[Frontiers in Bioscience 12, 4972-4985, September 1, 2007]

Mobile group II intron targeting: applications in prokaryotes and perspectives in eukaryotes

Xiaoxia Cui, Greg Davis

Sigma-Aldrich Biotechnology, Research and Development, 2909 Laclede Avenue, St. Louis, MO 63103, USA

TABLE OF CONTENTS

1. ABSTRACT

Mobile group II introns are ribozymes and use a novel mechanism--target DNA-primed reverse transcription--to proliferate in DNA. Group II introns are a unique mobile element for their high sequence-specific, yet readily flexible target site recognition. Both the intron RNA and the intron-encoded protein (IEP) are involved in target site recognition, and the specificity is determined primarily by base pairing between the intron RNA and DNA target. Therefore, the intron RNA can be modified according to the desired target sequence for specific gene disruption. Group II intron knockout technology is mature in bacteria and is currently being developed in eukaryotes. This technology has great potential to revolutionize fields such as functional genomics, gene therapy, and cell line engineering.

2. INTRODUCTION

All mobile elements can potentially be used to disrupt genes and obtain knockout phenotypes. However, DNA transposons and most retroelements insert randomly and are thus of limited use for rapid gene-specific disruptions. Some retrotransposons insert site-specifically into target sequences recognized and cleaved by one or more proteins. Although possible, it is not trivial to change targeting specificity determined by proteins, which typically involves sophisticated protein engineering. An ideal gene knockout tool should be equipped with both high specificity and feasibility to alter specificity for targeting any given gene. Mobile group II introns are such a remarkable tool.

Group II introns are both ribozymes and retroelements. The intron RNA catalyzes a series of

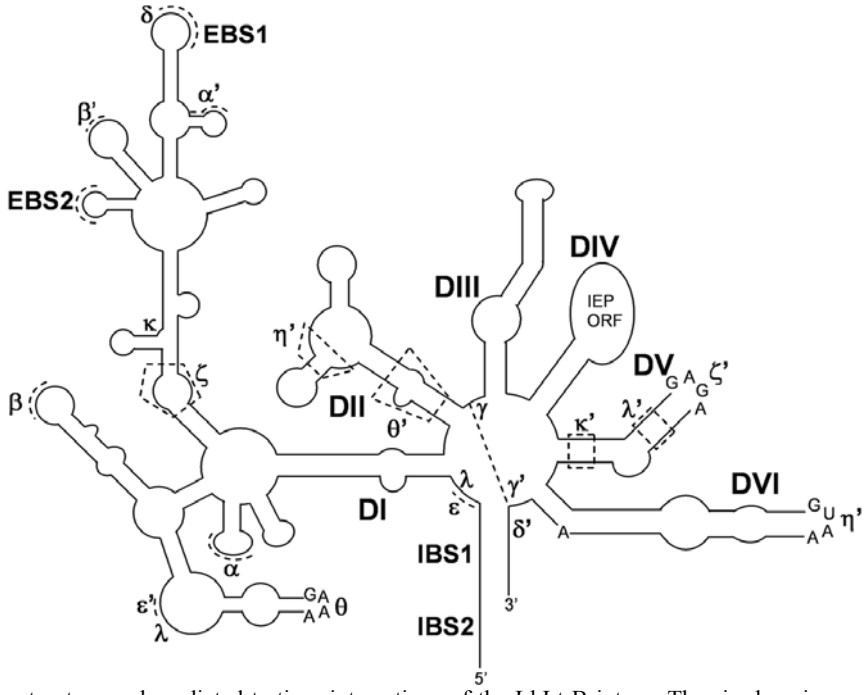

Figure 1. Secondary structure and predicted tertiary interactions of the Ll.LtrB intron. The six domains are labeled as I to VI. IBS1 and 2 sequences in the 5' exon and EBS1 and 2 sequences in the intron are labeled. Tertiary interactions are indicated as pairs in Greek letters. This figure is derived from Figure 1 of Cui *et al.* (37).

reactions for its own splicing and insertion into DNA sites (1). As a retroelement, a group II intron propagates via an RNA intermediate and reverse transcription in a process termed target DNA-primed reverse transcription (TPRT). Besides group II introns, other retroelements include long terminal repeat (LTR) and non-LTR retrotransposons, retroviruses, and telomerases (2). Like group II introns, non-LTR retrotransposon and telomerase reverse transcriptases (RTs) both use the TPRT mechanism. In addition, RTs encoded by group II introns and non-LTR retrotransposons are evolutionarily related, whereas the telomerase RT may have evolved from non-LTR RTs (3).

Among non-LTR retrotransposons, the insect R2 element inserts site-specifically into the 28S rRNA gene, and the insertion is catalyzed by the R2-encoded protein (4). Some other non-LTR retroelements also insert sequence-specifically into either rRNA genes or telomeres (2, 5). The human LINE-1 element inserts nearly randomly into the human genome, with minimal sequence preference (2). In all cases, the retrotransposon-encoded proteins are believed to recognize and cleave target sites (4). In contrast, group II introns are unique because they insert with high sequence specificity, yet the specificity is controlled mainly by base pairing between the intron RNA and a target site. Therefore, changing intron specificity is achieved by modifying intron sequences involved in base pairing with targets (6, 7). The group II intron knockout technology, developed in *Escherichia coli*, has been successfully adapted in a variety of bacterial species and proven to be a powerful genetic tool (8-15). It is obvious that once adapted in eukaryotic systems, this technology will significantly benefit the fields of functional genomics, gene therapy, and cell line engineering.

This review provides a general description of group II intron characteristics and targeting mechanisms and summarizes their applications in bacteria and potential uses in eukaryotes.

3. GROUP II INTRON DISTRIBUTION, STRUCTURE, AND SPLICING

Group II introns are widely distributed in fungal and plant organelles (16) and bacterial and archaeal genomes (17-20), and are identified by characteristic secondary and tertiary structures rather than primary sequence homology (21). Not all group II introns maintain the ability to splice. Some that do are also mobile, that is, capable of inserting into DNA sequences. The signature secondary structure of group II introns consists of six stem-loop domains denoted DI to DVI (Figure 1). A number of tertiary interactions are also conserved. Subtle variations in secondary and tertiary structures were used to further categorize group II introns into subclasses, IIA, IIB, and IIC (22). Here we focus on subclass IIA because most of the best-studied introns, such as the yeast mtDNA aI1 and aI2 introns and the *Lactococcus lactis* LtrB (Ll.LtrB) intron, belong to this group. Another well-studied intron, Rmlnt1 from the soil bacterium *Sinorhizobium meliloti*, is a group IIB intron.

Group II introns catalyze their own splicing. In IIA introns, exon-binding sequences (EBS) 1 & 2 and δ

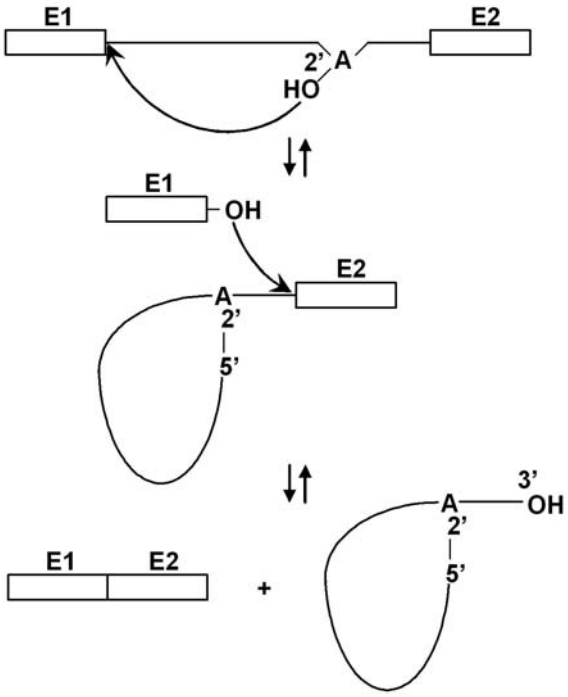

Figure 2. The splicing mechanism of group II introns. The 2' hydroxyl group of the bulged A in domain VI attacks the 5' splice site and form a 2', 5'-phosphodiester bond, releasing the 5' exon. The 3' hydroxyl group of the 5' exon then attacks the 3' splice site and generates splicing products: lariat intron and ligated exons. Both steps are reversible. The first transesterification step is rate limiting, and the second step is rapid and drives the reaction to completion (28).

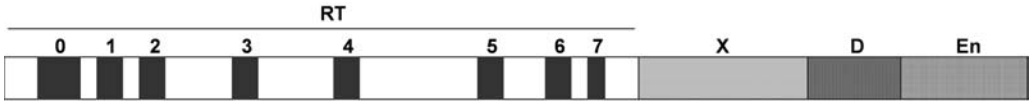

Figure 3. Schematic of the Ll.LtrB intron-encoded protein, LtrA, showing general features of group II intron IEPs. Four domains are present. The RT domain is at the N-terminus of the protein, with seven conserved motifs, RT1 to 7, found in the fingers and palm of retroviral RTs, and RT0, a motif characteristic of non-LTR RTs (31). Domain X, believed to be involved in splicing and historically called the maturase domain (39), follows the RT domain and functions in RNA splicing. Further downstream are the DNA-binding (D) and DNA endonuclease (En) domains, which are essential for mobility.

base pair with the exon sequences directly flanking the intron, namely intron binding sequences (IBS) and δ' during splicing to assist splice site selection (23). Tertiary interactions are essential to the formation of the catalytically active RNA conformation (24, 25) (Figure 1). Group II introns use the same splicing pathway by which nuclear introns are spliced by the spliceosome complex (Figure 2). In fact, segments of a group II intron and certain snRNAs are functionally exchangeable (26, 27).

The splicing reaction consists of two transesterification reactions. First, the bulged A residue in DVI attacks the phosphodiester bond at the 5' splice site to form a 2', 5'-phospodiester bond and to release the 5' exon. Second, the 3' hydroxyl of the 5' exon attacks the phosphodiester bond at the 3' splice site, resulting in ligated exons and an intron lariat. Both transesterification reactions are reversible, and the first is rate limiting. The reversibility may provide a proofreading mechanism for 5' splice site selection (28).

The spliceosome contains five snRNAs, which are believed to form its catalytic core, and over 200 proteins. Although U2 and U6 snRNAs can function together to perform simple reactions resembling splicing *in vitro*, proteins are indispensable for the splicing reaction (29). On the contrary, efficient and accurate group II intron splicing occurs *in vitro* completely in the absence of proteins if high concentrations of salt and magnesium are provided (1).

However, various proteins are involved in group II intron splicing *in vivo*. The best-studied example is the intron-encoded protein (IEP), which functions intron-specifically to help form and stabilize active RNA structures (1, 30). Most functional IEPs have four distinctive domains, such as the yeast and lactococcal IEPs, and are associated with several enzymatic activities. A reverse transcriptase (RT) domain is located at the N-terminus of an IEP (31), followed by domain X (a.k.a. the maturase domain) and the C-terminal DNA-binding domain (D) and DNA endonuclease domain (En) (Figure 3). The maturase

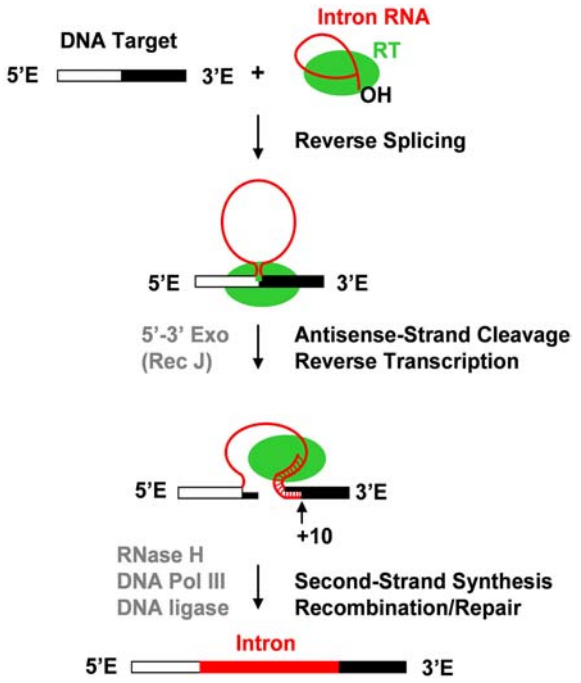

Figure 4. The homing mechanism of the Ll.LtrB intron. An RNP particle consisting of lariat intron and the IEP recognizes the target sequence, and the intron RNA reverse splices into the top strand of the insertion site. The IEP cleaves the antisense-strand and reverse transcribes the intron RNA using the 3' end of the cleaved antisense DNA as primer (51, 55, 58, 61). The 5'-3' exonuclease activity of host RecJ likely helps by chewing back the bottom strand. After reverse transcription, the intron RNA is degraded by the host RNase H activity, and the second strand is then synthesized by DNA polymerase III, which is the main polymerase for DNA replication in *E. coli*. Finally, DNA ligase seals the nicks on both strands to complete the insertion (56).

activity refers to the ability of an IEP to facilitate splicing (32-35). Reverse transcriptase, DNA-binding and DNA endonuclease activities play key roles in mobility (36). The N-terminus of the Ll.LtrB IEP (the LtrA protein) binds with high affinity to the intron DIV (34, 37), where the LtrA ORF is located, and self-regulates translation (38). Other regions of the LtrA protein make weaker contacts with the rest of the intron. The interactions between the LtrA protein and the Ll.LtrB intron RNA are believed to help form and maintain catalytically active RNA conformation for splicing (35). After splicing, the excised intron RNA stays bound to the IEP to form the ribonucleoprotein complex (RNP) that functions in intron mobility.

Not all group II IEPs contain all four domains. Many are degenerate, containing mutated RT and/or X domains, indicating a loss of function in mobility (39). Some, such as the RmInt1IEP, have no En domain but still support intron mobility (20). For a complete list of group II IEPs, refer to the following website: http://www.fp.ucalgary.ca/group 2introns/.

Some group II introns do not encode a protein in DIV and require assistance from maturase-like general splicing factors, such as MatK (39-41) and nMat proteins (42) from plants, in splicing. Unlike IEPs, these proteins are likely to function in splicing multiple ORF-less introns.

In addition, there are also non-maturase-related protein factors, involved in various aspects of splicing of introns from different organisms, such as an RNA helicase Mss116p in yeast (43, 44), CRS and CAF proteins in maize (45-47), and Raa proteins in algae chloroplast (48-50). Many of these splicing factors have other cellular functions in addition to aiding group II intron splicing.

It is worth noting that mobile group II introns identified so far rely on their own IEPs in splicing and mobility, and general splicing factors function mostly in ORF-less introns.

4. HOMING MECHANISMS

The process of a group II intron inserting into a double-strand DNA target (51) that resembles its cognate intron-less allele is referred to as homing, also termed targeting or mobility.

4.1. The Ll.LtrB intron

The homing mechanism for the Ll.LtrB intron is the best studied to date (Figure 4). After splicing, the excised lariat intron RNA stays bound to the IEP in an RNP complex. The RNP is the functional entity in homing and binds DNA non-specifically while searching for IEP recognition sites (52). The IEP first interacts with a few bases in the distal 5' exon region via major groove and phosphate-backbone interactions, which triggers the unwinding of the double-strand target DNA. The target DNA unwinding enables base pairing between the EBS and δ sequences of the intron RNA and target DNA for reverse splicing of the intron into one strand, designated the top strand (53, 54). The IEP also interacts with the 3' exon (for the Ll.LtrB intron, most importantly, a +5 T) and cleaves the bottom strand (36). If base pairing between the intron and target site does not meet the targeting requirement, the RNP complex will move along the DNA molecule to search for the next IEP recognition site. The IEP recognition site and base pairing region together span at least 15 bp, resulting a diversity of 4^{15}, that is, 1.1×10^9. On the other hand, IEP recognition sites are prevalent. For example, in *E. coli*, there is on average one IEP recognition site per 130 bp of genomic sequence predicted by the most current target site prediction algorithm (Perutka and Lambowitz, unpublished). The EBS and δ sequences are minimally involved in intron structure, and their main function is to provide base pairing during splicing and targeting. Following IEP/target DNA interaction, EBS and δ sequences in the intron RNA base pair with sequences flanking the insertion site in the DNA target. Modifying the EBS and δ sequences according to the target sequence changes the intron specificity (30, 53, 55), enabling retargeting of the Ll.LtrB intron to desired sites. These

features allow the intron to target with high specificity and remarkable flexibility.

After cleaving the bottom strand, the IEP uses the 3' end of the bottom strand as a primer to reverse transcribe the intron RNA, a process referred to as target DNA-primed reverse transcription (TPRT). The intron RNA is degraded after the completion of the RT reaction, and the gap in the top strand is repaired (1).

It is believed that host factors carry out the steps after cDNA synthesis. Smith *et al.* tested the Ll.LtrB intron mobility in a number of *E. coli* mutant strains and suggested possible players in steps following the RT reaction (56). Most of the tested mutations impair mobility. Following cDNA synthesis, RNase H1 and the 5'-3' exonuclease activity of DNA polymerase I likely remove the intron RNA. DNA polymerase III (Pol III), the major polymerase in bacterial DNA replication, probably synthesizes the second strand, in agreement with the strong link between DNA replication and mobility in *E. coli* (13). DNA ligase is most likely required to seal the nicks following the polymerase activities. The repair polymerases (II, IV, and V) are also involved in mobility. The function of these enzymes in mobility appears to be redundant, and maximal inhibition of mobility is only observed when all three were mutated.

There are also a few proteins whose disruption increased mobility efficiency. RNases E and I are likely involved in intron degradation. Exo III may degrade cDNA synthesis. RecJ may be involved in 5'-3' resection of the bottom strand of exon 1 so that newly synthesized cDNA can extend into the exon region. RecQ is a DNA helicase that possibly unwinds the exon 1 to facilitate RecJ's exonuclease activity.

4.2. Other introns

Before the Ll.LtrB intron was discovered, the TPRT mechanism was first observed in the yeast aI1 and aI2 introns(57, 58). However, the inability to express active recombinant IEPs made it difficult to adapt the yeast introns for targeting. Unlike the Ll.LtrB intron, the aI2 intron can home via a number of pathways. In the major homing pathway, the intron RNA inserts into the top strand through full or partial reverse splicing. The partially synthesized cDNA invades the intron-containing allele and uses that as a template to complete intron cDNA synthesis. In addition, some 5' exon sequences are also synthesized and provide homology for recombination with the bottom strand of the target. A characteristic of this pathway is coconversion of upstream exon sequences. Another pathway the aI2 intron uses is identical to the Ll.LtrB homing pathway, involving complete reverse splicing, synthesis of a full-length of cDNA, and repairing the gap in the top strand. This pathway is independent of homologous recombination. Finally, the aI2 intron also uses an RT-independent homing pathway, where target-site cleavage stimulates homologous recombination between the intron-containing and intronless alleles, characterized by bidirectional coconversion of flanking exon sequences. The DNA target site sequence, intron, and the environment that retrohoming takes place are contributing factors to which pathway is chosen (59).

The Rmnlt1 intron does not have an endonuclease domain for bottom strand cleavage and yet has efficient homing. Martinez-Abarca *et al.* demonstrated that the Rmnlt1 intron inserts into ssDNA at a replication fork and uses the lagging primer to initiate reverse transcription (60)**.** Interestingly, an Ll.LtrB intron with a mutant IEP defective in bottom strand cleavage employs a similar strategy. Reverse transcription proceeds using a nascent strand, preferably the leading strand, at a DNA replication fork as the primer but at a lower frequency than the normal DNA endonuclease-dependent pathway (13).

In vitro, the aI2 intron RNPs reverse splice into single-stranded DNA more efficiently than double-stranded targets. In the absence of the aI2 IEP and with high concentrations of magnesium, the lariat aI2 RNA can reverse splice into an RNA target site but not DNA (61). Single-stranded DNA target sites are observed *in vivo*, such as in the Rmnlt1 intron homing and LtrB intron retrotransposition events (see below). However, so far reverse splicing into RNA target *in vivo*, although theoretically feasible, has not been observed.

4.3. Intron difference and targeting

The Ll.LtrB intron is readily retargeted to nearly any given gene. As mentioned above, using current target site prediction algorithms (6), a potentially good protein recognition site is present at roughly every 130 bp in the *E. coli* genome. At this frequency it is possible that small gene sequences (<400 bp) may not contain an efficient target site. However, this limitation might be overcome by using other group II introns with IEPs recognizing different target sites (Karberg & Lambowitz, unpublished), expanding the range of sequences that are targetable.

It is worth noting that the Ll.LtrB intron has been applied successfully in various bacteria and shows promise in eukaryotes (see section 6). It is entirely possible that this intron will work in all organisms. However, it is also probable that an organism requires a group II intron that evolved within its own or some similar cellular environment for optimal targeting. Thus, it is important that new introns continue to be identified and characterized for mobility.

5. SPECIFICITY

High specificity is greatly desired in any method of genome modification. As mentioned previously, intron homing requires a protein recognition sequence and IBS/EBS and δ/δ' base pairing, spanning from -25 to +10. However, the stringency of each nucleotide within the target site is different and none are essential (12). One disfavored nucleotide in the IEP recognition region can be compensated by other residues. Therefore, the IEP recognition is flexible. On the other hand, single-base mismatches in the base-pairing region affect the K_d as well as the k_{cat} for reverse splicing (62). The more stringent base-pairing requirement likely provides high specificity to

target site recognition. In addition to actual nucleotide sequence, other factors, such as sequence context and secondary structures can also influence target site determination. Intron specificity is not easily defined but can be partially reflected by retrotransposition frequency.

Retrotransposition is referred to as insertion of group II introns into ectopic sites, which generally resemble homing sites, but typically miss either the 5' exon recognition site for unwinding dsDNA or the 3' site for bottom strand cleavage or both. Retrotransposition has been studied in yeast mitochondria, *L. lactis*, and *E. coli.* Dickson *et al.* demonstrated that the yeast aI1 intron retrotransposed into two known sites at an efficiency of 10^{-5} of that of homing, although the two target sites are fairly similar to the homing site (63). The study determined that the retrotransposition is independent of homologous recombination and that the intron inserted into DNA. It was not clear whether the target was dsDNA or ssDNA.

Studies on retrotransposition in both *L. lactis* and *E. coli* were done with the Ll.LtrB intron. The wild-type intron was used to target both genomes lacking the exact wild-type target site. A kanamycin-resistance marker was inserted in the intron to select for all retrotransposition events, thus representing all nonspecific targeting. The rate of retrotransposition is around 10^{-4} of the homing rate in *L. lactis* (64) and 10^{-8}, in *E. coli* (65). Alignment of the retrotransposition sites revealed that in addition to missing critical protein recognition site, retrotransposition sites did not have good EBS2/IBS2 pairings.

In *E. coli*, multiple retrotransposition pathways were identified. The Ll.LtrB intron can insert into dsDNA then use either lagging or leading strand primers in reverse transcription or insert into ssDNA at a replication fork and use the lagging strand primer to prime reverse transcription. These pathways are independent of the DNA endonuclease domain. The Ll.LtrB intron can also retrotranspose via the TPRT pathway used in homing. Cellular environment is a major determinant of which pathway to take (65).

Given its relative low efficiency, it is not surprising that retrotransposition is hardly detectable when homing sites are present. Zhong *et al.* (12) and Perutka *et al.* (6) demonstrated that retargeted introns predominantly inserted into homing sites. Most of the targeted strains had single insertions as expected. There were only a few occasions in which an extra insertion was observed. No strains with more than two insertions were obtained. In conclusion, group II introns are remarkably specific, and undesired insertion events are rare.

6. OVERVIEW OF PROKARYOTIC TARGETING

Group II introns have been successfully used in gene targeting in a number of bacterial species, both Gram-negative and Gram-positive. It has been especially useful for bacteria lacking efficient methods of genome modification.

6.1. Gram-negative strains

Karberg *et al.* first reported that the Ll.LtrB intron could be retargeted to *E. coli* chromosomal genes (9). First, the thymidylate synthase gene (*thyA*) was disrupted by introns selected from a randomized library by using a plasmid-based assay, conferring the bacterium resistance to trimethoprim and requirement of thymine for growth. The targeting efficiency was between 0.077-4.5%, and the insertion was specific to the *thyA* gene. Next, four other genes, *lacZ*, *trpE*, *dadA*, and *proA*, were disrupted with efficiencies up to 10%. In addition to *E. coli*, the *thyA* gene was disrupted in two other gram-negative bacteria, *Shigella flexneri* and *Salmonella typhimurium*, leading to the same phenotypic changes. Finally, a *thyA* intron was used to generate double-strand breaks in the target site, and stop codons were introduced into the *thyA* ORF through stimulated homologous recombination.

Retargeted group II introns were first dubbed as "targetrons" in Zhong *et al.* (12), where a retrotransposition-activated selection marker (RAM) was engineered in the intron DIV and used to select for targeting events. A RAM consists of an antibiotic resistance marker disrupted by a self-splicing group I intron. The group I and group II introns are in the same orientation for transcription, whereas the marker is transcribed from its own promoter in the opposite orientation so that it can only be expressed after group I intron splices out of the transcript and the intron inserts into DNA target site. With this method, a gene without a selectable phenotype could be targeted with almost 100% efficiency. The RAM marker was further modified to have flanking Flp recombinase sites so that multiple insertions in the same cell could be made by using the same selection marker. A library of RAM-targetrons with randomized IBS and δ sequences were transformed to make single insertions throughout the *E. coli* genome. Alignment of the randomly targeted sequences revealed more information on the target recognition rules and an interesting bias for insertion near the replication origin, suggesting a link between integration and DNA replication.

Yao *et al.* further used the library to "fish" introns targeting specific *E. coli* genes. Impressively, "fished" introns could be used without modification to obtain single gene disruptions. Some fished introns inserted into target sites lacking the +5T that is critical for bottom strand cleavage and were shown to utilize nascent replication primers for reverse transcription initiation, confirming the connection between targeting and DNA replication (11).

In order to apply group II introns to more bacterial species, Yao and Lambowitz used a broad-host-range vector, capable of replicating in at least 33 Gram-negative bacteria and one Gram-positive species, to express the intron and sucessfully disrupt genes in *E. coli, Pseudomonas aeruginosa, and Agrobacterium tumefaciens*. Theoretically, the same vector can be used in all bacteria that support its replication (15).

An important breakthrough in group II intron targeting was the development of a computer algorithm to design introns without tedious selection. The algorithm was based on a large database of previously retargeted group II introns. Targetrons designed with the algorithm were then used to knockout a series of *E. coli* DNA and RNA helicases with efficiencies of 1-80% (6), making it feasible to commercialize the targetron technology (http://www.sigmaaldrich.com/targetron).

During commercialization, a series of randomly chosen non-essential genes were disrupted to validate the performance of the kanamycin-RAM-containing introns for general gene disruption. Among the tested genes, *cynX*, *cynS*, and *cynT* genes were knocked out individually, which confirmed that only *cynS* and *cynT* were responsible for the cyanate-sensitive phenotype. In addition, *loxP* sites were inserted to flank the kanamycin-RAM marker. Expression of Cre recombinase removed the kanamycin-resistance gene, allowing subsequent disruption of another gene using kanamycin resistance selection. Interestingly, Flp-recombinase sites drastically reduced mobility when flanking the kan-RAM, while the *loxP* sites did not have a significant effect. Additionally, a 34-kb non-essential region on the *E. coli* chromosome was deleted between the *loxP* sites in two targeted introns, which can be used as a general method to obtain large chromosomal deletions (66).

6.2. Gram-positive strains

Besides Gram-negative strains, the Ll.LtrB intron was also used successfully for gene targeting in a number of Gram-positive strains. Frazier *et al.* disrupted several genes in the Ll.LtrB intron's original host, *L. lactis* (14). Not surprisingly, targeting in *L. lactis* was very efficient, and chromosomal *MleS* (encoding malate decarboxylase) and *tetM* (encoding tetracyclin resistance) genes were targeted easily without selection. Using the targeting intron as a carrier, the *tetM* gene and *abiD* gene, conferring phage resistance, were delivered separately to the *MleS* gene. Significantly, the splicing of the Ll.LtrB intron inserted in the *MleS* gene was conditional, controllable by the availability of the LtrA protein.

The next Gram-positive strain tested was the intestinal bacterium *Clostridium perfringens,* being developed as a vaccine vector (8). A designed Ll.LtrB intron disrupted the alpha toxin gene (*plc*) at 0.5% efficiency without selection. Knockout clones failed to form a white halo on an egg yolk BHI agar plate, characteristic of alpha toxin expression. Disruption of the alpha toxin makes *C. perfringens* safer for using in vaccine development. The insertion was specific and stable over 15 days of incubation and subculturing without any indication of reversion. Because of the relatively high targeting efficiency, no selection was necessary to obtain the knockout strain. Therefore, it is possible to disrupt multiple genes in the same cells consecutively. It is also worth noting that group II intron targeting drastically reduced the time needed to disrupt a gene, compared to conventional homologous recombination methods available in *C. perfringens*, which almost always require an antibiotic resistance marker.

Finally, the Ll.LtrB intron was also adapted to the well-studied Gram-positive and clinically important bacterium, *Staphylococcus aureus*. Acquired antibiotic resistance is a serious and increasing clinical problem for this pathogen, and creating site-specific mutations by using homologous recombination is often difficult. Yao *et al.* efficiently disrupted *hsa* (encoding a member of the HU family of bacterial histone-like proteins) and *seb* (encoding enterotoxin B) genes with designed targetrons without selection (10). More interestingly, with the essential *hsa* gene disrupted, cells were able to grow at a permissible temperature (32°C), at which the inserted intron spliced, but not at a non-permissible temperature (43°C). This temperature dependence is likely caused by temperature sensitivity of the LtrA protein and can potentially be used to generate conditional knockouts in other genes as well.

6.3. Discussion

One major advantage of group II intron targeting in prokaryotes is that many re-targeted introns are so efficient that they do not require antibiotic selection to isolate gene-specific mutants (6). This is extremely beneficial for several situations: (1) when organisms are naturally resistant to common antibiotics (for example, many *Pseudomonads*), (2) there are government regulatory restrictions on insertion of new antibiotic resistance genes into bioterrorism-related organisms, and (3) multiple knockouts are desired, and the selection process can be expedited by not requiring the removal of antibiotic markers between sequential knockout experiments. The combined advantages of high insertion efficiency and specificity of group II introns under non-selective conditions are unparalleled among chromosomal mutagenesis techniques.

The TargeTron Gene Knockout System attracted interest from researchers working with a wide variety of bacteria. Besides the eight successful species mentioned above, targetrons are currently being actively adapted into at least ten other species. There seems to be minimal bacterial species-specific factors required. Targetrons are likely to work in all bacteria. Generally, for targeting, the intron RNPs are expressed from DNA plasmids, construction of which can be the most time-consuming part of the adaptation process.

An expression-free option of TargeTron would be ideal to use in various bacterial species. One could potentially transform bacteria with retargeted RNP complexes, thus bypassing the expression step.

Insertion of the intron and/or selection marker into the genome could potentially affect downstream gene expression (e.g. the 'polar effect'). One way to avoid this is to use double-strand break-stimulated homologous recombination to introduce smaller mutations instead. Another approach, often used in DNA transposon mutagenesis, is to use a group II intron carrying a promoter to drive downstream gene expression and compensate for any polar effect.

7. APPROACHES IN EUKARYOTIC TARGETING

Gene targeting in higher eukaryotes is most commonly achieved by using homologous recombination to disrupt a gene by means of insertion of a selection marker (67). Homologous recombination can also be used to repair a subtle mutation, such as the point mutation causing sickle-cell anemia, in gene therapy to correct genetic defects. Unfortunately, most genes do not have a selectable phenotype, and high efficiency is critical for this approach to avoid introducing antibiotic resistance markers. Most eukaryotes lack efficient homologous recombination for convenient gene targeting. Introducing double strand breaks in the homologous locus stimulates homologous recombination up to thousands of folds (68, 69). Nevertheless, excess double strand breaks are dangerous, potentially lethal, to the cells, and it is crucial to make only site-specific double strand breaks (70).

The high sequence specificity of group II introns will greatly benefit eukaryotic gene targeting. Group II introns can disrupt genes through insertion. At the same time, they can carry an additional sequence as cargo within the intron for delivery to the desired locus. Group II introns can also generate double strand breaks and stimulate homologous recombination, as shown in prokaryotes (9). The specificity demonstrated so far *in vitro* and in bacteria suggests nonspecific cleavages in eukaryotes may also be rare. Once a workable efficiency is reached in eukaryotes, group II introns would have dramatic influence on the fields of functional genomics, gene therapy and cell line engineering.

As in prokaryotes, RNPs are the functional entity in group II intron targeting. RNPs can be introduced into the cells either as purified complexes or DNA constructs that express both the intron RNA and IEP for forming the complex. Each means has advantages and disadvantages.

The hurdles to express RNPs include achieving high-level expression of both the intron RNA and IEP, proper nuclear localization, and efficient splicing of the intron RNA precursor. Transfected DNA can also potentially integrate randomly into chromosomal loci and may be detrimental to the cells. On the other hand, plasmids are easily transfected at high efficiencies and remain in the cells for days. RNPs can potentially be made over a period of time and in different cell cycles and nuclear states, which bear unknown effects on RNP function.

On the other hand, when RNPs are reconstituted *in vitro* and delivered into the cells, splicing is bypassed. RNPs turn over faster and will not randomly integrate into the genome. However, the challenge of the RNP strategy is delivery, localization, and potentially shorter exposure time to targets since degraded RNPs are not replenished. Group II introns were shown to function in eukaryotic cells by targeting plasmids cotransfected with RNPs (7). It was very inefficient and required nested PCR to detect insertion products. *In vitro*, RNPs can target the human genome, either in the form of purified DNA or nuclear lysate, which suggests that the complexity of the human genome is not limiting, and the chemical environment in the nucleus is supportive of targeting reactions (71). Delivery of sufficient amounts of RNPs to the nucleus remains the primary task, which could potentially be aided by emerging new technologies for transfection.

8. THE PROMISE OF EUKARYOTIC TARGETRONS

Once group II introns are adapted to work in eukaryotes, many research fields would be greatly advanced by obtaining an invaluable tool to modify genomes.

With a selection marker inserted in DIV, a group II intron can be randomized at IBS and EBS sequences and potentially be used to generate a knockout library in cultured cells, which can be used for phenotypic studies. If this is done in mouse ES cells, targeted and characterized ES cells can be used to make knockout mice as disease models and for drug tests. The advantage over random DNA transposons would be that if a RAM marker is used, only RNA-mediated products would express the marker, and random plasmid integration events would not be selected.

Currently, transgenes are constructed through random integrations. With group II introns, one can potentially insert any desired gene into a specific place either by inserting a copy of the gene into group II intron DIV or using a group II intron to generate double-strand breaks and inserting via homologous recombination. This would be good for controlling the variable position-specific effects on transgene expression.

Additionally, when a group II intron inserts into the sense strand of a gene, the intron can potentially splice out from the mRNA or pre-mRNA if the IEP is provided, making the knockouts conditional and reversing knockout phenotypes.

As we mentioned above, group II introns could potentially assist repairing point mutations and small defects in a gene. In addition to repairing disease-causing genes, similar procedures could be used to make small modifications of genes whose functions can then be studied in the natural context, minimizing potential artificial effects resulting from expressing mutated genes from a plasmid or a random chromosomal location.

Cell line engineering is focused on efficient expression of therapeutically valuable molecules. Stable cell lines are important for maintaining long-term expression with fermentation to lower cost. Most of the time, cell line engineering involves overexpression of a gene of interest by randomly integrating the expression cassette into the genome and subsequently using enrichment selection methods to multiply insertion copy number at the vicinity of the location of the original insertion. Group II introns might assist in a number of ways. First, when a good expression location is identified,

group II introns can deliver a recombinase site into the position, and any desired gene can be recombined into the recombinase site with the help of the corresponding recombinase. Second, when a gene is known to inhibit production or affect activity, group II introns can be used in site-specific disruption of the gene. For example, knocking out the gene coding for a fucosyltransferase, *Fut8*, by using homologous recombination in Chinese Hamster Ovary cells for producing defucosylated therapeutic antibodies required screening over 45,000 clones for a single disruption (72). Potentially, group II intron targeting could be much less labor-intensive.

9. COMPARISON WITH OTHER AVAILABLE TECHNOLOGIES

Direct gene modification tools can be divided into two general categories: homologous recombination-based, which is sequence specific and transposon-based random insertion. More recently, small RNAs are widely used to regulate genes at the transcriptional level in eukaryotes.

9.1. Prokaryotic technologies for gene disruption

Most bacteria are not easily mutated by using homologous recombination. However, in *E. coli*, expression of the lambda phage red protein was shown to enhance homologous recombination between a chromosomal gene and PCR products bearing only 40-50 bp homology (73). This lambda-red method has been applied to closely related enteric bacteria, such as *Salmonella enterica* (74) and *Yersinia pestis* (75), but still retains a narrow host range in terms of stimulating homologous recombination directly in the knockout host. Additionally, in organisms with large numbers of repeat sequences, such as LINE elements in the human genome, increasing the global homologous recombination rate may cause genome instability. The most successful method for extending lambda-red recombination to other organisms is to mutate a gene harbored on BAC vectors in *E. coli* and then use large homology ends (> 1 kb) to increase the frequency of recombination in the desired knockout host strain (76). This method, recently termed "recombineering", typically requires a genome-scale clonal library of the organism of interest and has also been applied in engineering mouse knockouts (77).

DNA transposons have been a powerful mutagenesis tool for bacteria. A major benefit of transposon mutagenesis is that the inserted transposon sequences can be used as a base from which to explore the surrounding genomic enviroment by PCR and DNA sequencing. Genome-wide transposition gene mapping was achieved in various bacteria (78-80). The DNA transposition complexes can be reconstituted *in vitro* and electroporated into cells (81), greatly increasing transposon applications in a wide variety of bacterial species. However, to obtain a specific gene disruption by transposon mutagensis involves labor- intensive library screening.

9.2. Eukaryotic tools

As in bacteria, transposons have proven extremely useful in generating random mutations in higher eukaryotes (82) but not for gene-specific knockouts. While used widely for mutagenesis of flies, worms, and plants, active transposons for vertebrate applications have only recently been developed (83). In a manner similar to transposon mutagenesis, retroviruses can also integrate randomly into the genome and be used for genetic studies. A major distinction in the application of DNA transposon or retroviruses for euakaryotic mutagenesis is the range of cell types to which they are applicable. Retroviruses are extremely efficient at entering certain cell types, even non-dividing cells, and have been widely adopted for uses in gene trapping (84, 85). While this efficient delivery is extremely advantageous, it is limited to certain cell types by virus-specific tropisms. By contrast, transposons can be used with nearly any target tissue and are being applied in hopes of increasing the discovery of new oncogenes related to solid tumors not found in previous retroviral studies (86).

It is often desirable to deliver heterologous DNA to the chromosome within mobile elements. Retroviruses possess a strict upper limit in the amount of heterologous DNA they can tolerate between LTRs and still produce efficient titers. For instance, attempts at adding DNA encoding receptor components for an ecdysone-inducible expression system (for a total of ~8 kb) within a lentiviral construct reduced titers ~200-fold (87). In contrast, recent work with a eukaryotic DNA transposon, piggyBac, has shown that transposition activity can be unaffected when harboring up to 9.1kb of heterologous DNA (88). Thus, there are several unique benefits and drawbacks in the use of either DNA transposons and retorviruses for random mutagenesis.

Homologous recombination has been widely applied to create mouse knockouts. This is in large part due to the relative ease of integrating homologous sequences in mouse ES cells where recombination events can occur at >1% (89). Thus, a general tool for stimulating homologous recombination in various mammalian cell types is needed. However, as mentioned previously, it is desirable to localize stimulated homologous recombination to avoid cell toxicity and unwanted recombination events at other genomic locations.

Site-specific endonucleases, most noticeably, zinc finger nucleases (ZFNs) have been used to generate double-strand breaks to stimulate homologous recombination in eukaryotes and shown great potential. Urnov *et al.* reported that using expressed ZFNs single copy human genes can be targeted with up to 20% efficiency. Both alleles of the IL2Rγ gene were disrupted in the same cells at a stunning 7% rate (90). However, it is not yet proven that ZFNs are specific enough that off-target cleavages can be avoided (91). In addition, changing specificity of a ZFN is still labor-intensive and time-consuming. Similarly, meganucleases were shown to be effective in generating double-strand breaks in eukaryotic genomes, which face the same problems of specificity and difficulty in protein engineering (92-94).

The recent explosion in the use of RNA interference to study mammalian functional genomics is a testament to the difficulty of existing methods for manipulating mammalian chromosomes and the need for efficient alternatives. For example, synthetic small interfering RNAs are extremely easy to implement requiring only a simple transfection step. This has expedited scientific inquiry greatly by allowing rapid verification of a given genetic hypothesis prior to embarking on a lengthy and laborious chromosomal gene knockout procedure. Additionally the simplicity of implementing synthetic siRNA gene knockdowns has enabled economical high throughput experiments targeting hundreds of genes (95). As with chromosomal gene targeting, it is desirable to limit off-target effects when possible. For instance, when screening a large library of siRNA's directed towards kinases, most siRNA hits silenced a transcription factor directly instead of identifying a desired secondary TF-regulating kinase (96). Recently, more details behind off-target mechanisms and simple chemical modifications that significantly limit off-target effects have been identified (97, 98). While gene silencing by synthetic siRNAs is temporary, an often-desired drug-like effect, a more stable knockdown can be achieved by creating stable cell lines using lentiviral delivery of shRNAs (99).

10. SUMMARY

Group II introns are the only mobile element with the combined advantages of high specificity, efficiency, and convenient flexibility in re-targeting to different chromosomal locations. Group II introns are already a useful tool in prokaryotic gene targeting and hold much promise for future applications in eukaryotes.

11. ACKNOWLEDGEMENT

We thank Alan Lambowitz for comments on the manuscript.

12. REFERENCES

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

2. Kazazian, H. H., Jr.: Mobile elements: drivers of genome evolution. *Science*, 303, 1626-32(2004)

3. Eickbush, T. H.: Telomerase and retrotransposons: which came first? *Science*, 277, 911-2(1997)

4. Eickbush, T. H.: Mobile introns: retrohoming by complete reverse splicing. *Curr Biol*, 9, R11-4(1999)

5. Pardue, M. L., S. Rashkova, E. Casacuberta, P. G. DeBaryshe, J. A. George, K. L. Traverse: Two retrotransposons maintain telomeres in Drosophila. *Chromosome Research*, 13, 443-53(2005)

6. Perutka, J., W. Wang, D. Goerlitz & A. M. Lambowitz: Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *J Mol Biol*, 336, 421-39(2004)

7. Guo, H., M. Karberg, M. Long, J. P. Jones, 3rd, B. Sullenger & A. M. Lambowitz: Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science*, 289, 452-7(2000)

8. Chen, Y., B. A. McClane, D. J. Fisher, J. I. Rood & P. Gupta: 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, 7542-7(2005)

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

10. Yao, J., J. Zhong, Y. Fang, E. Geisinger, R. P. Novick & A. M. Lambowitz: Use of targetrons to disrupt essential and non-essential genes in *Staphaphylococcus aureus* reveals temperature-sensitivity of group II intron splicing. *RNA*, 12, 1271-81(2006)

11. Yao, J., J. Zhong & A. M. Lambowitz: Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. *Nucleic Acids Res*, 33, 3351-62(2005)

12. Zhong, J., M. Karberg & A. M. Lambowitz: Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res*, 31, 1656-64(2003)

13. Zhong, J. & A. M. Lambowitz: Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *EMBO J*, 22, 4555-65(2003)

14. Frazier, C. L., J. San Filippo, A. M. Lambowitz & D. A. Mills: Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation of stable insertions without selection. *Appl Environ Microbiol*, 69, 1121-8(2003)

15. Yao, J. & A. M. Lambowitz: Gene targeting in gram-negative bacteria by use of a mobile group II intron ("Targetron") expressed from a broad-host-range vector. *Appl Environ Microbiol*, 73, 2735-43(2007)

16. Lambowitz, A. M. & M. Belfort: Introns as mobile genetic elements. *Annu Rev Biochem*, 62, 587-622(1993)

17. Ferat, J. L. & F. Michel: Group II self-splicing introns in bacteria. *Nature*, 364, 358-61(1993)

18. Dai, L. & S. Zimmerly: ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA*, 9, 14-9(2003)

19. Dai, L., N. Toor, R. Olson, A. Keeping & S. Zimmerly: Database for mobile group II introns. *Nucleic Acids Res*, 31, 424-6(2003)

20. Toro, N.: Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ Microbiol*, 5, 143-51(2003)

21. Michel, F. & J. L. Ferat: Structure and activities of group II introns. *Annu Rev Biochem*, 64, 435-61(1995)

22. Toor, N., G. Hausner & S. Zimmerly: Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA*, 7, 1142-52(2001)

23. Muller, M. W., R. J. Schweyen & C. Schmelzer: Selection of cryptic 5' splice sites by group II intron RNAs in vitro. *Nucleic Acids Res*, 16, 7383-95(1988)

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

25. Robart, A. R. & S. Zimmerly: Group II intron retroelements: function and diversity. *Cytogenet Genome Res*, 110, 589-97(2005)

26. Hetzer, M., G. Wurzer, R. J. Schweyen & M. W. Mueller: Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature*, 386, 417-20(1997)

27. Shukla, G. C. & R. A. Padgett: A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell*, 9, 1145-50(2002)

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

29. Valadkhan, S.: snRNAs as the catalysts of pre-mRNA splicing. *Curr Opin Chem Biol*, 9, 603-8(2005)

30. Matsuura, M., R. Saldanha, H. Ma, H. Wank, J. Yang, G. Mohr, S. Cavanagh, G. M. Dunny, M. Belfort & A. M. Lambowitz: A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev*, 11, 2910-24(1997)

31. Xiong Y, E. T.: Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, 9, 3353-62(1990)

32. Noah, J. W. & A. M. Lambowitz: Effects of maturase binding and Mg2+ concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*, 42, 12466-80(2003)

33. Saldanha, R., B. Chen, H. Wank, M. Matsuura, J. Edwards & A. M. Lambowitz: RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*, 38, 9069-83(1999)

34. Wank, H., J. SanFilippo, R. N. Singh, M. Matsuura & A. M. Lambowitz: A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Mol Cell*, 4, 239-50(1999)

35. Matsuura, M., J. W. Noah & A. M. Lambowitz: Mechanism of maturase-promoted group II intron splicing. *Embo J*, 20, 7259-70(2001)

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

37. Cui, X., M. Matsuura, Q. Wang, H. Ma & A. M. Lambowitz: 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, 211-31(2004)

38. Singh, R. N., R. J. Saldanha, L. M. D'Souza & A. M. Lambowitz: Binding of a group II intron-encoded reverse transcriptase/maturase to its high affinity intron RNA binding site involves sequence-specific recognition and autoregulates translation. *J Mol Biol*, 318, 287-303(2002)

39. Mohr, G., P. S. Perlman & A. M. Lambowitz: Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Res*, 21, 4991-7(1993)

40. Vogel, J., T. Hubschmann, T. Borner & W. R. Hess: Splicing and intron-internal RNA editing of trnK-matK transcripts in barley plastids: support for MatK as an essential splice factor. *J Mol Biol*, 270, 179-87(1997)

41. Young, N. D. & C. W. dePamphilis: Purifying selection detected in the plastid gene matK and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. *Mol Biol Evol*, 17, 1933-41(2000)

42. Mohr, G. & A. M. Lambowitz: Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. *Nucleic Acids Res*, 31, 647-52(2003)

43. Huang, H. R., C. E. Rowe, S. Mohr, Y. Jiang, A. M. Lambowitz & P. S. Perlman: The splicing of yeast mitochondrial group I and group II introns requires a DEAD-box protein with RNA chaperone function. *Proc Natl Acad Sci U S A*, 102, 163-8(2005)

44. Seraphin, B., M. Simon, A. Boulet & G. Faye: Mitochondrial splicing requires a protein from a novel helicase family. *Nature*, 337, 84-7(1989)

45. Ostheimer, G. J., R. Williams-Carrier, S. Belcher, E. Osborne, J. Gierke & A. Barkan: Group II intron splicing

factors derived by diversification of an ancient RNA-binding domain. *EMBO J*, 22, 3919-29(2003)

46. Jenkins, B. D. & A. Barkan: Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. *EMBO J*, 20, 872-9(2001)

47. Till, B., C. Schmitz-Linneweber, R. Williams-Carrier & A. Barkan: CRS1 is a novel group II intron splicing factor that was derived from a domain of ancient origin. *RNA*, 7, 1227-38(2001)

48. Rivier, C., M. Goldschmidt-Clermont & J. D. Rochaix: Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in Chlamydomonas reinhardtii. *EMBO J*, 20, 1765-73(2001)

49. Perron, K., M. Goldschmidt-Clermont & J. D. Rochaix: A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. *EMBO J*, 18, 6481-90(1999)

50. Perron, K., M. Goldschmidt-Clermont & J. D. Rochaix: A multiprotein complex involved in chloroplast group II intron splicing. *RNA*, 10, 704-11(2004)

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

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

53. Guo, H., S. Zimmerly, P. S. Perlman & A. M. Lambowitz: Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *EMBO J*, 16, 6835-48(1997)

54. Singh, N. N. & A. M. Lambowitz: Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *J Mol Biol*, 309, 361-86(2001)

55. Eskes, R., J. Yang, A. M. Lambowitz & P. S. Perlman: Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell*, 88, 865-74(1997)

56. Smith, D., J. Zhong, M. Matsuura, A. M. Lambowitz & M. Belfort: Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes Dev*, 19, 2477-87(2005)

57. Moran, J. V., S. Zimmerly, R. Eskes, J. C. Kennell, A. M. Lambowitz, R. A. Butow & P. S. Perlman: Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol Cell Biol*, 15, 2828-38(1995)

58. Zimmerly, S., H. Guo, P. S. Perlman & A. M. Lambowitz: Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, 82, 545-54(1995)

59. Eskes, R., L. Liu, H. Ma, M. Y. Chao, L. Dickson, A. M. Lambowitz & P. S. Perlman: Multiple homing pathways used by yeast mitochondrial group II introns. *Mol Cell Biol*, 20, 8432-46(2000)

60. Martinez-Abarca, F., A. Barrientos-Duran, M. Fernandez-Lopez & N. Toro: The RmInt1 group II intron has two different retrohoming pathways for mobility using predominantly the nascent lagging strand at DNA replication forks for priming. *Nucleic Acids Res*, 32, 2880-8(2004)

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

62. Xiang Q, Q. P., Michels WJ, Freeland K, Pyle AM.: Sequence specificity of a group II intron ribozyme: multiple mechanisms for promoting unusually high discrimination against mismatched targets. *Biochemistry*, 37, 3839-49(1998)

63. Dickson, L., H. R. Huang, L. Liu, M. Matsuura, A. M. Lambowitz & P. S. Perlman: Retrotransposition of a yeast group II intron occurs by reverse splicing directly into ectopic DNA sites. *Proc Natl Acad Sci U S A*, 98, 13207-12(2001)

64. Ichiyanagi, K., A. Beauregard, S. Lawrence, D. Smith, B. Cousineau & M. Belfort: Retrotransposition of the Ll.LtrB group II intron proceeds predominantly via reverse splicing into DNA targets. *Mol Microbiol*, 46, 1259-72(2002)

65. Coros, C. J., M. Landthaler, C. L. Piazza, A. Beauregard, D. Esposito, J. Perutka, A. M. Lambowitz & M. Belfort: Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Mol Microbiol*, 56, 509-24(2005)

66. Spears, M. C., E. R. Eastlund, D. K. Douglas, K. R. Kayser & G. D. Davis: Operon deletion in *Escherichia coli* by site-specific placement of Cre/loxP sites using retargeted group II introns. *Submitted* (2007)

67. Capecchi, M. R.: Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet*, 6, 507-12(2005)

68. Jasin, M.: Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet*, 12, 224-8(1996)

69. Porteus, M. H. & D. Baltimore: Chimeric nucleases stimulate gene targeting in human cells. *Science*, 300, 763(2003)

70. Bowater, R. & A. J. Doherty: Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining. *PLoS Genet*, 2, e8(2006)

71. Cui, X.: RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells. *Dissertation*, University of Texas (2005)

72. Yamane-Ohnuki, N., S. Kinoshita, M. Inoue-Urakubo, M. Kusunoki, S. Iida, R. Nakano, M. Wakitani, R. Niwa, M. Sakurada, K. Uchida, K. Shitara & M. Satoh: Establishment of FUT8 knockout Chinese hamster ovary cells: an ideal host cell line for producing completely defucosylated antibodies with enhanced antibody-dependent cellular cytotoxicity. *Biotechnol Bioeng*, 87, 614-22(2004)

73. Datsenko, K. A. & B. L. Wanner: One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97, 6640-5(2000)

74. Worlock, A. J. & R. L. Smith: ZntB is a novel Zn2+ transporter in Salmonella enterica serovar Typhimurium. *J Bacteriol*, 184, 4369-73(2002)

75. Perry, R. D., A. G. Bobrov, O. Kirillina, H. A. Jones, L. Pedersen, J. Abney & J. D. Fetherston: Temperature regulation of the hemin storage (Hms+) phenotype of Yersinia pestis is posttranscriptional. *J Bacteriol*, 186, 1638-47(2004)

76. Gust, B., G. L. Challis, K. Fowler, T. Kieser & K. F. Chater: PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci U S A*, 100, 1541-6(2003)

77. Liu, P., N. A. Jenkins & N. G. Copeland: A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res*, 13, 476-84(2003)

78. Salama, N. R., B. Shepherd & S. Falkow: Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *J Bacteriol*, 186, 7926-35(2004)

79. Liberati, N. T., J. M. Urbach, S. Miyata, D. G. Lee, E. Drenkard, G. Wu, J. Villanueva, T. Wei & F. M. Ausubel: An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc Natl Acad Sci U S A*(2006)

80. Benton, B. M., J. P. Zhang, S. Bond, C. Pope, T. Christian, L. Lee, K. M. Winterberg, M. B. Schmid & J. M. Buysse: Large-scale identification of genes required for full virulence of *Staphylococcus aureus*. *J Bacteriol*, 186, 8478-89(2004)

81. Pajunen, M. I., A. T. Pulliainen, J. Finne & H. Savilahti: Generation of transposon insertion mutant libraries for Gram-positive bacteria by electroporation of phage Mu DNA transposition complexes. *Microbiology*, 151, 1209-18(2005)

82. Miskey, C., Z. Izsvak, K. Kawakami & Z. Ivics: DNA transposons in vertebrate functional genomics. *Cell Mol Life Sci*, 62, 629-41(2005)

83. Izsvak, Z., Z. Ivics & R. H. Plasterk: Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol*, 302, 93-102(2000)

84. Zambrowicz, B. P., G. A. Friedrich, E. C. Buxton, S. L. Lilleberg, C. Person & A. T. Sands: Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature*, 392, 608-11(1998)

85. Couldrey, C., M. B. Carlton, P. M. Nolan, W. H. Colledge & M. J. Evans: A retroviral gene trap insertion into the histone 3.3A gene causes partial neonatal lethality, stunted growth, neuromuscular deficits and male sub-fertility in transgenic mice. *Hum Mol Genet*, 8, 2489-95(1999)

86. Dupuy, A. J., N. A. Jenkins & N. G. Copeland: Sleeping beauty: a novel cancer gene discovery tool. *Hum Mol Genet*, 15 Suppl 1, R75-9(2006)

87. Galimi, F., E. Saez, J. Gall, N. Hoong, G. Cho, R. M. Evans & I. M. Verma: Development of ecdysone-regulated lentiviral vectors. *Mol Ther*, 11, 142-8(2005)

88. Ding, S., X. Wu, G. Li, M. Han, Y. Zhuang & T. Xu: Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*, 122, 473-83(2005)

89. Jasin, M. & F. Liang: Mouse embryonic stem cells exhibit high levels of extrachromosomal homologous recombination in a chloramphenicol acetyltransferase assay system. *Nucleic Acids Res*, 19, 7171-5(1991)

90. Urnov, F. D., J. C. Miller, Y. L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory & M. C. Holmes: Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435, 646-51(2005)

91. Porteus, M. H. & D. Carroll: Gene targeting using zinc finger nucleases. *Nat Biotechnol*, 23, 967-73(2005)

92. Arnould, S., P. Chames, C. Perez, E. Lacroix, A. Duclert, J. C. Epinat, F. Stricher, A. S. Petit, A. Patin, S. Guillier, S. Rolland, J. Prieto, F. J. Blanco, J. Bravo, G. Montoya, L. Serrano, P. Duchateau & F. Paques: Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol*, 355, 443-58(2006)

93. Chames, P., J. C. Epinat, S. Guillier, A. Patin, E. Lacroix & F. Paques: In vivo selection of engineered homing endonucleases using double-strand break induced

homologous recombination. *Nucleic Acids Res*, 33, e178(2005)

94. Perez, C., V. Guyot, J. P. Cabaniols, A. Gouble, B. Micheaux, J. Smith, S. Leduc, F. Paques & P. Duchateau: Factors affecting double-strand break-induced homologous recombination in mammalian cells. *Biotechniques*, 39, 109-15(2005)

95. Chi, J. T., H. Y. Chang, N. N. Wang, D. S. Chang, N. Dunphy & P. O. Brown: Genomewide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci U S A*, 100, 6343-6(2003)

96. Lin, X., X. Ruan, M. G. Anderson, J. A. McDowell, P. E. Kroeger, S. W. Fesik & Y. Shen: siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucleic Acids Res*, 33, 4527-35(2005)

97. Jackson, A. L., J. Burchard, D. Leake, A. Reynolds, J. Schelter, J. Guo, J. M. Johnson, L. Lim, J. Karpilow, K. Nichols, W. Marshall, A. Khvorova & P. S. Linsley: Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing. *RNA*, 12, 1197-205(2006)

98. Jackson, A. L., J. Burchard, J. Schelter, B. N. Chau, M. Cleary, L. Lim & P. S. Linsley: Widespread siRNA "off-target" transcript silencing mediated by seed region sequence complementarity. *RNA*, 12, 1179-87(2006)

99. Moffat, J., D. A. Grueneberg, X. Yang, S. Y. Kim, A. M. Kloepfer, G. Hinkle, B. Piqani, T. M. Eisenhaure, B. Luo, J. K. Grenier, A. E. Carpenter, S. Y. Foo, S. A. Stewart, B. R. Stockwell, N. Hacohen, W. C. Hahn, E. S. Lander, D. M. Sabatini & D. E. Root: A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, 124, 1283-98(2006)

Abbreviations: RNP: ribonucleoprotein, IEP: intron-encoded protein, ORF: open reading frame, RT: reverse transcriptase/ reverse transcription, TPRT: target DNA primed reverse transcription, LTR: long terminal repeat, EBS: exon binding sequence, IBS: intron binding sequence, ssDNA: single strand deoxyribonucleic acid, dsDNA: double strand deoxyribonucleic acid, DI-DVI: the intron domain I though DVI, ZFN: zinc finger nuclease

Key Words: functional Genomics, Gene Therapy, Cell Line Engineering, Gene Knockout, Ribozyme, Gene Targeting, Homologous Recombination, Splicing, Reverse Splicing, Targetron, Reverse Transcriptase, Reverse Transcription, Double-Strand Break, Transposons, Review

Send correspondence to: Dr Xiaoxia Cui, Sigma-Aldrich Biotechnology, 2909 Laclede Ave., St. Louis, MO 63103, Tel: 314-771-5765, Ext. 2035, Fax: 314-286-7645, E-mail: xcui@sial.com