A conserved intronic U1 snRNP-binding sequence promotes trans-splicing in *Drosophila*

Jun-Li Gao,1,3 Yu-Jie Fan,1,3 Xiu-Ye Wang,1,3 Yu Zhang,1 Jia Pu,1 Liang Li,1 Wei Shao,1 Shuai Zhan,1 Jianjiang Hao,2 and Yong-Zhen Xu1

1Key Laboratory of Insect Developmental and Evolutionary Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China; 2Poochon Scientific, Frederick, Maryland 21704, USA

Unlike typical cis-splicing, trans-splicing joins exons from two separate transcripts to produce chimeric mRNA and has been detected in most eukaryotes. Trans-splicing in trypanosomes and nematodes has been characterized as a spliced leader RNA-facilitated reaction; in contrast, its mechanism in higher eukaryotes remains unclear. Here we investigate *mod(mdg4)*, a classic trans-spliced gene in *Drosophila*, and report that two critical RNA sequences in the middle of the last 5′ intron, TSA and TSB, promote trans-splicing of *mod(mdg4)*. In TSA, a 13-nucleotide (nt) core motif is conserved across *Drosophila* species and is essential and sufficient for trans-splicing, which binds U1 small nuclear RNP [snRNP] through strong base-pairing with U1 snRNA. In TSB, a conserved secondary structure acts as an enhancer. Deletions of TSA and TSB using the CRISPR/Cas9 system result in developmental defects in flies. Although it is not clear how the 5′ intron finds the 3′ introns, compensatory changes in U1 snRNA rescue trans-splicing of TSA mutants, demonstrating that U1 recruitment is critical to promote trans-splicing in vivo. Furthermore, TSA core-like motifs are found in many other trans-spliced *Drosophila* genes, including *lola*. These findings represent a novel mechanism of trans-splicing, in which RNA motifs in the 5′ intron are sufficient to bring separate transcripts into close proximity to promote trans-splicing.

[Keywords: trans-splicing, *mod(mdg4)*, U1 snRNP, pseudo-5′ splice site, RNA motif, *Drosophila*]

Supplemental material is available for this article.

Received January 19, 2015; revised version accepted February 25, 2015.
snRNP [Bruzik and Steitz 1990]. However, in vitro purification using cell extracts from the parasitic nematode *Ascaris lumbricoides* revealed that the protein components of SL RNP are specific except for core Sm proteins [Denker et al. 2002]. It has been speculated that U5 snRNA and its associated proteins may identify and align the 5′ SSs in the SL RNA for the spliceosomal chemical reactions [Denker et al. 1996; Maroney et al. 1996].

Trans-splicing is found less commonly in many higher eukaryotes, including in several chloroplast and mitochondria genes in plants [Koller et al. 1987; Pereira de Souza et al. 1991], in *mod(mdg4)* and *lola* in flies [Horiuchi et al. 2003; McManus et al. 2010; Zhang et al. 2010; Shao et al. 2012]. In *Drosophila*, ~80 trans-spliced genes have been identified and confirmed using interspecies hybrids, including two classic trans-spliced genes, *mod(mdg4)* and *lola* [McManus et al. 2010]. The *mod(mdg4)* gene functions in establishing and maintaining an open chromatin conformation and therefore is being considered as an enhancer of position effect variegation [Dorn et al. 1993; Gerasimova et al. 1995; Dorn and Krauss 2003]. In *mod(mdg4)*, alternative 3′ exons from multiple loci on both DNA strands are trans-spliced to a 5′ transcript containing four common exons [Fig. 1A] and result in at least 31 isoforms [Labrador et al. 2001; Dorn and Krauss 2003; Gabler et al. 2005; McManus et al. 2010]. Trans-splicing of *mod(mdg4)* is conserved in insects, including silkworms [Shao et al. 2012], mosquitos [Krauss and Dorn 2004], and cotton bollworms [Cai et al. 2012]. The *lola* gene encodes a transcription factor that is involved in neuron and stem cell differentiation [Neumuller et al. 2011; Southall et al. 2014]. The 5′ transcripts of *lola*, containing four alternative exons and four common exons, are trans-spliced to multiple 3′ exons to generate at least 80 mRNAs and 20 protein isoforms [Horiuchi et al. 2003; McManus et al. 2010].

**Figure 1.** Two intronic RNA sequences are critical for trans-splicing of *mod(mdg4)*. (A) Schematic of the *mod(mdg4)* gene locus in *Drosophila melanogaster*. Exons and introns from 3′ transcripts are designated according to FlyBase, and the sequence between exon 4 and the first 3′ exon N is defined as the last 5′ intron (intron 4). (B) Schematic of a trans-splicing system in *Drosophila* S2 cells. (C) Truncation assays of intron 4 reveal that two regions are important for trans-splicing activity. Relative trans-spliced products are averages from six isoforms that were normalized to loading controls and the full-length intron 4. (D) Replacing intron 4 of *mod(mdg4)* with other introns abolishes trans-splicing activity. Sequences of other genes include exon 1–intron 1 (without the 3′SS) from *tubulin* (*tub*) and *rp49*. (E) Trans-splicing activity of tiled 101-nt deletions in intron 4. Relative trans-spliced products of *mod(mdg4)* V forms were quantitated and normalized to the intact intron 4 (mean ± SEM; *n* = 3). For shorter tiled deletions, see Supplemental Figure S1C. (F) TSA RNA is sufficient to promote trans-splicing, while TSB RNA enhances the activity. TSA and TSB RNAs are fragments of the 330–500 nt and 650–800 nt in *mod(mdg4)* intron 4, respectively. Relative enhancement by TSB was normalized to TSA alone (mean ± SEM; *n* = 3). [Purple boxes] Chromosomal 5′ exons; [blue boxes] 3′ exons from the same DNA strand; [green boxes] 3′ exons from the opposite strand; [brown boxes] exon 4 on plasmids; [white boxes] internal cis-spliced intron in 3′ exons (Hygro) hygromycin B. Boxes and labels are used similarly in other places unless otherwise indicated. ∗∗∗|P < 0.001; |∗∗|P < 0.01; |∗| P < 0.05.
Trans-splicing in higher eukaryotes and a few events in nematodes are not promoted by SL RNAs; we refer to these events as non-SL trans-splicing. Such non-SL trans-splicing was first observed in in vitro assays using HeLa nuclear extract, where exons from two transcripts with mutually complementary intron sequences were spliced together [Konarska et al. 1985; Solnick 1985]. In the absence of complementary sequences, trans-splicing occurred at lower efficiency, which was further extensively characterized using a 5’SS-containing RNA oligo and a second RNA with a 3’SS [Konforti and Konarska 1995]. Therefore, current models favor long complementary sequences bringing two separate transcripts together to promote in vivo non-SL trans-splicing. This mechanism was proposed to facilitate trans-splicing of RNAi regulatory gene ERI-6/7 in C. elegans through base-pairing in the UTRs [Fischer et al. 2008] and facilitate trans-splicing of identified genes in Giardia intestinalis and Bombyx mori through long base-pairing between introns [Kamikawa et al. 2011; Shao et al. 2012]. Furthermore, this model of trans-splicing has been successfully applied to develop gene therapy strategies to replace mutated exons in several cancer cell lines [Puttera et al. 1999; Gruber et al. 2013].

However, it remains unclear whether all trans-splicing events in higher eukaryotes are promoted by complementary intronic sequences and how trans-splicing is promoted between transcripts that do not have obvious complementary sequences. To address these questions, we investigated trans-splicing of mod(mdg4) in Drosophila S2 cells and identified two critical RNA sequences (TSA and TSB) in the last 5’ intron of mod(mdg4). Mutagenesis analyses reveal that TSA is required and sufficient for trans-splicing and that TSB functions as an enhancer. Flies with deletions of TSA and TSB, prepared using the CRISPR/Cas9 system, exhibited defects in viability and embryonic development. Furthermore, two highly conserved core motifs across Drosophila species were found in TSA and TSB, respectively. The TSA core motif contains a pseudo-5’SS and recruits U1 snRNP through strong base-pairing with U1 snRNA to promote trans-splicing. A similar motif is also found in lola, allowing for mutual trans-splicing between lola and mod(mdg4).

Results
Identification of RNA elements required for trans-splicing of mod(mdg4)

Based on the current model that long complementary intronic sequences promote trans-splicing between two transcripts [Konarska et al. 1985; Fischer et al. 2008; Kamikawa et al. 2011], we searched the mod(mdg4) gene locus in Drosophila melanogaster [Dmel] for regions of complementarity between the last 5’ intron and introns in the 3’ transcripts by BLASTN [E-value < 10]; however, no obvious long base-paired regions were found. The trans-splicing of mod(mdg4), which results from one common 5’ transcript spliced to multiple 3’ transcripts, led us to hypothesize that critical RNA sequences would be located in the last common exon and intron [exon and intron 4] (Fig. 1A). To address this, plasmid-borne constructs that carry exon 4 with various lengths of intron 4 were transfected into Drosophila S2 cells; we then tested for trans-spliced products between the exogenous exon 4 and the endogenous 3’ exons of mod(mdg4) (Fig. 1B). In the presence of the full-length intron 4 [1979 nt], all tested 3’ exons were efficiently trans-spliced to the exogenous exon 4 [Fig. 1C, lane 1]. When intron 4 was shortened from its 3’ end to 1500 or 1000 nt, similar levels of trans-splicing products were still detected; however, it was reduced ~50% when intron 4 was shortened to 500 nt and was undetectable when shortened to ≤200 nt or when the 5’SS was mutated [Fig. 1C; Supplemental Fig. S1A]. In addition, replacing the exogenous sequence with other exons and introns from the 5’ transcript of mod(mdg4) resulted in no detectable trans-splicing [Supplemental Fig. S1B]. Similarly, intronic sequences from either Drosophila tubulin or rp49 did not generate detectable trans-spliced products [Fig. 1D]. However, the presence of intron 4 was sufficient for trans-splicing of truncated exon 4 and trans-splicing of exons from tubulin or rp49 [Fig. 1D; Supplemental Fig. S1B]. Taken together, these results demonstrate that the last common intron, but not other sequences, is required and sufficient for trans-splicing of mod(mdg4).

TSA RNA is sufficient for trans-splicing, whereas TSB enhances

To further define RNA sequences that facilitate trans-splicing, we performed two rounds of tiled deletions of intron 4. In the first round, deletions of nucleotides 250–350, 300–400, and 350–450 nearly abolished the trans-spliced products, and deletions of nucleotides 600–700, 650–750, 700–800, and 750–850 significantly decreased trans-splicing activity [Fig. 1E], whereas deletions of other regions did not, suggesting that two regions in intron 4 (300–500 nt and 650–800 nt) are important to the trans-splicing reaction, consistent with the above truncation assay. In the second round, the 300- to 500-nt region of intron 4 was further examined by shorter tiled deletions. Deletions of the 345- to 375-nt and 420- to 450-nt regions dramatically decreased trans-splicing activity [Supplemental Fig. S1C]. Taken together, we designated the 330- to 500-nt region as “TSA RNA” and the 650- to 800-nt region as “TSB RNA.”

We next asked whether TSA or TSB RNA is sufficient to promote trans-splicing in S2 cells. The exogenous tubulin sequence [exon1–intron1] did not promote trans-splicing with 3’ exons of mod(mdg4); however, the addition of TSA generated a trans-spliced mRNA product between the exogenous tubulin exon and the endogenous 3’ exon of mod(mdg4) [Fig. 1F]. In contrast, TSB alone did not promote trans-splicing but enhanced trans-splicing fourfold when present together with TSA [Fig. 1F]. We conclude that TSA is sufficient to promote trans-splicing and functions independently of other 5’ common sequences in mod(mdg4) and that TSB functions as an enhancer element to the TSA-dependent trans-splicing reaction.
Conserved motifs in TSA and TSB across Drosophila species are critical

To identify functional motifs in TSA and TSB, we aligned intron 4 sequences from homologous mod(mdg4) genes in other available Drosophila genomes and found two highly conserved regions [Supplemental Fig. S2]. The first conserved region is a nearly invariant 13-nt RNA motif (GGUA/GAGUAGAGUG) located in the previously described TSA; the second forms highly conserved secondary structures in TSB and varies from 79 to 95 nt long across Drosophila species [Fig. 2A; Supplemental Fig. S3]. Therefore, we designate these two conserved regions as the “TSA core” and “TSB core.”

We generated a series of mutations to investigate the function of the TSA and TSB cores. In the TSA core, six single G-to-C mutations exhibited various reductions of trans-splicing activity in which the G369C mutant provided the strongest inhibition [Fig. 2B]. Importantly, when all six guanosines were mutated to cytosines (6c), trans-splicing activity was almost totally abolished [Fig. 2B]. Moreover, the TSA core alone or with additional short sequences was sufficient to promote trans-splicing between the exogenous tubulin exon and endogenous 3′ exons of mod(mdg4) [Supplemental Fig. S1D]. These results demonstrate that the highly conserved TSA core is a critical element for trans-splicing and suggest that a new mechanism promotes trans-splicing in Drosophila.

The 79-nt-long Dm-TSB core contains 48 nt that are invariant in other Drosophila species. Using the mfold server [Zuker 2003], TSB core sequences were predicted to form secondary structures (Fig. 2C). The structures of stems I and III in the TSB core are required for enhancing trans-splicing activity. Mutated sites used in D are indicated by arrows here. For structures of TSB cores in other Drosophila species, see Supplemental Figure S3. [D] Secondary structures of stems I and III in the TSB core are required for enhancing trans-splicing activity. Mutated sites used in D are indicated by arrows here. For structures of TSB cores in other Drosophila species, see Supplemental Figure S3. [E] Validation of trans-splicing at the protein level. The coding sequence (CDS) of EGFP is split into two halves (EG and FP) followed by intronic sequences; trans-spliced products were detected by Western blot using anti-EGFP antibody. (EG-500) Exon EG with 1–500 nt of intron 4; (EG-1000) exon EG with 1–1000 nt of intron 4; (6c) all six Gs are mutated to Cs in the TSA core; (In-Srsf6) 1–5048 nt of Srsf6 intron 4. Cis-splicing constructs were used as controls.
to form stable conserved secondary structures containing four stems and two loops [Fig. 2C; Supplemental Fig. S3]. The most notable difference between TSB core secondary structures across Drosophila species is the length of stem II. To address the contribution of these structures in the TSB core, we generated mutations at conserved nucleotides to disrupt base-pairing in stems, restore base-pairing by compensatory mutations, and change sequences in loops [Fig. 2C]. When base-pairing in stem I or III was disrupted (nucleotides in one chain were mutated to cytosines), the enhancement by TSB RNA was significantly reduced from 4.9-fold by the wild-type sequence to 1.7-fold and 2.7-fold by stem I and III mutants, respectively [Fig. 2D]. Importantly, the reduced enhancement was partially rescued by compensatory mutations that restore base-pairing in stem I or III [Fig. 2D, lanes 3–6]. This indicates that the structures of stems I and III are critical for TSB’s enhancement to trans-splicing activity. Disruption of stem IV base-pairing significantly reduced the enhancement, but compensatory mutations of stem IV could not rescue the impairment [Fig. 2D, lanes 7,8], implying the importance of the primary sequence rather than its structure; alternatively, the restoration construct did not fold as predicted. Last, trans-splicing activity was not significantly altered when the loop I or II sequence was mutated [Fig. 2D, lanes 9,10]. These results demonstrated that the secondary structures of the TSB core are critical to enhance the trans-splicing activity of mod(mdg4).

To verify the effect of these findings on protein expression and confirm that the chimeric mRNAs detected above were not artificially generated by homology-driven template switching during RT–PCR, we split the EGFP coding sequence (CDS) into two halves and separately fused them with intronic sequences from mod(mdg4) on two plasmids [Fig. 2E, top]. The upstream exon [EG] was followed by various intron 4 sequences, while the downstream exon [FP] was fused with 3′ intron V. Thus, EGFP can be expressed only when the two exons are trans-spliced. Consistent with the above mRNA analyses, EGFP was detected when either 1–1000 nt (TSATSB) or 1–500 nt (TSATSA) of intron 4 were included. The intron 4 sequence containing both TSA and TSB provided higher expression of EGFP, and the TSB core mutant EG-1000 (6c) significantly decreased EGFP expression, whereas replacement of intron 4 with an intron from the Drosophila Srs6 gene did not produce EGFP [Fig. 2E, lanes 1–4]. These results again argue that the TSB core motif is the critical element in trans-splicing. In addition, cis-splicing constructs were tested and showed more efficient EGFP expression [Fig. 2E, lanes 8,9].

Lack of a 5′ SS in 3′ transcripts of mod(mdg4) determines alternative trans-splicing

The mod(mdg4) gene has at least 31 identified 3′ exons, which are located in several regions transcribed from both DNA strands [Yu et al. 2014]. Using the splice site-forecasting software Human Splicing Finder (HSF) [Desmet et al. 2009], we found no strong 5′ SS sequences in introns between the alternatively trans-spliced exons except the identified internal cis-spliced introns [Supplemental Fig. S4]. One possible explanation is that the presence of a 5′ SS would promote cis-splicing between 3′ exons and inhibit its trans-splicing to the 5′ exon. To test this, we cloned 3′ intron–exon sequences of mod(mdg4) and analyzed their trans-spliced products with endogenous exon 4 in S2 cells [Fig. 3A]. Two 3′ exons, U and V with their own introns or the introns of tubulin or rp49, were trans-spliced to the endogenous exon 4 in the absence of a 5′ SS [Fig. 3A, top]; however, trans-splicing was abolished and resulted in high levels of cis-splicing when a 5′ SS (CAAG/GUAAGU) was added [Fig. 3A, bottom]. This demonstrates that the presence of a 5′ SS in the 3′ intron inhibits trans-splicing. Furthermore, in the absence of a strong 5′ SS, trans-splicing still occurred at high levels when 3′ introns and exons from mod(mdg4) were replaced by control genes [Fig. 3B], suggesting that the primary sequence of the 3′ introns and exons is not an important factor for trans-splicing of mod(mdg4).

TSA and TSB are required for fly development

Benefiting from the recently developed CRISPR/Cas9 system [Ren et al. 2013], we performed two fly experiments to investigate in vivo functions of TSA and TSB.

We generated flies with deletions of TSA [TSAX−/−, Δ340–487 nt] or TSB [TSB−/−, Δ646–799 nt] [Fig. 4A]. Neither TSA−/− nor TSB−/− adults showed visible phenotypes; however, only 44.6% of TSA−/− and 33.5% of TSB−/− embryos hatched, showing a semilethal phenotype [Fig. 4B, lanes 2,3]. Since the embryonic development of TSB−/− flies unexpectedly exhibited more defects than TSA−/− flies, we further crossed deletion flies with wild-type flies in both directions and investigated the heterozygous embryos. The hatching rate of TSB−/− was similar to that of the wild type when the paternal fly was TSB−/− but was significantly reduced when the maternal fly was TSB−/−. In
contrast, hatching rates of TSA+/− heterozygotes were significantly affected in crosses of both directions (Fig. 4B, lanes 4–7). Moreover, RT–PCR analyses revealed that all tested mod(mdg4) isoforms were obviously decreased in the adult internal reproductive organs from TSA−/− flies and exhibited strong accumulation of unspliced premRNA, whereas trans-spliced isoforms in TSB−/− fly samples exhibited subtle and inconsistent changes depending on the isoforms (Fig. 4C). These results reveal that deletion of TSA significantly reduces trans-spliced isoforms of mod(mdg4) in vivo and affects survival of the progeny of both male and female flies, whereas deletion of TSB does not obviously affect the survival of the progeny of male flies, but a serious reduction in the survival of female flies was observed.

It has been reported that a transgenic fragment containing the entire 5′ transcribed region of mod(mdg4) could partially rescue fly lethality caused by transposon insertion at the common region of mod(mdg4) [Buchner et al. 2000]. To address whether the presence of TSA or TSB is required for the rescue, we first generated a deletion allele, mod(mdg4)Δcom, in which the 5′ common region of the locus was deleted using the CRISPR/Cas9 system, causing recessive lethality; we then generated transgenic flies with various 5′ common regions of mod(mdg4) inserted at the attP40 site on chromosome II for genomic rescue crosses (Fig. 4D, top). The wild-type and ΔTSC [deletion of a control region in intron 4] transgenic alleles could partially rescue the lethality of mod(mdg4)Δcom; however, the transgenic allele without the TSA or TSB sequence (ΔTSA or ΔTSB allele) could not (Fig. 4D, bottom), demonstrating that TSA and TSB are critical elements for trans-splicing of mod(mdg4) in vivo.

TSA RNA associates with U1 snRNP components

The TSA core motif contains a notable 5′SS-like sequence [GGUAG/AGUAGAGUG], and Dm-TSA RNA could form a secondary structure similar to SL RNAs from trypanosome and nematode [Supplemental Fig. S5A], raising the possibility that TSA could function like SL RNA to facilitate trans-splicing in Drosophila. However, constructs in which the TSA sequence was replaced with SL RNAs (SL from T. brucei or SL1 and SL2 from C. elegans) promoted little trans-splicing, which was not further decreased by mutations in either the 5′SSs or Sm-binding sites (Supplemental Fig. S5B), indicating that TSA RNA in Drosophila is not an analog of SL RNAs in lower eukaryotes.
To gain insights into how TSA and TSB facilitate trans-splicing, we identified associated factors by affinity purification of in vitro transcribed RNA with MS2 RNA-binding sequences. After incubation with S2 cell lysate, RNA-associated proteins were purified and identified by mass spectrometry (Fig. 5A). In comparison with TSB and the control (MS2 RNA-binding sequence only) RNAs, TSA RNA specifically purified components of U1 snRNP, including U1-70K, U1A, SmB, D1, D2, D3, E, and G, in which peptides from U1-70K were the most abundant (Fig. 5B, Supplemental Table S1). In contrast, TSB RNA specifically associated with several identified proteins but no U1 proteins [Supplemental Table S1]. To confirm the TSA RNA and U1 snRNP interaction, we analyzed associated proteins by Western blot. Consistent with the mass spectrometry data, wild-type TSA RNA purified both U1-70K and SmD2, but the TSA-6c mutant, the control, and TSB RNA did not (Fig. 5C). Similarly, TSA RNA specifically pulled down U1 snRNA, whereas the TSA-6c mutant and TSB RNA did not (Fig. 5D). These results demonstrate that both RNA and protein components of U1 snRNP are specifically associated with TSA in vitro.

The TSA core contains a pseudo-5′SS and forms 9 base pairs (bp) with the 5′ end of U1 snRNA (Fig. 5E, left), and mutations in the TSA core significantly reduced trans-splicing activity. Therefore, we hypothesized that base-pairing between the TSA core and U1 snRNA is a critical interaction for trans-splicing of mod(mdg4). To test this, we constructed U1 snRNA mutants that restore [U1-3g] or enhance [hyper-U1] base-pairing with TSA core mutants and predicted that such compensatory mutations of U1 snRNA would rescue the decreased trans-splicing activity of TSA core mutants in S2 cells. Consistent with the previous assays, both the TSA core mutants 3c and 6c decreased trans-splicing of mod(mdg4) (Fig. 5E, lanes 4, 6). However, additional expression of U1 snRNA with compensatory mutations, U1-3g and hyper-U1, significantly rescued the decreased trans-splicing of the 3c and 6c mutants (Fig. 5E, lanes 6–8), demonstrating that the TSA core motif recruits U1 snRNP through base-pairing with the 5′ end of U1 snRNA to promote trans-splicing.

lola has a TSA core-like motif and is mutually trans-spliced with mod(mdg4)

Trans-splicing of lola gene transcripts occurs between the last 5′ exon (exon 8) and alternative 3′ exons (Fig. 6A; Horiuchi et al. 2003; McManus et al. 2010). Aligning available intron 8 sequences of lola from five Drosophila species, we found nine conserved regions (Supplemental Fig. S6) in which one region (2191CAGGUA/GAGUAUGCU) was...
similar to the TSA core in mod(mdg4) containing a 5′SS with strong base-pairing (12 bp) to the 5′ end of U1 snRNA [Fig. 6B]. Deletion of a fragment containing this conserved region (Δ2150–2250 nt) dramatically decreased trans-spliced products of lola in S2 cells [Fig. 6C], implying the existence of a TSA core-like motif in lola. Considering similar gene structures [Figs. 1A, 6A], we hypothesized that trans-splicing may occur between lola and mod(mdg4), that the 5′ exons of lola could be trans-spliced to the 3′ exons of mod(mdg4), and vice versa. Indeed, we detected multiple lola-mod(mdg4) or mod(mdg4)-lola chimeric mRNAs that were generated by trans-splicing between endogenous and exogenous transcripts [i.e., chromosomal exon 4 of mod(mdg4) efficiently trans-spliced to the chromosomal or plasmid-borne 3′ exons AA and G of lola] [Fig. 6D]. Meanwhile, mod(mdg4)-lola chimeric trans-splicing activity was significantly reduced when using the mod(mdg4) 6c mutant [Fig. 6D, lane 3], demonstrating that trans-splicing of lola and mod(mdg4) requires the TSA core motif and shares a mechanism similar to trans-splicing within the mod(mdg4) gene locus.

Discussion

Trans-splicing increases genome complexity by generating chimeric mRNAs through intermolecular splicing reactions between two distinct pre-mRNA transcripts, which can be transcribed from distant genomic loci (Lasda and Blumenthal 2011). It has been reported that trans-splicing undergoes two transesterification steps similar to those the cis-splicing does (Yu et al. 1993; MacMorris et al. 2007) except leaving a Y-structured intron rather than a lariat intron (Murphy et al. 1986; Kamikawa et al. 2011). One of the key questions in trans-splicing is how those two nascent transcripts are brought into close proximity to allow subsequent catalysis of splicing. Several mechanisms have been proposed for trans-splicing events in higher eukaryotes, such as close cellular transcription sites, SR protein interactions, and complementary intronic sequences (Konarska et al. 1985; Bruzik 1996; Furuyama and Bruzik 2002; Gruber et al. 2013). In this study, through an effective trans-splicing detection system in Drosophila S2 cells and fly strains with precise deletions generated using the CRISPR/Cas9 system, two critical RNA motifs were successfully identified in the last common intron of the 5′ transcript of mod(mdg4). We demonstrate that one of the conserved motifs, the TSA core, strongly binds to U1 snRNP, providing a novel mechanism to bring separate transcripts together and promote trans-splicing in higher eukaryotes.

We obtained multiple lines of evidence to support that the intronic TSA core motif plays a fundamental role in trans-splicing of mod(mdg4). First, this 13-nt RNA motif is one of the few highly conserved regions in the nearly 2-kb length of mod(mdg4) intron 4s in all Drosophila species [Supplemental Fig. S2], and similar sequences are also observed in mod(mdg4) homologous genes in other insects [data not shown], showing a significant evolutionary importance of this motif. Second, TSA RNA is required and sufficient to promote trans-splicing in our trans-splicing detection system. Containing a pseudo-5′SS sequence, the TSA core motif binds U1 snRNP through strong base-pairing with the 5′ end of U1 snRNA. This association can be disrupted by mutations within this sequence, which significantly decreased trans-splicing activity, and trans-splicing activity was restored by compensatory mutations in U1 snRNA. Last, in vivo
deletion of a region containing the TSA core results in developmental defects of flies. The second conserved motif found in intron 4 of mod(mdg4), the TSB core, has a conserved secondary structure and associates with chromatin proteins and splicing factors to enhance trans-splicing of mod(mdg4).

Therefore, we propose a model in which two intronic RNA sequences, TSA and TSB, are critical to promote trans-splicing of mod(mdg4) [Fig. 7]. First, U1 snRNP binds to TSA RNA in the last 5′ intron of mod(mdg4) through strong base-pairing between the conserved TSA core and U1 snRNA. Second, TSB RNA forms a highly conserved secondary structure that is required for enhancing the trans-splicing activity of mod(mdg4). Recruitment of splicosomal factors by TSA and TSB RNAs might introduce a bridging interaction between the 5′ and 3′ trans-spliced introns; however, it remains unclear how this happens. TSB association with chromatin proteins implies that TSB-specific enhancement of trans-splicing could be regulated by transcription and could explain the observed regulation of trans-spliced mod(mdg4) isoform abundance by Brahma, the ATPase subunit of the SWI/SNF chromatin remodeling complex, and the local density of RNA polymerase II (Yu et al. 2014).

Introns often contain many pseudo-5′SSs (Zhang et al. 2005); however, splicosomal A complex formation on these sites is inefficient due to unproductive U2 snRNP recruitment [Dhir et al. 2010]. There are 10 and 16 pseudo-5′SSs predicted with high scores in the last 5′ introns of mod(mdg4) and lola, respectively (Supplemental Fig. S7). Why is the TSA core motif so important for trans-splicing of these two genes? First, we found that among those pseudo-5′SS-containing sequences, the TSA core motif in each intron forms the strongest base-pairing with U1 snRNA: 9 bp in mod(mdg4) and 12 bp in lola (Supplemental Fig. S7). Second, TSA and TSB core motifs are highly conserved across Drosophila species, whereas other pseudo-5′SS sequences are not (Supplemental Figs. S2, S6). In addition, deletion of the TSA core significantly, but not totally, decreased trans-splicing activity in both S2 cells and flies, suggesting that other pseudo-5′SS sequences might function as backup and inefficiently facilitate trans-splicing in the absence of the optimal sequence.

Additionally, as another determinant of trans-splicing, the absence of a strong 5′SS in the trans-spliced 3′ introns is necessary to avoid competition from cis-splicing. We found that adding a 5′SS to the alternative 3′ transcripts of mod(mdg4) resulted in internal cis-splicing and abolished trans-splicing [Fig. 3A]. Consistent with this, trans-spliced genes in Drosophila have been classified into three categories: genes with multiple 3′-terminal exons, genes with multiple first exons, or genes with very large introns (McManus et al. 2010), where the lack of strong splice sites is common.

After recruitment of splicosomal factors by TSA and TSB, there are several possible models to explain the follow-up trans-splicing of mod(mdg4). Binding of U1 snRNP to nascent transcripts prevents shortening and the use of cryptic polyadenylation sites (telescripting) [Kaida et al. 2010]. The presence of a strong pseudo-5′SS in intron 4 of mod(mdg4) could increase the abundance of mod(mdg4) transcripts containing intron 4 available for trans-splicing. However, this is unlikely to be the mechanism, as there are no significant changes in transcript levels when introns either contain a mutated TSA core motif or completely lack the sequence [Fig. 4C]. In addition, recursive cis-splicing has been well characterized in Drosophila and is often used for excising long introns in which the key sequence is a ratcheting point (Y)n-NCAG/GUAAGU, similar to the TSA core motif [Burnette et al. 2005; Venables et al. 2012]. However, our data do not support a recursive splicing model for mod(mdg4). First, if the trans-splicing reaction occurs between the TSA core motif and downstream 3′SS, the newly generated 3′SS of D. melanogaster is a suboptimal 3′SS [AAG/] and could be an even worse in other Drosophila species, such as in Drosophila viridis (GGG/) [Fig. 2A]. Second, compensatory mutations in U1 snRNA rescue the decreased trans-splicing activity of TSA core motif mutants in which the 5′SS is totally mutated [Fig. 5E], and recursive splicing cannot occur.

Trans-splicing of mod(mdg4) and trans-splicing of lola share a similar mechanism, both using a conserved intronic TSA core motif to promote the reaction [Fig. 6]. In order to verify whether this is a common mechanism of trans-splicing in Drosophila, we searched the last 5′
introns of other trans-splicing events in *Drosophila* (McManus et al. 2010) and found that 16 of the 126 introns have pseudo-5′SS-containing sequences, which form ≥9 bp with the 5′ end of U1 snRNA (i.e., TSA core-like motifs) [Supplemental Table S2]. Under such strict searching criteria, this result implies that the TSA core-mediated mechanism of trans-splicing could represent a general mechanism for a class of trans-splicing events in *Drosophila*.

Materials and methods

Strains and plasmids

For plasmids used in *Drosophila* S2 cells, sequences from *mod* (*mdg4*), *lola*, or other genes in *D. melanogaster* were inserted into modified pMT/V5-His B vector [Invitrogen] with hygromycin B and P copia promoter as described (Yang et al. 2013). Plasmids were transfected into S2 cells by Effectene transfection reagent [Qiagen] and induced by 0.5 mM CuSO₄ at 12 h after transfection; cells were collected 24 h later for isolation of RNA and proteins. The sequence of U1 snRNA, including a 1-kb region, was amplified from genomic DNA using primers listed in Supplemental Table S3. The CDS of EGFP was amplified from vector pEGFP-N1 [Invitrogen] and split after nucleotide G489.

RT–PCR and sequence analyses

Total RNA was extracted using Trizol [Invitrogen], and reverse transcription was performed by a first strand cDNA synthesis kit (RevertAid, Thermo) using either oligo(dT) or specific primers. Trans-spliced products were amplified by Ex Taq DNA polymerase (TaKaRa), confirmed by Sanger sequencing, and quantified using a Tanon-2500 gel analysis system. All sequences from the *Drosophila* species were downloaded from FlyBase and aligned using DNAman version 4.0 (LynnBioSoft). Potential splice sites were identified according to scores obtained by the online software HSF [Desmet et al. 2009].

RNA affinity purification and mass spectrometry

Along with 6xMS2 RNA-binding sites [Zhou et al. 2002], studied RNAs were in vitro transcribed by T7 RNA polymerase [Promega] and gel-purified. MS2-MBP protein was expressed in *Escherichia coli* and purified by Superdex 200 10/300 GL column (GE) [Tange et al. 2005]. Ninety picomoles of RNA was mixed with 90 picomoles of MS2-MBP (Zhou et al. 2002), studied RNA and proteins. The sequence of U1 snRNA, including a ±1-kb region, was amplified from genomic DNA using primers listed in Supplemental Table S3. The CDS of EGFP was amplified from vector pEGFP-N1 [Invitrogen] and split after nucleotide G489.

Western and Northern analyses

Antiserum against *Drosophila* U1-70K was generated by immunizing rabbits with the peptide DGGKIDSKRVLVDV (amino acids 164–177). Anti-SNRPD2 (SmD2) antibody was purchased from Abcam [ab155030]. Western blots of anti-EGFP and anti-tubulin were probed using mAb 7G9 [Abmart] and mAb DM1A [Sigma], respectively. Northern analysis was performed by transferring RNAs from 8 M urea-PAGE gel to Hybond N membrane [Amersham]. Membranes were then probed by DIG-labeled [Roche] antisense DNAs complementary to the full-length U snRNAs in *D. melanogaster*.

Fly strains and mutagenesis

Culture and crosses of *D. melanogaster* were carried out on standard medium and at standard temperature. Mutagenesis of the *Drosophila* mod(*mdg4*) gene were performed using the recently developed CRISPR/Cas9 system [Ren et al. 2013]. Briefly, target sequences of two guide RNAs [sgRNA] (listed in Supplemental Table S3) were selected for each genomic deletion. Each pair of sgRNA plasmids was co-injected into embryos of transgenic line nanos-Cas9(attP40) by Core Facility of *Drosophila Resource and Technique*, Shanghai Institute for Biological Sciences. G0 flies were then crossed with the *Tm3/Tm6* balancer strain [Bloomington no. 5906], subsequently, G1 male flies were crossed with the same balancer strain followed by genomic DNA screening for deletion strains. Obtained flies were then backcrossed at least five generations into a controlled uniform homogeneous genetic background [Bloomington no. 5905] to eliminate potential off-target events. To generate constructs for genomic rescue of lethality caused by the deletion of *mod*(*mdg4*) common region, a 6690-bp wild-type fragment containing a sequence from part of the upstream gene *tinman* to nucleotide 1564 of *mod*(*mdg4*) intron 4 was amplified and cloned into the pUAST[attB] vector. Three parallel deletion constructs—ΔTSA [340–487 nt], ΔTSB [646–799 nt], and a control, ΔTSA [1162–1310 nt]—were constructed by overlapping PCR based on the above wild-type plasmid. Genomic rescue constructs were microinjected and incorporated into the *attP40* site on chromosome II. Obtained flies were then crossed in crosses to test their ability to rescue the lethality of the homozygous *mod*(*mdg4*)/com*om* strain. For each cross, total offspring from six parallel vials were counted based on their genotypes.

Acknowledgments

We thank C. Query and B. Kosmyna at Albert Einstein College of Medicine and L. Rabinow at Centre de Neurosciences de Université Paris Sud XI for critical reading and discussion of the manuscript, H. Cheng at Shanghai Institutes for Biological Sciences for pMS2-MBP plasmid, and other members in the Xu laboratory for critical reading and discussion of the manuscript. This work was supported by grants from National Natural Science Foundation of China (NSFC) (31270842 and 31472045) and National Basic Research Program of China (2012CB114101) to Y.-Z.X., and NSFC (31400653) to W.S.


A conserved intronic U1 snRNP-binding sequence promotes *trans*-splicing in *Drosophila*


*Genes Dev.* 2015 29: 760-771
Access the most recent version at doi:10.1101/gad.258863.115

---

**Supplemental Material**
http://genesdev.cshlp.org/content/suppl/2015/04/01/gad.258863.115.DC1.html

**References**
This article cites 59 articles, 27 of which can be accessed free at:
http://genesdev.cshlp.org/content/29/7/760.full.html#ref-list-1

**Creative Commons License**
This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genesdev.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

**Email Alerting Service**
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.