In vivo effects on intron retention and exon skipping by the U2AF large subunit and SF1/BBP in the nematode *Caenorhabditis elegans*

Long Ma, Zhiping Tan, Yanling Teng, et al.

*RNA* 2011 17: 2201-2211 originally published online October 27, 2011
Access the most recent version at doi:10.1261/rna.027458.111

**Supplemental Material**  
http://rnajournal.cshlp.org/content/suppl/2011/10/18/rna.027458.111.DC1.html

**References**  
This article cites 55 articles, 22 of which can be accessed free at:  
http://rnajournal.cshlp.org/content/17/12/2201.full.html#ref-list-1

**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.

---

To subscribe to *RNA* go to:  
http://rnajournal.cshlp.org/subscriptions

Copyright © 2011 RNA Society
In vivo effects on intron retention and exon skipping by the U2AF large subunit and SF1/BBP in the nematode Caenorhabditis elegans

LONG MA,1,6 ZHIPING TAN,2,6 YANLING TENG,1 SEBASTIAN HOERSCH,3,4 and H. ROBERT HORVITZ5,7
1State Key Laboratory of Medical Genetics, School of Biological Sciences and Technology, Central South University, Changsha, Hunan 410078, China
2Center for Clinical Gene Diagnosis and Therapy, The Second Xiangya Hospital, State Key Laboratory of Medical Genetics, Central South University, Changsha 410078, China
3Koch Institute for Integrative Cancer Research, MIT, Cambridge, Massachusetts 02139, USA
4Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
5Department of Biology, Howard Hughes Medical Institute, MIT, Cambridge, Massachusetts 02139, USA

ABSTRACT
The in vivo analysis of the roles of splicing factors in regulating alternative splicing in animals remains a challenge. Using a microarray-based screen, we identified a Caenorhabditis elegans gene, tos-1, that exhibited three of the four major types of alternative splicing: intron retention, exon skipping, and, in the presence of U2AF large subunit mutations, the use of alternative 3′ splice sites. Mutations in the splicing factors U2AF large subunit and SF1/BBP altered the splicing of tos-1. 3′ splice sites of the retained intron or before the skipped exon regulate the splicing pattern of tos-1. Our study provides in vivo evidence that intron retention and exon skipping can be regulated largely by the identities of 3′ splice sites.

Keywords: C. elegans; SFA-1; UAF-1; alternative splicing

INTRODUCTION
RNA splicing removes noncoding introns of eukaryotic pre-mRNAs and joins neighboring exons to generate functional coding mRNAs (Reed 2000; Maniatis and Tasic 2002). Alternative splicing generates multiple transcript isoforms from a single pre-mRNA and is believed to be a major molecular mechanism responsible for the generation of the biological complexity of metazoans (Smith and Valcarcel 2000; Graveley 2001; Maniatis and Tasic 2002). In humans, most genes generate alternatively spliced isoforms (Johnson et al. 2003; Pan 2008; Wang et al. 2008). In the nematode Caenorhabditis elegans, ~20% of the genes are alternatively spliced (Zahler 2005; Ramani et al. 2011).

Four basic types of alternative splicing have been identified: the use of alternative 5′ splice sites, the use of alternative 3′ splice sites, intron retention, and exon skipping (cassette exons) (Nilsen and Graveley 2010). Different genes undergo different numbers or types of alternative splicing events. For example, the splicing of the Drosophila sex-determination genes dsx and transformer involves exon skipping and the use of alternative 3′ splice sites, respectively (Boggs et al. 1987; Burtis and Baker 1989). In C. elegans, egl-15 transcripts contain mutually exclusive exons that are included in different isoforms at distinct developmental stages (Goodman et al. 2003). The human KCNMA1 (SLO) gene pre-mRNA uses alternative 5′ splice sites, alternative 3′ splice sites, and exon skipping to generate >500 splice isoforms (Navaratnam et al. 1997; Rosenblatt et al. 1997). More dramatically, the Drosophila Dscam pre-mRNA has been suggested to generate >30,000 isoforms by combining constitutive exons with different cassette exons (Schmucker et al. 2000).

The 5′ splice site of an intron is recognized by the U1 small nuclear ribonucleoprotein particle (snRNP) (Madhani and Guthrie 1994). SF1/BBP (splicing factor one/branch-point binding protein) and the large and small subunits of U2AF (U2 auxiliary factor) function together to recognize the 3′ splice site of an intron (Krainer and Maniatis 1985; Zamore and Green 1991; Arning et al. 1996; Abovich and Rosbash 1997; Merendino et al. 1999; Wu et al. 1999; Zorio et al. 1999).

6These authors contributed equally to this work.
7Corresponding author.
E-mail horvitz@mit.edu.
Article published online ahead of print. Article and publication date are at http://www.mjaonline.org/cgi/doi/10.1261/rna.027458.111.
and Blumenthal 1999a). Alternative RNA splicing is generally achieved by interactions between splicing factors and cis-regulatory sequences in exons and introns. Splicing factors that bind exonic or intronic cis-regulatory nucleotide sequences include arginine-serine-rich RNA-binding SR proteins (Zahler et al. 1992; Fu 1995; Hertel et al. 1997; Long and Caceres 2009) and hnRNP RNA-binding proteins (Smith and Valcarcel 2000). Although >140 splicing factors have been identified in mammals (Zhou et al. 2002), it remains largely unknown how these factors interact to generate alternative splice isoforms of different genes.

By screening for essential C. elegans genes that can affect the rubberband Unc phenotype of unc-93(e1500) animals, we isolated mutations in the C. elegans orthologs of the U2AF large subunit (UAF-1) and SF1/BBP (SFA-1) (Ma and Horvitz 2009). Because these mutations cause conditional lethality—e.g., uaf-1(n4588) and sfa-1(n4562) cause temperature-sensitive lethality and maternal-effect sterility, respectively—they provide novel opportunities to study the in vivo regulation of splicing. To further analyze the functions of these splicing factors, we searched for target genes of UAF-1 by a whole-genome tiling microarray analysis that could detect differential splicing of genes between wild-type and uaf-1(n4588) animals (see Materials and Methods). We identified 11 candidate genes (Supplemental Table S1). We examined the splicing of all 11 candidate genes by RT-PCR and found that only the splicing of K07B1.6 was dramatically altered in uaf-1(n4588) animals (Fig. 1A) (see below). Because K07B1.6 provides a sensitive readout for studying alternative splicing and the biological function of K07B1.6 remains to be identified (see below), we named K07B1.6 “tos-1” (target of splicing). We subcloned DNA fragments from the RT-PCR products of the tos-1 transcripts in wild-type animals (Fig. 1A, wild-type) and determined the sequences of DNA inserts from 46 independent clones. We identified four major tos-1

**RESULTS**

**Splicing of K07B1.6 in wild-type animals involves intron retention and exon skipping**

uaf-1(n4588) is a temperature-sensitive mutation in the gene that encodes the C. elegans U2AF large subunit (UAF-1) and SF1/BBP (SFA-1) (Ma and Horvitz 2009). To study the in vivo function of uaf-1, we used a whole-genome tiling microarray to seek genes with altered splicing in uaf-1(n4588) animals (see Materials and Methods). We identified 11 candidate genes (Supplemental Table S1). We examined the splicing of all 11 candidate genes by RT-PCR and found that only the splicing of K07B1.6 was dramatically altered in uaf-1(n4588) animals (Fig. 1A) (see below). Because K07B1.6 provides a sensitive readout for studying alternative splicing and the biological function of K07B1.6 remains to be identified (see below), we named K07B1.6 “tos-1” (target of splicing). We subcloned DNA fragments from the RT-PCR products of the tos-1 transcripts in wild-type animals (Fig. 1A, wild-type) and determined the sequences of DNA inserts from 46 independent clones. We identified four major tos-1

**FIGURE 1.** uaf-1 mutations alter the splicing pattern of tos-1. (A) RT-PCR experiments examining tos-1 splice isoforms in different genetic backgrounds. Genotypes are labeled on top. Splice isoforms of tos-1 are illustrated on the right. (Arrows) Positions of PCR primers. (B) Major tos-1 splice isoforms identified by subcloning and sequence determination of RT-PCR products. The base-pair length of each isoform is indicated on the right. (Red boxes) Exons. Splicing at the cryptic 3′ splice site in intron 1 caused an altered exon 2′ (blue). (C) RT-PCR experiments examining the recognition of the cryptic 3′ splice site compared to that of the endogenous 3′ splice site of tos-1 intron 1. Splice isoforms are illustrated on the right. (D) Comparison of the endogenous and cryptic 3′ splice sites from tos-1 intron 1 and unc-93(e1500) intron 8 and exon 9. The position of each nucleotide is labeled on top.
splice isoforms (Fig. 1A,B), with isoform 4 being the most abundant (Figs. 1A,B, 2A). Isoforms 1, 2, and 5 were less abundant (Figs. 1A,B, 2A) (for isoform 3, see below). The nature of these isoforms indicates that in wild-type animals, the splicing of tos-1 involves intron 1 retention and exon 3 skipping. The combination of these two types of alternative splicing generates isoforms 1, 2, 4, and 5 (Fig. 1B). Quantification of the RT-PCR products indicated that in wild-type animals isoform 4 constitutes >80% of all transcripts (Fig. 2A).

**uaf-1 mutations alter the splicing of tos-1**

The *uaf-1(n4588)* mutation dramatically affected the splicing pattern of tos-1 [Fig. 1A, *uaf-1(n4588)*]. We analyzed the tos-1 splice isoforms in *uaf-1(n4588)* animals by subcloning and sequence determination. In *uaf-1(n4588)* animals, isoform 4 was dramatically reduced, while isoform 1, which was 6% of all transcripts, and a new isoform, 3, which was absent in wild-type animals, were increased to 32% and 33% of all transcripts, respectively (Figs. 1B, 2A). Isoform 3 was generated by the recognition of a cryptic 3' splice site in intron 1 not recognized in wild-type animals, suggesting that *uaf-1(n4588)* caused altered recognition between the endogenous 3' splice site and this cryptic site.

Previously, we isolated four intragenic mutations of *uaf-1(n4588)*—*uaf-1(n4588 n5120)*, *uaf-1(n5123)*, *uaf-1(n4588 n5125)*, and *uaf-1(n4588 n5127)*—that partially restored the recognition of the endogenous 3' splice site of *unc-93(e1500)* intron 8 in *uaf-1(n4588)* animals (Ma and Horvitz 2009). We found that these *uaf-1* mutants similarly partially restored the splicing of tos-1 to levels intermediate between that of wild-type and *uaf-1(n4588)* animals (Figs. 1A, 2A).

To examine how different *uaf-1* alleles affect the recognition of the cryptic 3' splice site of tos-1, we performed RT-PCR experiments to specifically visualize the recognition of the normal and cryptic 3' splice sites of intron 1 (Fig. 1C). Our results indicated that the recognition of the cryptic 3' splice site occurred in *uaf-1(n4588)* and *uaf-1(n4588 n5127)* animals but was not (or was only very weakly) detectable in the wild type and in other *uaf-1* mutant animals (Figs. 1C, 2B). We previously showed from a mutagenesis analysis that the *n4588* (T180I) mutation caused UAF-1 to favor a G instead of a T at position −4 of 3’ splice sites for the splicing of *unc-93(e1500)* exon 9 (Ma and Horvitz 2009). A comparison of the normal and cryptic 3’ splice sites from *unc-93(e1500)* and tos-1 indicates that both normal sites have a T nucleotide at position −4, while both cryptic sites have a G nucleotide at this position (Fig. 1D). Therefore, the altered preference for a G nucleotide at position −4 of 3’ splice sites by UAF-1(T180I) is seen in both *unc-93(e1500)* and tos-1 animals (see below).

**The biological function of tos-1 is unknown**

Isoform 4, the major splice isoform of tos-1, encodes a predicted protein of 61 amino acids (Supplemental Fig. S1). Minor splice isoforms 1 and 2 encode proteins of 157 amino acids and 163 amino acids, respectively (Supplemental Fig. S1). Splice isoform 5 encodes a predicted protein of 61 amino acids identical to that encoded by isoform 4 (Supplemental Fig. S1). The new splice isoform 3 in *uaf-1(n4588)* animals encodes a predicted protein of 127 amino acids (Supplemental Fig. S1). We identified no homologs of TOS-1 in species other than nematodes using the BLAST search algorithm. In the *Caenorhabditis elegans* genome, the gene *D1086.19* encodes a protein of unknown function that is 30% identical to the protein encoded by tos-1 isoform 4.

We isolated a tos-1 deletion mutation, *n5384Δ*, which removed part of exon 1, all of intron 1, and part of exon 2 (Supplemental Fig. S2). *n5384Δ* is predicted to delete 64 amino acids after amino acid 25 for isoforms 1 and 2 and would cause a frameshift after amino acid 25 if the truncated transcript is expressed, suggesting that *n5384Δ* is likely a null allele of tos-1. Previously, we proposed that an unknown
gene with splicing regulated by uaf-1 and sfa-1 might be required for the expression of the rubberband Unc phenotype of unc-93(e1500) animals (Ma and Horvitz 2009). The n5384Δ mutation did not cause an obviously abnormal phenotype and failed to suppress the rubberband Unc phenotype of unc-93(e1500) animals (L Ma and HR Horvitz, unpubl.). Similarly, reducing the expression of uaf-1(n4588) unc-93(e1500) uaf-1(n4588) unc-93(e1500); tos-1(n5384) animals (L Ma and HR Horvitz, unpubl.). Furthermore, uaf-1(n4588) unc-93(e1500); tos-1(n5384) animals were suppressed for the Unc phenotype as well as uaf-1(n4588) unc-93(e1500) animals (Ma and Horvitz 2009), suggesting that the function of tos-1 is not required for the suppression of unc-93(e1500) by uaf-1(n4588).

sfa-1(n4562) dramatically alters the splicing of tos-1

The mutation n4562 causes a nonsense mutation in and is a loss-of-function allele of the sfa-1 gene (Ma and Horvitz 2009). sfa-1 encodes the C. elegans ortholog of the splicing factor SF1/BBP (Mazroui et al. 1999; Ma and Horvitz 2009). sfa-1(n4562) does not alter recognition between the endogenous 3′ splice site in unc-93(e1500) intron 8 and the cryptic 3′ splice site in unc-93(e1500) exon 9 (Ma and Horvitz 2009). We examined whether sfa-1(n4562) altered the splicing of tos-1. As shown in Figures 2A and 3A, sfa-1(n4562) dramatically increased isoforms 2 and 5 and reduced isoform 4 of tos-1. Subcloning and sequence determination of the tos-1 RT-PCR fragments did not identify isoform 3, suggesting that sfa-1(n4562) did not cause recognition of the cryptic 3′ splice site in intron 1. This result was confirmed by RT-PCR analysis to specifically visualize the recognition of the cryptic 3′ splice site and the endogenous 3′ splice site of intron 1 (Fig. 3B).

We also examined the splicing of tos-1 in animals with the expression of uaf-1, sfa-1, or uaf-2 reduced by RNAi feeding. uaf-2 encodes the U2AF small subunit, which interacts with UAF-1 and SFA-1 to regulate 3′ splice site recognition (Zorio and Blumenthal 1999b). As shown in Figure 3, C and D, reducing expression of uaf-1, sfa-1, or uaf-2 by RNAi all caused a similarly weak increase of isoforms 2 and 5 and did not cause recognition of the cryptic 3′ splice site of intron 1, suggesting that uaf-1, sfa-1, and uaf-2 similarly regulate tos-1 alternative splicing in vivo. Compared with the major tos-1 isoform, isoform 4, isoform 2 results from intron 1 retention and exon 3 skipping, and isoform 5 results from exon 3 skipping, suggesting that reducing expression of uaf-1, sfa-1, or uaf-2 increases intron 1 retention and exon 3 skipping.

**Effects of uaf-1 mutations and sfa-1(n4562) on tos-1 intron 1 retention and exon 3 skipping**

Because the splicing of tos-1 involves intron 1 retention, exon 3 skipping, and, in the case of uaf-1 mutations, alternative 3′ splice site selections, an analysis of these different splicing events might reveal how each type of alternative splicing is regulated by uaf-1 and sfa-1. Figure 2B presents a quantitative analysis of the recognition of the cryptic 3′ splice site. We similarly examined intron 1 retention and exon 3 skipping. As shown in Figure 4A, ~12% of tos-1 transcripts have intron 1 retention in wild-type animals. uaf-1(n4588) and uaf-1(n4588 n5127) increased intron 1 retention to 50% and 46% of all transcripts, respectively. A weaker uaf-1 mutation, uaf-1(n4588 n5125), caused a twofold increase
mutations, such as uaf-1 of intron 1 retention to 26%, while other weak
FIGURE 4. Quantitative analysis of 3 skipping and RT-PCR analysis of exon 3 skipping in animals. (A) The molar ratio of all tos-1 splice isoforms with intron 1 retention, presented as a percentage of all isoforms combined. Error bars: standard deviation. (*) p < 0.05. (B) The molar ratio of all tos-1 splice isoforms with exon 3 skipping, presented as a percentage of all isoforms combined. Error bars: standard deviation. (*) p < 0.05. (C) RT-PCR analysis of exon 3 skipping in tos-1(n5384A) animals. Splice isoforms are indicated on the right. (Arrows) PCR primers.

of intron 1 retention to 26%, while other weak uaf-1 mutations, such as uaf-1(n4588 n5120) and uaf-1(n5123), did not cause an apparent increase of intron 1 retention. sfa-1(n4562) caused the most significant increase of intron 1 retention, to almost 60%, which was about a fivefold increase compared with the wild type (Fig. 4B).

It has been reported that with a reporter transgene a reduced splicing of the first intron often results in a reduced splicing of the second intron (Zhang and Blumenthal 1996). With the availability of the tos-1(n5384A) allele, we tested whether the deletion of intron 1 (n5384A) (Supplemental Fig. S1) would affect exon 3 skipping. As shown in Figure 4, B and C, exon 3 skipping was increased to ~32% of all transcripts in tos-1(n5384A) animals, which was a fourfold increase compared with wild-type animals. This result suggests that the sequences deleted by tos-1(n5384A) are important for reducing exon 3 skipping. However, whether cis-elements other than those required for the splicing of intron 1 are involved remains to be investigated (see below).

3’ splice sites regulate the alternative splicing of tos-1
To investigate the roles of 3’ splice sites in regulating tos-1 alternative splicing, we used site-directed mutagenesis to modify 3’ splice sites and analyzed these modified 3’ splice sites in transgenic animals. In C. elegans, TTTTCAG is the major consensus 3’ splice site and is strongly recognized in both in vitro and in vivo experiments (Hollins et al. 2005; Ma and Horvitz 2009). 3’ splice sites of lower frequency in the genome are likely to be recognized less efficiently. In tos-1, both intron 1 and intron 2 contain 3’ splice sites (Fig. 5A) that are less frequent in the genome (Kent and Zahler 2000a,b), while intron 3 contains the consensus TTTTCAG (Fig. 5A). Interestingly, the 3’ splice site of intron 3 was similarly strongly recognized in all uaf-1 and sfa-1 mutants tested (see above).

We replaced the 3’ splice sites of intron 1 (GGTTTCAG) and intron 2 (TCTCAAG) with the consensus site TTTTCAG in a tos-1 reporter gene (Fig. 5A) and generated transgenic animals in a wild-type background. Substituting the 3’ splice site of intron 1 with TTTTCAG caused the apparent disappearance of isoforms 1 and 2 and hence a failure of intron 1 retention (Fig. 5B,C, Tg-1). Similarly, substituting the 3’ splice site of intron 2 with TTTTCAG led to the disappearance of isoforms 2 and 5 and exon 3 skipping (Fig. 5B,D, Tg-2). As expected, when the 3’ splice sites in both intron 1 and intron 2 were substituted with TTTTCAG, isoforms 1, 2, and 5 were barely detectable, and both intron 1 retention and exon 3 skipping were greatly reduced (Fig. 5B–D, Tg-3).

We also mutated the conserved AG nucleotides at positions −1 and −2 of the 3’ splice site of intron 1 to TC, which should eliminate splicing at this site (Fig. 5A), and examined the splicing of the transgene. As shown in uaf-1(n4588 n5120), uaf-1(n5123), and uaf-1(n4588 n5127) did not cause apparent changes in exon 3 skipping. Interestingly, uaf-1(n4588 n5125) appeared to reduce exon 3 skipping to 3% of all transcripts. sfa-1(n4562) dramatically increased exon 3 skipping to 85% of all transcripts, which was more than a 10-fold increase over the frequency seen in the wild type (Fig. 4B).

We also analyzed the proportions of exon 3 skipping (Fig. 4B). In wild-type animals, exon 3 skipping was found in 8% of transcripts. uaf-1 mutations have differing effects on exon 3 skipping. uaf-1(n4588) caused a more than twofold increase of exon 3 skipping (to 18% of all transcripts), while

C. elegans alternative splicing
Ma et al.

Figure 5B (Tg-4), isoforms 4 and 5 disappeared, while isoforms 1 and 2 were still detectable. We observed 100% intron 1 retention (Fig. 5C) and, interestingly, increased exon 3 skipping (Fig. 5D). This result suggests that lack of intron 1 splicing increases exon 3 skipping, which is consistent with the conclusion obtained from analyzing the tos-1(n5384Δ) mutant (Fig. 4C). In agreement with this notion, animals carrying transgene Tg-1 (which substituted the endogenous 3’ splice site of intron 1 with TTTTCAG) showed a much more efficient splicing of intron 1 (Fig. 5B,C), which resulted in a reduced exon 3 skipping (Fig. 5B,D). Interestingly, these effects were not as strong as those that resulted from substituting the 3’ splice site preceding exon 3 with the consensus sequence TTTTCAG (Fig. 5A,B,D, Tg-2), which caused a complete disappearance of exon 3 skipping. Therefore, splicing at a strong 3’ splice site, such as TTTTCAG, might be less affected by the splicing of preceding introns.

That the splicing at the consensus 3’ splice site TTTTCAG is not affected by the splicing of other introns is consistent with our previous study of unc-93(e1500) splicing, in which a substitution with the consensus sequence TTTTCAG of the weak 3’ splice site at either intron 8 or exon 9 completely abolished the effects of surrounding sequences on splicing at these sites (Ma and Horvitz 2009).

To examine whether the G nucleotide at position −4 of the cryptic 3’ splice site in intron 1 is important for the recognition of this site by UAF-1(T180I) (see above), we replaced this nucleotide with a T, A, or C and tested the effects of these substitutions on the splicing of tos-1 transgenes in uaf-1(n4588) animals (Fig. 6A). Substituting the G with a T, A, or C all caused altered splicing of the tos-1 transgenes in uaf-1(n4588) background (Fig. 6B). A G-to-T substitution (Tg-5) increased intron 1 retention (Fig. 6C) and exon 3 skipping (Fig. 6D) and abolished the recognition of the cryptic 3’ splice site (Fig. 6E). A G-to-A (Tg-6) or a G-to-C (Tg-7) substitution similarly abolished the recognition of the cryptic 3’ splice site but did not affect intron 1 retention and exon 3 skipping (Fig. 6C–E). These results confirmed and extended our previous analysis (Fig. 1D; Ma and Horvitz 2009) that a G at position −4 of the cryptic 3’ splice site is important for the recognition by UAF-1(T180I). It is interesting to note that the G-to-T substitution appeared to cause a decrease in isoform 4, while the G-to-A substitution caused an increase of isoform 4 (Fig. 6B), implying the presence of a potential regulatory sequence for intron 1 splicing where these nucleotide reside.

We also tested the splicing of Tg-1, Tg-2, and Tg-3 in uaf-1(n4588) mutants (Fig. 6B–E). As in wild-type animals, when the weak 3’ splice site of intron 1 was substituted with the consensus TTTTCAG in Tg-1, intron 1 retention was abolished (Fig. 6B,C). When the weak 3’ splice site of intron 2 was substituted with the consensus TTTTCAG in Tg-2, exon 3 skipping was abolished (Fig. 6B,D). Both intron 1 retention and exon 3 skipping were abolished when both weak 3’ splice sites in introns 1 and 2 were substituted with TTTTCAG in Tg-3 (Fig. 6B–D). No recognition of the cryptic 3’ splice site was detected in Tg-1 (Fig. 6B,E), suggesting that the presence of the consensus TTTTCAG at the endogenous 3’ splice site of intron 1 suppressed the recognition of this cryptic site. This result is consistent with our mutagenesis analysis of the unc-93(e1500) transcript (Ma and Horvitz 2009), in which substituting either the weak endogenous 3’ splicing site in intron 8 or the cryptic 3’ splice site in exon 9 with the consensus TTTTCAG abolished splicing at the neighboring
weak site and caused splicing exclusively at the consensus site.

**DISCUSSION**

**tos-1 splicing could be used to analyze the in vivo regulation of alternative splicing**

Genetic analysis of *C. elegans* unc-52 mutants identified the splicing factors SMU-1, SMU-2, and MEC-8 as regulators of unc-52 exon 17 skipping (Lundquist et al. 1996; Spike et al. 2002; Spartz et al. 2004). The mutually exclusive splicing of exons of the *C. elegans* genes egl-15 and let-2 at different developmental stages has been used to identify regulatory splicing factors (Kuroyanagi et al. 2006; Ohno et al. 2008). We have shown that altered splicing of unc-93(e1500) exon 9 can serve as a reporter for studying the in vivo recognition of alternative 3' splice sites (Ma and Horvitz 2009). Splicing of tos-1 involves three of the four basic classes of alternative splicing events that can be easily detected and are affected by mutations of *uaf-1* and *sfa-1*. We suggest that tos-1 could serve as a sensitive and efficient in vivo reporter for studying splicing regulation.

**uaf-1 can affect multiple aspects of alternative splicing**

Numerous splicing factors and intronic and exonic cis-regulatory elements act in alternative splicing (Maniatis and Tasic 2002; Zhou et al. 2002; Wang and Burge 2008; Barash et al. 2010). However, the in vivo function of the U2AF large subunit in regulating alternative splicing in animals remains to be fully understood. In this study, we analyzed the effects of a series of *uaf-1* mutations on the pattern of tos-1 alternative splicing. We found that *uaf-1(n4588)* caused increased intron 1 retention, increased exon 3 skipping, and recognition of the cryptic 3' splice site (Supplemental Table S2). Intragenic suppressors of *uaf-1(n4588)* exhibited variable effects on these splicing events. For example, *uaf-1(n4588 n5127)* caused increased intron 1 retention, increased exon 3 skipping, and recognition of the cryptic 3' splice site but did not cause increased exon 3 skipping, while *uaf-1(n4588 n5125)* caused increased intron 1 retention, decreased exon 3 skipping, and displayed a weak recognition of the cryptic 3' splice site (Supplemental Table S2). Two other *uaf-1(n4588)* intragenic suppressors, *uaf-1(n4588 n5120)* and *uaf-1(n5123)*, did not obviously alter the recognition of the tos-1 endogenous 3' splice sites (Supplemental Table S2) or tos-1 splicing (see above). The strength of each *uaf-1* mutation in recognizing the cryptic 3' splice site of tos-1 correlated with its strength in altering the splicing of unc-93(e1500) exon 9 (Ma and Horvitz 2009). These results suggest that in animals, UAF-1 might play important roles in regulating intron retention and exon skipping and possibly in the choice of alternative 3' splice sites.

**FIGURE 6.** 3' splice site substitution and transgenic experiments identified the G nucleotide at position −4 of the intron 1 cryptic 3' splice site as important for recognition by UAF-1(T180I). (A) Illustration of transgene structure and substitutions at the cryptic 3' splice site. (Arrows) PCR primers for detecting transgene-specific transcripts. (B) Splice patterns of the transgenes indicated in a *uaf-1(n4588)* background as detected by RT-PCR. Splice isoforms are indicated on the right. (C) Quantitative analysis of intron 1 retention of each transgene obtained using the NIH Image J software. (D) Quantitative analysis of exon 3 skipping of each transgene obtained using the NIH Image J software. (E) Percentage of the recognition of the intron 1 cryptic 3' splice site of each transgene calculated as the percentage of isoform 3 in the total of isoforms 3, 4, and 5. (*) p < 0.05.
n4588 is likely both a loss-of-function and an altered-function mutation of uaf-1

We previously postulated that n4588 might cause a loss of function or an altered function of UAF-1, or both (Ma and Horvitz 2009). In this study, we found that uaf-1(n4588) caused increased intron 1 retention and exon 3 skipping, which is similar to the effect of reducing uaf-1 expression by RNAi (Fig. 3C) (see above). However, uaf-1(n4588) also caused recognition of the intron 1 cryptic 3’ splice site, which was not detectable in either sfa-1(n4562) mutants or in animals treated with RNAi targeting uaf-1, sfa-1, or uaf-2. This result suggests that the recognition of the cryptic 3’ splice site by uaf-1(n4588) per se might be caused by an altered function of UAF-1. We propose that n4588 is a mutation that causes both a loss of function and an altered function of UAF-1.

sfa-1 might regulate alternative splicing of a subset of genes in C. elegans

We previously failed to detect an effect of sfa-1(n4562) on the splicing of the unc-93(e1500) transcript. Our findings are consistent with the conclusions of Guth and Valcarcel (2000) and Tanackovic and Kramer (2005), who suggested that SF1/BBP might be required for the splicing of a subset of transcripts. In this study, we found that splicing of tos-1 is dramatically altered by sfa-1(n4562), providing in vivo evidence that SFA-1 can affect alternative splicing in C. elegans. In addition, sfa-1(n4562) did not cause recognition of the cryptic 3’ splice site in tos-1 intron 1 (Fig. 3B; Supplemental Table S2), which is consistent with our previous finding that sfa-1(n4562) did not cause recognition of a cryptic 3’ splice site in exon 9 of unc-93(e1500) (Ma and Horvitz 2009).

Reducing the expression of uaf-1, sfa-1, or uaf-2 by RNAi caused a slight increase of isoforms 2 and 5, similar to that caused by the sfa-1(n4562) mutation. The similarity of sfa-1(n4562) and sfa-1(RNAi) in affecting tos-1 alternative splicing is consistent with our previous conclusion that n4562 is a loss-of-function allele of sfa-1 (Ma and Horvitz 2009). In addition, the similarity of uaf-1(RNAi), sfa-1(RNAi), and uaf-2(RNAi) in affecting tos-1 splicing is consistent with the notion that UAF-1, SFA-1, and UAF-2 can interact to regulate the recognition of 3’ splice sites. However, it remains to be investigated whether UAF-1, SFA-1, and UAF-2 always function together to regulate splicing of the same set of genes in vivo.

Intron retention and exon skipping could be regulated by 3’ splice sites in C. elegans

The 3’ splice sites of intron 1 (GGTTCAG) and intron 2 (TCTCAAG) are relatively uncommon in C. elegans introns (Kent and Zahler 2000a,b) and are likely weakly recognized by the U2AF splice factors compared with the consensus site TTTTCAG, based on in vitro binding assays (Hollins et al. 2005). Because intron 1 retention and exon 3 skipping are caused by altered recognition at these two sites and splicing at the 3’ splice site of intron 3 (a consensus sequence TTTTCAG) was not altered by uaf-1 and sfa-1 mutations (Supplemental Table S2), we tested whether the identities of 3’ splice sites (Fig. 5) play important roles in regulating intron retention and exon skipping. We found that substituting the less frequent endogenous 3’ splice sites in intron 1 and intron 2 with the consensus sequence TTTTCAG caused a persistent recognition of the substituted sites, the abolishment of intron 1 retention and exon 3 skipping, and a dramatic change of the splicing patterns of the tos-1 transgenes. Our nucleotide substitution analysis indicates that the G nucleotide at position −4 of the intron 1 cryptic 3’ splice site is important for the recognition of this site by UAF-1(T180I) (Fig. 6), confirming and extending our previous nucleotide substitution analysis of the unc-93(e1500) cryptic 3’ splice site (Ma and Horvitz 2009).

In C. elegans and other animals, alternative splicing is regulated by exonic and intronic cis-regulatory elements and trans-splicing factors that bind these cis-elements (Maniatis and Tasic 2002; Zahler 2005; Kuroyanagi et al. 2006). The contribution of 3’ splice sites in intron retention and exon skipping remains to be fully understood. Our study suggests that in addition to the activities of the U2AF factors and SF1/BBP, the sequences of 3’ splice sites play important roles in intron retention and exon skipping in C. elegans. By changing 3’ splice sites or the activities of the U2AF factors and SF1/BBP, we could dramatically alter the splicing pattern of tos-1. Our study also provides an in vivo system, which includes a sensitive endogenous reporter, tos-1, and mutations affecting UAF-1 and SFA-1 that can be used for analyzing other trans-splicing factors in the regulation of alternative splicing in C. elegans.
MATERIALS AND METHODS

Strains

*C. elegans* strains were grown at 20°C as described (Brenner 1974), except where otherwise specified. N2 (Bristol) was the wild-type strain. Mutations used in this study include LGII: *uaf-1*(*n*4588, *n*5120, *n*5123, *n*5125, *n*5127) (Ma and Horvitz 2009), unc-93(*e*1500) (Greenwald and Horvitz 1980); LGIV: *sfa-1*(*n*4562) (Ma and Horvitz 2009); and LGV: *tos-1*(*n*5384Δ) (this study). The translocation *n*T1 IV/V with the dominant *gfp* marker *qls51* (Belfiore et al. 2002) was used to balance the *sfa-1* locus.

Whole-genome tiling microarray-based screen and gene expression analysis

Total RNA was prepared from synchronized L1 wild-type and *uaf-1*(*n*4588) animals using TRIzol as described by the manufacturer (Invitrogen). Double-stranded cDNA was prepared using a SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen) and was labeled and hybridized to a *C. elegans* whole-genome tiling microarray according to the manufacturer’s instructions (Arraymetrix). Each double-stranded cDNA sample was analyzed using one microarray chip each for wild-type and *uaf-1*(*n*4588) animals. Raw data were processed using the Tiling Analysis Software (TAS v1.1.02; Affymetrix) using quantile normalization, with both samples normalized together. Probe intensity data were then overlaid with gene annotations based on genomic coordinates (*C. elegans* assembly WS120/ce2, 2004). Genes with at least three probes exhibiting an expression difference of fourfold or more between *uaf-1*(*n*4588) and wild-type animals were examined using the Integrated Genome Browser (http://www.bioviz.org) with TAS output and annotated public domain exon structure data. We focused on genes with annotated splice isoforms that were altered in *uaf-1*(*n*4588) animals. We identified 11 such genes (Supplemental Table S1) as candidates and examined their mRNA expression in both wild-type and *uaf-1*(*n*4588) animals using RT-PCR experiments (see below). Only one gene (*tos-1*) was confirmed to exhibit apparent altered splicing in *uaf-1*(*n*4588) animals.

RT-PCR and quantitative analysis of splice isoforms

Total RNA was prepared using TRIzol according to the manufacturer’s instructions (Invitrogen), treated with RNase-Free DNase I (New England Biolabs), and incubated for 10 min at 75°C to inactivate DNase I. First-strand cDNA was synthesized with random hexamer primers (New England Biolabs) using the SuperScript III First-Strand Synthesis Kit (Invitrogen). RT-PCR was performed using an Eppendorf cycler with the following conditions: 30 sec at 94°C, and 1 min at 72°C, 30 cycles. PCR primers for full-length *tos-1* cDNA were 5’-ATCTACGGATTCGAGTCGTCACCATC-3’ and 5’-TCAAGCGC TATCCCTCAGTGACTTC-3’. PCR primers for analyzing the recognition of the intron 1 endogenous and cryptic 3’ splice sites were 5’-ATCTACGGATTCGAGTCGCATCACC-3’ and 5’-GAAGAA ATCTTCCAGTCCGAGG-3’. PCR primers for analyzing transcript-specific transcripts were 5’-ATGATCTACGGATTCG-3’ and 5’-GGAATTCGATGTTACCGTGATC-3’. Two or three independent samples for each genotype of animals at mixed stages were analyzed, and the proportions of each splice isoform were quantified from each biological replicate using the NIH Imagejal software to measure the intensity of each isoform separated by agarose gels. We confirmed that our PCR assay could quantitatively amplify each *tos-1* splice isoform by using ~10<sup>6</sup> copies of purified cDNA for each isoform as a template and obtaining similar molar amounts of each isoform in the final product (Supplemental Fig. S3). For transgene analysis in a wild-type background, one stable line was analyzed for each of the transgenes *Tg-1*, *Tg-2*, and *Tg-4* (Fig. S5), and two stable lines were analyzed for each of the transgenes *Tg-wt* and *Tg-3* (Fig. S5). For transgene analysis in a *uaf-1*(*n*4588) background, one stable line for *Tg-3* and two stable lines for each of the transgenes *Tg-wt*, *Tg-1*, *Tg-2*, *Tg-5*, *Tg-6*, and *Tg-7* were analyzed (Fig. S5). Proportions of intron 1 retention or exon 3 skipping were calculated as the combined ratios of all transcripts with intron 1 retention or exon 3 skipping, respectively.

Identification of splice isoforms

RT-PCR samples were resolved with 2.5% agarose gels, and all visible PCR bands were isolated using the QIAquick Gel Extraction Kit (QIAGEN) as a mixture. The PCR product was subcloned to a pGEM-T Easy vector (Promega), and the sequences of each individual insert of 40 or more clones were determined. The nucleotide sequences of each splice isoform and the amino acid sequences of their encoded protein products are listed in Supplemental Figure S1.

Isolation of deletion alleles

Genomic DNA pools from EMS-mutagenized animals were screened for deletions using PCR as described (Jansen et al. 1997; Liu et al. 1999). Deletion mutant animals were isolated from frozen stocks and backcrossed to the wild type at least three times. *tos-1*(*n*5384Δ) removes nucleotides 16,301–16,640 of cosmid K07B1.

RNA interference

Young adult animals [wild-type or *unc-93*(*e*1500)] were fed HT115(DE3) bacteria containing plasmids directing the expression of dsRNAs targeting *uaf-1*, *sfa-1*, *uaf-2*, or *tos-1* on NGM plates with 1 mM IPTG and 0.1 µg/mL ampicillin (Timmons et al. 2001). F<sub>1</sub> progeny of wild-type or *unc-93*(*e*1500) animals were examined for obviously abnormal phenotypes or suppression of locomotion defects, respectively. We generated the DNA construct expressing dsRNA targeting *tos-1* (see below). RNAi feeding for *uaf-1* and *sfa-1* was performed as described (Ma and Horvitz 2009). The *uaf-2* RNAi clone was derived from an RNAi library (Kamath et al. 2003), and the sequence of the insert was determined to confirm the presence of the *uaf-2* coding region.

Plasmids

We cloned a full-length *tos-1* genomic DNA (from ATG to TGA) into a pGEM-TA easy vector (Promega) and subcloned a PstI/Apal fragment (670 bp) from this construct into pPD129.36 using PstI and Apal sites. This construct contains sequences from the first exon to a portion of the third intron of *tos-1* and was used for the RNAI feeding experiment. To examine the effects of different 3’ splice sites on *tos-1* intron 1 retention and exon 3 skipping,
Ma et al.

a genomic sequence of *los-1* containing a 1.3-kb promoter region, the whole coding sequence without the stop codon, and a 0.6-kb 3′ region was subcloned into a pGEM-TA easy vector (Promega). A short multiple cloning sequence of the pGEM-TA vector was inserted between the coding region and the 3′ region, which was used for detecting transgene-specific transcripts. Mutations in transgenes for splice site substitution analysis were introduced as described with primers containing corresponding mutations (Chiu et al. 2004). PCR was performed using Eppendorf Cyclers, and DNA products were resolved using agarose gels. DNA sequence determination was performed with an ABI Prism 3100 Genetic Analyzer and an ABI 3730XL DNA Analyzer.

**Transgene experiments**

Germline transgene experiments were performed as described (Mello et al. 1991). Transgene mixtures generally contained 50 μg/ml 1-kb DNA plus ladder (Invitrogen), 20 μg/ml *Arabidopsis* genomic DNA, 5 μg/ml pPD93.97 as a coinjection marker, and 5 μg/ml the transgene of interest. pPD93.97 drives the expression of GFP in body-wall muscles.

**SUPPLEMENTAL MATERIAL**

Supplemental material is available for this article.

**ACKNOWLEDGMENTS**

We thank members of the Horvitz laboratory, especially Daniel Denning, Nicolas Paquin, Takashi Hirose, and Dengke Ma, for critical suggestions. We thank Na An and Yaoyu Zhong for strain maintenance and Rita Droste, Fei Wang, and Liange Huang for technical support. We thank Andy Fire for providing the pPD plasmids we used to construct our RNAi plasmids. Some nematode strains used in this study were provided by the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources. Some genomic sequences and polymorphisms were identified by the WUSTL Genome Center. H.R.H. is the David H. Koch Professor of Biology at MIT and an Investigator of the Howard Hughes Medical Institute. This study was supported by a Natural Science Foundation of China grant (30971639) and a National Basic Research Program of China grant to L.M. (2011CB510005) and NIH Grant GM24663 to H.R.H.

Received March 23, 2011; accepted September 27, 2011.

**REFERENCES**


C. elegans alternative splicing


www.rnajournal.org 2211