA----------	tccgaaatttcca-ctttcta	+1	GTGCGCCCAGATAGG	1
GAACACATCCATAA(C)		+4	---CGCCCAGATAGG	1
GAACACAT(C)		+7	-------CAGATAGG	1
GAACACATCC(AT)		+275	ATGGTTGTGGACTTA	1
GAACACATCCATAAC	t	+346	AGCGATGCCGAGAAT	1

B ***E. coli* colony PCR- *lig4*$^{-}$**

5' Exon	Insertion		Intron	Freq
GAACACATCCATAAC		+1	GTGCGCCCAGATAGG	1
GAACACATCCATAAC	tggaaaggaaggaaccg	+1	GTGCGCCCAGATAGG	1
GAACACATCCATAAC	atccatatcggc-cccagatgacaccctg	+1	GTGCGCCCAGATAGG	1
GAACACATCCATAA-	tatc	+1	GTGCGCCCAGATAGG	1
GAACACATCCATAA-	tatt	+1	GTGCGCCCAGATAGG	1
GAACACATCCAT---		+1	GTGCGCCCAGATAGG	1
GAACACATCCA----	gcccagataggtt	+1	GTGCGCCCAGATAGG	1
GAACACATCCA----	ctact	+2	-TGCGCCCAGATAGG	1
GAACACATCCATA--		+3	--GCGCCCAGATAGG	1
GAACACATCCATAAC	caccaccaaccacca	+3	--GCGCCCAGATAGG	1
GAACACATCCATAAC	catt	+4	---CGCCCAGATAGG	1
GAACACATCCATAA(C)		+4	---CGCCCAGATAGG	1
GAACACATCCA----	gccgcct	+5	----GCCCAGATAGG	1
GAACACATCCAT---	ggttatggtttcgcgt-cacctatctcgcg	+7	------CCAGATAGG	1
GAACACAT(CCA)		+7	------CCAGATAGG	1
GAACACATCCAT(A)	cgacct	+9	--------AGATAGG	1
GAACACATCCATA		+10	---------GATAGG	1
GAACACATCCATAAC	acttac	+39	TACTACTCTGTAAGA	1
GAACACATCCATA(AC)		+40	ACTACTCTGTAAGAT	1
GAACACATCC(ATA)		+95	ATACGGGAACAGAGC	1
GAACACATCCATAAC	atccataatggatga	+102	AACAGAGCACGGTTG	1
GAACACATCCA----	gtt	+260	CAAAGAAAGGTAAGT	1
GAACACATCCATA(AC)		+300	ACCACATTTGTACAA	1
ACGTCGATCGTGAAC-12		+343	GAAAGCGATGCCGAG	1

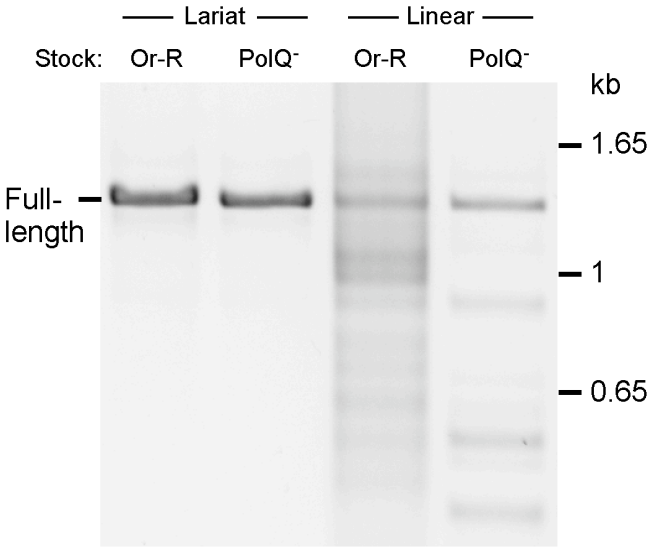
C ***E. coli* colony PCR- *ku70*$^{-}$**

5' Exon	Insertion		Intron	Freq
CCCACGTCGATC(GTG)-15		+1	GTGCGCCCAGATAGG	1
GAACACATCCA(T)		+2	-TGCGCCCAGATAGG	1
GAACACATCCAT---	g	+3	--GCGCCCAGATAGG	1
GAACACATCCATAA(C)		+4	---CGCCCAGATAGG	2
GAACACAT(C)		+7	------CCAGATAGG	1
GAACACATCCATA--	taggatactcgatc-cgccag	+9	--------AGATAGG	1
GAACAC---------		+14	----------G	1
GAACACATCCAT(A)	catct	+39	TACTACTCTGTAAGA	1
GAACACATCCATAAC	ccatagacaagaaagc-tatttgttacatcaa	+155	GAGTCGCAATGTTAA	1
GAACACATCCATAA-	aggtggaggccaatt-tactttcc	+265	AAAGGTAAGTTATGG	1
GAACACATCCATAAC	tgaacaaaaatttttt-tatctgaa	+271	AAGTTATGGTTGTGG	1
GAACACATCCATAA-	t	+347	GCGATGCCGAGAATC	2
GAACA(C)		+354	CGAGAATCTGAATTT	1
GAACACATCCA----	actcttaaatttta	+471	GAGTACTCGTAGTAG	1
TGCAACCCACGTC(GA)-20		+602	GAATCAGTAAAAATT	1
TCGATCGTG(AACACA)-9		+773	AACACAAGTGAATGT	1

Figure 2.10 Colony PCR sequences of 5'-integration junctions resulting from retrohoming of Ll.LtrB linear intron RNPs in *D. melanogaster* wild-type, *lig4*$^{-}$ and *ku70*$^{-}$ embryos

Retrohoming assays were done using linear RNPs as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. After incubating the embryos for 1 h at 30°C, nucleic acids were extracted and transformed into *E. coli* HMS174(DE3) for plating assays of retrohoming efficiency, described in Table 2.3, experiment 4, and colony PCR. The sequences of 5′ junctions were obtained from colony PCR products in randomly selected Tet^R + Amp^R colonies as described in Figure 2.3. Intron, exon and insertion sequences are depicted as in Figure 2.3. (A) Sequences of 5'-junction products from colony PCR of linear intron RNP retrohoming assays in wild-type or (B) *lig4*$^-$ or (C) *ku70*$^-$ embryos. Intron, exon and insertion sequences are depicted as in Figure 2.5. The frequency of occurrence of a sequence (Freq) is indicated to the right.

A



B Or-R

5' Exon	Insertion	Intron	Freq
GAAAGGCTCAGT(CGA)-285		+122 CGATGAGTTACCTAA	2
GAACACATCCATAA-	gg	+304 CATTTGTACAATCTG	1
GAACACATCCATAA-	gt	+347 GCGATGCCGAGAATC	1
AAACGAAAGGCTCAG-289	gcgcctttaggtcacgattat-gtcacgatgagtg	+666 ACGTCCAGATATTTA	1
GAACACATCCATAAC	ctaa	+701 AAATGGCAATGATAG	1
GAACACATCCATAA-	gtcg	+789 GACTAATACGACTCA	1
GAACACA--------	gaatgagtttcacgatcgactgtat-catcgtctgagtttgacacatccgtcaaact	+791 CTAATACGACTCACT	1
GAACA----------	tcggtcctg	+845 GAGCCGTATACTCCG	1
GAACACATCCATAA-	tca	+881 GAAGAGGGTGGTGCA	1

C *polQ*-

5' Exon	Insertion	Intron	Freq
GAACACATCCATAAC		+1 GTGCGCCCAGATAGG	12
GAGTAGGGAACTGC(C)-317		+689 CGACGCGTTGGGAAA	2
GAGAGTAGGGAAC(TG)-319		+704 TGGCAATGATAGCGA	1
GCAACCCACGTC(GAT)-19		+711 GATAGCGAAACAACG	3
TGAACACATC(CATAA)-2		+769 CATAAACACAAGTGA	1
GTCGATCGTG(AACAC)-10		+773 AACACAAGTGAATGT	10
CACGTCGATC(GTGAA)-13		+818 GTGAATTTTTACGAA	1
CAGTCGAAAGACTGG-278	atccacg	+818 GTGAATTTTTACGAA	1
^ g			
GTAGCGCCGATGGTA-353	acgtctaacgtcaaacggt-tcagttc	+818 GTGAATTTTTACGAA	1
GAATTTGTGAT(TGCA)-31		+892 TGCAAACCAGTCACA	14

Figure 2.11: 5' junctions resulting from retrohoming of Ll.LtrB linear intron RNPs in *D. melanogaster* wild-type and *polQ*- embryos

Retrohoming assays were done using lariat and linear RNPs as described in Figures 2.1, 2.2 and Materials and Methods section 2.5.4. (A) Direct PCR analysis of 5'-

integration junctions for retrohoming assay samples listed in Table 2.3, experiment 9 using forward primer, P1, located 388-bp upstream of the target site in pBRR3-ltrB, and reverse primer, LtrB933a, located 933-nt downstream of the 5' end of the intron bearing the antisense sequence (Materials and Methods section 2.5.5). (B) Sequences of 5'-junction products from PCR shown in panel A, cloned into the TOPO-TA pCR2.1 vector (Invitrogen) (Materials and Methods section 2.5.5) in wild-type embryos or (C) *polQ*$^{-}$ embryos. Intron, exon and insertion sequences are depicted as in Figure 2.5. The frequency of occurrence of a sequence (Freq) is indicated to the right.

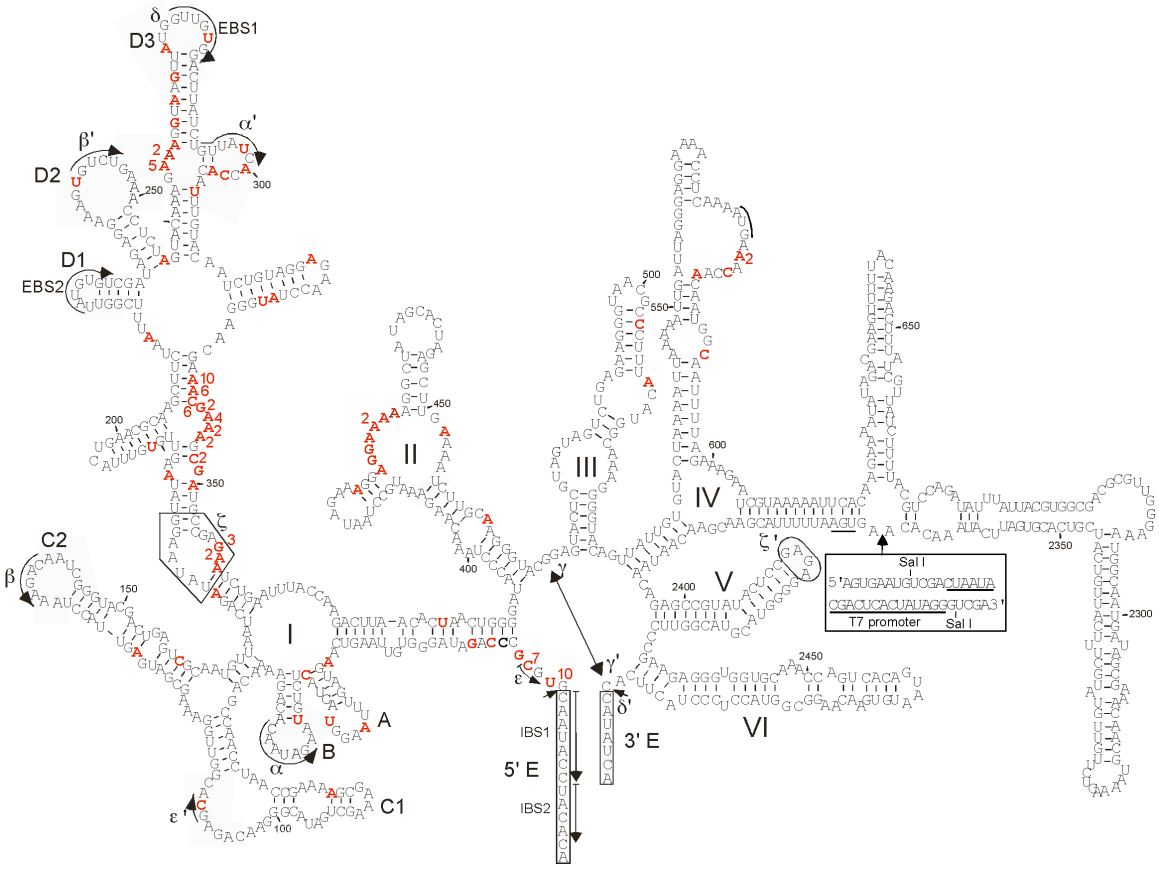


Figure 2.12: Location of 5' ends of truncated introns inserted by retrohoming of GG-Lin RNPs in *D. melanogaster* embryos

The figure shows a secondary structure model of the Ll.LtrB-ΔORF intron RNA. Intron RNA domains DI-VI, DI subdomains A, B, C1, C2, D1, and D2, and 5' and 3' exons (5'E and 3'E, respectively) are labeled. Sequence elements involved in RNA tertiary interactions (Greek letters and EBS and IBS sequences) are boxed and/or underlined. Also underlined are the phage T7 promoter inserted in DIV and the AUG start and UGA stop codons of the LtrA ORF (most of which is deleted in the Ll.LtrB-ORF construct). Nucleotide position numbers are those in the full-length wild-type Ll.LtrB intron. Nucleotide residues corresponding to the 5' ends of truncated introns inserted via retrohoming of GG-Lin RNPs (Figures 2.3-2.5) are in red with the numbers indicating repeats of the same 5' end.

Chapter 3: Biochemical characterization of template-switching and DNA polymerase activities of a group II intron reverse transcriptase

3.1 Introduction

The Ll.LtrB group II intron-encoded protein (IEP), LtrA, is a multifunctional protein which has reverse transcriptase (RT), maturase, and DNA endonuclease activities (Matsuura et al., 1997, Lambowitz & Zimmerly, 2010). The retrohoming of the linear form of the Ll.LtrB intron in the eukaryotic cell yields products with 5'-integration junctions characteristic of NHEJ-mediated repair. However, it is possible that LtrA, upon completion of cDNA synthesis of the integrated linear intron, contributes to formation of these 5' junctions through template-switching and/or non-templated nucleotide addition (NTNA) activities. These activities have been described for other non-LTR RTs as well as retroviral RTs (Kennell et al., 1994, Bibillo & Eickbush, 2002, Golinelli & Hughes, 2002a). The Mauriceville retroplasmid RT found in the mitochondria of *Neurospora crassa*, has been shown to undergo NTNA activity, as well as template switching to a second template RNA molecule (Kennell et al., 1994, Chen & Lambowitz, 1997, Chiang & Lambowitz, 1997). The R2 retroelement RT, which mediates integration of the R2 retrotransposon at specific chromosomal sites encoding 28S rRNA in many insect phyla, is capable of template switching to RNA or DNA as well as NTNA (George et al., 1996, Bibillo & Eickbush, 2004, Kurzynska-Kokorniak et al., 2007). The HIV-1 RT relies on template switching on the RNA genome as a means of completing minus strand DNA synthesis, a necessary step toward integration of proviral double-stranded DNA (Kulpa et al., 1997, Golinelli & Hughes, 2002b, a). In order to characterize the extent to which LtrA is able to template switch and incorporate non-templated nucleotides, I performed RT and DNA polymerase assays on oligonucleotide substrates mimicking the 5'-linear

intron retrohoming junction. In addition, I characterized LtrA polymerase activity on 3'-junction oligonucleotide mimics.

3.2 LtrA template-switching and polymerization activities at the Ll.LtrB linear intron integration 5' junction

3.2.1 LtrA can template switch from the 5' end of the intron RNA to the 3' end of the 5' exon

Although Lig4-independent NHEJ was the most likely explanation for the minority of Lig4-independent events observed during retrohoming of the linear intron RNP, a possible alternative mechanism was template switching of LtrA from the 5' end of the intron RNA directly to the 3' end of the 5'-exon DNA. Such template switching would generate a continuous cDNA linking the two sequences without the need for an additional ligation step. This possibility was deemed unlikely because group II intron RTs appeared to have low DNA-dependent DNA polymerase activity *in vitro* (Smith et al., 2005). However, template-switching reactions and NTNA by group II intron RTs had not been investigated systematically, and 5'-junction formation by this mechanism remained a possibility.

To determine if a template-switching mechanism could be responsible for the manner of 5'-integration junctions observed during linear intron retrohoming *in vivo*, I carried out biochemical assays using small artificial substrates that recreate the situation at the 5' junction just prior to completion of intron cDNA synthesis (Figure 3.1). The initial substrate consists of a 60-nt RNA template whose 5' end corresponds to that of the Ll.LtrB intron (Ll.LtrB RNA), with an annealed 45-nt DNA primer representing the nascent cDNA (primer c). LtrA initiates reverse transcription of the intron RNA template from the annealed DNA primer and extends it to the 5' end of the Ll.LtrB RNA template, where it could then jump to a second 40-nt DNA or RNA template with the nucleotide

sequence of exon 1 (E1 DNA or RNA, black and red, lanes 5 and 6, respectively). The 3' end of the Ll.LtrB RNA has an aminoblock to impede LtrA from switching to a second molecule of the initial template.

Figure 3.1 lanes 5 and 6 show that LtrA efficiently extends the primer to the end of the intron RNA template, yielding major labeled products of ~60-nt, which were resolved as a doublet, along with smaller amounts of larger products of the size expected for template switching to the exon 1 DNA or RNA (100-nt) or to a second molecule of Ll.LtrB RNA (120-nt). Controls show that no labeled products were detected after incubating LtrA with primer c in the presence or absence of the exon 1 RNA or DNA (lanes 2-4).

To further characterize the labeled products, the individual bands (a-n in lanes 1 and 2) were excised from the gel and cloned and sequenced, as described in Materials and Methods (section 3.5.4). The sequencing confirmed that the major ~60-nt products (Figure 3.1, bands a and b in lane 5 and h and i in lanes 6) correspond to cDNAs extending to or near the 5' end of the intron RNA, with the doublet reflecting the addition of extra nucleotide residues, mostly A-residues, to the 3' end of the cDNA upon reaching the end of the RNA template (Figures 3.2 and 3.3). Such NTNA is a common property of DNA polymerases and RTs (Clark et al., 1987, Clark, 1988, Peliska & Benkovic, 1992, Hu, 1993, Kennell et al., 1994, Bibillo & Eickbush, 2002, Golinelli & Hughes, 2002b). The first set of larger products (Figure 3.1, 90-110 nts; band c-e in lane 5 and j-l in lane 6) correspond to products generated by template switching from the 5' end of the intron to the 3' end of exon 1 DNA and RNA in lanes 5 and 6, respectively (Figures 3.2 and 3.3), as well as products generated by template switching to the 3' end and internal regions of the Ll.LtrB RNA, despite the 3'-terminal aminoblock (Figures 3.4 and 3.5). The majority of template switches to DNA occurs seamlessly (Figure 3.2), but non-

templated nucleotide residues, mostly A-residues, are found at some of the template-switching junctions, as well as at the 3' end of some of the cDNAs (Figures 3.2-3.5). The second set of larger products (Figure 3.1, 120-140 nts; bands f and g in lane 5 and m and n in lane 6) corresponds to products generated by a template switch to a second molecule of Ll.LtrB RNA or multiple template switches to exon 1 DNA and RNA, in lanes 5 and 6, respectively (Figures 3.2-3.5). These template-switching products have characteristics similar to the smaller products, including non-templated nucleotide residues, mostly A-residues, incorporated at a minority of template-switching junctions and at the 3' ends of the cDNAs.

In other experiments, I found that annealing a complementary DNA oligonucleotide to the exon 1 RNA or DNA substrate did not reduce the formation of the 100-nt product formed upon template switching. This lack of effect was observed regardless of whether the complementary DNA oligonucleotide created a blunt-end duplex or a 9-nt 5' overhang that might be present as a result of the initial staggered double-strand break (Figure 3.1, lanes 7-9). Native gel analysis showed that all exon 1 RNA and DNA templates used in these experiments are fully annealed to the complementary oligonucleotide under the reaction conditions (Figure 3.7). Template switching to the exon 1 duplex DNA to the 5' overhang substrate mimicking the staggered double-strand break was confirmed by sequencing (Figure 3.1, 100-110 nts; band p in lane 9). Some template switching to the complementary DNA oligonucleotide with a 5'overhang was also seen (Figure 3.6).

3.2.2 Template switching to the 5' exon occurs under near-physiological salt concentrations

The results above were obtained under reaction conditions found to be optimal for reverse transcription of artificial template-primer substrates by LtrA *in vitro* (450 mM

NaCl, 5 mM $MgCl_2$), the high salt concentration helping to stabilize and prevent aggregation of the free protein (Saldanha et al., 1999). However, similar results were obtained for template-switching reactions under near-physiological salt conditions (100 or 200 mM KCl, 5 mM $MgCl_2$) (Ashburner et al., 2005). Although the RT activity of LtrA indicated by the intensity of the major ~60-nt products bands was lower under these conditions, the gel profiles show roughly equal levels of template switching to exon 1 RNA and DNA (Figure 3.8). I found the sequence of cDNAs produced from reactions done in 100 mM KCl, 5 mM $MgCl_2$ had similar template-switching junctions and patterns of NTNA as reactions done in high salt (Figure 3.9).

Together, these biochemical assays show that LtrA can template jump from the 5' end of the intron RNA to exon 1 and surprisingly that template switching is equally efficient regardless of whether the exon 1 template is RNA or DNA. However, the junctions differ from those generated during retrohoming of linear intron RNPs in *D. melanogaster* embryos in that most lack any non-templated nucleotide residues and when they occur they are short and mostly A-residues. As the 5'-junction sequences of linear intron retrohoming *in vivo* do resemble those expected for canonical and Lig4-independent NHEJ, these findings suggest that template switching and NTNA by LtrA does not contribute substantially to the retrohoming of linear intron RNPs.

3.3 LtrA polymerase and template-switching activities at the Ll.LtrB intron retrohoming product 3' junction

3.3.1 Assays with exon 2 substrates

The finding that LtrA could template jump with similar efficiency to DNA and RNA substrates was surprising in view of its low DNA-dependent DNA polymerase activity in other assays (Smith et al., 2005). This observation reopened questions

regarding LtrA activity at the 3'-retrohoming junction, namely the issue of how efficiently LtrA could copy the 5'-DNA overhang generated by endonuclease cleavage in exon 2 before copying the intron RNA. It was suggested that host DNA repair polymerases might contribute to this step *in vivo* (Smith et al., 2005).

To address this question, I used artificial DNA or RNA substrates corresponding to exon 2 (denoted E2 DNA or RNA) with an annealed DNA primer (primer e2) corresponding the 3' end generated by En cleavage of the bottom strand (Figure 3.10). The structure thus resembles that generated by En cleavage of the target DNA during retrohoming in which LtrA must copy a 9-nt 5' overhang before extending into the intron RNA.

As expected from previous studies showing that LtrA has low DNA-dependent DNA polymerase activity (Smith et al., 2005), the 5' overhang of the physiologically relevant exon 2 DNA template was copied inefficiently by LtrA, while that of the exon 2 RNA template was copied efficiently (Figure 3.10, lanes 3 and 4, respectively). The lane for exon 2 RNA, but not exon 2 DNA, also shows larger bands resulting from template switching to a second and third exon 2 template (Figure 3.10, lane 4 bands c and d). Native gel analysis showed that similar amounts of DNA and RNA template used in 3'-junction assays are annealed to primer e2 under the reaction conditions (Figure 3.7).

After cloning and sequencing of the cDNA product bands, I found that the majority of 5'-overhang products for exon 2 DNA (band a) extend to the 5' end of the overhang and lack non-templated nucleotide residues at their 3' end (Figure 3.11, band a). In contrast, all 5'-overhang products for exon 2 RNA (band b) had NTNA at the 3' end of the cDNA of up to 7 residues (Figure 3.11, band b). Notably, I found that unlike the 5'-junction substrate, all the non-templated sequences added with this substrate began with three C-residues. The products resulting from template switching to the second and

third copies of the exon 2 RNA template had similar profiles of non-templated nucleotide residues, mostly C-residues, both at template-switching junctions and at the 3' end of the cDNAs (Figure 3.11, bands c and d). The different sequence composition of non-templated nucleotide residues with these substrates compared to those for the 5' junction, where the non-templated nucleotides were mostly A-residues, presumably reflect differences in the terminal nucleotide residues of the template or cDNA product strand, as has been observed for other RTs and DNA polymerases (Hu, 1993, Magnuson et al., 1996, Golinelli & Hughes, 2002b). The 3'-terminal nucleotide residues of the cDNA product could influence NTNA by engaging in base-stacking interactions that favor some incoming nucleotides over others for non-templated addition to the 3' end of cDNAs (Golinelli & Hughes, 2002b).

3.3.2 Assays with exon 2 substrates under near-physiological salt conditions

Experiments with exon 2 DNA and RNA substrates were also done under near-physiological salt conditions (100 or 200 mM KCl, 5 mM $MgCl_2$) to test the generality of template utilization in 3'-junction assays (Figure 3.12). As with 5'-junction experiments, overall LtrA polymerase activity was reduced under low salt compared with standard high salt conditions (450 mM NaCl, 5 mM $MgCl_2$). More efficient cDNA synthesis opposite the exon 2 RNA template compared with the exon 2 DNA template was evident in each salt condition tested.

I also analyzed the sequences of cDNA products resulting from a single template switch to exon 2 RNA in 100 mM KCl, 5 mM $MgCl_2$. I found addition of up to two non-templated A-residues between template switches and at the 3' ends of cDNAs in a minority of products (Figure 3.13). This NTNA sequence composition differs from those observed with this substrate in high salt conditions where all cDNA products had NTNA

initiating with three C-residues between template switches and at the 3' end of cDNAs (Figure 3.7). In contrast with low salt conditions, reactions done in high salt also had substantially longer tracts of NTNA of up to 7 residues (Figure 3.7).

3.3.3 Assays with Ll.LtrB-exon 2 DNA substrates

Retrohoming of both lariat and linear intron RNPs involve steps at which the reverse transcriptase must transition between RNA and DNA templates. Because these transitions have not been well characterized, I next tested the ability of LtrA to initiate DNA synthesis from the staggered double-strand break and continue into the 3' end of the intron. In these assays I used the exon 2 DNA substrate with appended sequences corresponding to the 3' end of the intron. In one substrate the appended intron sequence was RNA (Ll.LtrB RNA-E2 DNA), as it would be *in vivo*, while in the other substrate the appended intron sequence was DNA (Ll.LtrB DNA-E2 DNA) (Figure 3.8, lanes 6 and 5, respectively). LtrA exhibits similar activity on both substrates to yield major products of the size expected for copying the 5'-exon 2 overhang followed by extension to the end of the appended intron sequence, as well as slightly larger products expected for NTNA to the 3' end of the cDNA (Figure 3.8, bands e-h). The identities of these cDNAs were confirmed by sequencing, which showed that the non-templated nucleotides at the 3' ends of these products were mostly A-residues (Figure 3.14). The gel lanes for both substrates also show light bands of 160-170 nt, the size expected for template switching to a second Ll.LtrB-E2 DNA template (Figure 3.8, lanes 5 and 6). Notably, the activity on both the DNA and chimeric RNA/DNA substrates is comparable to that for the exon 2 DNA without an appended intron sequence (Figure 3.8, compare lanes 5 and 6 with lane 3), suggesting that the reaction is rate-limited by the ability of LtrA to initiate on the exon 2 DNA template. After copying the exon 2 DNA segment, LtrA appears to copy the

appended intron RNA or DNA sequence equally well, with few premature stops and most of these occurring in the RNA rather than the DNA segment of the chimeric substrate.

3.3.4 Time-course assays with Ll.LtrB-exon 2 DNA substrates

To confirm inferences about the rate-limiting step of LtrA activity on the Ll.LtrB-exon 2 DNA substrates, I performed time-course experiments. In these assays LtrA showed similar rate for synthesis of full-length products opposite both the DNA and chimeric RNA/DNA substrates (Figure 3.15). Importantly, the time-course experiments show that even at the earliest time point (1 min), the major bands for both substrates correspond to full-length product without smaller intermediates, except for light bands corresponding to premature stops in the RNA segment of the chimeric substrate. This finding suggests that the rate-limiting step for both substrates is primer-dependent initiation on the exon 2 DNA and that after initiation, LtrA extends rapidly to the 5' ends of both the DNA and chimeric RNA/DNA templates. The time courses also show that NTNA for both substrates occurs more slowly than templated DNA synthesis, as observed previously for the HIV-1 RT (Golinelli & Hughes, 2002b). Together, these findings suggest that the ability of LtrA to copy DNA templates is limited by inefficient primer-dependent initiation and not by an inherent lack of DNA-dependent DNA polymerase activity.

3.4 DISCUSSION

3.4.1 LtrA activity at the 5' junction

First, I used model oligonucleotide substrates to investigate how LtrA's biochemical activities might contribute to events at the 5' junction during linear intron retrohoming. By using an RNA oligonucleotide corresponding to the 5' end of the linear intron RNA with an annealed DNA primer representing the cDNA, I found that LtrA

readily extends the primer to the end of the RNA and then adds up to three non-templated nucleotide residues in some cases. Most importantly, I found that upon reaching the 5' end of the intron RNA, LtrA is capable of template switching to either DNA or RNA oligonucleotides representing the 5' exon and extends opposite these substrates to the end of the molecule, allowing further addition of non-templated nucleotides. Template switching to these exon 1 oligonucleotides is not blocked by the addition of a complementary oligonucleotide with or without a 5' overhang, as would be formed during the staggered cleavage by LtrA during retrohoming. As both the exon 1 DNA and RNA oligonucleotides are shown to be double-stranded under these reaction conditions (Figure 3.7), this finding indicates that LtrA can displace a complementary strand, as described for other RTs (Kurzynska-Kokorniak et al., 2007, Liu et al., 2008). However, these model duplex substrates are relatively short, and it is possible that LtrA would have greater difficulty displacing a complementary strand *in vivo*.

I observed one instance of template switching from the exon 1 RNA to the Ll.LtrB RNA induced by complementarity (Figure 3.5 band k). It appears that incomplete cDNA synthesis opposite an exon 1 RNA template generated a 3'-terminal AT in the cDNA which base-paired with an internal 5'-…AU…-3' in the Ll.LtrB RNA, which was then copied to the end of the molecule.

NTNA may aid in the ability of LtrA to template switch by generating 3' overhangs on the cDNA product with complementarity to the acceptor oligonucleotide. Base pairing could stabilize the cDNA-acceptor oligonucleotide interaction, at which point LtrA would resume templated cDNA synthesis. Studies with the HIV-1 RT have shown that complementarity generated by NTNA can induce specific template-switching reactions *in vitro* (Golinelli & Hughes, 2002a). In my experiments, template switching by base-pairing of non-templated nucleotides present at the 3' end of a cDNA product is not

apparent from the cDNA sequence, since these products cannot be distinguished from a seamless template switch and subsequent templated cDNA extension.

Evidence of template switching without initial extension of primer c is observed in several cDNA clones (Figures 3.4-3.6). This presumably results from an interaction of the Ll.LtrB RNA/primer c hybrid with LtrA prior to switching to a second molecule to be used as a template, as control reactions including primer c without the annealed Ll.LtrB RNA did not form reacted products (Figure 3.1 lanes 3 and 4).

Considering the cDNA sequence data acquired from LtrA template switching to exon 1 DNA *in vitro*, it is unlikely that LtrA activity is a major contribution to the formation of 5'-linear intron retrohoming junctions *in vivo*. A minority of template-switching reactions formed *in vitro* contain additional inserted nucleotides. Further, these insertions are composed entirely of A-residues in 10 of 12 instances characterized, and never exceed two nucleotides in length (Figures 3.2 and 3.9). In contrast, the majority of 5'-linear intron retrohoming junctions formed in plasmid targeting experiments in wild-type embryos have additional nucleotides inserted at the junction. Only 13 of 93 insertion events are mono- or di-nucleotides and in no cases are these one or two A-residues.

Furthermore, the presence of inserted nucleotides apparently templated from sequences proximal to the DSB is sometimes observed at the 5'-junction insertions formed *in vivo* (see underlined residues in Figures 2.3, 2.4 and 2.7). Such insertions have been attributed to eukaryotic repair polymerases involved in end-joining reactions (Jager et al., 2000, Adams et al., 2003, McVey et al., 2004, Yu & McVey, 2010). In my experiments, LtrA has not been found to insert nucleotides prior to switching templates that appear to be templated from proximal sequences. Nor has such an activity been described for other retroviral or retroelement RTs studied to date. Thus, although LtrA may be responsible for template switches characterized by small insertions, it is unlikely

to have contributed to the formation of 5'-retrohoming junctions containing extensive insertion sequences (up to 86 nts, Figure 2.4). Retrohoming experiments described in Table 2.3 and Figure 2.10 indicate PolQ is the repair polymerase that contributes to the majority of additional nucleotides inserted at the 5' junction during NHEJ repair.

In biochemical assays template switching to internal sites within exon 1 RNA oligonucleotides was observed, yet no instances of template switching to an internal site was seen for exon 1 DNA oligonucleotides. Exon 1 deletions associated with linear intron retrohoming *in vivo* are likely the result of 5' to 3' exon 1 DNA resection by eukaryotic factors at the DSB and/or base-pairing between 5'-exon DNA sequences and intron cDNA at microhomologies and subsequent "trimming" of DNA overhangs (Wyman & Kanaar, 2006). Reactions exploring LtrA activities at the 5' junction under near-physiological KCl conditions show little effect on polymerase or template-switching activities of LtrA. Thus, the intracellular environment is not expected to alter the activity of this enzyme in a manner that would drastically increase NTNA following reverse transcription to generate the 5' junctions with large insertions observed during linear intron retrohoming *in vivo*.

3.4.2 LtrA activity at the 3' junction

My experiments using model oligonucleotide substrates to probe LtrA function at the 3' retrohoming junction used a 40-nt exon 2 DNA or RNA template with an annealed DNA primer that ends at the exon 2 +9 position corresponding to the site of DNA endonuclease cleavage. I found that LtrA extends an annealed DNA primer opposite an RNA substrate much more efficiently than opposite a DNA substrate. This result confirms previous findings for LtrA on cognate Ll.LtrB RNA and DNA substrates (Smith et al., 2005). In high salt conditions, most of the reacted primer is extended beyond the

RNA template due to NTNA, and in all cases, these non-templated nucleotides begins with three C-residues. This phenomenon is likely sequence dependent, as the sequence present at the 3' end of the cDNA has been seen to influence the rate of NTNA of a particular dNTP relative to others for retroviral RTs (Golinelli & Hughes, 2002b). Following the addition of CCC, NTNA displays a predominance of purine residues, as has been observed for LtrA in 5'-junction assays and other retroelement RTs (Kennell et al., 1994, Chiang & Lambowitz, 1997, Kurzynska-Kokorniak et al., 2007).

Assays using oligonucleotides in which a 20-nt RNA or DNA sequence corresponding to the 3' intron was appended to the exon 2 DNA gave similar results. The extent and rate of reaction to reach the 5' end of the intron sequence were comparable to that with the 40-nt exon 2 DNA substrate without the intron RNA or DNA sequence (Figure 3.8). Time-course assays reveal similar rates for completion of cDNA synthesis regardless of whether the appended intron sequence was DNA or RNA, with full-length product appearing in equal amounts at the earliest time point (1 min) in both cases (Figure 3.15). Notably, few products of intermediate size resulting from incomplete extension are evident, with most appearing within the RNA portion of the chimeric substrate. Initiation of LtrA polymerase activity opposite the exon 2 DNA portion of the substrate appears to limit the overall reaction. Once initiated, LtrA exhibits equal activity in copying the template regardless of whether it is composed entirely of DNA or has an RNA component at the 5' end. Thus, although LtrA exhibits inefficient primer-dependent initiation opposite DNA substrates, it is not compromised in its ability to copy DNA templates.

The latter finding, that LtrA can efficiently copy DNA templates, agrees with what was observed in 5'-junction assays where initiation of cDNA synthesis began opposite the Ll.LtrB RNA substrate and template switching occurred to exon 1 DNA or

RNA with equal efficiency (Figure 3.1 lanes 5 and 6). These findings indicate that template switching enables more efficient initiation on DNA templates than does initiation using a base-paired oligonucleotide.

My experiments show that LtrA can initiate cDNA synthesis at the 3' exon 5' overhang produced upon Ll.LtrB intron retrohoming. However, due to inefficient primer-dependent initiation opposite DNA substrates, I cannot exclude that, *in vivo*, LtrA dissociates and the 5' overhang is copied by a cellular DNA polymerase prior to LtrA resuming on the reverse-spliced intron RNA.

3.5 Materials and Methods

3.5.1 Recombinant plasmids

pMAL-LtrA, used for expression of the LtrA protein for biochemical assays, contains the LtrA ORF (Mills et al., 1996) cloned downstream of a tac promoter and Φ10 Shine-Dalgarno sequence between BamHI and HindIII of the protein-expression vector pMAL-c2 (New England Biolabs, Ipswich, MA). The latter is a derivative of pMal-c2x (New England Biolabs, Ipswich MA) with a TEV protease-cleavage site in place of the factor Xa site (Kristelly et al., 2003). LtrA is expressed from this plasmid as a fusion protein with an N-terminal tag containing a maltose-binding protein domain, enabling LtrA purification via an amylose-affinity column.

3.5.2 Preparation of LtrA protein

The LtrA protein was expressed in *E. coli* BL21(DE3) from the plasmid pMAL-LtrA. A starter culture of cells was grown in LB medium overnight at 37°C and used to inoculate ultra yield flasks containing 0.5 L of LB medium, which were autoinduced by growing at 37°C for 3 h, followed by 18°C for 24 h (Studier, 2005). Cells were harvested by centrifugation (Beckman JLA-8.1000; 4,000 x g, 15 min, 4°C), resuspended in 1 M

NaCl, 20 mM Tris-HCl pH 7.5, 20% glycerol, and 0.1 mg/ml lysozyme (Sigma-Aldrich, St. Louis, MO) and kept for 15 min on ice. Cells were lysed by 3 freeze-thaw cycles on dry ice followed by sonication (Branson 450 Sonifier, Branson Ultrasonics, Danbury CT; three or four 10 sec bursts on ice at an amplitude of 60%, with 10 sec between bursts). After pelleting cell debris (Beckman JA-14 rotor, 10,000 rpm, 30 min, 4°C), nucleic acids were precipitated from the supernatant with 0.4% polyethylenimine (PEI) and constant stirring for 20 min at 4°C, followed by centrifugation (Beckman JA-14 rotor, 14,000 rpm, 30 min, 4°C). Proteins were then precipitated from the supernatant by adding ammonium sulfate to 50% saturation with constant stirring for 1 h at 4°C. The precipitated protein was pelleted (Beckman JA-14 rotor, 14,000 rpm 30 min, 4°C) and dissolved in 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol. The protein was applied to a 10-ml amylose column (Amylose High-Flow resin; New England Biolabs, Ipswich, MA), which was washed with 3 column volumes of 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol and eluted with 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol containing 10 mM maltose. Fractions containing MalE-LtrA were incubated with TEV protease (80 μg/ml) for 18 h at 4°C. These fractions were further purified from the TEV protease by FPLC through a Ni-NTA column loaded with 40 mM imidazole, washed with 3 column volumes of 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol, 40 mM imidazole, and eluted in 500 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol, 300 mM imidazole. Monomeric LtrA was further purified by FPLC through a column with heparin Sepharose (New England Biolabs). The purified protein was then concentrated to 30 μM and exchanged into 100 mM NaCl, 20 mM Tris-HCl pH 7.5, 10% glycerol by dialysis.

3.5.3 LtrA biochemical assays

Biochemical assays were done by incubating purified LtrA protein with small artificial oligonucleotide substrates (see below) in 20 µl in 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 1 mM dithiothreitol (DTT) and 200 µM dNTPs. The reaction components were assembled on ice with substrate added last and then incubated at 30°C for 30 min. Reactions were terminated by phenol-CIA extraction. For 3'-junction time-course assays, all reagents excluding protein were added to the 200-µl reactions. The temperature was then raised to 30°C, and reactions were started by adding LtrA. Time-course reactions were terminated by adding EDTA pH 7.5 to 20-µl portions of the reaction to a final concentration of 25 mM, followed by phenol-CIA extraction. Portions of the reaction product (3 µl) were added to an equal volume of gel loading buffer II (95% formamide, 18 mM EDTA and 0.025% each of SDS, xylene cyanol, and bromophenol blue; Ambion, Austin, TX), denatured at 98°C for 7 min, and loaded onto a pre-warmed denaturing 10 or 15% polyacrylamide gel. The wet gels were placed between cellophane and saran wrap, exposed to a phosphor screen, and scanned with a phosphorimager.

5'-junction assays used Ll.LtrB RNA [LtrB5'S20Anchor6,5 RNA] (5'-GUGCGCCCAGAUAGGGUGUUCUCGUUGGCAAUGGUGUCCAACUUGUGCUGCCAGUGCUCG) with annealed primer c (5'- CGAGCACTGGCAGCACAAG-deoxyuridine-TGGACACCATTGCCAACGAGAACAC); exon 1 DNA (5'-TGTGATTGCAACCCACGTCGATCGTGAACACATCCATAAC) or RNA (5'-UGUGAUUGCAACCCACGUCGAUCGUGAACACAUCCAUAAC). Oligonucleotides complementary to exon 1 DNA or RNA were: exon 1 AS (5'-GTTATGGATGTGTTCACGATCGACGTGGGTTGCAATCACA) and exon 1 AS +9 (5'-AATGATATGGTTATGGATGTGTTCACGATCGACGTGGGTTGCAATCACA).

3'-junction assays used exon 2 DNA (5'-CATATCATTTTAATTCTACGAATCTTTATACTGGCAAAC), exon 2 RNA (5'-CAUAUCAUUUUUAAUUCUACGAAUCUUUAUACUGGCAAAC), Ll.LtrB DNA-E2 DNA (5'-GCGGTACCTCCCTACTTCACCATATCATTTTTAATTCTACGAATCTTTATACTGGCAAAC), Ll.LtrB RNA-E2 DNA chimeric oligonucleotide (5'-RNA[GCGGUACCUCCCUACUUCAC]DNA[CATATCATTTTTAATTCTACGAATCTTTATACTGGCAAAC]), and annealed primer e2 (5'-CATCTGGCGGCTGTTCTCG-deoxyuridine-TGGACACCATTGCCAACGAGGTTTGCCAGTATAAAGATTCGTAGAATTAAA).

DNA and RNA oligonucleotides were obtained from Integrated DNA Technologies (IDT; Coralville, IA) and purified in a denaturing 10% (w/v) polyacrylamide gel by visualizing the oligonucleotide using UV shadowing, excising the band, freezing in an Eppendorf tube at -80°C for 10 min, crushing in the tube, adding 600 µl of 500 mM NH_4Cl, 0.1 mM EDTA, 10 mM MOPS pH 6.5 and 0.1% SDS and shaking at 4°C overnight. The oligonucleotides were separated from acrylamide fragments using Costar Spin-X centrifuge tube filters, 0.45 µm pore size (Corning Inc, Lowell, MA), then ethanol precipitated in the presence of linear acrylamide carrier (58 µg/ml) and dissolved in nuclease-free water. DNA primers were 5'-end labeled with [γ-^{32}P]-ATP (10 Ci/mmol; Perkin-Elmer) by using phage T4 polynucleotide kinase (New England Biolabs) according to the manufacturer's protocol. For annealing of primers or complementary strands, oligonucleotides were mixed at 20x the concentration used in RT assays, then heated to 82°C and slowly cooled to 25°C for 45 min in 1x annealing buffer (100 mM Tris-HCl pH 7.5 and 5 mM EDTA). The efficiency of annealing was assessed by electrophoresis in a non-denaturing 6% polyacrylamide gel containing Tris-borate-EDTA

(90 mM Tris, 90 mM boric acid, 2 mM EDTA) at 30ºC (Sambrook, 1989).

3.5.4 Cloning and sequence of cDNA products

For cloning and sequencing of cDNAs synthesized with the Ll.LtrB RT, the cDNA products were gel-purified from a denaturing 10% (w/v) polyacrylamide gel slices by excising the band, freezing in an Eppendorf tube at -80ºC for 10 min, crushing in the tube, adding 600 µl of 500 mM NH_4Cl, 100 µM EDTA, 10 mM MOPS pH 6.5 and 0.1% SDS, and incubating at 4°C overnight. The oligonucleotide was separated from gel fragments by using Costar Spin-X centrifuge tube filters, 0.45 µm pore size (Corning Inc, Lowell, MA), ethanol precipitated in the presence of linear acrylamide carrier (58 µg/ml), and dissolved in nuclease-free water. The cDNAs were circularized using CircLigase I or II (Epicentre Biotechnologies, Madison, WI), treated with exonuclease I (Epicentre Biotechnologies), and linearized with uracil-DNA excision enzyme mix (Epicentre Biotechnologies), all according to manufacturer's instructions with excision buffer at 0.5x concentration to keep the EDTA concentration low enough for PCR. For the experiment in Figures 3.1 and 3.4, the linearized products were PCR amplified by using Phusion High Fidelity PCR Master Mix with HF buffer (New England Biolabs, Ipswich, MA) with the primers Anchor 6 complement (5'- CTTGTGCTGCCAGTGCTCG) and Anchor 5 (5'-TGGACACCATTGCCAACGAG). For the experiment in Figures 3.6 and 3.8, the linearized products were PCR amplified similarly with the primers Anchor 4 complement (5'-CGAGAACAGCCGCCAGATG) and Anchor 5 (see above). PCRs were done in 50 µl of reaction medium Phusion High Fidelity PCR Master Mix with HF buffer (New England Biolabs) with the following cycling conditions: 98°C initial denaturing for 2 min, 25 cycles of 98°C for 10 s, 60°C for 10 s, 72°C for 5 s, and a final extension at 72°C for 7 min. PCR products were resolved in a 2% agarose and gel purified with

MinElute Gel Extraction Kit (Qiagen) prior to cloning into the TOPO-TA pCR2.1 vector (Invitrogen) according to the manufacturer's protocol. Random colonies were picked and the cloned PCR products were amplified by colony PCR using Phusion High Fidelity PCR Master Mix with HF buffer with primers M13 F(-20) (5'-GTAAAACGACGGCCAGT) and M13 R(-26) (5'-CAGGAAACAGCTATGAC), then sequenced using the M13 Rev(-24) (5'-GGAAACAGCTATGACCATG) primer.

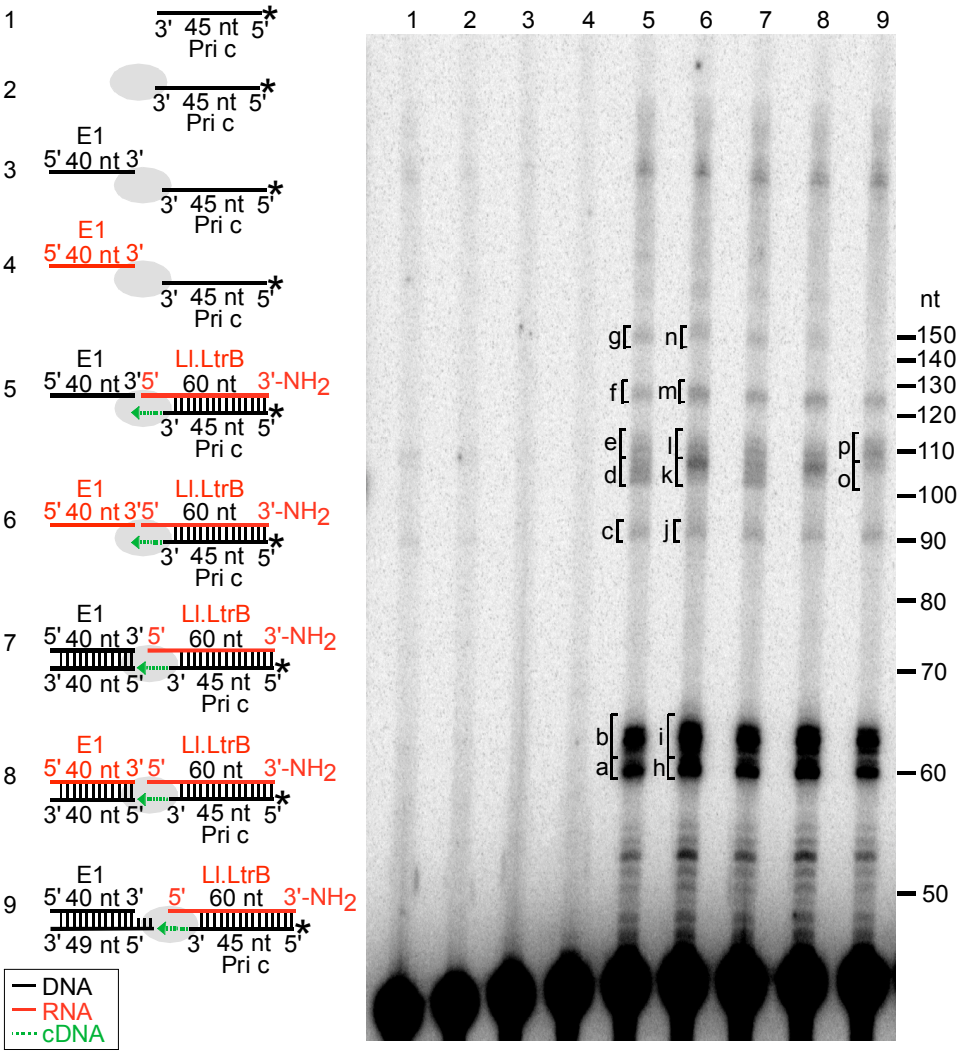


Figure 3.1: Biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 DNA or RNA

The LtrA protein (40 nM) was incubated with small artificial substrates (44 nM 5'-^{32}P-labeled DNA primer c (Pri c; 45 nt) by itself or annealed to 40 nM Ll.LtrB RNA (60 nt) and 40 nM exon 1 (E1; 40 nt) DNA or RNA as diagrammed to the left) in reaction medium containing 200 µM dNTPs, 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, and 1 mM dithiothreitol (DTT) for 30 min at 30°C. The reaction was terminated by phenol-CIA extraction, and the products were analyzed in a denaturing 15% polyacrylamide gel. Lanes (1) and (2) ^{32}P-labeled Pri c incubated without and with LtrA, respectively; (3) and (4) LtrA incubated with ^{32}P-labeled Pri c and E1 DNA or RNA, respectively; (5) and (6) LtrA incubated with Ll.LtrB RNA with annealed ^{32}P-labeled Pri

c and E1 DNA or RNA, respectively; (7-10) LtrA incubated with Ll.LtrB RNA with annealed ^{32}P-labeled Pri c and E1 DNA or RNA with annealed complementary DNA oligonucleotides to create blunt or recessed 3' ends, as shown in the schematics to the left. Bands excised for sequencing are indicated in the gel. DNA and RNA oligonucleotides are shown in black and red, respectively; LtrA is shown as a gray oval; the direction of cDNA synthesis is indicated by the dotted green arrow. The numbers to the right of the gel indicate the nucleotide position of the 5'-^{32}P labeled 10-bp ladder (Invitrogen).

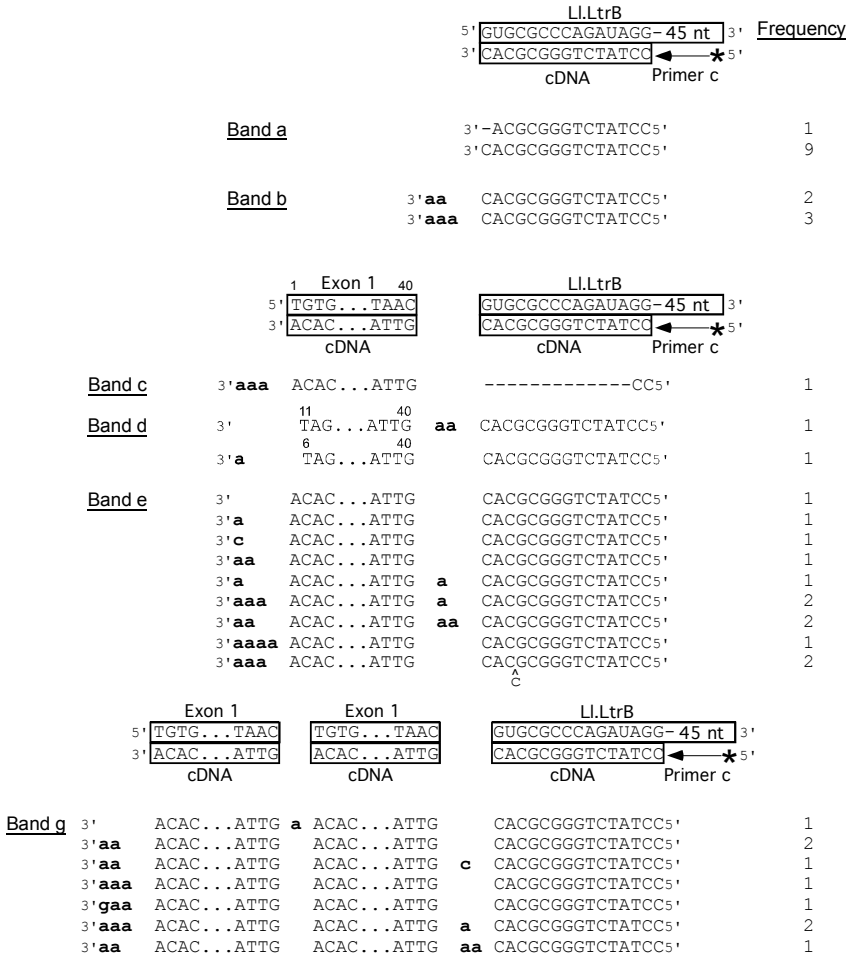


Figure 3.2: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 DNA

Products were obtained from the indicated gel bands in Figure 3.1 lane 5. These products result from extension of primer c to the 5' end of the Ll.LtrB RNA in the Ll.LtrB RNA template/DNA primer c substrate and subsequent template switching to exon 1 DNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Portions of the cDNA product sequences not shown in the figure included one G to A transition. * denotes ^{32}P-label at 5' end of primer c.

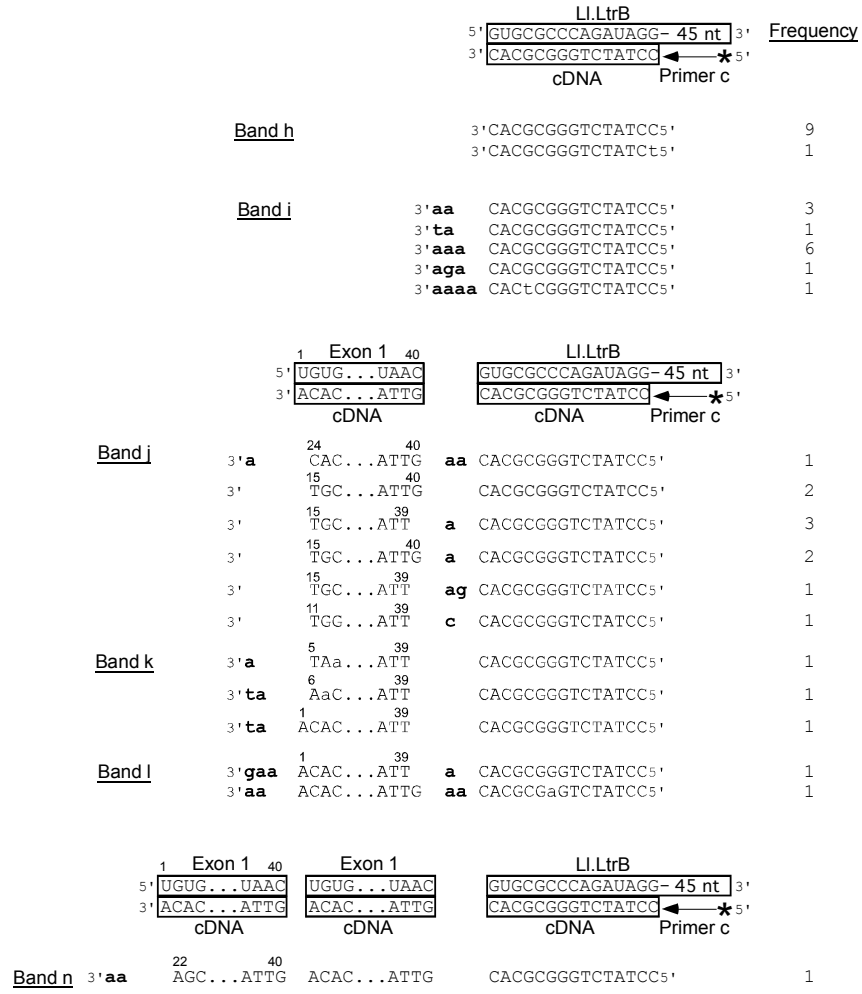


Figure 3.3: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 RNA

Products were obtained from the indicated gel bands in Figure 3.1 lane 6. These products result from extension of primer c to the 5' end of the Ll.LtrB RNA in the Ll.LtrB RNA template/DNA primer c substrate and subsequent template switching to exon 1 RNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Portions of the cDNA product sequences not shown in the figure included two A to G transitions. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer c.

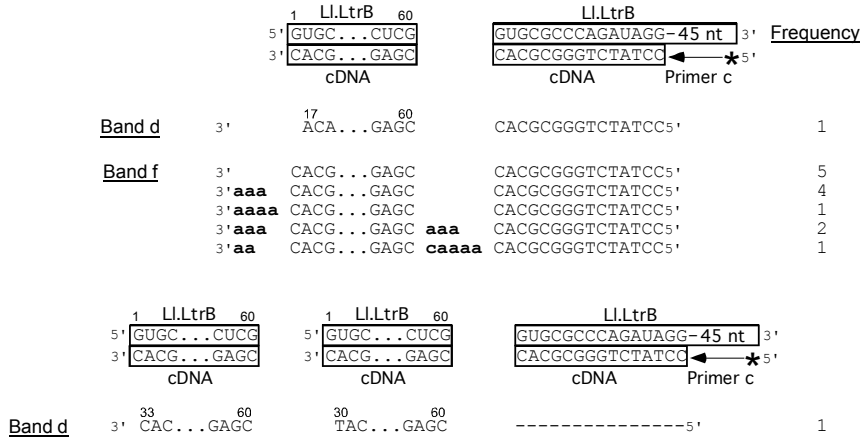


Figure 3.4: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to Ll.LtrB RNA in exon 1 DNA template switching assays

Products obtained from the indicated gel bands in Figure 3.1 lane 5, with or without extension of primer c in the Ll.LtrB RNA template/DNA primer c substrate and subsequent template switching to Ll.LtrB RNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer c.

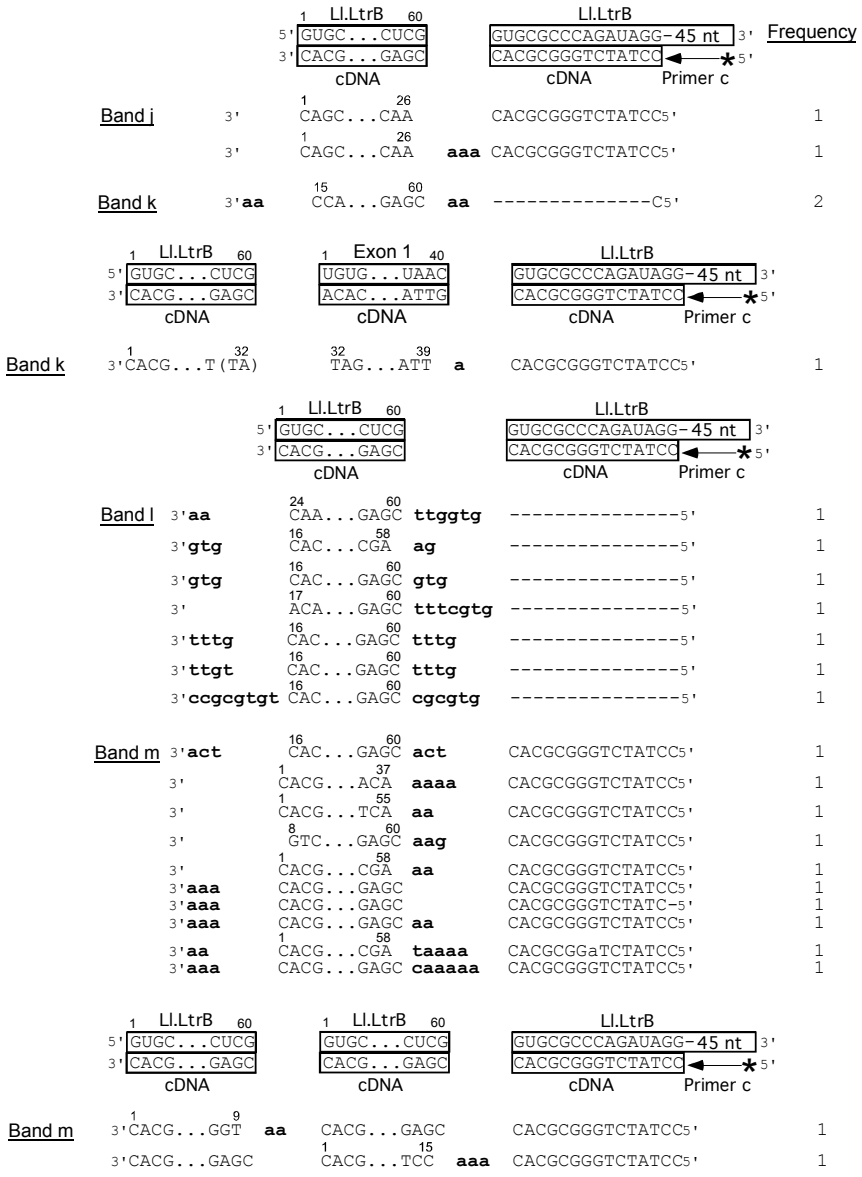


Figure 3.5: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to Ll.LtrB RNA in exon 1 RNA template-switching assays

Products obtained from the indicated gel bands in Figure 3.1 lane 6, with or without extension of primer c in the Ll.LtrB RNA template/DNA primer c substrate, and subsequent template switching to Ll.LtrB RNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally

determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. Microhomology at a template-switching junction is shown in parentheses. * denotes ^{32}P-label at 5' end of primer c.

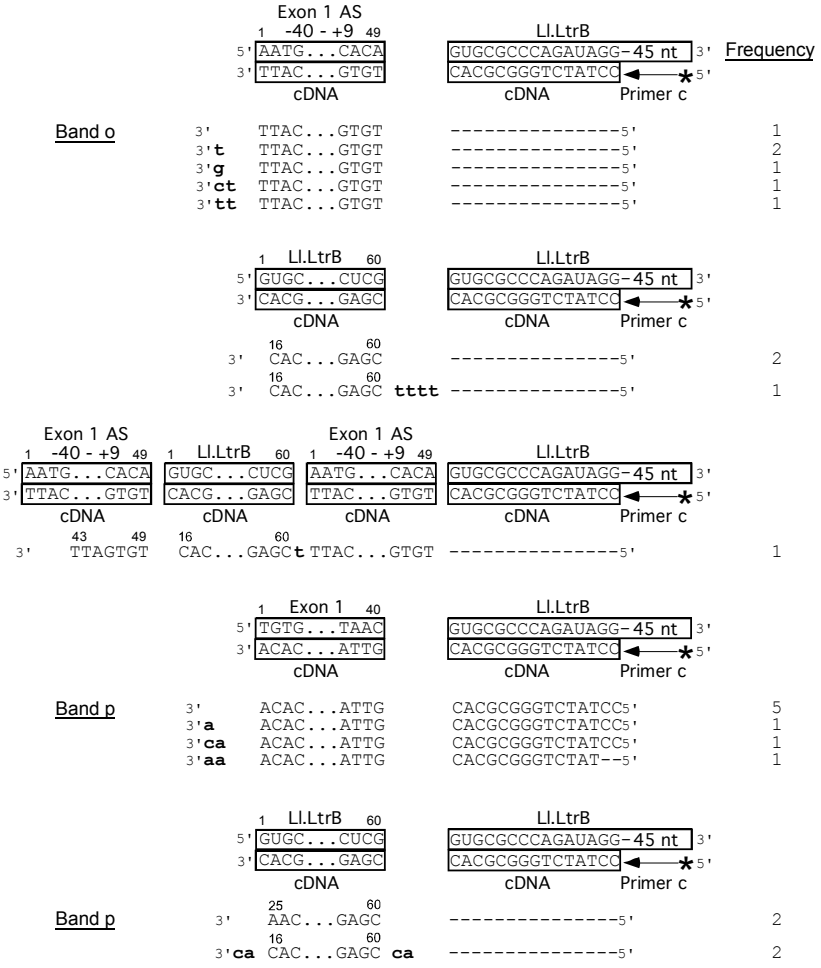


Figure 3.6: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 DNA, exon 1 AS +9 DNA, and Ll.LtrB RNA

Products obtained from the indicated gel bands in Figure 3.1 lane 9 with or without extension of primer c in the Ll.LtrB RNA template/DNA primer c substrate, and subsequent switching from the Ll.LtrB RNA template/DNA primer c substrate to exon 1 DNA, exon 1 AS +9 DNA, and the Ll.LtrB RNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer c.

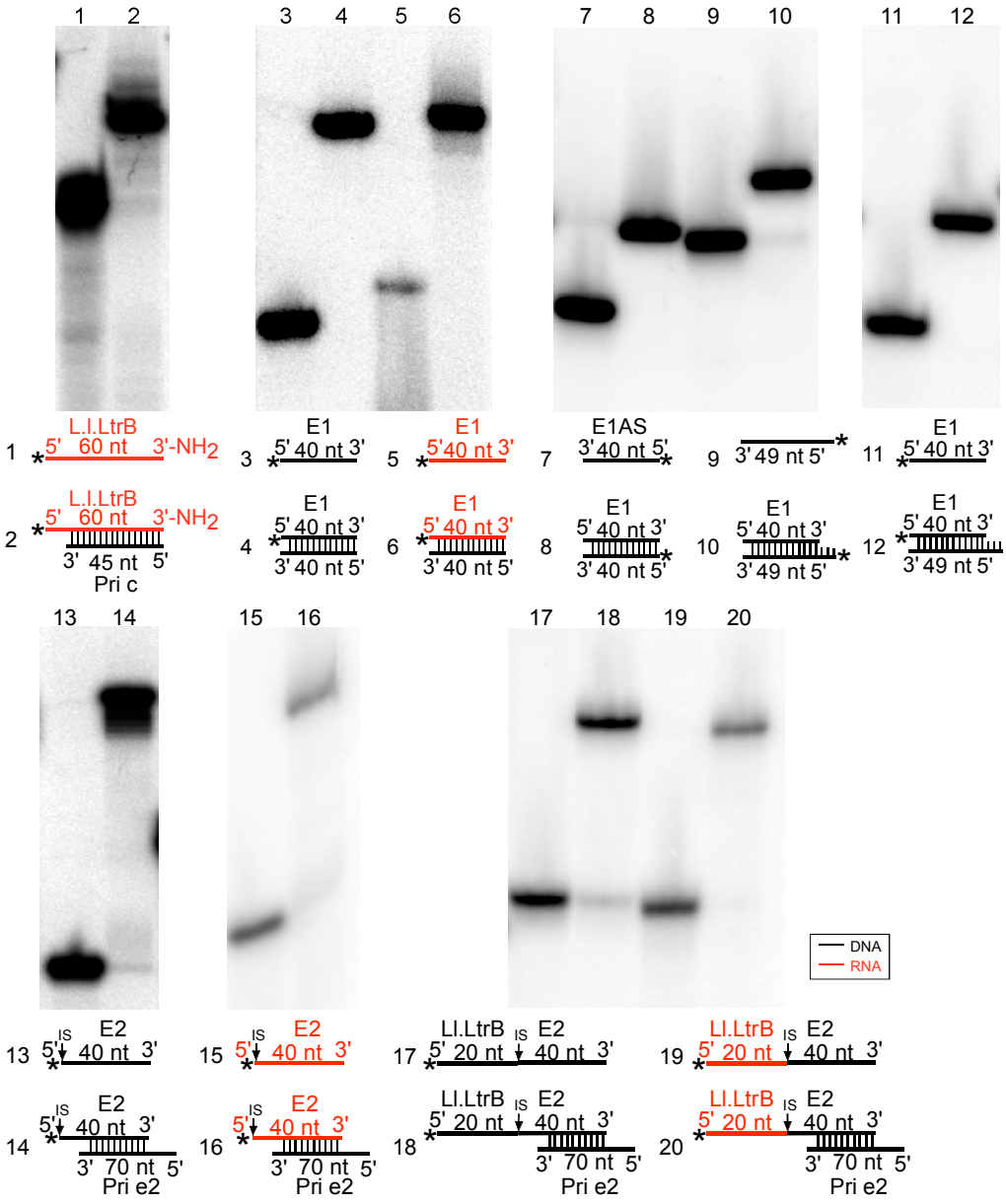


Figure 3.7: Native gel analysis of annealed oligonucleotides used in 5'- and 3'-intron integration biochemical assays

5'-^{32}P-labeled oligonucleotides by themselves or combined with the indicated complementary oligonucleotide were annealed as described in Materials and Methods (section 3.5.3) before a 1:20 dilution into 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5 and incubation for 30 min at 30°C. The samples were then mixed 6:1 with 30°C non-denaturing loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol and 1.5%

Ficoll 400) and assessed by electrophoresis in a non-denaturing 6% polyacrylamide gel containing Tris-borate-EDTA (90 mM Tris, 90 mM boric acid, 2 mM EDTA) at 30ºC (Sambrook, 1989). Gels were soaked for 15 min in 25% isopropanol, 20% glycerol and 10% acetic acid to prevent cracking during drying, dried, exposed to a phosphor screen, and scanned with a phosphorimager. DNA and RNA oligonucleotides are shown in black and red, respectively, diagrammed below the corresponding gel. Lanes (1) 40 nM ^{32}P-labeled Ll.LtrB RNA (2) 40 nM ^{32}P-labeled Ll.LtrB RNA incubated with 44 nM DNA primer c (Pri c) (3) 40 nM ^{32}P-labeled exon 1 DNA (4) 40 nM ^{32}P-labeled exon 1 DNA incubated with 40 nM exon 1 DNA AS (5) 40 nM ^{32}P-labeled exon 1 RNA (6) 40 nM ^{32}P-labeled exon 1 RNA incubated with 40 nM exon 1 DNA AS (7) 40 nM ^{32}P-labeled exon 1 DNA AS (8) 40 nM ^{32}P-labeled exon 1 DNA AS incubated with 40 nM exon 1 DNA (9) 40 nM ^{32}P-labeled exon 1 DNA AS +9 (10) 40 nM ^{32}P-labeled exon 1 DNA AS +9 incubated with 40 nM exon 1 DNA (11) 40 nM ^{32}P-labeled exon 1 DNA (12) 40 nM ^{32}P-labeled exon 1 DNA incubated with 40 nM exon 1 DNA AS +9 (13) 40 nM ^{32}P-labeled exon 2 DNA (14) 40 nM ^{32}P-labeled exon 2 DNA incubated with 44 nM primer e2 (Pri e2) (15) 40 nM ^{32}P-labeled exon 2 RNA (16) 40 nM ^{32}P-labeled exon 2 RNA incubated with 44 nM primer e2 (Pri e2) (17) 40 nM ^{32}P-labeled Ll.LtrB DNA-Exon 2 DNA (18) 40 nM ^{32}P-labeled Ll.LtrB DNA-Exon 2 DNA incubated with 44 nM primer e2 (Pri e2) (19) 40 nM ^{32}P-labeled Ll.LtrB RNA-Exon 2 DNA chimeric oligonucleotide (20) 40 nM ^{32}P-labeled Ll.LtrB RNA-Exon 2 DNA chimeric oligonucleotide incubated with 44 nM primer e2 (Pri e2).

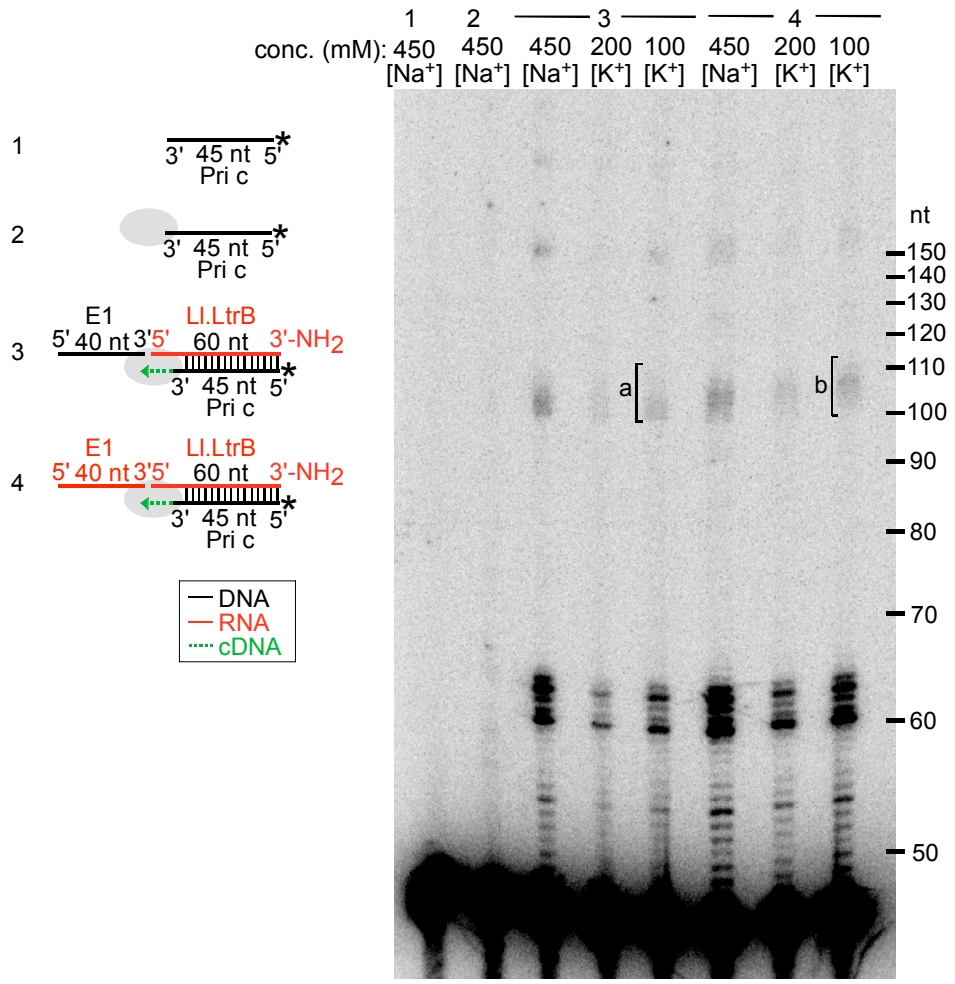


Figure 3.8: Biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 DNA or RNA under near-physiological KCl concentrations

The LtrA protein (40 nM) was incubated with small artificial substrates (44 nM 5'-^{32}P-labeled DNA primer c (Pri c; 45 nt) by itself or annealed to 40 nM Ll.LtrB RNA (60 nt) and 40 nM exon 1 (E1; 40 nt) DNA or RNA as diagrammed to the left) in reaction medium containing 200 μM dNTPs, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, and 1 mM dithiothreitol (DTT) in addition to 450 mM NaCl, 200 mM KCl, or 100 mM KCl for 30 min at 30°C. The reaction was terminated by phenol-CIA extraction, and the products were analyzed in a denaturing 10% polyacrylamide gel. Lanes (1) and (2) ^{32}P-labeled Pri c incubated without and with LtrA in 450 mM NaCl, respectively; (3-5) LtrA incubated with ^{32}P-labeled Pri c and E1 DNA in 450 mM NaCl, 200 mM KCl, and 100 mM KCl, respectively; (6-8) LtrA incubated with Ll.LtrB RNA with annealed ^{32}P-labeled Pri c and E1 RNA in 450 mM NaCl, 200 mM KCl, and 100 mM KCl, respectively. Bands excised

for sequencing are indicated in the gel. DNA and RNA oligonucleotides are shown in black and red, respectively; LtrA is shown as a gray oval; the direction of cDNA synthesis is indicated by the dotted green arrow. The numbers to the right of the gel indicate the nucleotide position of the 5'-^{32}P labeled 10-bp ladder (Invitrogen).

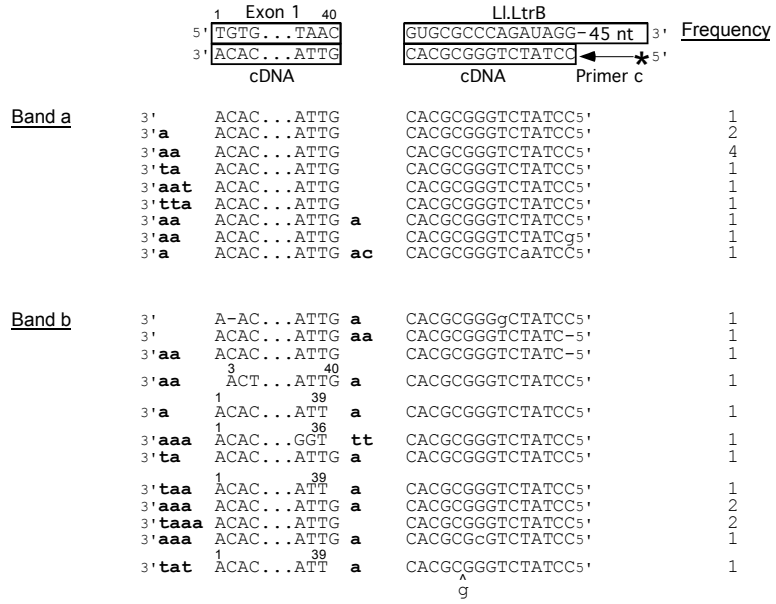


Figure 3.9: Sequence analysis of cDNA products from biochemical assays of LtrA template switching from the 5' end of linear intron RNA to exon 1 DNA or RNA under near-physiological KCl concentrations

Products were obtained from the indicated gel bands (a and b) in the gel in Fig 3.8 lanes 3 and 6, respectively. These products result from extension of primer c to the 5' end of the Ll.LtrB RNA in the Ll.LtrB RNA template/DNA primer c substrate and subsequent template switching to exon 1 DNA or RNA. Bands were excised from the gel, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequence (boxed) is shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes the ^{32}P-label at 5' end of primer c.

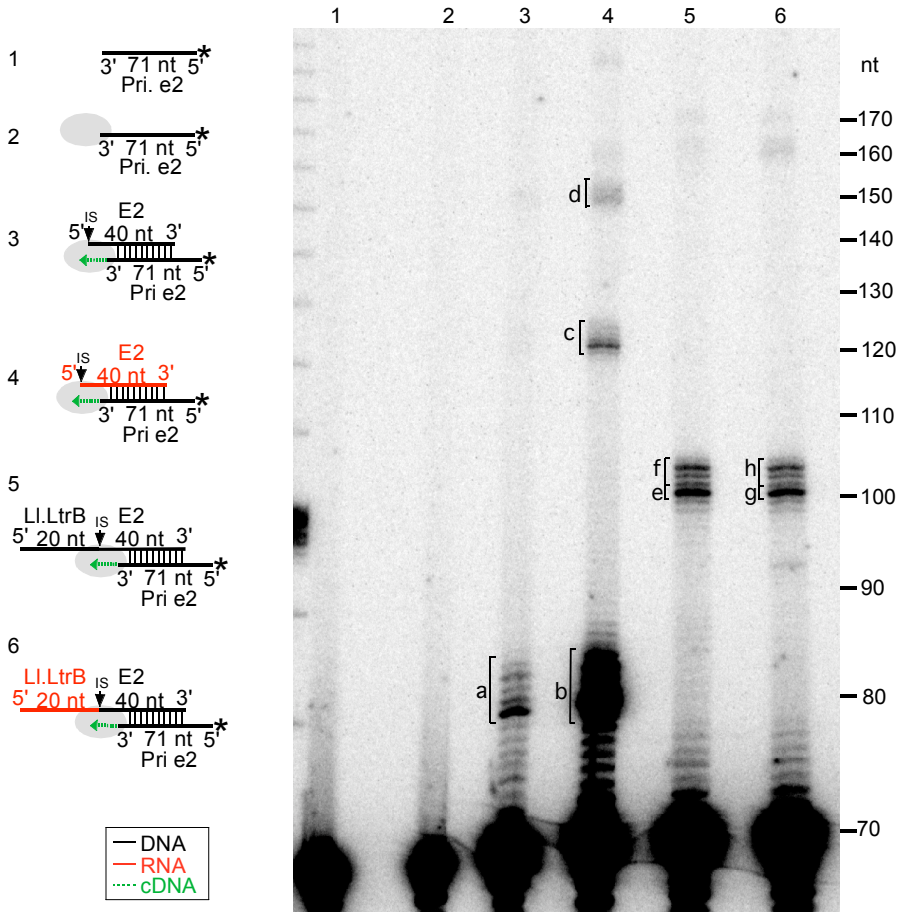


Figure 3.10: Biochemical assays of LtrA DNA polymerase and template-switching activities at the 3'-intron integration junction

The LtrA protein (40 nM) was incubated with small artificial substrates (44 nM 5'-^{32}P-labeled primer e2 (Pri e2; 70 nt)) by itself or annealed to (40 nM E2 RNA or DNA (40 nt), Ll.LtrB-E2 DNA or chimeric Ll.LtrB RNA-E2 DNA (60 nt) oligonucleotides, as diagrammed to the left) in reaction medium containing 200 μM dNTPs, 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, 1 mM dithiothreitol (DTT) for 30 min at 30°C. The reaction was terminated by phenol-CIA extraction, and the products were analyzed in a denaturing 10% polyacrylamide gel. Lanes (1) and (2) ^{32}P-labeled DNA primer e2 incubated without and with LtrA, respectively; (3) and (4) LtrA incubated with E2 DNA or RNA template with annealed ^{32}P-labeled Pri e2, respectively; (5) and (6) LtrA incubated with Ll.LtrB DNA-E2 DNA or chimeric Ll.LtrB RNA-E2 DNA template with annealed ^{32}P-labeled Pri e2, respectively. Bands excised for sequencing are indicated in the gel. DNA and RNA oligonucleotides are shown in black and red, respectively; the LtrA is shown as a gray oval; the direction of cDNA synthesis is indicated by the dotted

green arrow; the arrow indicates the intron-insertion site (IS). The numbers to the right of the gel indicate the nucleotide position of the 5'-^{32}P labeled 10-bp ladder (Invitrogen).

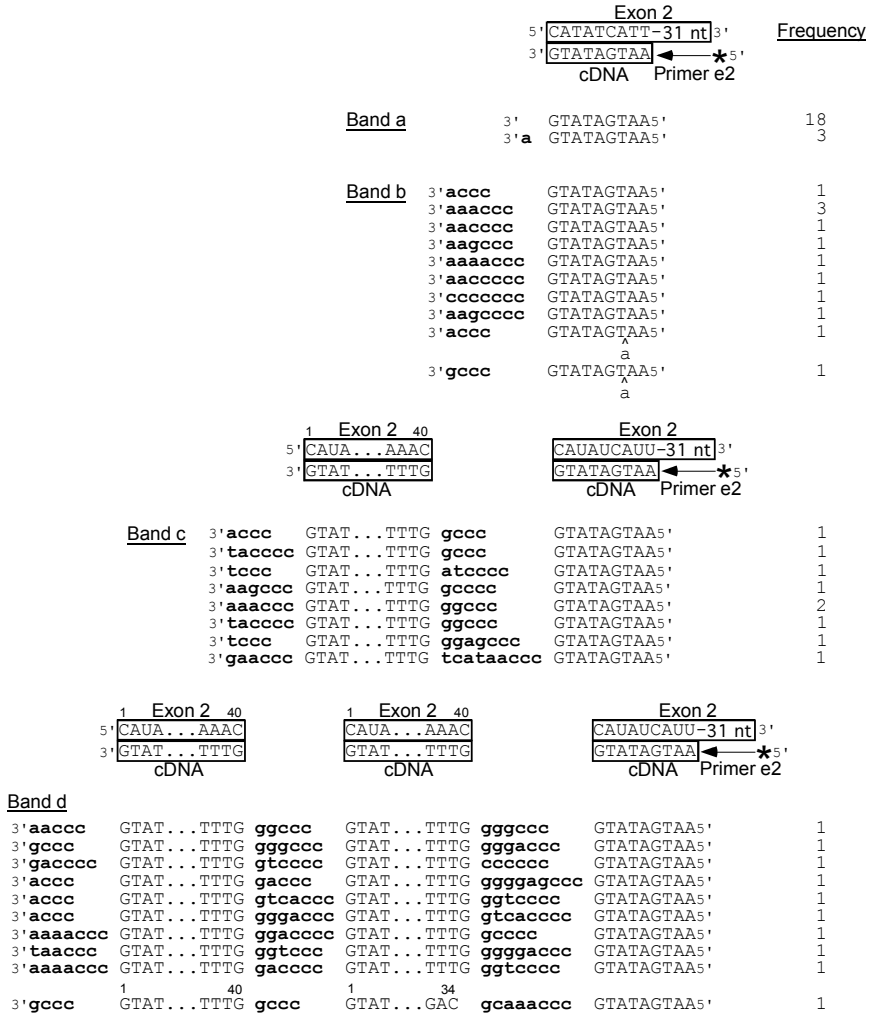


Figure 3.11: Sequence analysis of cDNA products from biochemical assays of LtrA DNA polymerase and template-switching activities on 3'-exon substrates

The bands indicated in the gel (a-d) in Figure 3.10 were excised, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequences (boxed) are shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer.

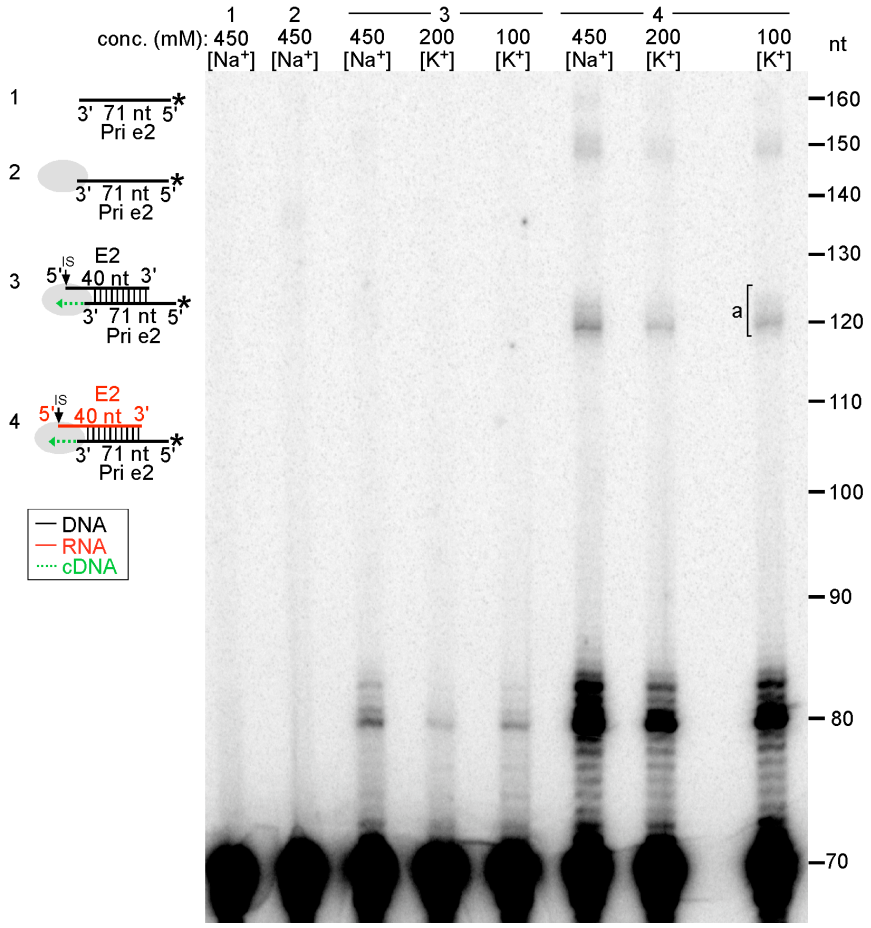


Figure 3.12: Biochemical assays of LtrA DNA polymerase and template-switching activities on 3'-exon substrates under near-physiological KCl concentrations

The LtrA protein (40 nM) was incubated with small artificial substrates (44 nM 5'-^{32}P-labeled primer e2 (Pri e2; 70 nt)) by itself or annealed to 40 nM E2 RNA or DNA (40 nt) as diagrammed to the left) in reaction medium containing 200 μM dNTPs, 5 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5, 1 mM dithiothreitol (DTT) in addition to 450 mM NaCl, 200 mM KCl, or 100 mM KCl for 30 min at 30°C. The reaction was terminated by phenol-CIA extraction, and the products were analyzed in a denaturing 10% polyacrylamide gel. Lanes (1) and (2) ^{32}P-labeled DNA primer e2 incubated without and with LtrA in 450 mM NaCl, respectively; (3-5) LtrA incubated with E2 DNA template with annealed ^{32}P-labeled Pri e2 in 450 mM NaCl, 200 mM KCl, and 100 mM KCl, respectively; (6-8) LtrA incubated with E2 RNA template with annealed ^{32}P-labeled Pri e2 in 450 mM NaCl, 200 mM KCl, and 100 mM KCl, respectively. DNA and RNA oligonucleotides are shown in black and red, respectively; LtrA is shown as a gray oval; the direction of cDNA synthesis is indicated by the dotted green arrow; the arrow

indicates the intron-insertion site (IS). The numbers to the right of the gel indicate the nucleotide position of the 5'-^{32}P labeled 10-bp ladder (Invitrogen).

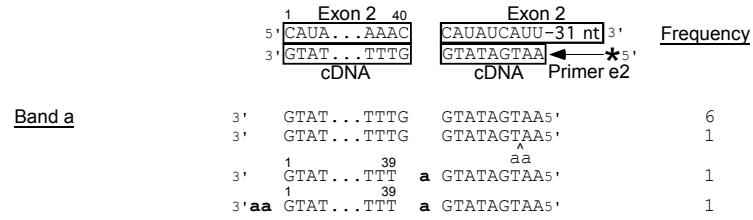


Figure 3.13: Sequence analysis of cDNA products from biochemical assays of LtrA DNA polymerase and template-switching activities on 3'-exon substrates under near-physiological KCl concentrations

Sequences of cDNA products indicated as band a in the gel in Figure 3.12. The template and expected cDNA product sequence (boxed) is shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer.

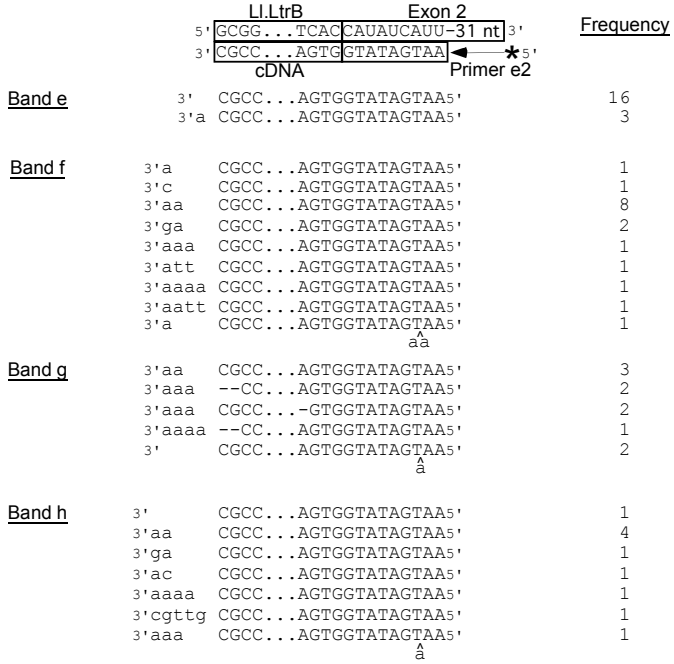


Figure 3.14: Sequence analysis of cDNA products from biochemical assays of LtrA DNA polymerase activities on Ll.LtrB-exon 2 DNA substrates

The bands indicated in the gel (e-h) in Figure 3.10 were excised, cloned, and sequenced, as described in Materials and Methods (section 3.5.4). The template and expected cDNA product sequence (boxed) is shown above each set of experimentally determined cDNA product sequences. Mutant nucleotide residues are shown in lower case letters, and extra nucleotide residues inserted at template-switching junctions or the 3' ends of cDNAs are shown in bold lower-case letters. Numbers to the right indicate the frequency of each sequence. * denotes ^{32}P-label at 5' end of primer.

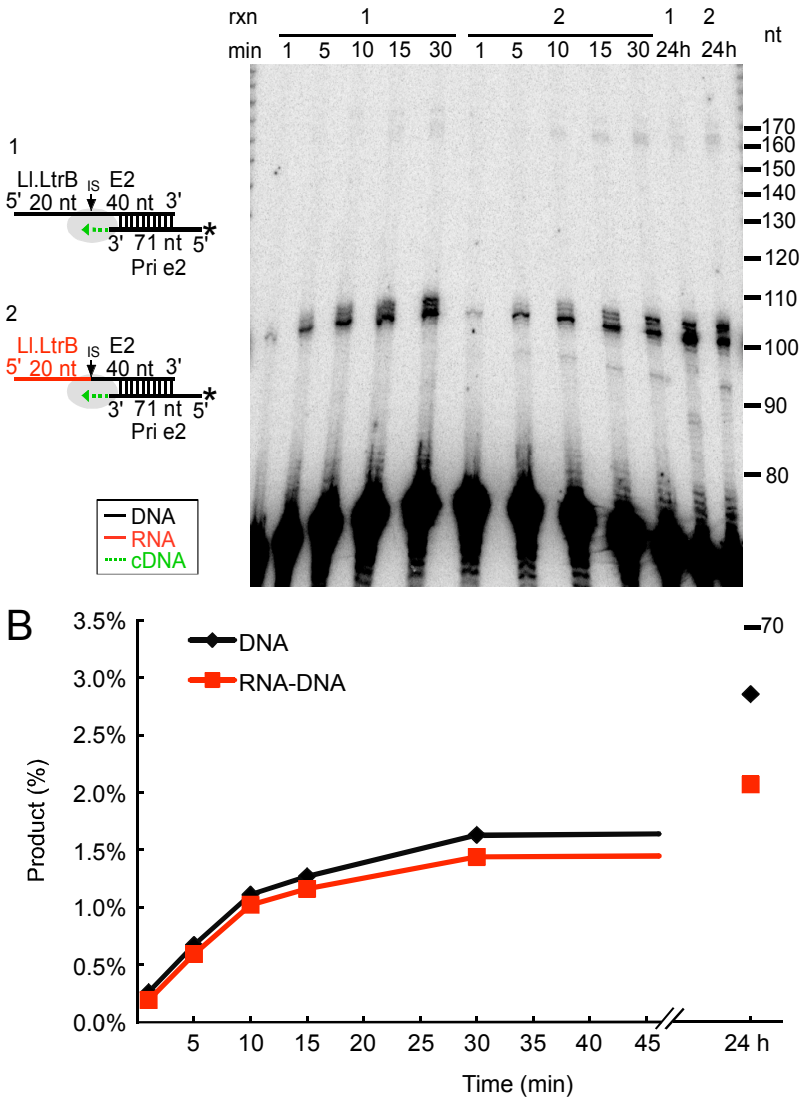


Figure 3.15: Time-course assays of LtrA DNA polymerase activity on Ll.LtrB-exon 2 DNA substrates

(A) Ll.LtrB DNA-E2 DNA or chimeric Ll.LtrB RNA-E2 DNA templates (40 nM) with annealed ^{32}P-labeled Pri e2 (44 nM) were incubated with LtrA protein (40 nM) in reaction medium containing 200 µM dNTPs, 450 mM NaCl, 5 mM $MgCl_2$, 20 mM Tris-HCl pH 7.5, and 1 mM dithiothreitol (DTT). The 200 µl reactions were initiated by adding LtrA protein, incubated at 30°C for times indicated in the Figure, and terminated by removal of 20 µl and addition of EDTA, pH 7.5 to 25 mM final followed by extraction with ice-cold phenol-CIA. The products were analyzed in a denaturing 10%

polyacrylamide gel. DNA and RNA are indicated in black and red, respectively; LtrA is shown as a gray oval; the direction of cDNA synthesis is indicated by the dotted green arrow; the arrow indicates the intron-insertion site (IS). The numbers to the right of the gel indicate the nucleotide position of the 5'-^{32}P labeled 10-bp ladder (Invitrogen). (B) The plot shows ^{32}P-labeled total product resulting from reverse transcription of DNA (black line) or RNA-DNA chimeric (red line) templates expressed as percent of total radioactivity as a function of time.

Chapter 4: Chromosomal gene targeting reactions in *Drosophila melanogaster* embryos using retargeted Ll.LtrB group II intron RNPs

4.1 Introduction

Recognition of a dsDNA target site by a group II intron RNP is primarily mediated by base-pairing between the intron RNA and the top DNA strand with only a small number of bases recognized by the IEP (Guo et al., 1997, Guo et al., 2000, Mohr et al., 2000, Singh & Lambowitz, 2001, Perutka et al., 2004). This feature together with their high insertion frequencies and specificity enabled the development of mobile group II introns into novel bacterial gene targeting vectors ("targetrons"), which can be programmed to insert efficiently into desired DNA targets simply by modifying the intron RNA (Karberg et al., 2001). Targetrons have been used to obtain gene knock-outs and knock-ins in diverse Gram-negative and Gram-positive bacteria, including medically and commercially important species in which gene targeting by other methods had previously been difficult or impossible (Frazier et al., 2003, Chen et al., 2005, Yao et al., 2006, Heap et al., 2007, Shao et al., 2007, Yao & Lambowitz, 2007, Rodriguez et al., 2008). Demonstration of Ll.LtrB linear and lariat intron RNP plasmid targeting reactions in *D. melanogaster* precellular blastoderm stage embryos prompted me to investigate the chromosomal gene targeting abilities of these RNPs. During embryonic development, 14 cycles of nuclear division occur in the absence of cytokinesis, forming an embryonic syncytium of up to 5,000 nuclei present in a single cell (Campos-Ortega JA, 1985). This attribute makes the precellular blastoderm stage embryo well suited for group II intron RNP-mediated gene targeting experiments, given that many nuclei are potentially accessible to RNPs at the site of microinjection.

4.2 DISRUPTION OF THE ENDOGENOUS *YELLOW* LOCUS IN *DROSOPHILA MELANOGASTER* EMBRYOS

4.2.1 Detection of retargeted Ll.LtrB intron RNP integration via PCR

The *yellow* (*y*) gene was chosen as a candidate for group II intron-mediated disruption because gene knockout in epithelial cells results in a cuticle lacking black pigmentation and having a yellow appearance (Wittkopp et al., 2002). *y* is a non-essential gene present on the X chromosome in which mutations result in a visual phenotype for an epithelial cell knockout (Bibikova et al., 2002). Since male flies have only one copy of the X chromosome per cell (i.e., are hemizygous), this knockout phenotype could in principle be observed from disruption of a single copy of the gene in male epithelial cells (Bibikova et al., 2002). Gene knockout of *y* has been demonstrated in both somatic and germline cells by various gene targeting methods including gene disruption via homologous recombination (HR) and expressed chimeric zinc-finger endonucleases (Rong & Golic, 2001, Bibikova et al., 2002).

The procedure I used for gene targeting reactions and downstream processing of embryos is outlined in Figure 4.1. Using a computer algorithm for efficient target site prediction (Perutka et al., 2004), targetrons were designed to insert at two different sites in the *y* gene (Figure 4.2). These targetrons are denoted Y18a and Y3776s, where the number indicates the nucleotide position of the targetron-insertion site relative to the first nucleotide of the ATG start codon and "a" or "s" indicate the antisense and sense strands, respectively.

In the first experiment (Figure 4.2 B), fifty *D. melanogaster* embryos at the pre-cellular blastoderm stage were injected with 100 mM Mg^{2+} plus 17 mM dNTPs, followed by Y18a lariat RNPs. The embryos were incubated at 30ºC for 1 h, and nucleic acids were then extracted and analyzed by nested PCR. The PCRs showed bands corresponding

to the expected 5'- and 3'-integration junctions (Figure 4.2 B), and precise integration at the correct site in the *y* gene was confirmed by sequencing the PCR products. A similar double-injection experiment showed precise integration of Y3776s lariat RNPs at their target site, detected by direct PCR and confirmed by sequencing of the 3'-integration junction in 40 pooled embryos (Figure 4.2 C). For both targetrons, efficient integration into the chromosomal target sites was not detected without injection of Mg^{2+}.

The decreased viability caused by double injection made it desirable to develop single-injection conditions. However, group II intron RNPs tend to aggregate when injected with the relatively high concentrations of Mg^{2+} required for efficient DNA integration. Thus, I tested whether substitution of polyamines might enable us to lower the amount of Mg^{2+} that must be injected with the RNPs. Polyamines had been shown previously to reduce the Mg^{2+} requirement for group II intron self-splicing reactions (Daniels et al., 1996). Figure 4.2 D shows that a single injection of linear Y3776s RNPs with 5 mM Mg^{2+} and a polyamine cocktail containing 5 mM putrescine, 3 mM spermidine, 1 mM spermine, corresponding to polyamine ratios in *D. melanogaster* larvae (Dion & Herbst, 1967), gave site-specific integration into the *y* gene. I detected site-specific integration by nested PCR of 30 pooled embryos and confirmed the correct 3'-integration junction by DNA sequencing. Viability as measured by hatch rate after injection of this polymine/Mg^{2+} mixture was similar to that for a single injection of distilled water. Thus, the use of polyamines to reduce the Mg^{2+} requirement for group II intron function provides a potential solution to the problem of double injection into *D. melanogaster* embryos, enabling the introduction of RNPs and required supplements in a single injection that gives chromosomal gene targeting.

4.2.2 Detection of Y18a integration using TaqMan qPCR

To investigate Y18a RNP integration efficiencies, I utilized TaqMan qPCR of genomic DNA extracted from flies at the embryonic and adult stages. This qPCR reaction utilizes a *y* specific primer, an Ll.LtrB intron specific primer and a probe spanning the 3' junction of Y18a insertion at the *y* target site. The amount of integration detected was normalized to the endogenous *y* gene present in each sample by qPCR with a primer-probe set located approximately 3-kb downstream of the target site on the sense strand of *y*. To assess the frequency of ectopic integration by the Y18a RNP, a primer-probe set within the Ll.LtrB intron sequence was included in the qPCR analysis.

TaqMan qPCR of genomic DNA extracted from the targeted embryos revealed that Y18a integration occurred at the *y* locus of microinjected embryos and this integration junction was stable up to 72 h (Table 4.1). Adults not subjected to microinjection were processed for genomic DNA extraction in parallel with experimental samples as a negative control for Y18a 3-junction DNA contamination during sample preparation. All samples were analyzed by qPCR in triplicate and quantities given are the result of normalization to a plasmid standard curve made during the same qPCR experiment and the average number of copies for each sample is given. In Table 4.1 experiment 1, I did not detect 3'-junction sequences in an adult fly processed for genomic DNA extraction in parallel, whereas a targeting percentage (Y18-*y* 3' junction: *y*) of 0.0092% of was observed in wild-type (Or-R) embryos. Gene targeting in wild-type embryos was increased to 0.12% using a double injection procedure (Table 4.1, experiment 2).

Following microinjection of RNP solution, developing larvae from microinjected embryos were grown to adulthood and crossed to corresponding y^- adults to assay germline knockout of *y*. After crossing, I extracted the genomic DNA of these adults and

analyzed them for Y18a insertion. The insertion of Y18a was detected in healthy, fertile *D. melanogaster* adults. All 7 adults (5 female, 2 male) from microinjected *lig4*$^{-}$ embryos described in Table 4.1 were fertile, and most were found to contain Y18a integration, in frequencies up to 0.008% (Table 4.2). I saw was no evidence of a *y* knockout phenotype of epithelial cells that form the adult cuticle. Subsequent crosses to *y*$^{-}$ flies did not reveal *y* knockout in the germline cells of the injected adults. Targeting frequencies as high as 0.021% were observed in wild-type adults recovered from experiment 2 (Table 4.2). Although in much lower abundance, both Y18a integration and Ll.LtrB sequences were detected in a negative control adult at 14 and 17 copies, respectively, reflecting low levels of contamination.

The presence of Ll.LtrB sequence in embryo and adult samples can be quite variable (Table 4.1 and 4.2). Although in all cases copies of Ll.LtrB exceed copies of the Y18a site-specific 3' junction at *y*, the ratio of Ll.LtrB:Y18a 3- junction sequence can range from 4 (Table 1, exp. 2, female A) to 711 (Table 2, exp. 1, female A). In experiment 3 it was observed that the ratio of Ll.LtrB to *y* copies is larger in embryos compared with adults, 2.5×10^{-3} and 7.3×10^{-5}, respectively (averages of embryos, Tables 4.1 and 4.2). This additional Ll.LtrB sequence may reflect residual *in vitro* transcription template non-specific reverse transcription of intron RNA present in injected embryos.

4.2.3 Integration of EGFP expression cassette- containing Y18a RNP

For general gene targeting applications using retargeted group II intron RNPs, it would be advantageous to track intron integration visually into genomic locations where insertion may not yield an observable phenotype at the time of screening. For this purpose, an LtrB intron harboring an EGFP expression cassette in DIV was made. Expression of EGFP is dependent upon Gal4 inducible promoter from the pUAST

construct, routinely used in *D. melanogaster* transgenesis (Brand & Perrimon, 1993). This promoter includes GAGA insulator elements and is ideal for use as a marker within the Ll.LtrB intron due to its minimal size and conditional expression.

Y18a EGFP linear intron RNPs in polyamine buffer were used for gene targeting in *D. melanogaster* embryos as described above. Targeting up to 0.011% was achieved as indicated by qPCR analysis of microinjected embryos expressing Gal4 from the ubiquitous promoter, Act5C or Cy (Table 4.3) (Roper & Brown, 2003). However, no mosaic expression of EGFP could be detected in embryos when observed 12-72 h post-injection.

The use of LtrB with the EGFP marker cassette was extended to include random genome targeting using linear intron EGFP RNPs with randomized EBS sequences. Construction of a randomized EBS RNP library with approximately 10^6 variants was microinjected into both Act5C-Gal4 and Cy-Gal4 expressing flies. No EGFP signal could be detected in embryos or larvae and analysis of genomic DNA extracted from these embryos or larvae by thermal asymmetric interlaced (TAIL) PCR did not yield any amplification of sequence flanking LtrB.

4.3 DISRUPTION OF THE ENDOGENOUS *WHITE* LOCUS IN *DROSOPHILA MELANOGASTER* EMBRYOS

4.3.1 Detection of W3711a integration using TaqMan qPCR

To investigate the general application of targetron-mediated gene disruption in *D. melanogaster*, I chose to design a targetron directed to the *white* (*w*) gene. The *w* gene is a member of the ABC transporter superfamily G responsible for transport of eye pigment precursors in the developing adult fly (Mackenzie et al., 1999, Shulenin et al., 2001,

Evans et al., 2008). This gene is present on the X chromosome and knockout yields a white eye phenotype in the adult fly.

The efficiency of W3711a integration was determined by qPCR analysis of genomic DNA extracted from embryos injected with W3711a RNPs in a polyamine buffer. This qPCR reaction utilizes a *w* specific primer, an LtrB specific primer and a probe spanning the 3' junction of LtrB insertion at the W3711a target site. The amount of integration detected was normalized to the endogenous *y* gene present in each sample, a gene near *w* on the same chromosome. As with Y18a experiments, the frequency of ectopic integration by the W3711a RNP was investigated by qPCR using a primer-probe set within the Ll.LtrB sequence. These microinjection experiments and qPCR analyses were done by James Burke under my supervision.

Table 4.4 summarizes the results of a gene targeting experiment using W3711a RNPs at high concentration (9.8 mg/ml based on A_{260}) in a polyamine solution. Site-specific integration of W3711a RNPs was detected in wild-type (Or-R) embryos 72 h post-injection at frequencies up to 0.4%. The W3711a 3' junction was not detected in negative control genomic DNA preparations processed in parallel. These results show that gene disruption by targetron integration in *D. melanogaster* embryos is possible at multiple genomic locations.

Analysis of Ll.LtrB sequences by qPCR in embryos containing site-specific W3711a integrations reveals the presence of Ll.LtrB copies that exceed those integrated at the *w* gene by more than 100-fold in some cases (Table 4.4). However, 400 copies were detected in the negative control genomic DNA preparation samples analyzed, reflecting low levels of contamination.

4.4 DISCUSSION

Experiments using microinjection of targetron RNPs into the posterior of precellular blastoderm stage *D. melanogaster* embryos show site-specific integration of two different *y*-directed targetrons and one *w*-directed targetron is readily detectable in embryos up to 72-h post-injection. Full-insertion of lariat Y18a is observed as well as the 3' junction of lariat and linear Y18a and Y3776s, respectively, the latter of which was achieved via introduction of RNPs in polyamine-containing solution using a single-injection procedure. I found integration frequencies of 0.12% using Y18a and 0.4% using W3711a RNPs in embryos. Further, Y18a integration was detected in healthy, fertile adult flies at efficiencies up to 0.021% of endogenous *y* containing the 3' junction of Y18a integration. Further, I show that linear targetrons harboring a genetic marker in domain IV can be directed toward a desired target site in the *D. melanogaster* embryo.

Although the amount of genomic integration by Y18a RNPs is detectable in adults by qPCR, it is possible that certain cells containing the genomic insertion do not develop normally. A Chk2-dependent sequestering of nuclei in the early embryo has been described previously, where damaged or incompletely replicated DNA induces chromosome segregation defects (Takada et al., 2003, Gao et al., 2009). These nuclei drop from the embryo cortex to the center of the syncytium and are not incorporated into cells that form the embryo proper. It is possible that targeted nuclei are undergoing such a process and only those that escape this nuclei "sinking" remain to the adult stages. Indeed, in recent descriptions of *Drosophila* hypomorphic alleles of *mre11* and *nbs*, genes which encode products that constitute portions of the DNA repair MRN complex, generate embryos with "large nuclei-free areas" and their interiors display "abnormally large and highly condensed nuclei", indicative of this process of nuclei sequestering in embryos with compromised DNA integrity (Gao et al., 2009). It may be of interest to

investigate whether such abnormalities are seen in developing embryos following active RNP injection.

As retroelements akin to retroviruses, group II introns have a highly evolved DNA integration mechanism, which enables targeted rather than random integration. For applications in gene therapy, a potential advantage of the group II intron-integration reaction is that it can be accomplished by introduction of assembled RNPs without adding exogenous DNA that could result in deleterious ectopic integrations. Retrohoming of linear and lariat to plasmid and chromosomal sites described in this work shows Ll.LtrB RNPs are highly active in the eukaryotic cell and broadens the repertoire of group II intron-mediated gene targeting reactions. These studies indicate that targetrons are capable of inserting at chromosomal locations within the rapidly dividing nuclei of *D. melanogaster* embryos, and persisting through adulthood, providing some evidence that this system is amenable to integration of microinjected group II intron RNPs. Future efforts to improve Ll.LtrB targetron components, investigate the utility of other group II introns, or optimize reactivity within the environment within a eukaryotic cell have the potential to greatly improve upon the gene targeting efficiencies currently observed and bring group II introns to the forefront eukaryotic gene targeting technology.

While further work will be required to develop the group II intron-based methods for routine use, my findings indicate that under appropriate conditions microinjected group II intron RNPs are highly active in eukaryotes, can access plasmid or chromosomal target sites, and can integrate into those target sites at reasonably high efficiency. Thus, there appear to be no insurmountable barriers for using group II introns for gene targeting in eukaryotes.

4.5 MATERIALS AND METHODS

4.5.1 Recombinant plasmids

pACD5C Y18a, pACD5C W3711a and pACD5G Y3776s have C or G at the δ' position in the 3' exon and were used for construction and *in vitro* transcription of targetrons Y18a, W3711a, and Y3776s, respectively (section 2.5.1)(Perutka et al., 2004). Y18a and Y3776s are Ll.LtrB-ΔORF introns in which the EBS2, EBS1 and δ sequences were modified to be complementary to IBS2, IBS1, and δ' sequences in DNA target sites in the *D. melanogaster yellow* (*y*) gene (NCBI accession number P09957). W3711a is a Ll.LtrB-ΔORF introns in which the EBS2, EBS1 and δ sequences were modified to be complementary to IBS2, IBS1, and δ' sequences in DNA target sites in the *D. melanogaster white* (*w*) gene (NCBI accession number X02974.2). The donor plasmid used for each targetron has a δ' residue in the 3' exon complementary to the retargeted δ residue in the intron RNA for optimal RNA splicing (Perutka et al., 2004). pIMP-1P, used for expression of the LtrA protein, is described in section 2.5.1.

For generation of Y18a 3'-junction and *y* standard curves used in qPCR assays, pCRIIY18aInt-y was constructed by cloning the Y18a 3'-junction PCR product amplified from Y18a microinjected *D. melanogaster* embryos using primers LtrB+880s, (5'-GAAGAGGGTGGTGCAAACCAGTCAC) and yellow-160s, (5'-CGCCACGGTCCACAGAAGAG) (section 4.3.4) into the TOPO TA cloning vector (pCRII-TOPO; Invitrogen) according to the manufacturer's protocol. Subsequently, the *y* exon 2 sequence was amplified from embryos using primers y2939s (5'-GATTCCGGCCACTCTGAC) and y3216a (5'- CGCCAGGTAGCTCGTATCTC) and inserted into the EcoRV site.

For generation of W3711a 3'-junction and *y* standard curves used in qPCR assays, pCRIIW3711aInt was constructed by cloning the W3711a 3'-junction PCR product

amplified from W3711a microinjected *D. melanogaster* embryos using primers LtrB+880s, (5'-GAAGAGGGTGGTGCAAACCAGTCAC) and yellow+3369s, (5'-ATCAGCCGGGCTCCGGAT) into the TOPO TA cloning vector (pCRII-TOPO; Invitrogen) according to the manufacturer's protocol.

pMod4, described in (Zhuang, 2009), is related to pACD3 but contains an alternate ΔORF construction based on the predicted intron secondary structure, with MluI and EcoRV sites introduced at the site of the ORF deletion site to facilitate the insertion of selectable markers. To make the intended deletion and add the two restriction sites, the 5' and 3' parts of the intron with an overlapping sequence were amplified separately by PCR of pLE12, which contains the full-length Ll.LtrB intron (Matsuura et al. 1997). Primers Ex1F (5'-AGCTTgaattcCTTAGAGAAAAATAATGCGGTGC) and Mod4R (5'-CGTGTGGGCGATAAAacgcgtGTCTTGTAAAAACTTCGTCT) were used to amplify the 5' part of Ll.LtrB intron, and primers Mod4F (5'-TTTATCGCCCACACgatatcTCCCGTTAAAAGCTAAATGT) and Ex2R (5'-AGCTTaagcttCTTTGCCGCTTTTTGTTTTCTC) were used to amplify the 3' part of the Ll.LtrB intron. The lower case letters indicate the appended restriction enzyme sites used for cloning. In the second round of PCR, the two first round PCR products were mixed and amplified by PCR with the outside primers Ex1F and Ex2R. The second round PCR product was digested by EcoRI and HindIII and inserted between the EcoRI and HindIII sites of pBSIISK+ (Stratagene) to yield plasmid pMod4. The *bsd* gene from pCMV/bsd (Invitrogen) was then inserted into this pMod4 to generate pMod4Bsd. For this step, the *bsd* gene was amplified by PCR, using pCMV/bsd as a template, and, BsdEV (5'-ATCGTgatatcTTACATAACTTACGGTAAATGGC-3') and BsdML (5'-ATGCAacgcgtCAGACATGATAAGATACATTGATG-3') as primers. The PCR product was digested by EcoRV and MluI and inserted into the pMod4 to generate pMod4Bsd.

To generate pMod4PproEGFP, in a first round of PCR the EGFP ORF was amplified from pEGFPmRNA (provided by Anastasois Pavlopoulos, bearing the same EGFP sequence as pMi{3xP3-EGFP} (Pavlopoulos et al., 2004)) using primers P-EGFPF1 (5'-CGCCCGGGGATCAATGGTGAGCAAGGGCG) and EGFPR1 (5'-GGAGgcttagcTTACTTGTACAGCTCGTCCA), with the BlpI site in lower case letters. Also in the first round PCR, the P transposase promoter including GAGA insulator elements and UAS sequences was amplified from pUAST (Brand & Perrimon, 1993) with primers PF1 (5'-GGAGgatatcAATTGGCCGCTCTAGCCCC) with the EcoRV site shown in lower case letters and P-EGFPR1 (5'-CGCCCTTGCTCACCATTGATCCCCGGGCG). In the second round of PCR the EGFP sequence was fused with the P transposase promoter via overlap PCR. The two first round PCR products were mixed and amplified by PCR with the outside primers PF1 and EGFPR1. The second round PCR product was digested by EcoRI and EoRV and inserted between these two sites in pMod4Bsd to yield plasmid pMod4PproEGFP. pACD5C Y18a was digested with XbaI and MluI, and the resulting fragment containing Y18a EBS and IBS sequences was cloned into the XbaI and MluI sites of pMod4PproEGFP to create pMod4PproEGFP Y18a.

4.5.2 Preparation of Ll.LtrB lariat and linear RNAs

The Y18a and Y3776s Ll.LtrB-ΔORF precursor RNAs were transcribed with phage T7 RNA polymerase (Megascript T7 Kit; Ambion, Austin, TX) from pACD5C or G, respectively, which had been linearized with NheI. The resulting precursor RNAs containing the Ll.LtrB-ΔORF intron and flanking exon sequences were self-spliced in 1.25 M NH_4Cl, 50 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5 for 3 h at 37°C, then ethanol-precipitated and dissolved in distilled water.

Linear Ll.LtrB-ΔORF intron RNAs were transcribed from DNA templates generated by PCR of plasmid pACD5G Y3776s or pMod4PproEGFP Y18a using the 5' primer T3LIS-3G (AATTAACCCTCACTAAAGGGTGCGCCCAGATAGGGTGTTAAGTCAAG), which appends a T3 promoter (underlined), and the 3' primer LtrB+940a (GTGAAGTAGGGAGGTACCGCCTTGTTC), which corresponds to the 3' end of the intron without the 3' exon. DNA oligonucleotides were obtained from Integrated DNA Technologies (IDT; Coralville, IA). The PCR products were purified by using the QIAquick PCR Purification Kit (Qiagen), and *in vitro* transcription was done with phage T3 RNA polymerase (Megascript T3 Kit; Ambion, Austin, TX).

4.5.3 Preparation of LtrA protein and RNPs

The LtrA protein was purified from pIMP-1P and used for RNP reconstitution as described in section 2.4.3.

RNPs in low salt buffer were resuspended in 20 μl of 10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0. RNPs in polyamine buffer were resuspended in 20 μl of 100 mM KCl, 5 mM putrescine dihydrochloride, 3 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochoride, and 5 mM $MgCl_2$.

4.5.4 Detection of Y18a, Y3776s, W3711a integration by PCR and qPCR

D. melanogaster precellular blastoderm stage embryos were collected and microinjected as described in Materials and Methods section 2.5.4. For experiments outlined in Figure 4.2, 30-50 embryos were injected with Y18a or Y3776s RNPs, which contain Ll.LtrB introns targeted to insert at sites in the *y* gene. Mg^{2+}, dNTPs, and polyamines were injected prior to or mixed with the RNPs, and after the injections, the embryos were incubated, as specified for individual experiments in the Figure 4.2 legend.

Nucleic acids were isolated as described in section 2.5.4 and targetron integrations were detected by PCR of one fifth of the extracted DNA using the following primers: Y18a, 5'-integration junction: primers LtrB+940a (5'-GTGAAGTAGGGAGGTACCGCCTTGTTC) and yellow+277a (5'-ACGATCTCCCCAAGGGCTTCAT) and nested primers LtrB+933a (5'-AGGGAGGTACCGCCTTGTTCACATTAC) and yellow+241a (5'-TACCATCACGCCAGCGGGGAA); Y18a, 3'-integration, LtrB+788s (5'-CGACTAATACGACTCACTATAGGGTC) and yellow-350s (5'-GCAAAGTTGGCCGATCTATGGGAAC) and nested primers LtrB+880s, (5'-GAAGAGGGTGGTGCAAACCAGTCAC) and yellow-160s, (5'-CGCCACGGTCCACAGAAGAG); Y3776s, 3'-integration junction: primers LtrB+788s (see above) and yellow+4325a (5'-ATGCCACCACCCAGATTGG) and nested primers LtrB+870s and yellow+4054a (5'-TGAGGTTTCTGTGGCAAGACAGGA). In each case, the PCR products were purified in a 1% agarose gel and sequenced to confirm integration at the expected target site.

For qPCR, adult fly samples are processed for nucleic acid extraction following the same procedure as embryos. All samples are precipitated in the presence of pellet paint non-fluorescent co-precipitant (EMD Chemicals Inc., Gibbstown, NJ) according to manufacturer's protocol and resuspended in 30 μl of nuclease-free water (IDT). Samples were diluted 1:10 for detection of *y* and LtrB sequences. Standard curves for Y18a 3'-junction and *y* quantification were made using known quantities of pCRIIY18aInt-y. Standard curves for LtrB quantification were made using known quantities of pACD5C Y18a. Detection of *y* sequences was done using 5' primer yellow+2983s (5'-CGGGTTCACCGGAGCTAA), 3' primer yellow+3042a (5'-CGCAATCTCCAGCTGTATTTGAG), and a TaqMan 5' 6-carboxyfluorescein (6-FAM)

fluorophore 3' non-fluorescent dihydrocyclopyrroloindole tripeptide minor groove binder quencher (MGB) probe yellow+3002s (5'-TCCGTATCCAGATTGG). Detection of Y18a 3'-junction sequences was done using 5' primer LtrB+906s (5'-AGTAATGTGAACAAGGCGGTACC), 3' primer yellow-13s (5'-AGAGCTAAGTGCAATGTTCCAGG), and a TaqMan 5'-6-FAM 3'-MGB probe yellow+18a-LtrB (5'-CCTACTTCACCCCTTTG). Detection of W3711a 3'-junction sequences was done using 5' primer LtrB+831s (5'-AACGAACAATAACAGAGCCGTATACTC), 3' primer white+3669s (5'-GATGCGACTGCTCAATGGCCAA), and a TaqMan 5'-6-FAM 3'-MGB probe white+3711a-LtrB (5'-CCTACTTCACCATCTCCT). Detection of LtrB sequences was done using 5' primer LtrB+831s (5'-AACGAACAATAACAGAGCCGTATACTC), 3' primer LtrB+930a (5'-GAGGTACCGCCTTGTTCACATTAC), and a TaqMan 5'-6-FAM 3'-MGB probe LtrBpr+867s (5'-TACGTACGGTTCCCGAAGA). qPCR was done using TaqMan Gene Expression Master Mix (Applied Biosystems) in a 20 µl total volume per well with 3 µl of template added. Reactions included 0.9 µM of each primer and 0.25 µM probe in standard 96-well plates, capped with MicroAmp 8-cap strips (Applied Biosystems) and run on an Applied Biosystems 7900HT Fast Real-Time PCR System in 9600 emulation mode for an initial denaturing of 50°C for 2 min, 95°C for 10 min, followed by 45 cycles of 95°C for 15 s and 60°C for 1 min. Samples were run in triplicate and data was analyzed using SDS 2.2 software (Applied Biosystems).

Table 4.1

Exp.	Sample	Injection	Injection buffer	Copies of *yellow*	Copies of Y18a integration	Copies of LtrB	Frequency of Y18 targeting
1	*lig4*$^{-}$ adult male	none	none	7.58×10^4	none detected	33	0.0000%
1	*lig4*$^{-}$ embryos-sample 1	Y18a- 3.4 mg/ml	polyamine	3.68×10^5	34	NT	0.0092%
1	Or-R embryos-sample 2	Y18a- 3.4 mg/ml	polyamine	2.55×10^5	0.23	330	0.00009%
2	Act5C-Gal4/CyO adult female	none	none	1.12×10^5	14	17	0.012%
2	Or-R embryos	Y18a- 0.44 mg/ml	1 M $MgCl_2$ + 17 mM dNTPs, HKM	2.33×10^5	270	NT	0.12%
2	Or-R embryos-sample 1	Y18a- 0.44 mg/ml	1.5 M $MgCl_2$ + 17 mM dNTPs, HKM	1.18×10^5	9	NT	0.0075%
2	Or-R embryos-sample 2	Y18a- 0.44 mg/ml	1.5 M $MgCl_2$ + 17 mM dNTPs, HKM	2.25×10^5	36	NT	0.016%
3	w^{1118} adult female	none	none	9.50×10^4	0	0	0%
3	w^{1118} embryos-sample 1	Y18a- 0.5 mg/ml	polyamine	8.99×10^4	0.8	147	0.0009%
3	w^{1118} embryos-sample 2	Y18a- 0.5 mg/ml	polyamine	1.78×10^5	0.6	644	0.0003%

Table 4.1: Detection of Y18a insertion at *yellow* in *D. melanogaster* embryos by qPCR

Microinjection of lariat Y18a RNPs, where indicated, was done as described in Materials and Methods 2.5.4. in polyamine buffer (100 mM KCl, 5 mM putrescine dihydrochloride, 3 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochoride, 5 mM MgCl2) or low salt buffer (40 mM HEPES pH 7.5, 10 mM KCl, 10 mM $MgCl_2$ (HKM)). Experiments using RNPs in low salt buffer included an injection of $MgCl_2$ and 17 mM dNTPs in solution followed within 5 min by the RNP injection. Embryos (50-80 total) were incubated for 1 h at 30°C then shifted to 18°C for 72 h; hatched larvae are separated for growing to adulthood. Genomic DNA from unhatched embryos was extracted and analyzed using qPCR (Materials and Methods sections 2.5.4 and 4.5.4). The frequency of Y18a targeting is determined by the copies of Y18a integration divided by copies of yellow detected. Exp., experiment; NT, not tested.

Table 4.2

Exp.	Sample	Injection	Buffer	Copies of *yellow*	Copies of Y18a integration	Copies of LtrB	Frequency of Y18 targeting
1	*lig4*$^{-}$ adult male	none	none	7.58×10^4	none detected	33	0.0000%
1	*lig4*$^{-}$ female A	Y18a- 3.4 mg/ml	polyamine	1.45×10^5	6	4267	0.0041%
1	*lig4*$^{-}$ female C	Y18a- 3.4 mg/ml	polyamine	2.09×10^5	14	359	0.0067%
1	*lig4*$^{-}$ male A	Y18a- 3.4 mg/ml	polyamine	4.26×10^5	34	NT	0.008%
1	Or-R female C	Y18a- 3.4 mg/ml	polyamine	4.59×10^5	20	NT	0.0044%
1	*lig4*$^{-}$ female D	Y18a- 3.4 mg/ml	polyamine	4.17×10^4	0.6	114	0.0014%
1	*lig4*$^{-}$ female E	Y18a- 3.4 mg/ml	polyamine	2.78×10^4	0	31	0%
1	*lig4*$^{-}$ male C	Y18a- 3.4 mg/ml	polyamine	4.70×10^4	0.6	76	0.0013%
2	Act5C-Gal4 adult female	none	None	1.12×10^5	14	17	0.012%
2	Or-R Female A	Y18a- 0.44 mg/ml	1.5 M $MgCl_2$ + 17 mM dNTPs, HKM	6.49×10^5	139	547	0.021%
3	W^{1118} adult female	none	none	9.50×10^4	0	0	0%
3	Or-R female B	Y18a- 0.5 mg/ml	polyamine	1.22×10^5	0	8.9	0%

Table 4.2: Detection of Y18a insertion at *yellow* in *D. melanogaster* adults by qPCR

Embryos are microinjected and incubated as described for Table 4.1. Hatched larvae were separated and grown to adulthood. Following crossing, adult genomic DNA was extracted and analyzed using qPCR (Materials and Methods sections 2.5.4 and 4.5.4). The frequency of Y18a targeting is determined as the copies of Y18a integration divided by copies of yellow detected. Exp., experiment; NT, not tested.

Sample	Injection	Buffer	Copies of *yellow*	Copies of Y18a integration	Frequency of Y18 targeting
Or-R adult female	none	none	1.89×10^5	none detected	0.0000%
Act5C-Gal4 embryos	Y18a EGFP linear 2.1 mg/ml	polyamine	6.33×10^5	12	0.0019%
Cy-Gal4 embryos	Y18a EGFP linear 2.1 mg/ml	polyamine	1.87×10^5	21	0.011%

Table 4.3: Detection of Y18a EGFP linear insertion at *yellow* in *D. melanogaster* embryos by qPCR

Embryos are microinjected and incubated as described for Table 4.1. The frequency of Y18a EGFP targeting is determined as the copies of Y18a EGFP integration divided by copies of yellow detected.

Sample	Injection	Injection buffer	Copies of *yellow*	Copies of W3711a integration	Copies of LtrB	Frequency of W3711a targeting
Or-R adult male	none	none	$8.06x10^3$	none detected	400	0.0000%
Or-R embryos- sample 1	W3711a- 9.8 mg/ml	polyamine	$1.68x10^3$	2.4	370	0.14%
Or-R embryos- sample 2	W3711a- 9.8 mg/ml	polyamine	$7.38x10^3$	30	4894	0.4%
Or-R embryos- sample 3	W3711a- 9.8 mg/ml	polyamine	$4.04x10^3$	6.5	2565	0.16%

Table 4.4: Detection of W3711a insertion into *white* in *D. melanogaster* embryos by qPCR

Microinjection of lariat W3711a RNPs, where indicated, was done as described in Materials and Methods 2.5.4. in polyamine buffer (100 mM KCl, 5 mM putrescine dihydrochloride, 3 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochoride, 5 mM MgCl2). Embryos (50-80 total) were incubated for 1 h at 30°C then shifted to 18°C for 72 h; hatched larvae are separated for growing to adulthood. Genomic DNA from unhatched embryos was extracted and analyzed using qPCR (Materials and Methods sections 2.5.4 and 4.5.4). The frequency of W3711a targeting is determined by the copies of W3711a integration divided by copies of *yellow*, a nearby gene on the same chromosome, detected.

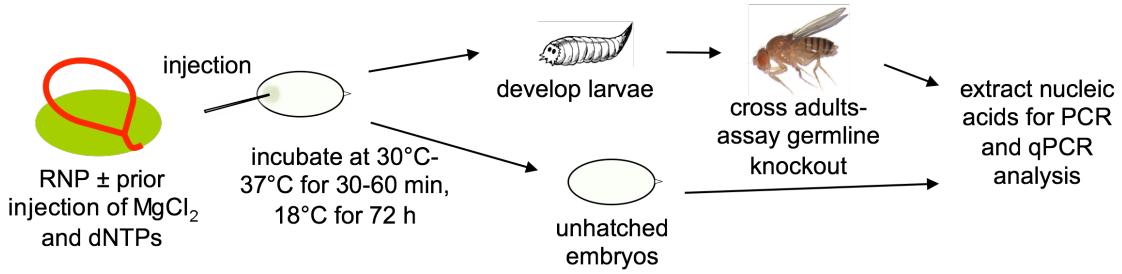


Figure 4.1: Procedure for group II intron RNP-mediated gene targeting analysis in *D. melanogaster*

D. melanogaster embryos were collected as described in Materials and Methods section 2.5.4 and are microinjected with approximately 0.3 nl of targetron RNPs. In specified experiments an injection of approximately 0.3 nl of solution containing $MgCl_2$ and 17 mM dNTPs is followed within 5 min by injection of RNPs. Embryos are incubated for 30 min or 1 h at 30 or 37˚C, where indicated. Unhatched embryos are recovered by pipetting off the slide with a micropipette tip and processed for nucleic acid extraction (Materials and Methods section 2.5.4). Hatched embryos (larvae) are transferred to an apple juice agar plate with yeast paste and allowed to develop at 18˚C. Larvae that develop into pupae are transferred to a vial, where, upon eclosion, are inspected for somatic cell knockout and crossed to the appropriate y^- partner. After crossing or upon death flies are processed for extraction of nucleic acids as for embryos (Materials and Methods section 2.5.4).

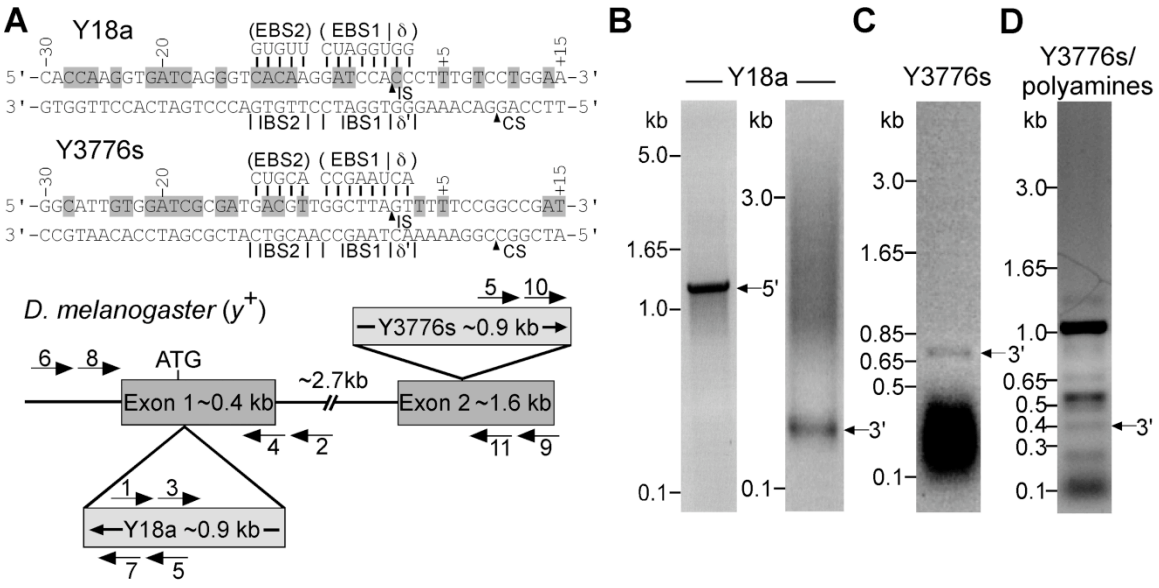


Figure 4.2: Site-specific integration of retargeted Ll.LtrB introns into chromosomal target sites in the *Drosophila melanogaster yellow* (*y*) gene

(A) DNA target site sequences and Ll.LtrB intron RNA base-pairing interactions for targetrons Y18a and Y3776s. Retargeted Ll.LtrB-ΔORF introns (targetrons) are denoted by a number that corresponds to the nucleotide position 5' to the Ll.LtrB intron-insertion site numbered from the A of the ATG initiation codon, followed by "a" or "s", indicating sense and antisense strand respectively. The DNA target sequences are shown from positions –30 to +15 from the intron-insertion site, with nucleotide residues that match those in the wild-type Ll.ltrB intron target sequence highlighted in gray in the top strand. The intron-insertion site (IS) in the top strand and the IEP-cleavage site (CS) in the bottom strand are indicated by arrowheads. Below is shown a schematic of the *Drosophila yellow* (*y*) gene (NCBI accession number P09957), with the targetron-insertion sites indicated, and a diagram of the PCRs used to detect site-specific targetron insertion. *y* gene exons are gray rectangles, and introns and flanking sequences are lines. PCR primers used to detect and sequence the targetron integrations are indicated by numbered arrows. (B) PCR analysis of targetron Y18a integration using a double-injection protocol. Fifty embryos were injected with a solution containing 100 mM $MgCl_2$ + 17 mM dNTPs followed by Y18a lariat RNPs (0.9 mg/ml). The embryos were incubated for 1 h at 30°C followed by 48 h at 18°C. Nucleic acids were isolated, and PCR

products corresponding to the 5' and 3' junctions of Y18a integrated at its chromosomal target site (1,156 bp and 238 bp, respectively) were detected by nested PCR using the following primer pairs: 5' junction, primers 1 and 2, followed by primers 3 and 4; 3' junction, primers 5 and 6, followed by primers 7 and 8 (Materials and Methods section 4.3.4). (C) PCR analysis of targetron Y3776s integration using a double-injection protocol. Forty embryos were injected with 100 mM $MgCl_2$ followed by Y3776s lariat RNPs (1.3 mg/ml), and the embryos were then incubated for 30 min at 37°C. Nucleic acids were isolated, and a 702-bp PCR product corresponding to the 3' junction of Y3776s integrated at its target site was detected by PCR using primers 5 and 9. The dark band at the bottom of the gel is RNA (Materials and Methods section 4.3.4). (D) PCR analysis of Y3776s integration using a single-injection protocol. Thirty embryos were injected with a solution containing 100 mM KCl, 5 mM putrescine dihydrochloride, 3 mM spermidine trihydrochloride, 1 mM spermine tetrahydrochoride, 5 mM $MgCl_2$, and Y3776s linear RNPs (0.5 mg/ml). After incubating the embryos for 30 min at 37°C, nucleic acids were isolated as described above, and a 348-bp PCR product (arrow) corresponding to the 3' junction for Y3776s integrated at its target site was detected by nested PCR using primers 5 and 9, followed by primers 10 and 11 (Materials and Methods section 4.3.4). Other bands in the gel are likely due to non-specific annealing of the primers. In (B–D), the PCR products were analyzed by electrophoresis of the PCR products were confirmed by sequencing across the integration junctions.in a 1% agarose gel, which was stained with ethidium bromide, and the identities

Bibliography

Adachi N, Ishino T, Ishii Y, Takeda S, Koyama H. 2001. DNA ligase IV-deficient cells are more resistant to ionizing radiation in the absence of Ku70: Implications for DNA double-strand break repair. *Proc Natl Acad Sci U S A 98*:12109-12113.

Adams MD, McVey M, Sekelsky JJ. 2003. Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science 299*:265-267.

Aizawa Y, Xiang Q, Lambowitz AM, Pyle AM. 2003. The pathway for DNA recognition and RNA integration by a group II intron retrotransposon. *Mol Cell 11*:795-805.

Anderson WF. 1998. Human gene therapy. *Nature 392*:25-30.

Ashburner M, Golic KG, Hawley RS. 2005. *Drosophila : a laboratory handbook*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Bailey JN, Klement JF, McAllister WT. 1983. Relationship between promoter structure and template specificities exhibited by the bacteriophage T3 and T7 RNA polymerases. *Proc Natl Acad Sci U S A 80*:2814-2818.

Baum C, von Kalle C, Staal FJ, Li Z, Fehse B, Schmidt M, Weerkamp F, Karlsson S, Wagemaker G, Williams DA. 2004. Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther 9*:5-13.

Beumer K, Bhattacharyya G, Bibikova M, Trautman JK, Carroll D. 2006. Efficient gene targeting in Drosophila with zinc-finger nucleases. *Genetics 172*:2391-2403.

Bibikova M, Golic M, Golic KG, Carroll D. 2002. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics 161*:1169-1175.

Bibillo A, Eickbush TH. 2002. The reverse transcriptase of the R2 non-LTR retrotransposon: continuous synthesis of cDNA on non-continuous RNA templates. *J Mol Biol 316*:459-473.

Bibillo A, Eickbush TH. 2004. End-to-end template jumping by the reverse transcriptase encoded by the R2 retrotransposon. *J Biol Chem 279*:14945-14953.

Bowater R, Doherty AJ. 2006. Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet 2*:e8.

Boyd JB, Sakaguchi K, Harris PV. 1990. mus308 mutants of Drosophila exhibit hypersensitivity to DNA cross-linking agents and are defective in a deoxyribonuclease. *Genetics 125*:813-819.

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development 118*:401-415.

Bunz F. 2002. Human cell knockouts. *Curr Opin Oncol 14*:73-78.

Campos-Ortega JA VH. 1985 *The embryonic development of Drosophila melanogaster*: Springer.

Capecchi MR. 1989. The new mouse genetics: altering the genome by gene targeting. *Trends Genet 5*:70-76.

Carter GP, Awad MM, Hao Y, Thelen T, Bergin IL, Howarth PM, Seemann T, Rood JI, Aronoff DM, Lyras D. 2011. TcsL is an essential virulence factor in Clostridium sordellii ATCC 9714. *Infect Immun 79*:1025-1032.

Cathomen T, Keith Joung J. 2008. Zinc-finger nucleases: the next generation emerges. *Mol Ther 16*:1200-1207.

Chan SH, Yu AM, McVey M. 2010. Dual roles for DNA polymerase theta in alternative end-joining repair of double-strand breaks in Drosophila. *PLoS Genet 6*:e1001005.

Chen B, Lambowitz AM. 1997. *De novo* and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3' CCA sequence. *J Mol Biol 271*:311-332.

Chen Y, McClane BA, Fisher DJ, Rood JI, 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.

Chiang CC, Lambowitz AM. 1997. The Mauriceville retroplasmid reverse transcriptase initiates cDNA synthesis de novo at the 3' end of tRNAs. *Mol Cell Biol 17*:4526-4535.

Clark JM. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res 16*:9677-9686.

Clark JM, Joyce CM, Beardsley GP. 1987. Novel blunt-end addition reactions catalyzed by DNA polymerase I of Escherichia coli. *J Mol Biol 198*:123-127.

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

Cousineau B, Smith D, Lawrence-Cavanagh S, Mueller JE, Yang J, Mills D, Manias D, Dunny G, Lambowitz AM, Belfort M. 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell 94*:451-462.

Cui X, Matsuura M, Wang Q, Ma H, Lambowitz AM. 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.

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

Dickson L, Huang HR, Liu L, Matsuura M, Lambowitz AM, Perlman PS. 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.

Dion AS, Herbst EJ. 1967. The localization of spermidine in salivary gland cells of *Drosophila melanogaster* and its effect on H3-uridine incorporation. *Proc Natl Acad Sci USA 58*:2367-2371.

Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW, Sands MS. 2007. AAV vector integration sites in mouse hepatocellular carcinoma. *Science 317*:477.

Dorsett Y, Tuschl T. 2004. siRNAs: applications in functional genomics and potential as therapeutics. *Nat Rev Drug Discov 3*:318-329.

Doyon Y, McCammon JM, Miller JC, Faraji F, Ngo C, Katibah GE, Amora R, Hocking TD, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Amacher SL. 2008. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol*.

Ducau J, Bregliano JC, de La Roche Saint-Andre C. 2000. Gamma-irradiation stimulates homology-directed DNA double-strand break repair in Drosophila embryo. *Mutat Res 460*:69-80.

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

Eskes R, Yang J, Lambowitz AM, Perlman PS. 1997. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell 88*:865-874.

Evans JM, Day JP, Cabrero P, Dow JA, Davies SA. 2008. A new role for a classical gene: white transports cyclic GMP. *J Exp Biol 211*:890-899.

Fedorova O, Mitros T, Pyle AM. 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, Su LJ, Pyle AM. 2002. Group II introns: highly specific endonucleases with modular structures and diverse catalytic functions. *Methods 28*:323-335.

Frazier CL, San Filippo J, Lambowitz AM, Mills DA. 2003. Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl Environ Microbiol 69*:1121-1128.

Gao G, Bi X, Chen J, Srikanta D, Rong YS. 2009. Mre11-Rad50-Nbs complex is required to cap telomeres during Drosophila embryogenesis. *Proc Natl Acad Sci U S A 106*:10728-10733.

George JA, Burke WD, Eickbush TH. 1996. Analysis of the 5' junctions of R2 insertions with the 28S gene: implications for non-LTR retrotransposition. *Genetics 142*:853-863.

Glanz S, Kuck U. 2009. Trans-splicing of organelle introns--a detour to continuous RNAs. *Bioessays 31*:921-934.

Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix JD. 1991. A small chloroplast RNA may be required for trans-splicing in Chlamydomonas reinhardtii. *Cell 65*:135-143.

Golinelli MP, Hughes SH. 2002a. Nontemplated base addition by HIV-1 RT can induce nonspecific strand transfer in vitro. *Virology 294*:122-134.

Golinelli MP, Hughes SH. 2002b. Nontemplated nucleotide addition by HIV-1 reverse transcriptase. *Biochemistry 41*:5894-5906.

Gorski MM, Eeken JC, de Jong AW, Klink I, Loos M, Romeijn RJ, van Veen BL, Mullenders LH, Ferro W, Pastink A. 2003. The Drosophila melanogaster DNA Ligase IV gene plays a crucial role in the repair of radiation-induced DNA double-strand breaks and acts synergistically with Rad54. *Genetics 165*:1929-1941.

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

Green MR. 1986. Pre-mRNA splicing. *Annu Rev Genet 20*:671-708.

Gregan J, Kolisek M, Schweyen RJ. 2001. Mitochondrial Mg(2+) homeostasis is critical for group II intron splicing in vivo. *Genes Dev 15*:2229-2237.

Guo H, Karberg M, Long M, Jones JP, 3rd, Sullenger B, Lambowitz AM. 2000. Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science 289*:452-457.

Guo H, Zimmerly S, Perlman PS, Lambowitz AM. 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.

Hagmann M, Bruggmann R, Xue L, Georgiev O, Schaffner W, Rungger D, Spaniol P, Gerster T. 1998. Homologous recombination and DNA-end joining reactions in zygotes and early embryos of zebrafish (Danio rerio) and Drosophila melanogaster. *Biol Chem 379*:673-681.

Hannon GJ. 2002. RNA interference. *Nature 418*:244-251.

Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The ClosTron: a universal gene knock-out system for the genus Clostridium. *J Microbiol Methods 70*:452-464.

Hu G. 1993. DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. *DNA Cell Biol 12*:763-770.

Ichiyanagi K, Beauregard A, Lawrence S, Smith D, Cousineau B, Belfort M. 2002. Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol 46*:1259-1272.

Jackson AL, Bartz SR, Schelter J, Kobayashi SV, Burchard J, Mao M, Li B, Cavet G, Linsley PS. 2003. Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol 21*:635-637.

Jacquier A, 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.

Jager U, Bocskor S, Le T, Mitterbauer G, Bolz I, Chott A, Kneba M, Mannhalter C, Nadel B. 2000. Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. *Blood 95*:3520-3529.

Jasin M. 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet 12*:224-228.

Johnson-Schlitz DM, Flores C, Engels WR. 2007. Multiple-pathway analysis of double-strand break repair mutations in Drosophila. *PLoS Genet 3*:e50.

Jones JP, 3rd, Kierlin MN, Coon RG, Perutka J, Lambowitz AM, Sullenger BA. 2005. Retargeting mobile group II introns to repair mutant genes. *Mol Ther 11*:687-694.

Karberg M, Guo H, Zhong J, Coon R, Perutka J, Lambowitz AM. 2001. Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nat Biotechnol 19*:1162-1167.

Kennell JC, Wang H, Lambowitz AM. 1994. The Mauriceville plasmid of Neurospora spp. uses novel mechanisms for initiating reverse transcription in vivo. *Mol Cell Biol 14*:3094-3107.

Kristelly R, Earnest BT, Krishnamoorthy L, Tesmer JJ. 2003. Preliminary structure analysis of the DH/PH domains of leukemia-associated RhoGEF. *Acta Crystallogr D Biol Crystallogr 59*:1859-1862.

Kulpa D, Topping R, Telesnitsky A. 1997. Determination of the site of first strand transfer during Moloney murine leukemia virus reverse transcription and identification of strand transfer-associated reverse transcriptase errors. *EMBO J 16*:856-865.

Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A, Eickbush TH. 2007. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol 374*:322-333.

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

Lambowitz AM, Zimmerly S. 2010. Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb Perspect Biol*.

Lieber MR, Ma Y, Pannicke U, Schwarz K. 2003. Mechanism and regulation of human non-homologous DNA end-joining. *Nat Rev Mol Cell Biol 4*:712-720.

Liu S, Abbondanzieri EA, Rausch JW, Le Grice SF, Zhuang X. 2008. Slide into action: dynamic shuttling of HIV reverse transcriptase on nucleic acid substrates. *Science 322*:1092-1097.

Liu YG, Mitsukawa N, Oosumi T, Whittier RF. 1995. Efficient isolation and mapping of Arabidopsis thaliana T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J 8*:457-463.

Lloyd A, Plaisier CL, Carroll D, Drews GN. 2005. Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci USA 102*:2232-2237.

Lombardo A, Genovese P, Beausejour CM, Colleoni S, Lee YL, Kim KA, Ando D, Urnov FD, Galli C, Gregory PD, Holmes MC, Naldini L. 2007. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol 25*:1298-1306.

Ma JL, Kim EM, Haber JE, Lee SE. 2003. Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol 23*:8820-8828.

Mackenzie SM, Brooker MR, Gill TR, Cox GB, Howells AJ, Ewart GD. 1999. Mutations in the *white* gene of *Drosophila melanogaster* affecting ABC transporters that determine eye colouration. *Biochim Biophys Acta 1419*:173-185.

Magnuson VL, Ally DS, Nylund SJ, Karanjawala ZE, Rayman JB, Knapp JI, Lowe AL, Ghosh S, Collins FS. 1996. Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *Biotechniques 21*:700-709.

Malik HS, Burke WD, Eickbush TH. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol 16*:793-805.

Marshall E. 2002. Clinical research. Gene therapy a suspect in leukemia-like disease. *Science 298*:34-35.

Martin W, Koonin EV. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature 440*:41-45.

Mastroianni M, Watanabe K, White TB, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM. 2008. Group II intron-based gene targeting reactions in eukaryotes. *PLoS One 3*:e3121.

Matsuura M, Noah JW, Lambowitz AM. 2001. Mechanism of maturase-promoted group II intron splicing. *EMBO J 20*:7259-7270.

Matsuura M, Saldanha R, Ma H, Wank H, Yang J, Mohr G, Cavanagh S, Dunny GM, Belfort M, Lambowitz AM. 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.

McVey M, Radut D, Sekelsky JJ. 2004. End-joining repair of double-strand breaks in Drosophila melanogaster is largely DNA ligase IV independent. *Genetics 168*:2067-2076.

Melnikova L, Biessmann H, Georgiev P. 2005. The Ku protein complex is involved in length regulation of Drosophila telomeres. *Genetics 170*:221-235.

Meng X, Noyes MB, Zhu LJ, Lawson ND, Wolfe SA. 2008. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol*.

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

Mills DA, McKay LL, Dunny GM. 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. *J Bacteriol 178*:3531-3538.

Moehle EA, Rock JM, Lee YL, Jouvenot Y, DeKelver RC, Gregory PD, Urnov FD, Holmes MC. 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci USA 104*:3055-3060.

Mohr G, Smith D, Belfort M, Lambowitz AM. 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*:559-573.

Mörl M, Niemer I, Schmelzer C. 1992. New reactions catalyzed by a group II intron ribozyme with RNA and DNA substrates. *Cell 70*:803-810.

Mörl M, Schmelzer C. 1990. Integration of group II intron bI1 into a foreign RNA by reversal of the self-splicing reaction in vitro. *Cell 60*:629-636.

Noah JW, Park S, Whitt JT, Perutka J, Frey W, Lambowitz AM. 2006. Atomic force microscopy reveals DNA bending during group II intron ribonucleoprotein particle integration into double-stranded DNA. *Biochemistry 45*:12424-12435.

Pâques F, Duchateau P. 2007. Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr Gene Ther 7*:49-66.

Pavlopoulos A, Berghammer AJ, Averof M, Klingler M. 2004. Efficient transformation of the beetle Tribolium castaneum using the Minos transposable element:

quantitative and qualitative analysis of genomic integration events. *Genetics 167*:737-746.

Peliska JA, Benkovic SJ. 1992. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. *Science 258*:1112-1118.

Perutka J, Wang W, Goerlitz D, Lambowitz AM. 2004. Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol 336*:421-439.

Peterson SE, Stellwagen AE, Diede SJ, Singer MS, Haimberger ZW, Johnson CO, Tzoneva M, Gottschling DE. 2001. The function of a stem-loop in telomerase RNA is linked to the DNA repair protein Ku. *Nat Genet 27*:64-67.

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

Porteus MH, Carroll D. 2005. Gene targeting using zinc finger nucleases. *Nat Biotechnol 23*:967-973.

Preston CR, Flores CC, Engels WR. 2006. Differential usage of alternative pathways of double-strand break repair in Drosophila. *Genetics 172*:1055-1068.

Qin PZ, Pyle AM. 1998. The architectural organization and mechanistic function of group II intron structural elements. *Curr Opin Struct Biol 8*:301-308.

Qiu YL, Palmer JD. 2004. Many independent origins of trans splicing of a plant mitochondrial group II intron. *J Mol Evol 59*:80-89.

Rago C, Vogelstein B, Bunz F. 2007. Genetic knockouts and knockins in human somatic cells. *Nat Protoc 2*:2734-2746.

Rodriguez SA, Yu JJ, Davis G, Arulanandam BP, Klose KE. 2008. Targeted inactivation of *Francisella tularensis* genes by Group II Introns. *Appl Environ Microbiol*.

Rodriguez-Trelles F, Tarrio R, Ayala FJ. 2006. Origins and evolution of spliceosomal introns. *Annu Rev Genet 40*:47-76.

Roitzsch M, Pyle AM. 2009. The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA 15*:473-482.

Romani AM, Maguire ME. 2002. Hormonal regulation of Mg2+ transport and homeostasis in eukaryotic cells. *Biometals 15*:271-283.

Rong YS, Golic KG. 2001. A targeted gene knockout in Drosophila. *Genetics 157*:1307-1312.

Roper K, Brown NH. 2003. Maintaining epithelial integrity: a function for gigantic spectraplakin isoforms in adherens junctions. *J Cell Biol 162*:1305-1315.

Russell DW, Hirata RK. 1998. Human gene targeting by viral vectors. *Nat Genet 18*:325-330.

Saldanha R, Chen B, Wank H, Matsuura M, Edwards J, Lambowitz AM. 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., and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. New York, NY: Cold Spring Harbor Laboratory Press.

San Filippo J, Lambowitz AM. 2002. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *J Mol Biol 324*:933-951.

Santiago Y, Chan E, Liu PQ, Orlando S, Zhang L, Urnov FD, Holmes MC, Guschin D, Waite A, Miller JC, Rebar EJ, Gregory PD, Klug A, Collingwood TN. 2008. Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. *Proc Natl Acad Sci USA 105*:5809-5814.

Schmelzer C, Schweyen RJ. 1986. Self-splicing of group II introns in vitro: mapping of the branch point and mutational inhibition of lariat formation. *Cell 46*:557-565.

Shao L, Hu S, Yang Y, Gu Y, Chen J, Yang Y, Jiang W, Yang S. 2007. Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. *Cell Res 17*:963-965.

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

Shulenin S, Schriml LM, Remaley AT, Fojo S, Brewer B, Allikmets R, Dean M. 2001. An ATP-binding cassette gene (ABCG5) from the ABCG (White) gene subfamily maps to human chromosome 2p21 in the region of the Sitosterolemia locus. *Cytogenet Cell Genet 92*:204-208.

Singh NN, Lambowitz AM. 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.

Smith D, Zhong J, Matsuura M, Lambowitz AM, Belfort M. 2005. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev 19*:2477-2487.

Sonoda E, Hochegger H, Saberi A, Taniguchi Y, Takeda S. 2006. Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. *DNA Repair (Amst) 5*:1021-1029.

Stellwagen AE, Haimberger ZW, Veatch JR, Gottschling DE. 2003. Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends. *Genes Dev 17*:2384-2395.

Studier FW. 2005. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif 41*:207-234.

Takada S, Kelkar A, Theurkauf WE. 2003. Drosophila checkpoint kinase 2 couples centrosome function and spindle assembly to genomic integrity. *Cell 113*:87-99.

Thummel CS. 1992. Mechanisms of transcriptional timing in Drosophila. *Science 255*:39-40.

Tissières A, Mitchell HK, Tracy UM. 1974. Protein synthesis in salivary glands of Drosophila melanogaster: relation to chromosome puffs. *J Mol Biol 84*:389-398.

Tuschl T, Zamore PD, Lehmann R, Bartel DP, Sharp PA. 1999. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev 13*:3191-3197.

Urnov FD, Miller JC, Lee YL, Beausejour CM, Rock JM, Augustus S, Jamieson AC, Porteus MH, Gregory PD, Holmes MC. 2005. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature 435*:646-651.

Verma IM, Somia N. 1997. Gene therapy -- promises, problems and prospects. *Nature 389*:239-242.

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

Wang H, Perrault AR, Takeda Y, Qin W, Iliakis G. 2003. Biochemical evidence for Ku-independent backup pathways of NHEJ. *Nucleic Acids Res 31*:5377-5388.

Wank H, SanFilippo J, Singh RN, Matsuura M, Lambowitz AM. 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.

Weterings E, van Gent DC. 2004. The mechanism of non-homologous end-joining: a synopsis of synapsis. *DNA Repair (Amst) 3*:1425-1435.

Wiesenberger G, Waldherr M, Schweyen RJ. 1992. The nuclear gene MRS2 is essential for the excision of group II introns from yeast mitochondrial transcripts in vivo. *J Biol Chem 267*:6963-6969.

Wittkopp PJ, True JR, Carroll SB. 2002. Reciprocal functions of the Drosophila yellow and ebony proteins in the development and evolution of pigment patterns. *Development 129*:1849-1858.

Wyman C, Kanaar R. 2006. DNA double-strand break repair: all's well that ends well. *Annu Rev Genet 40*:363-383.

Yang J, Mohr G, Perlman PS, Lambowitz AM. 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.

Yang J, Zimmerly S, Perlman PS, Lambowitz AM. 1996. Efficient integration of an intron RNA into double-stranded DNA by reverse splicing. *Nature 381*:332-335.

Yao J, Lambowitz AM. 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*:2735-2743.

Yao J, Zhong J, Fang Y, Geisinger E, Novick RP, Lambowitz AM. 2006. Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA 12*:1271-1281.

Yao J, Zhong J, Lambowitz AM. 2005. Gene targeting using randomly inserted group II introns (targetrons) recovered from an Escherichia coli gene disruption library. *Nucleic Acids Res 33*:3351-3362.

Yu AM, McVey M. 2010. Synthesis-dependent microhomology-mediated end joining accounts for multiple types of repair junctions. *Nucleic Acids Res 38*:5706-5717.

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*:1656-1664.

Zhong J, Lambowitz AM. 2003. Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J 22*:4555-4565.

Zhuang F. 2009. Group II Intron Mobility and its Gene Targeting Applications in Prokaryotes and Eukaryotes. *Microbiology*. Austin, TX: The University of Texas at Austin. pp 171.

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

Zimmerly S, Guo H, Eskes R, Yang J, Perlman PS, Lambowitz AM. 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 PS, Lambowitz AM. 1995b. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell 82*:545-554.

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

Zoraghi R, See RH, Gong H, Lian T, Swayze R, Finlay BB, Brunham RC, McMaster WR, Reiner NE. 2010. Functional analysis, overexpression, and kinetic characterization of pyruvate kinase from methicillin-resistant Staphylococcus aureus. *Biochemistry 49*:7733-7747.

Vita

Travis White was born in Waco, TX. He is the oldest child of Dallas and Carol White. After graduating from Grand Prairie High School in Grand Prairie, TX, he attended the University of Texas at Austin where he received the degree of Bachelor of Science in Microbiology in December 2003. He began his graduate studies at the University of Texas at Austin at the Institute for Cellular and Molecular Biology in September 2004. In August 2011 he received the degree of Doctor of Philosophy from the University of Texas at Austin. He has coauthored the following publications:

Mastroianni M, Watanabe K, **White TB**, Zhuang F, Vernon J, Matsuura M, Wallingford J, Lambowitz AM. 2008. Group II intron-based gene targeting reactions in eukaryotes. *PLoS One 3*:e3121.

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

Permanent address: 1401 S. Redondo Dr. Austin, TX 78721.

This dissertation was typed by the author.