Splicing does the two-step

Heidi Cook-Andersen and
Department of Reproductive Medicine, University of California, San Diego, La Jolla, California, 92093, USA

Miles F. Wilkinson
Department of Reproductive Medicine, University of California, San Diego, La Jolla, California, 92093, USA

Institute of Genomic Medicine, University of California, San Diego

Miles F. Wilkinson: mfwilkinson@ucsd.edu

Abstract

An intricate recursive RNA splicing mechanism that removes especially long introns (non-coding sequences) from genes has been found to be evolutionarily conserved and more prevalent than previously thought.

One of the biggest surprises in molecular biology was the discovery in 1977 that coding information in genes is interrupted by non-coding sequences known as introns. Much has since been learned about how introns are recognized and spliced out of precursor RNA to yield mature messenger RNA in which the remaining sequences — the exons — are stitched together. A lingering challenge has been to work out the way in which long introns are correctly recognized and spliced out, because they have a greater potential for splicing errors than do short introns.

One intriguing solution to this problem arrived 17 years ago, with the discovery that a long intron in the *Ultrabithorax* gene in the fruit fly *Drosophila melanogaster* is removed in a progressive, stepwise fashion, thereby reducing the size of the chunks that need to be defined for splicing[1]. However, subsequent studies identified only a handful of fly genes that undergo this 'recursive' splicing[2,3], and no examples were demonstrated in other species[4], casting doubt on the generality of the process. Two papers in this issue report that recursive splicing is actually quite widespread in fly genes[5] and that it is also used by genes expressed in the human brain[5,6].

Recursive splicing depends on juxtaposed 3′ and 5′ splice-site sequences, called recursive splice sites, in the middle of long introns (Fig. 1a). Duff *et al.*[5] (page 376) set out to identify recursive splice sites in *D. melanogaster* using deep-sequencing methods. Their screen yielded 197 functional recursive splice sites, many of which were highly conserved across several *Drosophila* strains. The authors identified a total of 115 fly genes that undergo recursive splicing, greatly expanding the range of this mechanism.

By evaluating the spliced-out intron segments (lariats), Duff *et al.* obtained evidence that recursive splicing is a sequential and largely obligate process for genes that have recursive

splice sites. They also found that recursive 3′ splice sites are typically richer in the long tracts of pyrimidines (the nucleotide bases cytosine and uracil) required for splicing than are non-recursive 3′ splice sites. This raises the possibility that their splicing depends more than that of typical introns on the polypyrimidine-tract-binding protein U2AF. Indeed, the authors found that recursive splicing is strikingly more sensitive to U2AF depletion than is canonical splicing. The physiological significance of this intriguing discovery remains to be determined.

Sibley et al.[6] (page 371) addressed the long-standing question of whether recursive splicing is evolutionarily conserved. Using two complementary approaches, they identified nine genes that undergo recursive splicing in the human brain. In contrast to sites in *Drosophila*, in which the majority of recursive introns are completely spliced out[1–3,5], all recursive splice sites identified in humans harboured an 'RS exon' that seems to be pivotal for removing the long intron and can be retained in some circumstances (Fig. 1b).

The authors identified two roles for the RS exon in recursive splicing in humans. First, it facilitates recognition of the recursive splicing site, presumably through the process of exon definition. This is a complex mechanism that defines splice sites on either side of an exon through recruitment of splicing-promoting proteins[7]. Second, it provides opportunities for quality control: RS exons are almost always spliced out of normal mRNAs, but the authors found that they are usually retained when the upstream exon is generated from an aberrant promoter sequence or from a potentially faulty splicing event. RS-exon inclusion is favoured in these instances because its 5′ splice site drives splicing more effectively than the 5′ splice site required to remove the RS exon.

RS-exon retention often leads to death of the mRNA, because RS exons typically contain in-frame premature-termination codons — sequences that cause the mRNA to be degraded by the nonsense-mediated RNA decay (NMD) pathway[8] (Fig. 1b). This is physiologically relevant because most RS-exon-containing mRNAs are probably 'garbage' transcripts. But a subset of these mRNAs may be functional; their formation might be induced when NMD is repressed, such as during particular stages of development and in response to stress[8].

Why do humans and *Drosophila* seem to use different mechanisms to splice out recursive exons? Species-specific splicing factors may be one explanation. Alternatively, differential RS-exon usage might result from known differences in how these two species define splice sites[7]. It could also be that the differences in these two species seem greater than is actually the case — for example, RS exons might participate in an intermediate step of *Drosophila* recursive splicing, being included in mature RNAs so infrequently that they are usually undetectable.

It was previously proposed that recursive splicing might increase the fidelity of splicing[1–3]. Sibley et al. examined this possibility using antisense oligonucleotide molecules to block recursive splice sites. They found that this had no obvious effect on the recursive splicing of two human genes, and only modestly inhibited recursive splicing of a zebrafish gene. These data suggest that recursive splicing is not required for the efficiency or accuracy of long-intron splicing. It is possible, however, that this experiment did not reveal a crucial role of

recursive splicing because blockade of the natural recursive splice site led to the use of other recursive splice sites that are not normally used.

Duff *et al.* performed extensive genomewide analyses of *Drosophila* (35 dissected tissues, 24 cell lines and 30 developmental stages) and found that recursive splicing occurs in about 6% of long introns in all tissues tested. By contrast, recursive splicing may exhibit some tissue specificity in humans. Sibley *et al.* found that genes with long introns tend to be expressed in the human nervous system, and they identified recursively spliced RNAs expressed in the human brain[6]. Duff *et al.* detected some selectivity for recursive splicing in the brain in a screen of 20 human tissues (including fetal brain and adult cerebellum), but this may partly reflect the difficulty of detecting recursively spliced RNAs in tissues that express such RNAs at low levels. It will be important to determine whether this specificity, if real, results from the tendency of recursively spliced genes to be expressed in the brain, or whether cells in the nervous system have factors that promote recursive splicing.

Many genes that have long introns, including those that undergo recursive splicing, are linked to neurological diseases and to autism[9–11]. Whether these conditions are sometimes triggered by errors in the multistep recursive RNA-splicing process will be an exciting avenue for future studies.

References

1. Hatton AR, Subramaniam V, Lopez AJ. Mol. Cell. 1998; 2:787–796. [PubMed: 9885566]

2. Burnette JM, Miyamoto-Sato E, Schaub MA, Conklin J, Lopez AJ. Genetics. 2005; 170:661–674. [PubMed: 15802507]

3. Grellscheid S-N, Smith CWJ. Mol. Cell. Biol. 2006; 26:2237–2246. [PubMed: 16508000]

4. Shepard S, McCreary M, Fedorov A. PloS ONE. 2009; 4:e7853. [PubMed: 19924226]

5. Duff MO, et al. Nature. 2015; 521:376–379. [PubMed: 25970244]

6. Sibley CR, et al. Nature. 2015; 521:371–375. [PubMed: 25970246]

7. Berget SM. J. Biol. Chem. 1995; 10:2411–2414. [PubMed: 7852296]

8. Karam R, Wengrod J, Gardner LB, Wilkinson MF. Biochim Biophys Acta. 2013; 1829:624–633. [PubMed: 23500037]

9. Lagier-Tourenne C, et al. Nature Neurosci. 2012; 15:1488–1497. [PubMed: 23023293]

10. Polymenidou M, et al. Nature Neurosci. 2011; 14:459–468. [PubMed: 21358643]

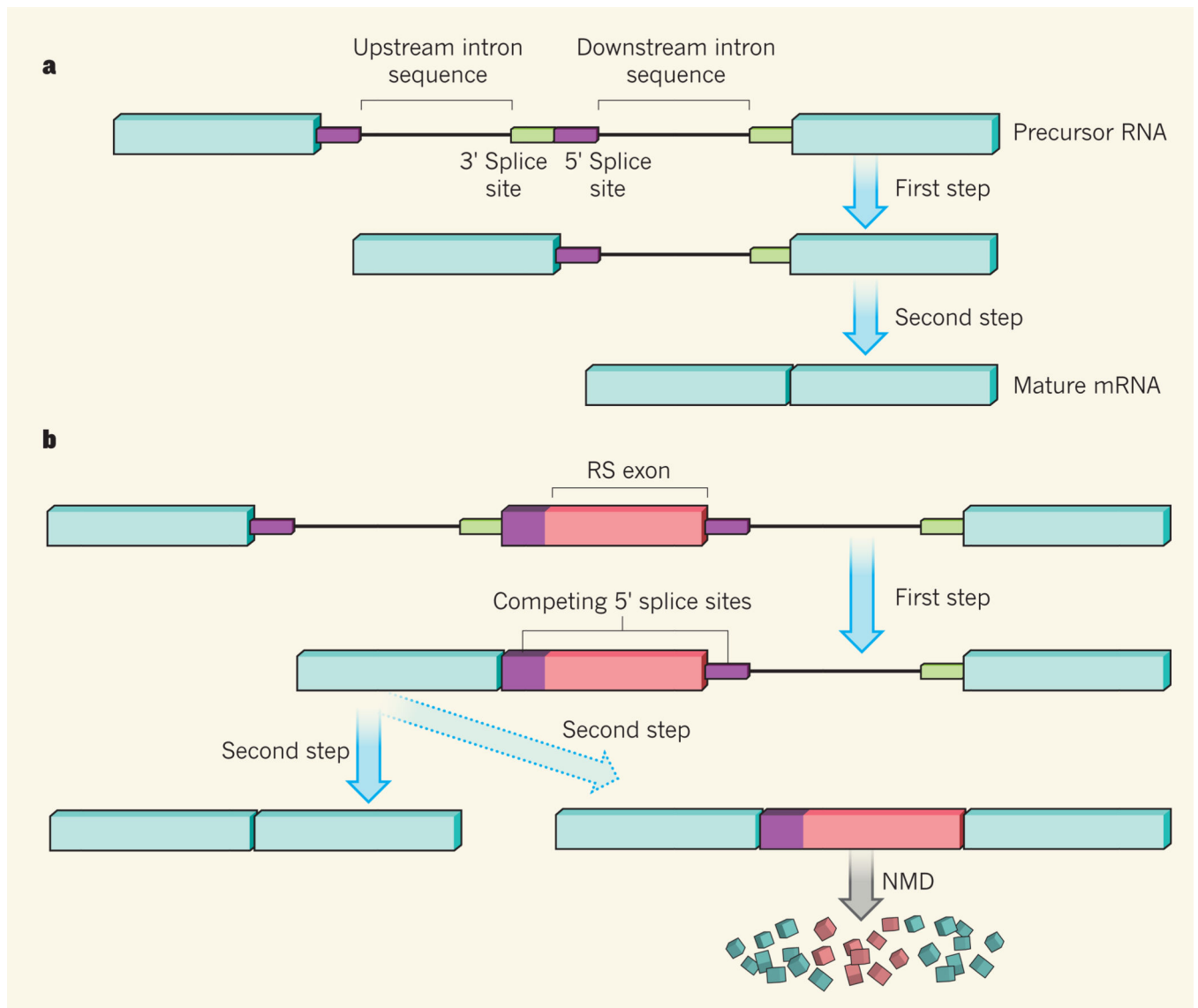11. King IF, et al. Nature. 2013; 501:58–62. [PubMed: 23995680]

Figure 1. Mechanisms of recursive splicing
a, In recursive splicing, long intron sequences of precursor RNA are removed in a stepwise process mediated by juxtaposed internal 3′ and 5′ splice sites. In the first step, the 3′ splice site is used to remove the upstream intronic sequences. The second step uses the 5′ splice site to remove the downstream intron sequences, forming a mature messenger RNA. Duff *et al.*[5] report that this recursive splicing process occurs in the fruit fly *Drosophila melanogaster* much more commonly than was previously thought. **b**, Sibley *et al.*[6] find that some recursively spliced messenger RNAs — including all those known in humans — contain a recursive splicing (RS) exon. The RS exon can be either completely removed or retained in the mature mRNA, depending on which of two competing 5′ splice sites is used in the second step. Most mRNAs that harbour RS exons are degraded by nonsense-mediated RNA decay (NMD).