6850–6862 *Nucleic Acids Research, 2012, Vol. 40, No. 14* *Published online 13 April 2012*
doi:10.1093/nar/gks322

The transition in spliceosome assembly from complex E to complex A purges surplus U1 snRNPs from alternative splice sites

Mark J. Hodson[1], Andrew J. Hudson[2], Dmitry Cherny[1] and Ian C. Eperon[1,*]

[1]Department of Biochemistry and [2]Department of Chemistry, University of Leicester, Leicester LE1 9HN, UK

Received February 13, 2012; Revised March 19, 2012; Accepted March 23, 2012

ABSTRACT

Spliceosomes are assembled in stages. The first stage forms complex E, which is characterized by the presence of U1 snRNPs base-paired to the 5′ splice site, components recognizing the 3′ splice site and proteins thought to connect them. The splice sites are held in close proximity and the pre-mRNA is committed to splicing. Despite this, the sites for splicing appear not to be fixed until the next complex (A) forms. We have investigated the reasons why 5′ splice sites are not fixed in complex E, using single molecule methods to determine the stoichiometry of U1 snRNPs bound to pre-mRNA with one or two strong 5′ splice sites. In complex E most transcripts with two alternative 5′ splice sites were bound by two U1 snRNPs. However, the surplus U1 snRNPs were displaced during complex A formation in an ATP-dependent process requiring an intact 3′ splice site. This process leaves only one U1 snRNP per complex A, regardless of the number of potential sites. We propose a mechanism for selection of the 5′ splice site. Our results show that constitutive splicing components need not be present in a fixed stoichiometry in a splicing complex.

INTRODUCTION

Little is known at a molecular level about the processes that connect the recognition of potential splice sites to the selection of particular splice sites during pre-mRNA splicing. Splicing is essential for mammalian gene expression, because almost all protein-coding genes contain introns, but it is quite unlike transcription and translation in that there is no stable complex that mediates catalysis and there are no high-resolution structures for the catalytic entity. Moreover, there are on average around seven significant alternative splicing events per multi-exon gene, generating an as yet uncounted number of mRNA and protein isoforms (1,2). This means that there is a very high level of flexibility in the processes recognizing the sites of action.

Both *in vivo* (3,4) and *in vitro* (5,6), spliceosomes assemble on each intron by the incorporation of five small ribonucleoprotein particles (snRNPs) in a series of consecutive steps. *In vitro*, native gel electrophoresis has shown that a simple transcript containing two exons separated by a short intron will be incorporated successively into complexes designated E, A, B and C (7,8). Complex E contains U1 snRNP, base-paired via its U1 snRNA component with the 5′ splice site, and weakly associated U2 snRNP (9). The U2 snRNP becomes base-paired with the branchpoint in complex A, and the U4/U6.U5 tri-snRNP is incorporated in complex B. Complex B undergoes substantial remodelling, including the loss of U1 and U4 snRNPs, to form a catalytically competent spliceosome that mediates the first transesterification reaction (10). The intermediate-containing complex C catalyses the second reaction, in which the exons are spliced together. The number of protein components in each complex has been catalogued in full only on a simple model substrate, MINX, where there appear to be about 100–140 proteins in the purified complexes (9,11). Not all of these have been identified.

Complex E is necessary for the subsequent assembly of complex A (12). It assembles only on substrates with splice sites and it is committed to splicing (9–14). There is a functional association in assembly between the 5′ and 3′ splice sites (14). Recent evidence suggests that this functional association reflects a physical link. Complex E on MINX contains the U1 snRNP, the U2 snRNP, the SMN complex and other proteins that might mediate or affect recognition of the splice sites (9). These include the SR proteins, a family of RNA-binding proteins involved in constitutive and alternative splicing that is characterized by the presence of one or two RRM-type RNA-binding domains and a C-terminal domain rich in RS dipeptides.

*To whom correspondence should be addressed. Tel: +44 116 2297012; Fax: +44 116 2297018; Email: eci@leicester.ac.uk

One of these, the protein SRSF1, is known to affect 5′ splice site selection and the efficiency of complex E formation (15) by stabilizing the binding of U1 snRNPs (16–18). U2 auxiliary factor (U2AF), a heterodimer, recognizes polypyrimidine tracts at the 3′ splice site. In addition, there are two proteins that might mediate interactions between the 5′ and 3′ splice sites. PRPF40A is an orthologue of yeast Prp40, which is part of the U1 snRNP and can interact with the branchpoint-binding protein, although in humans PRPF40A is associated with U2 snRNPs (9). DDX46 is an orthologue of yeast Prp5, which can bridge U1 and U2 snRNPs (19). The existence of these interactions is supported by experiments using hydroxyl radical probes attached either near the 3′ splice site (20) or at the 5′-end of U2 snRNA (21), which showed that the 5′ and 3′ splice sites are close in complex E, i.e. within ~2 nm. Thus, it appears that in complex E splicing components are bound to the splice sites and the sites are held in close proximity.

The foregoing picture suggests that the splice sites are selected before or during the process of complex E formation. However, for both 5′ and 3′ splice-sites, it has been found that a shift in preferences between alternative sites can be induced after complex E formation by the addition of a splicing activator that binds a nearby enhancer sequence (22,23). Particular sites are only committed to splicing after assembly of complex A. In the case of 5′ splice sites, this commitment process required a separate ATP hydrolysis step, although its role is not clear (22). These results appear to conflict with the evidence that the 5′ and 3′ sites are recognized and, apparently, held together in complex E. One possibility in regard to 5′ splice-sites is that the U1 snRNP forms only weak interactions with a 5′ splice site and that, tethered within complex E, a single U1 snRNP is able to dissociate and reassociate with candidate sites. A critical test of this is whether only one U1 snRNP is bound to the pre-mRNA in complex E. We have previously shown that both of two alternative 5′ splice sites are protected concurrently against digestion by RNase H, but it is not possible to exclude the explanation that only one U1 snRNP was bound and that it sequestered both sites within complex E. Determination of the number of U1 snRNPs present in complex E can be done using single molecule methods. We show here that most pre-mRNA molecules with two alternative 5′ splice sites are bound by two U1 snRNPs in complex E, but an ATP-dependent process eliminates one of the snRNPs during progression to complex A. We infer that the ATP-dependent commitment of sites involves destabilization of alternative complexes.

MATERIALS AND METHODS

Radioactive pre-mRNA, splicing *in vitro* and gel analysis of complexes

Transcripts were prepared from β-globin and adenovirus based templates as described (17,24,25). The 5′ splice site in transcript C, CAG/GUAAGU, was replaced in transcript M with CGGAU. Splicing reactions were done as described (26) in either commercial HeLa extracts (Cilbiotech) or extracts containing fusion proteins, made as described below. Native gel electrophoresis was as described (7) for 3.5–7 h. To stall complexes at complex A, the extract was incubated with 0.3 mM (one intron) or 0.5 mM (two introns) anacardic acid (Calbiochem). The anti-U6 2′-*O*-methyl oligonucleotide was incubated with extract at 1 μM for 15 min at 30°C prior to addition of RNA. Ribonuclease H treatment was as described (26).

Cloning and expression of mEGFP–U1A

The full reading frame of mEGFP–U1A was fused to the C-terminus of mEGFP, which carried a mutation (A206K) that prevents dimerization (27). A linker encoding $(gly)_3$ser was inserted between the C-terminus of mEGFP and N-terminus of U1A. Plasmid was transfected into HeLa cells and active nuclear extracts prepared as described (28). Quantification of the relative abundance of U1A and mEGFP–U1A (Figure 1B) was done using western blotting with a mouse monoclonal antibody (Ab55751 AbCam) and fluorescent anti-mouse secondary antibodies (Ir Dye 800 CW; LiCor) with excitation at 778 nm and detection at 795 nm using an Odyssey system (LiCor). The demonstration of activity and quantification of relative abundance shown in Figure 1A and B were done with the preparation of extract used for all the single molecule analyses. To conserve this preparation, Figures 1D and E were done with a different sample prepared likewise.

Immunoprecipitation and biotin affinity purification

Immunoprecipitation was done with protein G beads (Pierce) and a Living Colours Full-length A.v. polyclonal antibody to GFP (Clontech). Co-precipitated pre-mRNA was detected after gel electrophoresis with a phosphor imager (Figure 1C). Affinity purification was done with Neutravidin-agarose beads (Pierce) from splicing reactions containing biotinylated RNA at 50 nM. The U1A protein was detected by western blotting, using a mouse monoclonal antibody (29) with protein A/G-peroxidase (Thermo Fisher) and chemiluminescence (Figure 1D and E).

Fluorescent oligonucleotides

Oligonucleotides were synthesized and purified by Eurogentec, dissolved and stored following the manufacturer's instructions. Anti-U1-Cy5 (Cy5-UGCCAGGUAA GUAU-biotin, all 2′-*O*-methyl) was used for tethering U1snRNP. Bg-5′-Cy5 (Cy5-UAGACAACCAGCAGCC C-biotin, 2-*O*-methyl/LNA) was used for tethering globin-based RNAs. Bg-3′-Alexa488 (Alexa Fluor 488-A CCAAAAUGAUGAGAC) was used to create electrophoretic mobility shifts of spliced products containing the 3′-end of the transcript. Oligo-9 (Cy5-ACCUGCAG GCAUGCA-biotin, 2′-*O*-methyl) was used to tether Ad1CC RNA; Oligo-9 and Ad1-intron (Cy3-TGCAGC AAGCTTGACAAC, all 2′-*O*-methyl) were used for dual labelling.

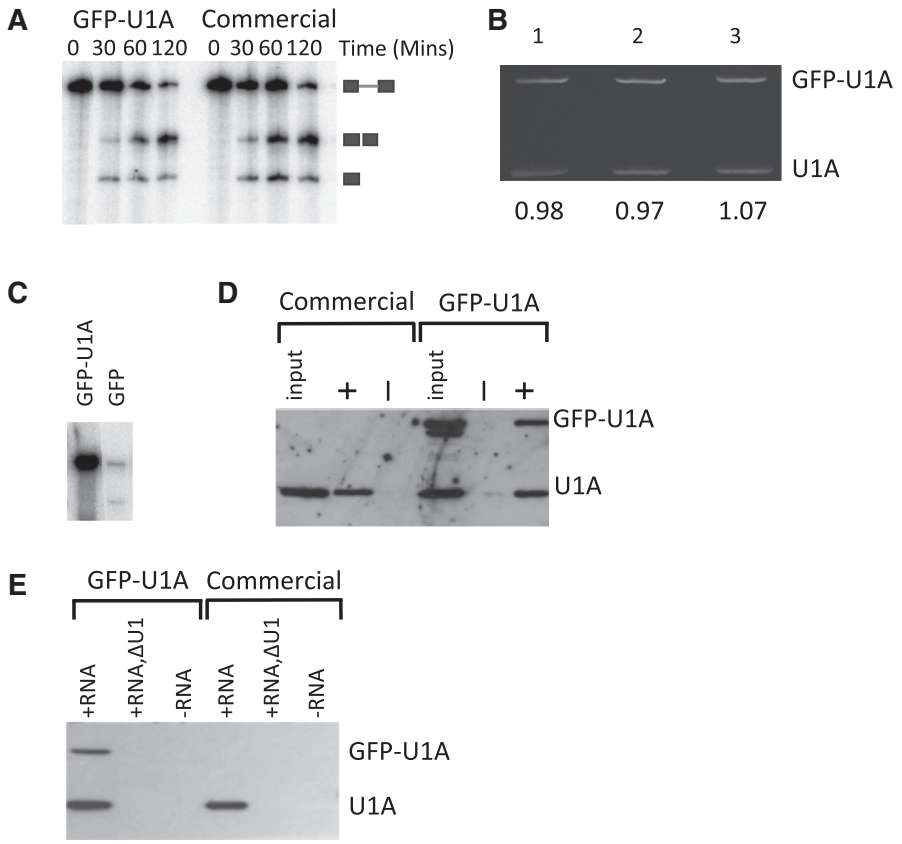

Figure 1. Fluorescent U1A protein and substrate pre-mRNA are functional. (**A**) Analysis by gel electrophoresis of time courses of splicing of pre-mRNA C in an extract prepared from HeLa cells expressing mEGFP–U1A and a commercial extract. (**B**) Analysis by western blotting of the relative levels of mEGFP–U1A and endogenous U1A proteins in the extract used for single molecule experiments. The number under each lane represents the measured ratio. The proteins were detected by antibodies to U1A and fluorescent secondary antibodies (emission at 795 nm). (**C**) Association of mEGFP–U1A with pre-mRNA. ^{32}P-labelled pre-mRNA was incubated in extracts expressing mEGFP–U1A or mEGFP and immunoprecipitated with anti-GFP antibodies. The RNA was analysed by gel electrophoresis and radioactivity is shown. The lanes were juxtaposed for illustration by removing intervening lanes of no relevance. (**D**) Analysis by western blotting of mEGFP–U1A incorporation into U1 snRNPs purified using a biotinylated oligonucleotide complementary to U1 snRNA. The extract made from cells expressing transfected mEGFP–U1A is compared with commercial HeLa nuclear extract; samples represent aliquots of the extract (input), the recovered proteins (+) and control samples in which no biotinylated oligonucleotide had been added to the extract (−). (**E**) Dependence of association of mEGFP–U1A with pre-mRNA on U1 snRNP. Biotinylated pre-mRNA (+RNA) was incubated in extract and recovered on avidin beads; U1A protein was detected by western blotting. The extract was treated also with oligonucleotides and RNase H to cleave the 5′-end of U1 snRNA (+RNA, ΔU1). Controls lacked biotinylated RNA (−RNA).

Cold transcription and oligonucleotide annealing

Uncapped cold transcripts were purified using S-300 columns (GE Healthcare) and phenol–chloroform extraction. RNA at 0.8–1.5 μM was incubated with 1 μM oligonucleotide in 30 μl reaction volumes in 10 mM HEPES, pH 8 and 100 mM NaCl. The mixture was heated to 80°C for 5 min and then cooled to 45°C on a hot block. The mixture was incubated on ice before being run on a 6% denaturing polyacrylamide gel and imaged using a Typhoon imager (GE Healthcare). A sample from each set with minimal free oligonucleotide was used for single molecule analysis.

Microscopy, acquisition of data and analysis of bleaching

Samples were prepared as described (30) except that the last wash contained an oxygen scavenger comprising 50 nM protocatechuate 3,4-dioxygenase (Sigma) and 2.5 mM protocatechuic acid (Spectrum Chemicals) (31). Imaging was done using prism-based TIRF microscopy as described (30). A cumulative image was calculated to identify co-localized Cy5 and mEGFP signals. Time series intensities of mEGFP fluorescence from these spots were analysed, and the numbers of apparent bleaching steps recorded. With an oligonucleotide containing only a single binding site for U1 snRNP, 88% of the spots showed single step bleaching and 12% showed bleaching in two steps (designated dimers). This result, together with the known fraction of U1A labelled with mEGFP (0.505), was used to calculate a probability of dimer formation (~0.27) assuming that unlabelled U1A showed the same behaviour. Calculation of the fractions of spots expected to bleach in one to four steps when two U1 snRNPS were bound was done by calculating the binomial distribution of such molecules containing two, three or four molecules of U1A, and for each class calculating the binomial

distribution of mEGFP-labelled U1A, where the probability was 0.505. To allow for possible heterogeneity of the complexes, the fractions of molecules expected to be bound by one or two U1 snRNPs were varied at intervals of 10%, and chi-squared tests of goodness of fit to the observed distributions were calculated. We assume that the probabilities for association of U1 snRNPs containing normal U1A or mEGFP–U1A depend only on their relative abundance. Inspection of the crystal structure of U1 snRNP containing a portion of U1A shows that the protein is far removed from the site of interaction with the 5′ splice site or other constraints (32), and we consider this assumption reasonable. We disregarded the effects of bleaching prior to recording [<10% of molecules (30)]. About one-quarter of all co-localized spots in each experiment did not show clear mEGFP bleaching steps and were not counted. We assumed that these spots would not modify the distributions of bleaching in discrete steps because the proportion did not depend on any variable not associated with the binding properties of U1 snRNPs. Moreover, such spots were found also in the absence of co-localized pre-mRNA, where U1 snRNPs had adsorbed non-specifically onto the surface.

Computational model of the diffusion of a random coil

RNA was modelled as a freely jointed chain with a Kuhn segment length of 6 nt. Conformational structures were simulated by a computer program using a random number algorithm (33) to select any values of the azimuthal and inclination angles in a 3D coordinate system centred on each joint in the chain. The locations of the two U1 snRNPs were identified in a sequence of different structures. The number of instances in which either U1 snRNP made contact with the 3′ splice site was determined, and the sequence was terminated after 5000 contacts. These represented on average ~1% of the structures tested. The distance between the downstream U1 snRNP and the 3′ splice site was fixed at 576 nt. The distribution of distances between the U1 snRNPs and the 3′ splice site was found for all simulated structures by counting frequencies in binned increments corresponding to a length of 0.01 nt. The variations tested included setting the Kuhn segment length at 25 nt and constraining the joints to take any of five orthogonal directions, excluding direct reversal.

RESULTS

To enable detection of single U1 snRNPs, nuclear extracts were made from HeLa cells expressing a protein in which mEGFP was fused to U1A, an integral U1 snRNP protein. U1A was chosen for several reasons. It binds stem-loop 1 of the snRNA, and a crystal structure of the snRNP containing the first of the two RNA-binding domains of U1A shows that this stem-loop is not folded into the main body of the snRNP and the U1A domain makes no contacts with the other proteins (32). Neither U1A nor its binding site on the U1 snRNA are required for splicing or splice site selection (34–36). GFP fusions of U1A have been used and validated as labels for the U1 snRNP (37,38). The extracts were shown to be functional in splicing assays (Figure 1A). The ratio of mEGFP–U1A to endogenous U1A in the extract used for single molecule experiments was measured by western blotting as 1.02:1 (Figure 1B). The fusion protein was shown to associate with pre-mRNA by immunoprecipitation (Figure 1C). Moreover, its recovery on avidin beads depended on the presence of a biotinylated substrate for the U1 snRNP, either in the form of a 2′-*O*-methyl oligonucleotide complementary to the 5′-end of U1 snRNA (Figure 1D) or biotinylated pre-mRNA (Figure 1E). In the latter case, mEGFP–U1A did not associate with the pre-mRNA if the 5′-end of U1 snRNA had been cleaved by RNase H. We conclude that the association of mEGFP–U1A protein with substrate RNA was dependent on inclusion into U1 snRNPs. U1 snRNPs are neither involved in the catalytic steps of splicing nor essential for them; their function is in 5′ splice site binding and recognition (39–42).

To establish the signal associated with binding of one snRNP, the extract containing mEGFP–U1A was incubated with a fluorescent biotinylated oligonucleotide complementary to the 5′-end of U1 snRNA, injected onto the surface of a streptavidin-coated silica slide, and time courses of mEGFP fluorescence recorded for spots with co-localized oligonucleotide and protein fluorescence. Most co-localized spots showed clear steps in mEGFP bleaching (Figure 2), 298 (88%) in one step and 39 (12%) in two; no spots showed bleaching in three steps. Bleaching in two steps may arise in part from the ability of U1A to form homodimers (43,44), since similar fractions were seen with non-co-localized GFP spots (72 spots and 9 spots, respectively). Around one-quarter of the spots observed could not be assigned with certainty to any class. Non-co-localized spots of mEGFP in this and other experiments arise from non-specific attachment to the surface, since they are observed at similar levels in the absence of any biotinylated oligonucleotide or RNA (data not shown). We have used the 88:12 distribution hereafter for the values expected from association of a single U1 snRNP with pre-mRNA.

Binding of U1 snRNPs in complex E

To determine the number of U1 snRNPs bound in complex E with a single 5′ splice site, we used a globin-derived pre-mRNA with one CAG/GUAAGU 5′ splice site (C, Figure 3A). This splices with high efficiency (Figure 3B). An otherwise identical transcript lacking the splice site (M, Figure 3A) showed only weak splicing to a cryptic site, showing that 5′ splice site recognition is limiting. The pre-mRNA was labelled by hybridization to an oligonucleotide that contained the fluorophore Cy5 at the 5′-end and a biotin group at the 3′-end (Figure 3C). The oligonucleotide did not interfere with splicing and remained bound quantitatively to the pre-mRNA not only during the splicing reactions but also during electrophoresis in 7 M urea (Figure 3D). The hybridized pre-mRNA formed complex E when incubated in nuclear extract in the absence of ATP, dependent on both a 5′ splice site and the 5′-end of U1 snRNA (data not shown); importantly, when incubated in the presence of

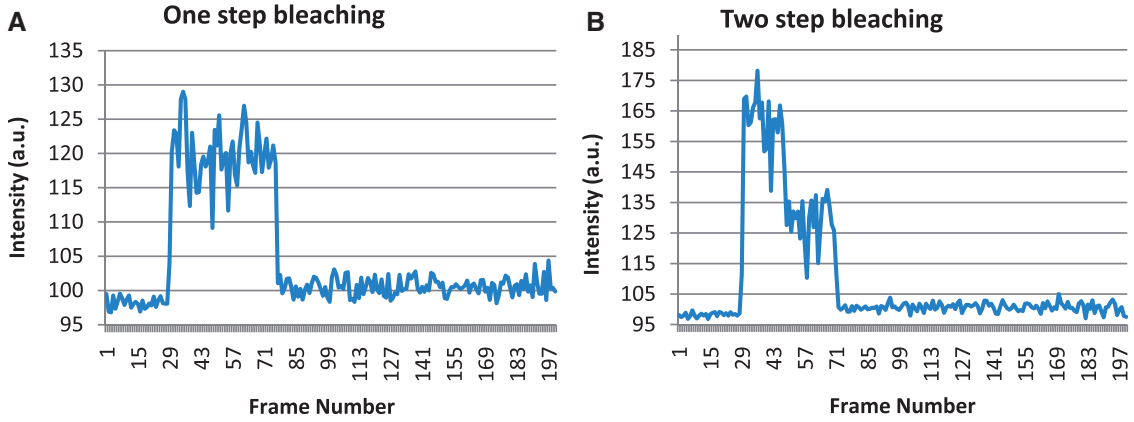

Figure 2. Representative time courses of fluorescence from mEGFP–U1A co-localized with single molecules of anti-U1-Cy5 biotinylated oligonucleotide. Emission was collected in frames of 200 ms. A factory offset in the emCCD produces intensity readings in arbitrary units around 100. The 488 nm laser exciting mEGFP was switched on at frame 27. Bleaching took place in one (**A**) or two (**B**) steps.

ATP the hybridized RNA spliced as normal (Figure 3D). After incubation of labelled transcript under these conditions, dilution and injection onto a streptavidin-coated silica slide, 87% of the co-localized Cy5 and mEGFP fluorophores that showed photobleaching of the mEGFP in steps (Figure 3E) did so in a single step, consistent with binding of a single U1 snRNP (Tables 1 and 2). To ensure that no U1 snRNPs bound in the steps following incubation in the extract, the transcript and extract were incubated separately, diluted and then combined. Importantly, no co-localization was detected. The mutant globin transcripts that lack the normal 5′ splice-site showed 8-fold reduced co-localization, consistent with the reduced efficiency of splicing from a cryptic site (a representative field of view is shown in Figure 3F). The last two experiments demonstrate that co-localization arises from binding rather than coincidental proximity and that the complexes detected had formed during the incubation in splicing conditions. We conclude that there is one stably-bound U1 snRNP in complex E on transcripts containing one 5′ splice site.

With a similar transcript containing two identical 5′ splice sites (C174C) or an adenovirus-based transcript with two such 5′ splice sites (Ad1CC), the fractions of co-localized spots showing photobleaching of mEGFP in two steps increased 3-fold (Table 1). The fractions exclude the possibility that these are homogeneous populations with only one U1 snRNP on each transcript (Table 2). Instead, the distributions observed are close to those expected if there were homogeneous populations with two U1 snRNPs bound. To take account of possible heterogeneity in the population, the distributions of spots bleaching in one or two steps were calculated assuming that 10%, 20%, 30%, etc. of the transcripts were bound by only one U1 snRNP. The numbers bleaching in three steps were too low to be included. The experimental data for both C174C and Ad1CC corresponded most closely to a distribution in which 70% of the transcripts were bound by two U1 snRNPs. We infer that the protection against oligonucleotide-directed RNase H cleavage of both alternative 5′ splice sites in most molecules of C174C observed previously (17) can be attributed directly to U1 snRNP occupancy. The relatively small fraction of molecules protected at only one site in those experiments is consistent with the existence of some heterogeneity.

To confirm that both of the 5′ splice sites were in principle functional, and that the sequence separating the two sites in C174C did not bind U1 snRNPs, individual sites were deleted. The remaining site in both M174C and C174M was functional (Figure 4) and the predicted distributions suggest that only one U1 snRNP was recruited in complex E (Table 1).

We conclude that it is possible for two U1 snRNPs to be bound to alternative 5′ splice sites in a single particle of complex E, notwithstanding the evidence that a 5′ and a 3′ splice site are held in close proximity within the complex.

Binding of U1 snRNPs in complex A

Kinetic trap experiments have shown that particular 5′ splice sites are committed to splicing in complex A (22). However, the presence of a second U1 snRNP in complex A as in complex E would be expected to leave open the possibility of shifting to another site. To test this, the assembly of splicing complexes on the various pre-mRNA substrates in nuclear extracts in the presence of ATP was stalled at complex A either with an oligonucleotide complementary to U6 snRNA (21) or with anacardic acid (45) (Figure 5). Strikingly, analysis of the fluorescence time courses of single molecule complexes by TIRF microscopy showed primarily single-step photobleaching of mEGFP, i.e. that only one U1 snRNP was bound to the transcripts irrespective of the number of 5′ splice sites (Tables 3 and 4). In the absence of a 3′ splice site (transcript C174CΔ3′), the results show that many molecules with two 5′ splice sites retain both U1 snRNPs.

Removal of surplus U1 snRNP during progression from complex E to complex A

If two U1 snRNPs are found in complex E but only one in complex A, it is likely that the redundant snRNP is

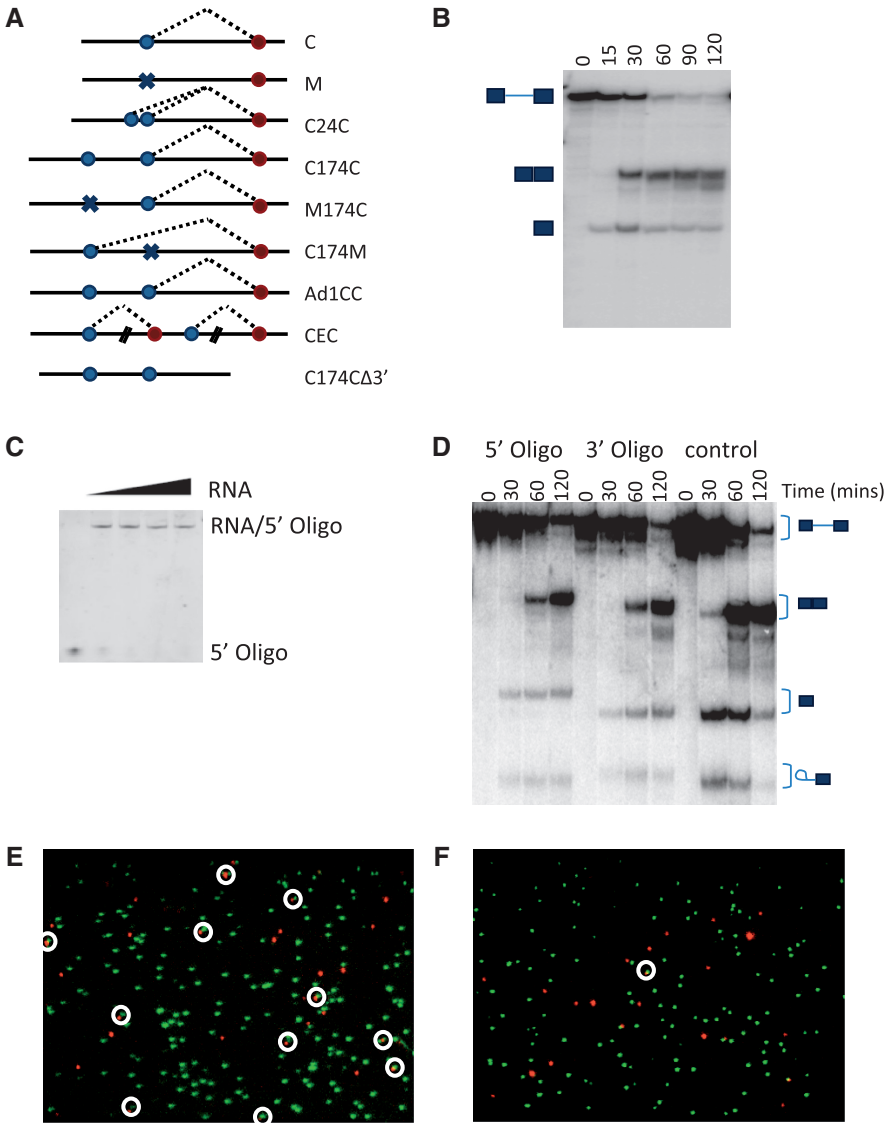

Figure 3. Detection of U1 snRNP binding to transcript C. (**A**) Diagrams of the pre-mRNA substrates and their patterns of splicing (dotted lines). Blue circles, 5′ splice sites; red circles, 3′ splice sites. Crosses indicate sites inactivated by mutation. In CEC, the introns are identical to those in the single intron constructs but are not shown in their entirety. (**B**) Time courses of splicing *in vitro* of transcripts C analysed by gel electrophoresis. (**C**) Quantitative hybridization of fluorescent biotinylated oligonucleotide to the 5′-end of the pre-mRNA C. An amount of 1 μM Cy5-labelled oligonucleotide was incubated with unlabelled pre-mRNA at 0.85–1.5 μM (l to r) before gel electrophoresis in 7 M urea. Cy5 fluorescence was detected. (**D**) Time courses of splicing after hybridization of tethering oligonucleotides to pre-mRNA C. 5′-oligo, pre-mRNA was incubated in splicing reactions at 50 nM after hybridization to oligonucleotide Bg-5′-Cy5, complementary to the 5′ 16 nt of the transcript, and the splicing of the fraction that was radiolabelled detected after electrophoresis; 3′ oligo, pre-mRNA was incubated in splicing reactions at 50 nM after hybridization to oligonucleotide Bg-3′-Alexa488, complementary to the 3′ 15 nt of the transcript; control, radiolabelled RNA spliced at 1 nM in the absence of oligonucleotide. RNA molecules with complementarity to the oligonucleotides show reduced mobility. (**E**) A representative TIRF microscope field showing co-localization of transcript C (red) and mEGFP–U1A (green) in the absence of ATP. Transcript C was annealed to the fluorescent biotinylated oligonucleotide Bg-5′-Cy5, and incubated with nuclear extract containing mEGFP–U1A. The mixture was diluted and injected into a chamber on a prepared silica slide. Fluorescence from the molecules attached to the surface was detected by TIRF. The images obtained in the Cy5 and GFP channels were superimposed and then offset slightly to show both colours. White circles show co-localized Cy5 and GFP signals. (**F**) A TIRF field showing Cy5 and mEGFP fluorescence from incubation of Cy5-labelled transcript M, as for transcript C in Figure 3E. The sole example of co-localization is circled.

destabilized or removed as complex A assembles. However, it is possible that assembly of complex A, in the presence of ATP, bypasses complex E. To show more directly that the processes forming complex A are able to destabilize the surplus snRNP, Ad1CC RNA was incubated in an extract containing the oligonucleotide complementary to U6 snRNA but in the absence of ATP, allowing complex E to form; after removal of a sample for TIRF analysis, ATP was added and complex A allowed to form. The results show that most transcripts were associated with two molecules of U1 snRNP before ATP was added but that afterwards the majority were

Table 1. Observed numbers of transcripts showing bleaching of co-localized mGFP-U1A in discrete steps after assembly was stalled at complex E by incubation in nuclear extract in the absence of ATP

Transcript	Bleaching in one step	Bleaching in two steps	Bleaching in three steps
C	93	12	2
C174C	75	35	4
M174C	95	16	2
C174M	96	14	3
Ad1CC	69	35	2
C24C	71	38	4

Table 2. Goodness of fit of the numbers of transcripts bleaching in one or two steps (Table 1) by the numbers predicted with different fractions of transcripts bound by one or two U1 snRNPs

1 U1 snRNP	2 U1 snRNPs	C	C174C	M174C	C174M	Ad1CC	C24C
1	0	9.86	10^{-10}	0.42	0.81	10^{-11}	10^{-13}
0.9	0.1	0.30	10^{-8}	0.80	0.50	10^{-7}	10^{-8}
0.8	0.2	0.08	10^{-4}	0.31	0.15	10^{-5}	10^{-5}
0.7	0.3	0.02	0.01	0.08	0.03	10^{-3}	10^{-4}
0.6	0.4	10^{-3}	0.06	0.02	0.01	0.02	0.01
0.5	0.5	10^{-4}	0.26	10^{-3}	10^{-3}	0.13	0.07
0.4	0.6	10^{-5}	0.68	10^{-4}	10^{-4}	0.42	0.27
0.3	0.7	10^{-6}	0.79	10^{-5}	10^{-5}	0.89	0.68
0.2	0.8	10^{-7}	0.36	10^{-6}	10^{-7}	0.61	0.80
0.1	0.9	10^{-8}	0.12	10^{-7}	10^{-8}	0.26	0.37
0	1	10^{-10}	0.03	10^{-9}	10^{-9}	0.08	0.13

The numbers of transcripts expected to bleach in one or two steps were calculated from the relative abundance of mEGFP-tagged and endogenous U1A proteins and the frequency of observing bleaching in one or two steps with U1 snRNPs captured on a short 5′ splice site oligonucleotide. The left-hand columns show the fractions of transcripts bound by one or two U1 snRNPs used for calculating the expected distributions and the other columns show for each data set in Table 1 the probability that a value of chi-square at least as high as that calculated would arise by chance if the null hypothesis were correct. Values < 0.01 are rounded to the nearest integer factor of ten. The result showing the highest probability is boxed in each case.

bound to only one snRNP (Tables 5 and 6). Thus, we conclude that the superfluous U1 snRNP on the upstream 5′ splice site is displaced during the transition from complex E to complex A and that this process requires an intact 3′ splice site.

Binding of U1 snRNPs to pre-mRNA containing multiple introns

The removal of surplus U1 snRNPs during formation of complex A might jeopardise U1 snRNPs present at the 5′ splice site of a separate upstream intron. This possibility was tested using transcripts containing a duplication of the intron and flanking exon sequences of transcript C (CEC; Figure 6A). All three exons were spliced in nuclear extracts (Figure 6B). The splicing complexes that assembled were significantly larger than those seen on transcript C after incubations of 2.5 and, most strikingly, 15 and 30 min (Figure 6C). If these large complexes contained two distinct spliceosomes, then cleavage with ribonuclease H in the central exon of CEC would produce complexes equivalent in size to the individual spliceosomes formed on transcript C. This was observed (Figure 6D). Single molecule analyses showed that the addition of ATP had only a small effect on the number of U1 snRNPs present (Tables 7 and 8). We conclude that the destabilization is restricted to surplus U1 snRNPs.

Concentrations of active U1 snRNPs affect splice site selection

Both C174C and Ad1CC pre-mRNAs are spliced predominantly to the intron-proximal 5′ splice site. To test whether this is connected to the observed binding of U1 snRNPs to both alternative 5′ splice sites in complex E, the concentration of active U1 snRNPs was reduced by adding a 2′-*O*-methyl oligonucleotide complementary to the 5′-end of U1 snRNA. This reduced the positional bias in splicing (Figure 7). This suggests that the bias towards the intron-proximal site usually is a result of high occupancy by U1 snRNPs in complex E rather than an intrinsic property of the assembly of complex A. The result is consistent with models that complex E is an intermediate in the assembly of complex A.

DISCUSSION

Previous evidence has demonstrated that the 5′ and 3′ splice sites are held in close proximity to each other and to the snRNAs in complex E (20,21) and that this complex is committed to splicing (9,12–14). Recent kinetic trap experiments have shown that the sites to be used for splicing are not committed until complex A (22). The purpose of our experiments was to test one model that might be able to reconcile these results, i.e. that a single U1 snRNP was held in complex E but that it was able to form only weak interactions with the 5′ splice sites and therefore could exchange between sites (Figure 8A). Our results show that this is not the case. Instead, complex E appears to be compatible with concurrent binding by at least two U1 snRNPs, whereas only one U1 snRNP is bound in complex A. ATP hydrolysis has been shown by Kotlajich *et al.* to be required for the commitment of splice sites, separately from the ATP hydrolysis involved in complex A assembly (22). Our findings suggest that this requirement might arise because an ATP-dependent process displaces any factors binding alternative splice sites. Moreover, further binding by U1 snRNPs to orphan 5′ splice sites is eliminated.

The method we have used involves the assembly of complexes, followed by rapid dilution, capture of complexes onto the surface of a silica slide and detection by total internal reflection fluorescence microscopy. This has the advantage that the complexes have been assembled under standard conditions. There is no artefactual binding after dilution, and the process is so rapid that we have not seen any evidence for losses by dissociation. The preservation of the complexes is unsurprising, since complexes E and A withstand purification processes

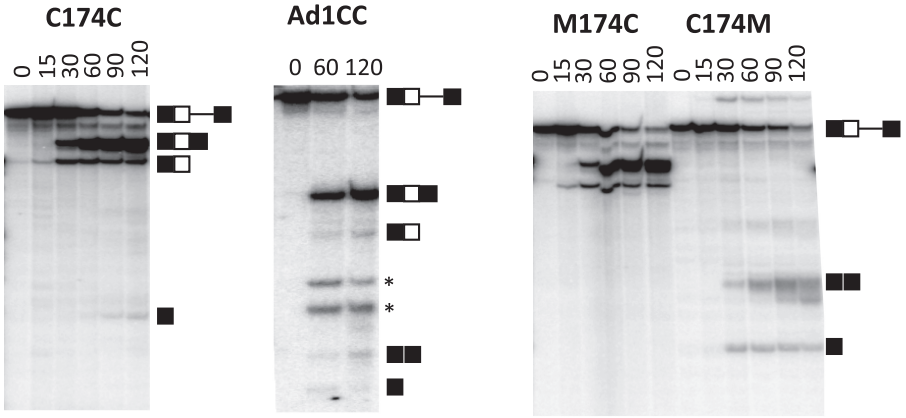

Figure 4. Time courses of splicing of pre-mRNA sequences containing two 5′ splice sites (C174C and Ad1CC). These splice almost wholly to the intron-proximal site. To confirm functionality of the individual sites, additional substrates were tested in which either of the two sites in C174C had been deleted (M174C and C174M). Asterisks mark intron lariats from use of the downstream site in Ad1CC.

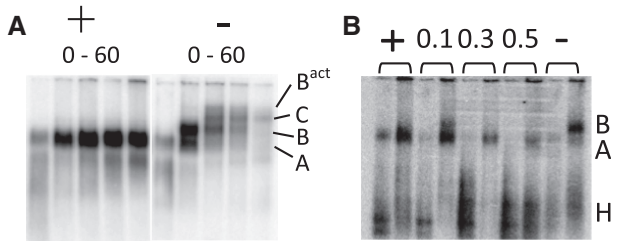

Figure 5. Stalling of spliceosome assembly at complex A. (**A**) Inhibition of assembly on transcript C with a 2′-*O*-methyl oligonucleotide complementary to U6 snRNA (21). After incubation at 30°C for 0, 5, 15, 30 and 60 min in a commercial nuclear extract, reactions were treated with heparin and subjected to electrophoresis on agarose gels. The identities of the complexes are marked on the right. +, presence of oligonucleotide at 1 μM; −, no oligonucleotide. (**B**) Inhibition of assembly on transcript C after inclusion of anacardic acid at various concentrations (0.1, 0.3 and 0.5 mM). Pre-mRNA was incubated in the reaction mixtures for 0 (left lane) and 30 min (right lane). −, negative control lacking anacardic acid; +, positive control with oligonucleotide complementary to U6 snRNA. 0.3 mM anacardic acid prevents assembly of complex B without increasing levels of complex H.

Table 3. Observed numbers of transcripts associated with bleaching of mGFP-U1A in discrete steps after incubation in the presence of ATP

Transcript	Bleaching in one step	Bleaching in two steps	Bleaching in three steps
C + α-U6	66	7	3
C + AA	69	7	1
C174C + α-U6	79	10	0
C174C + AA	86	12	5
C174CΔ3′ + α-U6	70	26	2
Ad1 CC + α-U6	50	6	0
Ad1 CC + AA	95	14	4

Assembly was stalled at complex A by the inclusion of either anacardic acid (+ AA) or an oligonucleotide complementary to U6 snRNA (+ α-U6).

using centrifugation velocity and column chromatography that can last for as long as 2 days. It is not possible at present to work with fluorescently labelled protein factors at functional concentrations to follow mammalian splicing in real time on the surface of the slide. U1 snRNPs, SR proteins, hnRNP proteins and other factors involved in splice site selection in mammalian pre-mRNA have to be present at around micromolar concentrations to be functional. If unbound fluorophores were present at these concentrations, proper visualization of fluorophores bound to the pre-mRNA on the surface would be impossible. In contrast, real-time observations have been made on yeast pre-mRNA splicing, under conditions where protein components at concentrations of ~20 nM supported splicing (6).

The use of nuclear extracts derived from transfected cells entails the presence of some endogenous U1A protein. In the current case, the concentrations of the fusion and endogenous proteins were approximately equal, and there was no effect on splice site selection. It is likely that feedback inhibition of U1A on polyadenylation of its own pre-mRNA ensures that the overall levels of U1A are not affected significantly (46). Moreover, the mEGFP–U1A protein only binds pre-mRNA via its attachment to U1 snRNA (Figure 1E). The predicted distributions of the bleaching steps assumed that the fractions of U1A and mEGFP–U1A followed a binomial distribution in which the probability that a snRNP contained mEGFP–U1A depended on the relative abundance of the two proteins. This seems reasonable, given the properties of U1A (see 'Results' section). A second assumption made in calculating the distributions of the bleaching steps expected if two U1 snRNPs were bound was that the probability of association of two molecules of U1A on one U1 snRNP, calculated from the control experiment when U1 snRNPs were recruited to a short oligonucleotide, was unaffected by the presence of the mEGFP moiety. If the endogenous U1A had a lower or higher propensity to associate, then the calculated fractions of molecules with two 5′ splice sites bound by two

Table 4. Goodness of fit of the numbers of transcripts bleaching in one or two steps (Table 3) by the numbers predicted with different fractions of transcripts bound by one or two U1 snRNPs

1 U1 snRNP	2 U1 snRNPs	C α-U6	C AA	C174C α-U6	C174C AA	C174CΔ3′ α-U6	Ad1CC α-U6	Ad1CC AA
1	0	0.53	0.45	0.82	0.94	10^{-5}	0.77	0.79
0.9	0.1	0.20	0.16	0.32	0.44	10^{-3}	0.37	0.53
0.8	0.2	0.06	0.05	0.10	0.14	0.02	0.16	0.16
0.7	0.3	0.02	0.01	0.02	0.03	0.14	0.06	0.04
0.6	0.4	10^{-3}	10^{-3}	0.01	0.01	0.48	0.02	0.01
0.5	0.5	10^{-3}	10^{-4}	10^{-3}	10^{-3}	0.99	0.01	10^{-3}
0.4	0.6	10^{-4}	10^{-4}	10^{-4}	10^{-4}	0.53	10^{-3}	10^{-4}
0.3	0.7	10^{-5}	10^{-5}	10^{-5}	10^{-5}	0.22	10^{-4}	10^{-5}
0.2	0.8	10^{-6}	10^{-6}	10^{-6}	10^{-6}	0.07	10^{-4}	10^{-7}
0.1	0.9	10^{-7}	10^{-7}	10^{-7}	10^{-7}	0.02	10^{-5}	10^{-8}
0	1	10^{-8}	10^{-8}	10^{-9}	10^{-9}	10^{-3}	10^{-6}	10^{-10}

See notes to Table 2.

Table 5. Observed numbers of transcripts associated with bleaching of mGFP-U1A in steps before and after addition of ATP

Ad1CC + αU6	Bleaching in one step	Bleaching in two steps	Bleaching in three steps
Before ATP	60	30	6
After ATP	72	15	6

Assembly of complex E was followed by addition of ATP and incubation to allow complexes to progress to complex A. Complex assembly beyond complex A was prevented by the inclusion of an oligonucleotide complementary to U6 snRNA (+ α-U6).

Table 6. Goodness of fit of the numbers of transcripts bleaching in one or two steps (Table 5) by the numbers predicted with different fractions of transcripts bound by one or two U1 snRNPs

1 U1 snRNP	2 U1 snRNPs	−ATP	+ATP
1	0	10^{-10}	0.13
0.9	0.1	10^{-6}	0.56
0.8	0.2	10^{-4}	0.85
0.7	0.3	10^{-3}	0.39
0.6	0.4	0.04	0.14
0.5	0.5	0.18	0.04
0.4	0.6	0.49	0.01
0.3	0.7	0.95	10^{-3}
0.2	0.8	0.59	10^{-4}
0.1	0.9	0.27	10^{-5}
0	1	0.09	10^{-6}

See notes to Table 2.

U1 snRNPs would be either an underestimate or an overestimate, respectively.

Our results showed that two U1 snRNPs were present on the majority of the transcripts with two 5′ splice sites in complex E. We have previously shown almost complete protection of both 5′ splice sites in C174C against ribonuclease cleavage in conditions allowing assembly of complex E (17). The protection was interpreted to be a consequence of U1 snRNP binding because it was eliminated by cleavage of the 5′-end of U1 snRNA and could be produced by pure U1 snRNPs. Our present results demonstrate directly that, in complex E, transcripts can be bound by two U1 snRNPs, and we conclude that for C174C, C24C and Ad1CC this is the case for most transcripts.

The mechanism by which the surpus U1 snRNP is removed from the alternative 5′ splice site during the subsequent assembly of complex A is unknown, but this process is likely to underlie the commitment to specific sites noted after complex A assembly (22). One possibility is that complex A itself displaces surplus U1 snRNPs as it forms, which is consistent with the requirement for ATP and a 3′ splice site. There are two arguments against this: Kotlajich *et al.* observed that commitment, unlike complex A formation, required ATP hydrolysis (22), and complexes do not make extensive contacts beyond the selected 5′ splice site (47) and are therefore unlikely to physically exclude a snRNP bound 174-nt upstream. Instead, the dependence on ATP and a 3′ splice site suggest that displacement is mediated by an RNA-dependent ATPase or helicase that acts in association with complex A formation. Two RNA helicases are known to associate with early complexes, both of which belong to the DEAD-box class of superfamily 2 helicases. UAP56 is recruited by interactions with the polypyrimidine-bound U2AF65 protein, and it is required for U2 snRNP to bind the branchpoint (48). However, its helicase activity is not required for assembly of complex A (49), suggesting that it would not be involved in displacing U1 snRNPs. The other, DDX46, interacts with U1 and U2 snRNPs, and in *Schizosaccharomyces pombe* complex A formation requires ATP hydrolysis by its orthologue (Prp5) (50). Interestingly, Prp5 interacts indirectly with U1-A protein (19), which is not required for splicing (see above). This suggests that one activity of Prp5 might be the displacement of U1 snRNPs.

None of the mechanisms discussed above for displacement of the surplus U1 snRNPs addresses the mechanism by which the U1 snRNP to be displaced is identified. Splicing of C174C and Ad1CC is almost exclusively

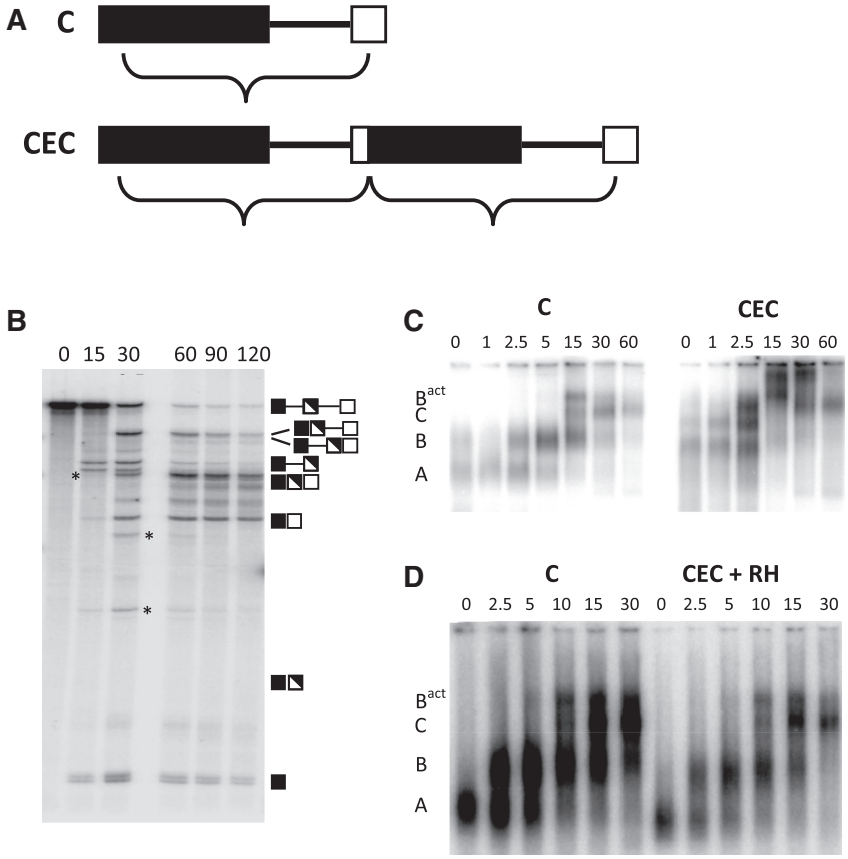

Figure 6. Splicing and assembly of complexes on a transcript containing two introns. (**A**) Diagrams showing the pre-mRNAs C and CEC. The 5′-exon of C is black and the 3′-exon is white. The bracket under C shows the region duplicated to form CEC, as shown by the brackets under CEC. (**B**) Time course of splicing of the double intron transcript (CEC). No intermediates or products diagnostic of exon skipping were detected. Reactions were done in commercial nuclear extract. Times are shown above each lane in minutes. Asterisks mark lariats containing introns. (**C**) Comparison of spliceosome assembly between transcripts C and CEC. After incubation at 30°C for the times shown (mins) in a commercial nuclear extract, reactions were treated with heparin and subjected to electrophoresis on agarose gels. Spliceosomal complexes formed on C are indicated on the left. The region of the gel showing complex H has been cropped. (**D**) Comparison of assembly, as in panel C, but with the addition of a DNA oligonucleotide and ribonuclease H after the times shown to cleave transcript CEC in the central exon before electrophoresis.

Table 7. Observed numbers of transcripts associated with bleaching of mGFP-U1A in steps for a transcript containing two introns

Transcript	Bleaching in one step	Bleaching in two steps	Bleaching in three steps
CEC −ATP	65	36	5
CEC + α-U6	69	31	3
CEC + AA	67	29	3

A transcript with two introns (CEC) was incubated under conditions permitting assembly of complex E (−ATP) or complex A (inclusion of either anacardic acid (+ AA) or an oligonucleotide complementary to U6 snRNA (+ α-U6).

Table 8. Goodness of fit of the numbers of transcripts bleaching in one or two steps (Table 7) by the numbers predicted with different fractions of transcripts bound by one or two U1 snRNPs

1 U1 snRNP	2 U1 snRNPs	CEC ATP	CEC α-U6	CEC AA
1	0	10^{-13}	10^{-8}	10^{-8}
0.9	0.1	10^{-8}	10^{-5}	10^{-5}
0.8	0.2	10^{-6}	10^{-3}	10^{-3}
0.7	0.3	10^{-4}	0.01	0.03
0.6	0.4	0.01	0.10	0.16
0.5	0.5	0.05	0.37	0.48
0.4	0.6	0.22	0.83	0.97
0.3	0.7	0.58	0.67	0.56
0.2	0.8	0.94	0.29	0.24
0.1	0.9	0.49	0.10	0.08
0	1	0.19	0.03	0.02

See notes to Table 2.

from the intron-proximal splice site, and we infer that it is the other site that loses the U1 snRNP during progression to complex A. A similar marked preference for an intron-proximal site has been observed in a number of experiments when the alternative sites have high levels of complementarity to U1 snRNA (17,51–53). Several lines of evidence suggested previously that the intron-proximal site is used whenever it is bound. In addition to the outcome of double occupancy, described here, these include: a correlation between U1 snRNP binding to the intron-proximal site, but not the distal site, measured by

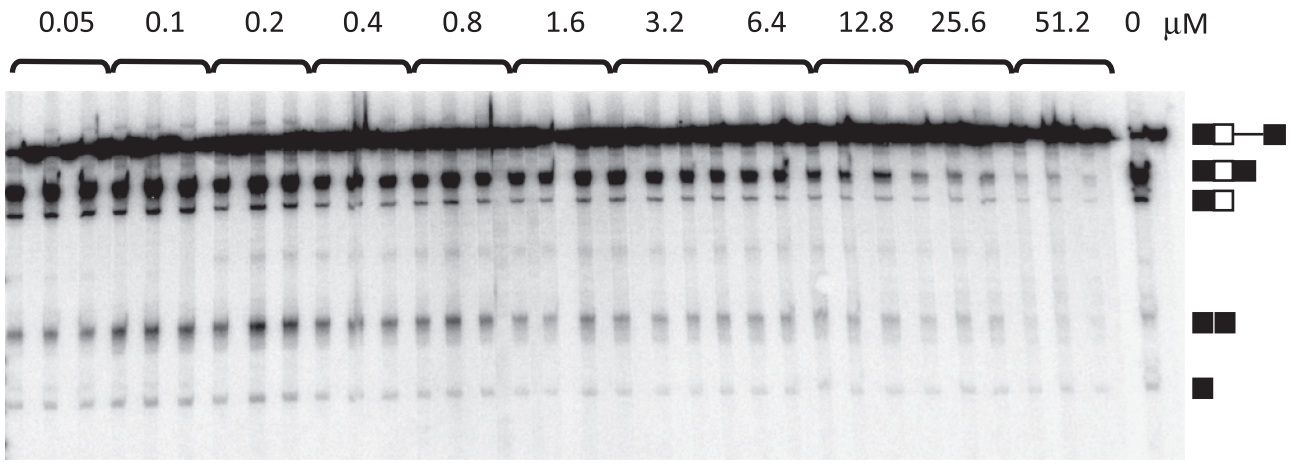

Figure 7. Induction of splicing at an intron-distal site of transcript G176G by addition of a 2′-*O*-methyl oligonucleotide complementary to U1 snRNA. G176G is identical to C174C with the exception that the splice sites contain a globin splice site sequence (AGG/GUGAGU) rather than the consensus sequence. Each concentration of oligonucleotide was tested in triplicate (bracketed), and in each reaction samples were analysed at 0 and 120 min. 0, no oligonucleotide.

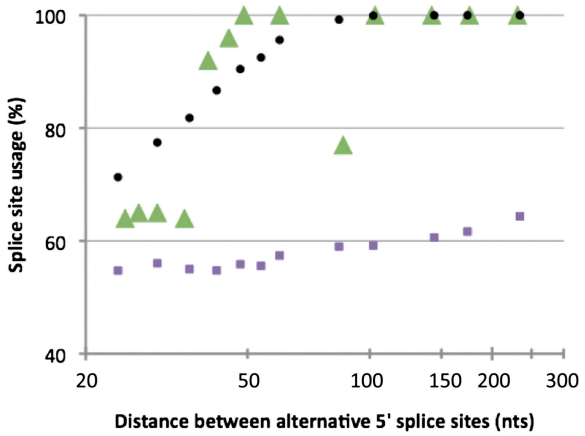

Figure 8. Comparison of the observed intron-proximal 5′ splice site usage with the expected probabilities of collision with the 3′ splice site. Green triangles, dependence of intron-proximal site use in HeLa cells on the distance between two candidate globin 5′ splice sites (56). Splice site usage at each site is shown, with distance (nt) on a logarithmic scale. Splicing *in vitro* shows the same effect with a transition to predominantly intron-proximal site use with separations >45 nt. Squares represent the predictions from a simulation based on the distribution at thermodynamic equilibrium of possible conformations of a homogeneous random coil, modelled as a freely-jointed chain. Circles represent the predictions likewise if the sequence between the splice sites is rigid.

psoralen cross-linking, and the fraction of splicing from the sites (24); activity of a distal site if the proximal is weak but disuse if both sites are strong (17); increases in U1 snRNP binding at both of two alternative 5′ splice sites when concentrations of SRSF1 are increased, with an increase in intron-proximal splicing (17,24). However, the mechanism selecting the intron-proximal site has remained obscure.

The ATP-dependent process that destabilizes surplus U1 snRNPs during complex A formation, described above, suggests one possible mechanism for selection. It is conceivable that a bound U1 snRNP initiates helicase action in a 5′-direction, displacing intron-distal snRNPs. However, both the candidate helicases belong to the DEAD-box family and are not likely to act processively (54). Moreover, such a mechanism does not explain the preference for one U1 snRNP rather than another. An obvious possibility is that the intron-proximal U1 snRNP is selected because it is the more likely of the two to contact the 3′ splice site if the RNA were to behave like a 3D random coil. While this is plausible and has often been suggested (53), it has never been tested. The simplest test is to determine how the use of the two sites depends on their separation, a test that was used previously to support a diffusional model for interactions between proteins bound to a enhancer sequence and a splice site. The dependence on distance based on diffusion can be predicted by both analytical calculations (55) and computer simulation based on a freely-jointed chain at thermodynamic equilibrium. These approaches predict a gradual increase in use of the intron-proximal site as the separation is increased (Figure 8). We have previously measured the use of two sites separated by a range of distances (56) (Figure 8). This shows that use of the intron-proximal site increases rapidly with separations beyond ~30 nt and is ~100% for separations of ≥45 nt. We conclude that the simple diffusional model does not describe the processes seen.

An alternative physical model is that the binding of U1 snRNPs alters the properties of the RNA. Exons appear more particulate than introns by electron microscopy (57). This is likely to be the result of binding by SR proteins. If the binding of a U1 snRNP stimulated assembly of SR and SR-related proteins on its 5′-side, it is possible that the exon would become more rigid. Both computer simulations and analytical calculations of diffusion in which we treated the exon as a rigid rod predict probabilities for contacts between the 5′ splice sites and the 3′ splice site that are similar to the experimental data for splice site usage (Figure 8), in that use of the intron-proximal site increases

with separation and reaches ~100%. The model does not account for the threshold separation at which distance starts to be effective, and 100% use of the intron-proximal site is seen at 45 nt rather than the 80 nt predicted by the model. The threshold may arise because of steric interference between the U1 snRNPs, preventing binding of SR proteins, and the earlier onset of exclusive use may arise because the form of the complex is not a rod or because of excluded volume effects. Nonetheless, we consider this model to be a better starting point for explanations of preferences for the intron-proximal site than any other yet suggested. Interestingly, it prompts the suggestion that exon-bound hnRNPs affect selection by introducing flexibility into the exon. The process of displacing the unused snRNP might be initiated by the formation of tight contacts between the other snRNP and the 3′ splice site.

It has generally been thought that constitutive splicing complexes would be identical in all splicing events, and that only the regulatory or non-participating proteins would vary in abundance on specific introns. Our findings show unambiguously that there is not a fixed stoichiometry: the number of U1 snRNPs in the first complex, complex E, depends on the candidate sites. This raises the possibility that complex E will contain various numbers of other molecules involved in the recognition of RNA sequences.

ACKNOWLEDGEMENTS

The authors thank Dr L.P. Eperon for cloning mEGFP–U1A, Prof. C.R. Bagshaw for assistance and advice in TIRF microscopy, Dr N. Owen for advice on immunoprecipitation and Ms L. Blackmore for help with imaging the western blots.

FUNDING

The European Commission [EURASNET-LSHG-CT-2005-518238; European Network of Excellence in Alternative Splicing; I.C.E.]; Biotechnology and Biological Sciences Research Council [BBS/B/0367X and a studentship to M.J.H.]; Wellcome Trust [088113/Z/08/Z to D.C.]. Funding for open access charge: University of Leicester; Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Pan,Q., Shai,O., Lee,L.J., Frey,B.J. and Blencowe,B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
2. Wang,E.T., Sandberg,R., Luo,S., Khrebtukova,I., Zhang,L., Mayr,C., Kingsmore,S.F., Schroth,G.P. and Burge,C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470–476.
3. Huranová,M., Ivani,I., Benda,A., Poser,I., Brody,Y., Hof,M., Shav-Tal,Y., Neugebauer,K.M. and Stanek,D. (2010) The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J. Cell. Biol.*, **191**, 75–86.
4. Listerman,I., Sapra,A.K. and Neugebauer,K.M. (2006) Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. *Nat. Struct. Mol. Biol.*, **13**, 815–822.
5. Will,C.L. and Luhrmann,R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell. Biol.*, **13**, 290–301.
6. Hoskins,A.A., Friedman,L.J., Gallagher,S.S., Crawford,D.J., Anderson,E.G., Wombacher,R., Ramirez,N., Cornish,V.W., Gelles,J. and Moore,M.J. (2011) Ordered and dynamic assembly of single spliceosomes. *Science*, **331**, 1289–1295.
7. Das,R. and Reed,R. (1999) Resolution of the mammalian E complex and the ATP-dependent spliceosomal complexes on native agarose mini-gels. *RNA*, **5**, 1504–1508.
8. Konarska,M.M. and Sharp,P.A. (1986) Electrophoretic separation of complexes involved in the splicing of precursors to mRNAs. *Cell*, **46**, 845–855.
9. Makarov,E.M., Owen,N., Bottrill,A. and Makarova,O.V. (2012) Functional mammalian spliceosomal complex E contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Res.*, **40**, 2639–2652.
10. Bessonov,S., Anokhina,M., Krasauskas,A., Golas,M.M., Sander,B., Will,C.L., Urlaub,H., Stark,H. and Luhrmann,R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA*, **16**, 2384–2403.
11. Agafonov,D.E., Deckert,J., Wolf,E., Odenwälder,P., Bessonov,S., Will,C.L., Urlaub,H. and Luhrmann,R. (2011) Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.*, **31**, 2667–2682.
12. Jamison,S.F., Crow,A. and Garcia-Blanco,M.A. (1992) The spliceosome assembly pathway in mammalian extracts. *Mol. Cell. Biol.*, **12**, 4279–4287.
13. Michaud,S. and Reed,R. (1991) An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes Dev.*, **5**, 2534–2546.
14. Michaud,S. and Reed,R. (1993) A functional association between the 5′ and 3′ splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev.*, **7**, 1008–1020.
15. Staknis,D. and Reed,R. (1994) SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.*, **14**, 7670–7682.
16. Cho,S., Hoang,A., Sinha,R., Zhong,X.Y., Fu,X.D., Krainer,A.R. and Ghosh,G. (2011) Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proc. Natl Acad. Sci. USA*, **108**, 8233–8238.
17. Eperon,I.C., Ireland,D.C., Smith,R.A., Mayeda,A. and Krainer,A.R. (1993) Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *EMBO J.*, **12**, 3607–3617.
18. Kohtz,J.D., Jamison,S.F., Will,C.L., Zuo,P., Lührmann,R., Garcia-Blanco,M.A. and Manley,J.L. (1994) Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature*, **368**, 119–124.
19. Shao,W., Kim,H.S., Cao,Y., Xu,Y.Z. and Query,C.C. (2012) A U1-U2 snRNP interaction network during intron definition. *Mol. Cell. Biol.*, **32**, 470–478.
20. Kent,O.A. and MacMillan,A.M. (2002) Early organization of pre-mRNA during spliceosome assembly. *Nat. Struct. Biol.*, **9**, 576–581.
21. Dönmez,G., Hartmuth,K., Kastner,B., Will,C.L. and Luhrmann,R. (2007) The 5′ end of U2 snRNA is in close proximity to U1 and functional sites of the pre-mRNA in early spliceosomal complexes. *Mol. Cell*, **25**, 399–411.
22. Kotlajich,M.V., Crabb,T.L. and Hertel,K.J. (2009) Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Mol. Cell. Biol.*, **29**, 1072–1082.

23. Lim,S.R. and Hertel,K.J. (2004) Commitment to splice site pairing coincides with A complex formation. *Mol. Cell*, **15**, 477–483.
24. Eperon,I.C., Makarova,O.V., Mayeda,A., Munroe,S.H., Cáceres,J.F., Hayward,D.G. and Krainer,A.R. (2000) Selection of alternative 5′ splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.*, **20**, 8303–8318.
25. Skordis,L.A., Dunckley,M.G., Yue,B., Eperon,I.C. and Muntoni,F. (2003) Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates SMN2 gene expression in patient fibroblasts. *Proc. Natl Acad. Sci. USA*, **100**, 4114–4119.
26. Eperon,I.C. and Krainer,A.R. (1994) In: Higgins,S.J. and Hames,B.D. (eds), *RNA Processing: A Practical Approach*. IRL Press, Oxford, pp. 57–101.
27. Zacharias,D.A., Violin,J.D., Newton,A.C. and Tsien,R.Y. (2002) Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science*, **296**, 913–916.
28. Lee,K.A., Bindereif,A. and Green,M.R. (1988) A small-scale procedure for preparation of nuclear extracts that support efficient transcription and pre-mRNA splicing. *Gene Anal. Tech.*, **5**, 22–31.
29. Kambach,C. and Mattaj,I.W. (1992) Intracellular distribution of the U1A protein depends on active transport and nuclear binding to U1 snRNA. *J. Cell. Biol.*, **118**, 11–21.
30. Cherny,D., Gooding,C., Eperon,G.E., Coelho,M.B., Bagshaw,C.R., Smith,C.W. and Eperon,I.C. (2010) Stoichiometry of a regulatory splicing complex revealed by single-molecule analyses. *EMBO J.*, **29**, 2161–2172.
31. Aitken,C.E., Marshall,R.A. and Puglisi,J.D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.*, **94**, 1826–1835.
32. Weber,G., Trowitzsch,S., Kastner,B., Lührmann,R. and Wahl,M.C. (2010) Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.*, **29**, 4172–4184.
33. Press,W.H., Flannery,B.P., Teukolsky,S.A. and Vettering,W.T. (1992) *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, 2nd edn. Cambridge University Press, Cambridge.
34. Heinrichs,V., Bach,M., Winkelmann,G. and Luhrmann,R. (1990) U1-specific protein C needed for efficient complex formation of U1 snRNP with a 5′ splice site. *Science*, **247**, 69–72.
35. Tarn,W.Y. and Steitz,J.A. (1995) Modulation of 5′ splice site choice in pre-messenger RNA by two distinct steps. *Proc. Natl. Acad. Sci. USA*, **92**, 2504–2508.
36. Will,C.L., Rümpler,S., Klein Gunnewiek,J., van Venrooij,W.J. and Luhrmann,R. (1996) In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. *Nucleic Acids Res.*, **24**, 4614–4623.
37. Sleeman,J., Lyon,C.E., Platani,M., Kreivi,J.P. and Lamond,A.I. (1998) Dynamic interactions between splicing snRNPs, coiled bodies and nucleoli revealed using snRNP protein fusions to the green fluorescent protein. *Exp. Cell Res.*, **243**, 290–304.
38. Lorković,Z.J. and Barta,A. (2008) Role of Cajal bodies and nucleolus in the maturation of the U1 snRNP in Arabidopsis. *PLoS One*, **3**, e3989.
39. Will,C.L. and Lührmann,R. (2011) Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.*, **3**, a003707.
40. Crispino,J.D., Blencowe,B.J. and Sharp,P.A. (1994) Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. *Science*, **265**, 1866–1869.
41. Crispino,J.D., Mermoud,J.E., Lamond,A.I. and Sharp,P.A. (1996) Cis-acting elements distinct from the 5′ splice site promote U1-independent pre-mRNA splicing. *RNA*, **2**, 664–673.
42. Tarn,W.Y. and Steitz,J.A. (1994) SR proteins can compensate for the loss of U1 snRNP functions in vitro. *Genes Dev.*, **8**, 2704–2717.
43. Klein Gunnewiek,J.M., Hussein,R.I., van Aarssen,Y., Palacios,D., de Jong,R., van Venrooij,W.J. and Gunderson,S.I. (2000) Fourteen residues of the U1 snRNP-specific U1A protein are required for homodimerization, cooperative RNA binding, and inhibition of polyadenylation. *Mol. Cell. Biol.*, **20**, 2209–2217.
44. Varani,L., Gunderson,S.I., Mattaj,I.W., Kay,L.E., Neuhaus,D. and Varani,G. (2000) The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein. *Nat. Struct. Biol.*, **7**, 329–335.
45. Kuhn,A.N., van Santen,M.A., Schwienhorst,A., Urlaub,H. and Luhrmann,R. (2009) Stalling of spliceosome assembly at distinct stages by small-molecule inhibitors of protein acetylation and deacetylation. *RNA*, **15**, 153–175.
46. Boelens,W.C., Jansen,E., Van,V.W., Stripecke,R., Mattal,I.W. and Gunderson,S.I. (1993) The human U1 snRNP-specific U1A protein inhibits polyadenylation of its own pre-mRNA. *Cell*, **72**, 881–892.
47. Chabot,B. and Steitz,J.A. (1987) Multiple interactions between the splicing substrate and small nuclear ribonucleoproteins in spliceosomes. *Mol. Cell. Biol.*, **7**, 281–293.
48. Fleckner,J., Zhang,M., Valcárcel,J. and Green,M.R. (1997) U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev.*, **11**, 1864–1872.
49. Shen,H., Zheng,X., Shen,J., Zhang,L., Zhao,R. and Green,M.R. (2008) Distinct activities of the DExD/H-box splicing factor hUAP56 facilitate stepwise assembly of the spliceosome. *Genes Dev.*, **22**, 1796–1803.
50. Xu,Y.Z., Newnham,C.M., Kameoka,S., Huang,T., Konarska,M.M. and Query,C.C. (2004) Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *EMBO J.*, **23**, 376–385.
51. Hicks,M.J., Mueller,W.F., Shepard,P.J. and Hertel,K.J. (2010) Competing upstream 5′ splice sites enhance the rate of proximal splicing. *Mol. Cell. Biol.*, **30**, 1878–1886.
52. Reed,R. and Maniatis,T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell*, **46**, 681–690.
53. Yu,Y., Maroney,P.A., Denker,J.A., Zhang,X.H., Dybkov,O., Lührmann,R., Jankowsky,E., Chasin,L.A. and Nilsen,T.W. (2008) Dynamic regulation of alternative splicing by silencers that modulate 5′ splice site competition. *Cell*, **135**, 1224–1236.
54. Jarmoskaite,I. and Russell,R. (2011) DEAD-box proteins as RNA helicases and chaperones. *Wiley Interdiscip. Rev. RNA*, **2**, 135–152.
55. Rippe,K. (2001) Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.*, **26**, 733–740.
56. Cunningham,S.A., Else,A.J., Potter,B. and Eperon,I.C. (1991) Influences of separation and adjacent sequences on the use of alternative 5′ splice sites. *J. Mol. Biol.*, **217**, 265–281.
57. Beyer,A.L., Bouton,A.H. and Miller,O.L. Jr (1981) Correlation of hnRNP structure and nascent transcript cleavage. *Cell*, **26**, 155–165.