

hnRNP A1 Proofreads 3' Splice Site Recognition by U2AF

Joao Paulo Tavanetz,^{1,2} Tobias Madl,^{4,5,6} Hamed Kooshapur,^{4,5} Michael Sattler,^{4,5} and Juan Valcárcel^{1,2,3,*}

¹Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain

²Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain

³Institució Catalana de Recerca i Estudis Avançats, Dr. Aiguader 88, 08003 Barcelona, Spain

⁴Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany

⁵Biomolecular NMR and Center for Integrated Protein Science Munich, Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

⁶Present address: NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

*Correspondence: juan.valcarcel@crg.es

DOI 10.1016/j.molcel.2011.11.033

SUMMARY

One of the earliest steps in metazoan pre-mRNA splicing involves binding of U2 snRNP auxiliary factor (U2AF) 65 KDa subunit to the polypyrimidine (Py) tract and of the 35 KDa subunit to the invariant AG dinucleotide at the intron 3' end. Here we use *in vitro* and *in vivo* depletion, as well as reconstitution assays using purified components, to identify hnRNP A1 as an RNA binding protein that allows U2AF to discriminate between pyrimidine-rich RNA sequences followed or not by a 3' splice site AG. Biochemical and NMR data indicate that hnRNP A1 forms a ternary complex with the U2AF heterodimer on AG-containing/uridine-rich RNAs, while it displaces U2AF from non-AG-containing/uridine-rich RNAs, an activity that requires the glycine-rich domain of hnRNP A1. Consistent with the functional relevance of this activity for splicing, proofreading assays reveal a role for hnRNP A1 in U2AF-mediated recruitment of U2 snRNP to the pre-mRNA.

INTRODUCTION

Splicing of mRNA precursors is an important step in eukaryotic gene expression whereby noncoding intronic sequences are excised and exonic sequences are spliced together to generate mature mRNAs (reviewed by Wahl et al., 2009). The process offers multiple opportunities for gene regulation, which are abundantly exploited during the development and cell differentiation of multicellular organisms (reviewed by Nilsen and Graveley, 2010). While the sequences defining the intron/exon boundaries are well defined in yeast, splice site sequences are much more degenerate in higher eukaryotes (reviewed by Keren et al., 2010), which may be related to the higher prevalence of alternative splicing in these organisms (Wang et al., 2008; Pan et al., 2008). The mechanisms by which the splicing machinery

discriminates between bona fide splice sites and many other related sequences are therefore of great interest to reconcile the accuracy required to maintain gene expression with the versatility necessary to allow alternative splicing regulation.

3' splice sites in higher eukaryotes are composed of three main sequence elements. The branch site is an invariable adenosine residue which forms a 2'-5' phosphodiester bond with the 5' end of the intron after the first catalytic step of the splicing reaction. The branchpoint adenosine is flanked by sequences that establish base-pairing interactions with U2 snRNA, the RNA component of the U2 snRNP ribonucleoprotein, during formation of prespliceosomal (A) complex (Parker et al., 1987; Nelson and Green, 1989; Wu and Manley, 1989; Zhuang and Weiner, 1989; Zhang, 1998). The 3' end of the intron contains an invariable AG dinucleotide within the consensus CAG/G ("/" indicates the intron-exon boundary). A pyrimidine-rich stretch (the polypyrimidine [Py] tract) is often located between the branch site and the 3' splice site AG, although recent results in fission yeast indicate that other sequence arrangements can also be functional (Sridharan and Singh, 2007; Sridharan et al., 2010). The length and pyrimidine richness of the Py tract often correlate with the efficiency of 3' splice site recognition (Zamore et al., 1992; Singh et al., 1995).

The branch site sequence is recognized initially by the branchpoint binding protein (BBP/SF1) (Berglund et al., 1997; Liu et al., 2001) and subsequently through base-pairing interactions with U2. The Py tract and 3' splice site AG are recognized by the 65 KDa and 35 KDa subunits of the U2 snRNP auxiliary factor (U2AF), respectively (Zamore et al., 1992; Wu et al., 1999; Zorio and Blumenthal, 1999; Merendino et al., 1999). An interaction between the amino-terminal region of BBP/SF1 and the carboxy-terminal U2AF homology motif (UHM) of U2AF65 facilitates cooperative binding of these proteins to their relatively loosely defined cognate sites (Berglund et al., 1998; Selenko et al., 2003). Similarly, the tight interaction between the two subunits of the U2AF heterodimer stabilizes U2AF65 binding to the Py tract through recognition of the 3' splice site AG by U2AF35 (Zamore and Green, 1989; Zhang et al., 1992; Wu et al., 1999; Zorio and Blumenthal, 1999; Merendino et al., 1999). The arrangement of protein and RNA contacts in the initial

complex is believed to enforce a particular RNA conformation in the 3' end of the intron that serves as a trigger for subsequent events in the splicing process (Kent et al., 2003).

U2AF65 contains two RNA recognition motifs (RRM1 and RRM2) that interact with Py tracts (Zamore et al., 1992; Singh et al., 1995; Banerjee et al., 2003, 2004), an amino-terminal arginine/serine-rich (RS) domain that facilitates base pairing between the branch site sequences and U2 snRNA (Lee et al., 1993; Valcárcel et al., 1996) and a carboxy-terminal UHM, a RRM-like domain that engages in interactions with the SF3b 155 protein component of U2 snRNP to facilitate U2 snRNP recruitment (Gozani et al., 1998). The domain structure of U2AF35 (Zhang et al., 1992) includes a UHM domain that has been shown to bind to the 3' splice site AG dinucleotide (Wu et al., 1999; Zorio and Blumenthal, 1999; Merendino et al., 1999). This interaction becomes essential for spliceosome assembly on AG-dependent introns characterized by relatively uridine-poor Py tracts that can only weakly recruit U2AF65 in the absence of the U2AF35/AG interaction (Wu et al., 1999; Merendino et al., 1999; Guth et al., 2001). We have previously shown that purified U2AF heterodimer fails to provide discrimination between uridine-rich RNAs followed by a bona fide 3' splice site and a similar RNA containing a 3' splice site mutation, while proofreading activities present in nuclear extracts allow U2AF to discriminate between the two RNAs. The protein DEK is necessary, but not sufficient, for this discriminatory activity (Soares et al., 2006).

Multiple hnRNP proteins associate with nascent pre-mRNA molecules and influence RNA metabolism from transcription to RNA decay (Reed and Maniatis, 1986; McAfee et al., 1996; Krecic and Swanson, 1999; Singh and Valcarcel, 2005). hnRNP A/B proteins contain two RRMs at the N terminus and a glycine-rich domain at the C terminus. hnRNP A1 and its N-terminal proteolytic fragment UP1 (lacking the glycine-rich domain) were initially characterized as single-stranded telomeric DNA-binding proteins (McKay and Cooke, 1992; Ishikawa et al., 1993; Ding et al., 1999; Dallaire et al., 2000), where they stimulate telomerase activity and regulate telomere length in vivo (LaBranche et al., 1998; Zhang et al., 2006). A classical function of hnRNP A1 in splicing is to antagonize SR proteins to promote the use of distal splice sites and exon skipping (Mayeda and Krainer, 1992; Mayeda et al., 1994; Caceres et al., 1994). One proposed mechanism relies on glycine-rich domain-mediated intermolecular contacts between hnRNP A1 molecules, which loop out the intervening RNA (Blanchette and Chabot, 1999). In other instances, hnRNP A1 binding to specific splicing silencer elements interferes with spliceosome assembly (Tange et al., 2001; Kashima and Manley, 2003; Cartegni et al., 2006).

A more general role for hnRNP A1 in the splicing reaction has also been proposed. Selection of RNA sequences from a random pool (SELEX) identified high-affinity hnRNP A1 binding sites composed of one or more copies of the motif UAGGG(A/U), which has similarities with 3' splice site sequences (Swanson and Dreyfuss, 1988). hnRNP A1 has been proposed to display binding preference for both 5' and 3' splice sites (Swanson and Dreyfuss, 1988; Buvoli et al., 1990), and one study identified clustering of high-affinity hnRNP A1 binding sites at 3' splice sites, with mutations at the 3' splice site AG dinucleotide

affecting hnRNP A1 binding (Burd and Dreyfuss, 1994). Interestingly, it has also been reported that hnRNP A1 interacts with U2 snRNP, therefore suggesting a role for hnRNP A1 in early stages of spliceosome assembly (Buvoli et al., 1992).

In this manuscript we report a function for hnRNP A1 in facilitating discrimination between pyrimidine-rich RNAs followed or not by a 3' splice site AG by U2AF. Our data indicate that U2AF and hnRNP A1 form a ternary complex on AG-containing pyrimidine-rich RNAs, while hnRNP A1 displaces U2AF from binding to pyrimidine-rich RNAs not followed by bona fide 3' splice site signals, an activity that requires the glycine-rich domain of hnRNP A1.

RESULTS

Previous work documented that the protein DEK is necessary but not sufficient to confer specific binding of U2AF to uridine-rich RNAs followed by a 3' splice site AG dinucleotide (Soares et al., 2006). To identify additional factors involved in 3' splice site discrimination by U2AF, chromatography in oligo(dT) cellulose columns was used to separate U2AF and strong interacting partners—retained in the column even in 1M KCl—from other nuclear extract components (Zamore and Green, 1989; Valcárcel, 1997). Flowthrough fractions are depleted of U2AF and cannot support pre-mRNA splicing reactions, whereas the U2AF heterodimer and other proteins retained in the column can be eluted either with 2.5M KCl (high-salt fraction) or 2M guanidine-HCl (guanidine eluate) (Zamore and Green, 1989; Guth et al., 2001; Tange et al., 2001) (Figure 1A).

To evaluate 3' splice site discrimination by U2AF, two uniformly α 32P-UTP-labeled uridine-rich RNAs, (U)13ACAGG (AG RNA) and (U)13ACCGG (CG RNA), were incubated with minimal U2AF heterodimer (U2AF RRM1 and RRM2 of U2AF65 and the UHM of U2AF35). These domains are known to be sufficient for heterodimer RNA binding (Singh et al., 1995; Guth et al., 2001; Soares et al., 2006). RNA and proteins were incubated either in buffer or in the presence of HeLa cell nuclear extracts or oligo-dT-depleted extracts, under splicing conditions for 30 min at 30°C, followed by ultraviolet (UV) light-induced photocrosslinking and immunoprecipitation with anti-U2AF65 antibodies. As observed previously (Soares et al., 2006), no preference for AG-RNAs or CG-RNAs was observed for the purified heterodimer (Figure 1B, lanes 1 and 2), while a clear preference for U2AF65 crosslinking to AG RNA was observed in the presence of HeLa nuclear extracts, both for endogenous U2AF65 and for the exogenously added minimal heterodimer (Figure 1C, lanes 1 and 2). Oligo(dT)-depleted extracts, which contain DEK but do not contain U2AF (Figure 1A), showed reduced binding discrimination by the minimal heterodimer compared to nondepleted extracts (Figure 1B, compare, for example, lanes 7 and 8 with 13 and 14), suggesting that factors retained in the column are important for U2AF binding discrimination. Consistent with this possibility, U2AF present in the guanidine eluate and in the high-salt fractions showed a preference for crosslinking to the AG-RNA, and these fractions were also sufficient to provide discrimination to exogenously added minimal heterodimer, (Figure 1C, lanes 3 and 4 and 5 and 6). These fractions contain both subunits of

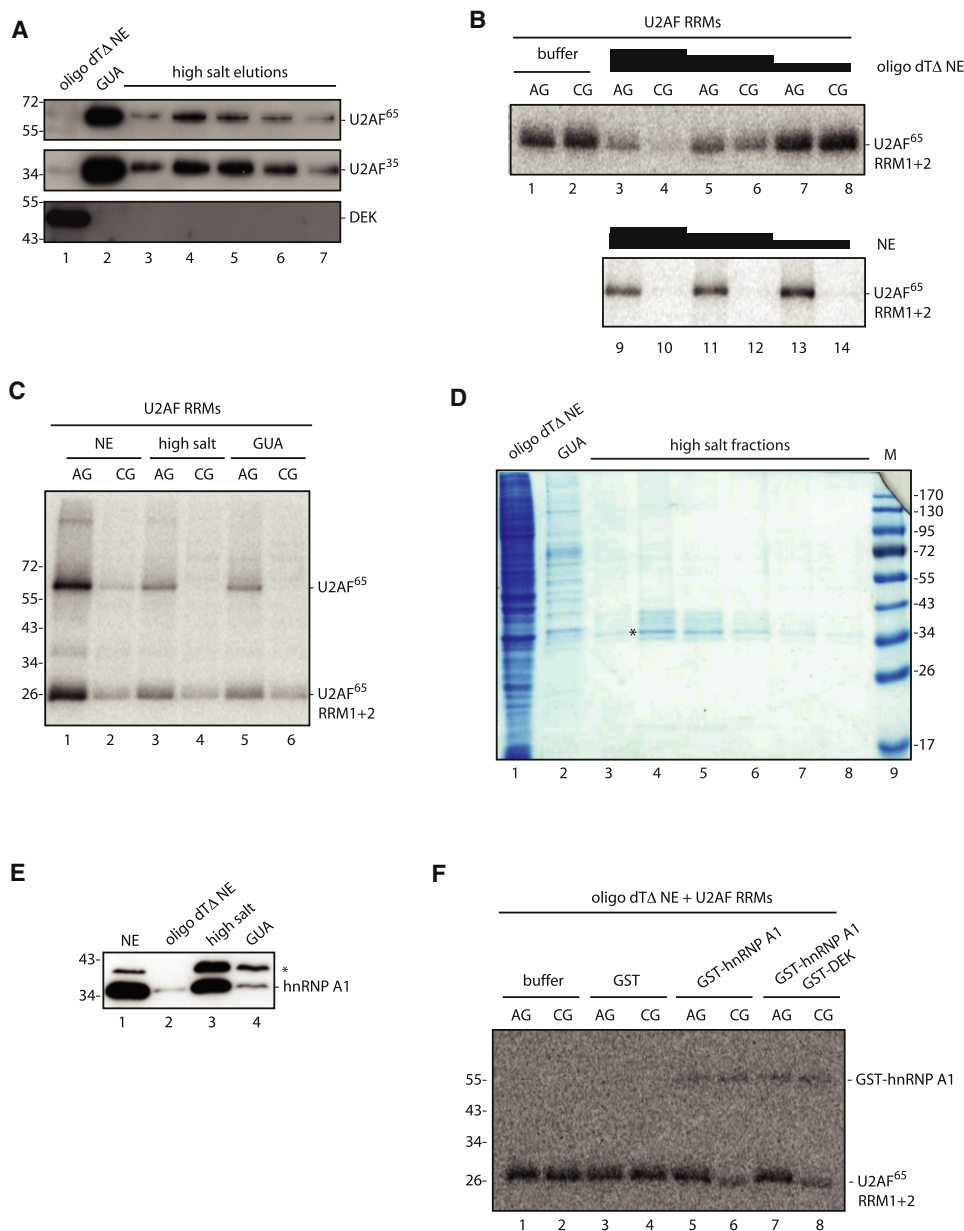


Figure 1. Identification of hnRNP A1 as a Factor Important for Proofreading of U2AF Binding to 3' Splice Sites

(A) Western blot analysis of U2AF and DEK levels after chromatographic depletion of HeLa nuclear extracts using oligo-dT cellulose. Oligo-dT-depleted nuclear extracts (2 μ l, 15 μ g of protein), guanidine eluate, and high-salt fractions (5 μ l) were analyzed using anti-U2AF⁶⁵, anti-U2AF³⁵, and anti-DEK antibodies, as indicated.

(B) UV crosslinking and immunoprecipitation of U2AF⁶⁵ RRM 1+2 in the presence of decreasing amounts of oligo(dT)-depleted and HeLa nuclear extracts (normalized by protein concentration), indicated by the height of the bars above the lanes.

(C) UV crosslinking assay using ³²P-uridine-labeled RNAs 5' GGG(U)₁₃AC-AG/CG followed by immunoprecipitation of U2AF⁶⁵. Nuclear extracts, guanidine eluate, and high-salt fraction were supplemented with purified minimal heterodimer (U2AF RRM5). The positions corresponding to endogenous and recombinant U2AF⁶⁵ are indicated.

(D) Coomassie staining of protein profiles of chromatographic fractions described in (A). Asterisk corresponds to the molecular mass of hnRNP A1.

(E) Western blot of fractions described in (A) using an anti-hnRNP A1 antibody.

(F) UV crosslinking and immunoprecipitation of U2AF⁶⁵ RRM 1+2 in the presence of oligo-dT-depleted extracts complemented with 2 ng/ μ l of the indicated recombinant proteins.

the U2AF heterodimer but do not contain DEK (Figure 1A), indicating that these fractions contain DEK-independent activities capable of providing AG versus CG discrimination to U2AF.

To identify candidate factors, the chromatographic fractions were analyzed by SDS polyacrylamide gel electrophoresis (Figure 1D), and gel slices were analyzed by mass spectrometry. The most conspicuous species (labeled with an asterisk in Figure 1D) corresponded to the multifunctional RNA binding protein hnRNP A1, whose levels were reduced more than 90% in oligo-dT-depleted extracts and which was readily detectable by western blot in the high-salt and guanidine eluates (Figure 1E). Importantly, AG versus CG discrimination, which was severely compromised in oligo(dT)-depleted extracts (Figures 1B and 1F, lanes 1–4), was restored by complementing these extracts with recombinant purified hnRNP A1 (Figure 1F, lanes 5 and 6), indicating that hnRNP A1 plays a role in 3' splice site discrimination by U2AF and that the lack of AG versus CG discrimination observed in oligo(dT)-depleted extracts is due, at least in part, to depletion of hnRNP A1. Addition of DEK did not enhance the effect of hnRNP A1 (Figure 1F, lanes 7 and 8).

Depletion of hnRNP A1 from Nuclear Extracts Compromises 3' Splice Site Discrimination by U2AF

In addition to U2AF and hnRNP A1, other factors present in nuclear extracts are likely to be depleted by oligo-dT chromatography. To obtain rigorous evidence for the role of hnRNP A1 in 3' splice site discrimination by U2AF, RNA affinity depletion of hnRNP A1 from nuclear extracts was carried out as previously described (Bonnal et al., 2005). SELEX experiments have identified high-affinity hnRNP A1 binding sequences composed of one or more copies of the motif UAGGG(A/U), which can bind to the protein either as DNA or RNA. 5'-biotinylated DNA oligonucleotides corresponding to this sequence were used to deplete hnRNP A1. Nuclear extracts adjusted to 0.6M KCl were subject to five sequential rounds of depletion by incubation with A1-WT oligo or a mutant derivative, A1-mut oligo, bound to streptavidin beads. Approximately 80% of the hnRNP A1 protein was depleted in A1-WT-depleted extracts (Figure 2A), a significant level of depletion considering hnRNP A1's abundance in nuclear extracts (Dreyfuss et al., 1993), and significant levels of hnRNP A1 were detected among the proteins specifically bound to the A1-WT beads (Figure 2A, lanes 6 and 7). Depletion was also specific, as more similar levels of U2AF65, U2AF35, and DEK were observed in both extracts (Figure 2A). Consistent with a role for hnRNP A1 in providing AG versus CG discrimination to U2AF, depletion of hnRNP A1 from nuclear extracts resulted in reduced AG versus CG discrimination by endogenous U2AF65 and by the minimal heterodimer (Figure 2B, compare lanes 5 and 6 with 7 and 8; see Figure S1 available online), an effect that was progressively reversed when recombinant hnRNP A1, purified by two different methods, was titrated in the depleted extracts (Figure 2C, compare lanes 1 and 2 with 3 and 4, 5 and 6, and 7 and 8; Figure S2D). In contrast, no increase in discrimination was observed upon addition of U2AF35 or DEK (Figures S2A and S2B).

Also consistent with this conclusion were titration experiments using oligonucleotides containing high-affinity hnRNP A1 binding sites. hnRNP A1 has been implicated in telomeric

maintenance by binding to single-stranded telomeric repeats (McKay and Cooke, 1992; Ishikawa et al., 1993). Because hnRNP A1 binds strongly and specifically to a single-stranded oligonucleotide containing two human telomeric repeats, but not to mutant telomeric DNA repeats (Zhang et al., 2006), telomeric DNA oligos were used as competitors in standard discrimination assays. Limiting amounts of nuclear extracts were preincubated with increasing amounts of both WT and mut telomeric DNA repeats, and AG versus CG discrimination assays were carried out. Titration of oligonucleotides containing hnRNP A1 binding sites progressively reduced AG versus CG discrimination by U2AF (Figure 2D, upper panel), while titration of the mutant oligonucleotide did not (Figure 2D, lower panel).

A Minimal Recombinant System Composed of Purified Components Provides 3' Splice Site Discrimination to U2AF

To test whether hnRNP A1 was sufficient to provide AG versus CG discrimination to U2AF, assays were carried out in a minimal recombinant system containing only purified components. Recombinant hnRNP A1, purified by two different methods, was sufficient to provide a significant level of discrimination to the minimal heterodimer (Figure 2E, compare lanes 1 and 2 with 5 and 6; Figure S2E). Addition of full-length recombinant DEK did not enhance this effect (Figure 2E, lanes 7 and 8).

The activity conferring AG versus CG discrimination to U2AF is temperature dependent, as discrimination in nuclear extracts is compromised when the assays are carried out on ice (Figure S3A). This was also the case for the activity present in the high-salt fraction and for assays carried out in the minimal system of purified hnRNP A1 and minimal heterodimer (Figures S3B and S3C), further arguing that the biochemical properties of hnRNP A1 contribute to the discrimination activity observed in nuclear extracts.

While recombinant hnRNP A1 was found to be sufficient to confer discrimination to the minimal heterodimer, it failed to provide discrimination to U2AF65 RRM 1+2 alone (Figure 2F, compare lanes 1 and 2 with 3 and 4), indicating that hnRNP A1 activity requires the presence of U2AF35. AG versus CG discrimination was also analyzed using the UP1 N-terminal proteolytic fragment of hnRNP A1 (Pandolfo et al., 1985) that lacks the C-terminal glycine-rich domain essential for some hnRNP A1 functions such as modulation of alternative splicing (Mayeda et al., 1994). AG versus CG discrimination was not provided by UP1, indicating that the glycine rich C-terminal domain is necessary for the discriminatory activity of hnRNP A1 (Figure 2F, compare lanes 1 and 2 with 5 and 6). Collectively, the results demonstrate that hnRNP A1 can provide 3' splice site discrimination to U2AF, a temperature-dependent activity that requires U2AF35 and the C-terminal, glycine-rich domain of hnRNP A1.

hnRNP A1 Mediates Proofreading of Other 3' Splice Site Mutations

The proofreading activity of hnRNP A1 was also investigated using uridine-rich RNAs containing different 3' splice site dinucleotide variants (AG, CG, GG, and UG). While the minimal heterodimer crosslinked to all the RNAs, crosslinking of endogenous U2AF65 and the minimal heterodimer was detected in

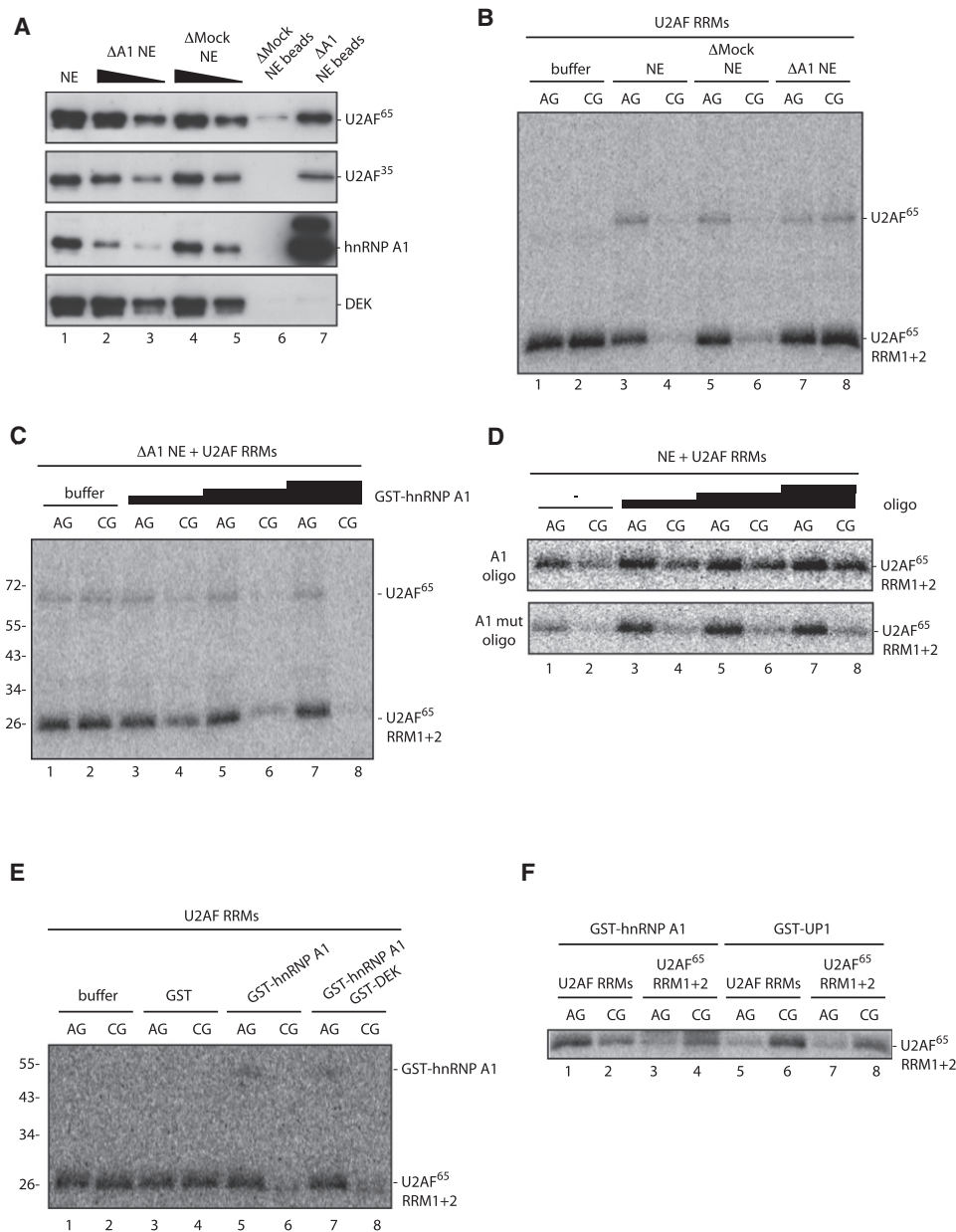


Figure 2. Depletion of hnRNP A1 from Nuclear Extracts Compromises U2AF Proofreading Activity

(A) Quantification of hnRNP A1 levels after depletion of HeLa nuclear extracts with biotinylated DNA oligos. Mock-depleted and hnRNP A1-depleted nuclear extracts (1 μ l and a 2-fold dilution) were analyzed by western blot using anti-hnRNP A1, anti-U2AF65, anti-U2AF35, and anti-DEK antibodies. Twenty-five percent of the bound material was loaded on the gel.

(B) UV crosslinking and immunoprecipitation of endogenous U2AF65 and U2AF65 RRM 1+2 in the presence of buffer or the indicated nuclear extracts supplemented with purified minimal heterodimer (U2AF RRM1+2) (see also Figure S1).

(C) UV crosslinking and immunoprecipitation assays as in (B) using hnRNP A1-depleted nuclear extracts supplemented with purified minimal heterodimer (U2AF RRM1+2) and complemented with increasing amounts of recombinant GST-hnRNP A1 (0.6, 1.2, and 2.4 ng/ μ l). Average levels of discrimination for these concentrations of hnRNP A1 were 61%, 82%, and 93% of those observed in nondepleted nuclear extracts.

(D) UV crosslinking and immunoprecipitation assays as in (B) using limited amounts of nuclear extracts preincubated for 1 hr at 30°C with wild-type or mutant telomeric DNA oligos (4, 12, and 36 μ M).

(E) UV crosslinking and immunoprecipitation of U2AF65 RRM 1+2 in a minimal recombinant system composed of purified components. The indicated GST recombinant proteins (7 ng/ μ l) were incubated with purified minimal heterodimer (U2AF RRM1+2) and RNAs before crosslinking and immunoprecipitation using anti-U2AF65 antibodies. See also Figures S2 and S3.

(F) Domains of hnRNP A1 and U2AF involved in AG versus CG discrimination. Assays were carried out as in (E), using the indicated recombinant purified proteins and RNAs.

nuclear extracts only with the AG RNA and not with the CG, GG, and UG RNA variants (Figure 3A). Efficient crosslinking of U2AF to the four RNAs was detected, however, in hnRNP A1-depleted nuclear extracts (Figure 3B), an effect that was reversed upon addition of full-length recombinant hnRNP A1 to the depleted extracts (Figure 3C). This result confirms the requirement of hnRNP A1 for proofreading of multiple 3' splice site mutations by U2AF. Using purified proteins and RNA, however, hnRNP A1 allowed U2AF to discriminate between AG and CG, but did not provide discrimination between AG and GG or UG RNA variants (Figure 3D). Collectively, the results of Figures 3A–3D indicate that hnRNP A1 is required for preventing U2AF binding to pyrimidine-rich RNAs containing different 3' splice site mutations, but for some mutations proofreading of U2AF requires additional factors.

hnRNP A1 Proofreads U2AF Binding In Vivo

Next we asked whether depletion of hnRNP A1 could also allow spurious binding of U2AF to uridine-rich sequences outside of 3' splice sites in living cells. HeLa cells were transfected with siRNAs against hnRNP A1 or with scrambled control siRNAs, leading to significant and specific depletion of hnRNP A1 (Figure 3E). Binding of U2AF to particular RNA regions was measured in total extracts of formaldehyde-crosslinked cells by immunoprecipitation with U2AF65 antibodies followed by DNA digestion, RNA isolation, and RT-PCR using specific oligonucleotide primers. To avoid detection of U2AF association at sites distant from the actual binding sites, nucleic acids were fragmented to an average size of 250 nucleotides by sonication previous to immunoprecipitation. To avoid any ambiguity in the association of U2AF with splicing signals, we chose to evaluate the binding of U2AF to uridine-rich sequences not followed by AG dinucleotides, present in the 5' or 3'UTR of the intronless gene *c-jun* (Figure 3F, upper panel). We defined uridine-rich stretches as sequences of at least four consecutive uridines within a context of at least ten pyrimidines. Binding of U2AF to regions of the transcript devoid of uridine stretches longer than three residues was used as control. Depletion of hnRNP A1 led to a significant specific increase in the association of U2AF with uridine-rich sequences (Figure 3F, lower panel, compare the increase in signal in the U2AF immunoprecipitates corresponding to regions A and F—which contain uridine-rich stretches—with regions B to E—which do not). Quantitative RT-PCR analyses revealed a 4- to 13-fold enrichment in U2AF association to uridine-rich-containing regions in the absence of hnRNP A1 (Figure 3G). Additional analysis confirmed enriched U2AF association with uridine-rich sequences located at 5'UTRs of three other intron-containing transcripts upon hnRNP A1 depletion (Figure 3H). We conclude that hnRNP A1 contributes to prevent spurious binding of U2AF to uridine-rich sequences and therefore restricts U2AF binding to 3' splice site regions, both in vitro and in vivo.

RNA- and U2AF35-Dependent Interaction between U2AF and hnRNP A1

To evaluate potential interactions between hnRNP A1 and U2AF, recombinant purified GST-hnRNP A1 was used for GST-pull-down assays in HeLa nuclear extracts. After extensive washes,

western blot analysis confirmed the presence of Sam68—a known hnRNP A1-interacting protein (Paronetto et al., 2007)—in GST-hnRNP A1 precipitates (Figure 4A, lane 4), as well as both U2AF subunits (Figure 4A, lanes 4 and 7). Reciprocal assays using recombinant GST-U2AF65 revealed, in addition to the expected U2AF65 partners U2AF35 and Sam68 (Figure 4A, lane 3), specific enrichment of hnRNP A1 in GST-U2AF65 precipitates (Figure 4A, lanes 9 and 10). This interaction was RNA dependent, because it was no longer detected after treatment with RNase A (Figure 4B, compare lanes 3 and 6) and no direct binding between these proteins was observed (data not shown and see below).

To further evaluate whether RNA-dependent interactions between hnRNP A1 and U2AF are relevant for AG versus CG discrimination, gel retardation assays were carried out using purified proteins and RNAs. Both the minimal U2AF heterodimer and hnRNP A1 bound similarly to AG- and CG-containing pyrimidine-rich RNAs in these assays (Figure 4C, lanes 3 and 4 and 9 and 10). In the presence of U2AF and hnRNP A1, however, distinct complexes were formed on AG-RNAs and CG-RNAs (Figure 4C, lanes 11 and 12). Because the complex formed on CG-RNAs has identical mobility to that formed with hnRNP A1 (compare lanes 10 and 12), the result suggests that U2AF is displaced from CG-RNAs by hnRNP A1 (also confirmed by NMR data, see below). Consistent with the requirement of the glycine-rich carboxy-terminal domain of hnRNP A1 (Figure 2F), UP1 did not displace U2AF from CG-RNAs (Figure 4C, lanes 7 and 8), while it bound similarly to both RNAs with affinities similar to that of full-length hnRNP A1 (Figure 4C, lanes 5 and 6). hnRNP A1-mediated displacement of U2AF from CG-RNAs was not observed in the absence of U2AF35 RRM either (compare lanes 11 and 12 in Figures 4C and 4D), indicating that both subunits of U2AF are required for AG versus CG discrimination.

The mobility of the complexes formed on AG-containing RNAs in the presence of both U2AF and hnRNP A1 is slightly slower than that of the complexes formed with U2AF alone. To determine whether the complex formed on AG RNAs contains U2AF and hnRNP A1, several approaches were used. Addition of a U2AF65-specific monoclonal antibody led to a supershift of U2AF-containing complexes, including the complex formed on AG-RNAs in the presence of hnRNP A1 (Figure 4E, compare lanes 3 and 4 with 9 and 10 and lane 7 with 13), confirming the presence of U2AF in these complexes. Antibodies against hnRNP A1 also supershifted partially hnRNP A1-containing complexes, including the complex formed on AG-RNAs in the presence of U2AF and hnRNP A1 (Figure 4F, compare lanes 5 and 6 with 11 and 12 and lanes 7 and 8 with 13 and 14). Because supershifts with a variety of anti-hnRNP A1 antibodies were not complete, two additional experiments were carried out to further document the presence of hnRNP A1 in the ternary complex. First, taking advantage of the GST tag present in the hnRNP A1 recombinant protein, glutathione beads were added after initial incubation of the RNA-protein mixes, followed by separation by centrifugation and electrophoretic fractionation of the supernatant. GST-hnRNP A1-containing complexes, including the complex formed in the presence of U2AF on AG-RNAs, were depleted by pull-down with glutathione agarose beads, while binding of U2AF complexes was not affected by the

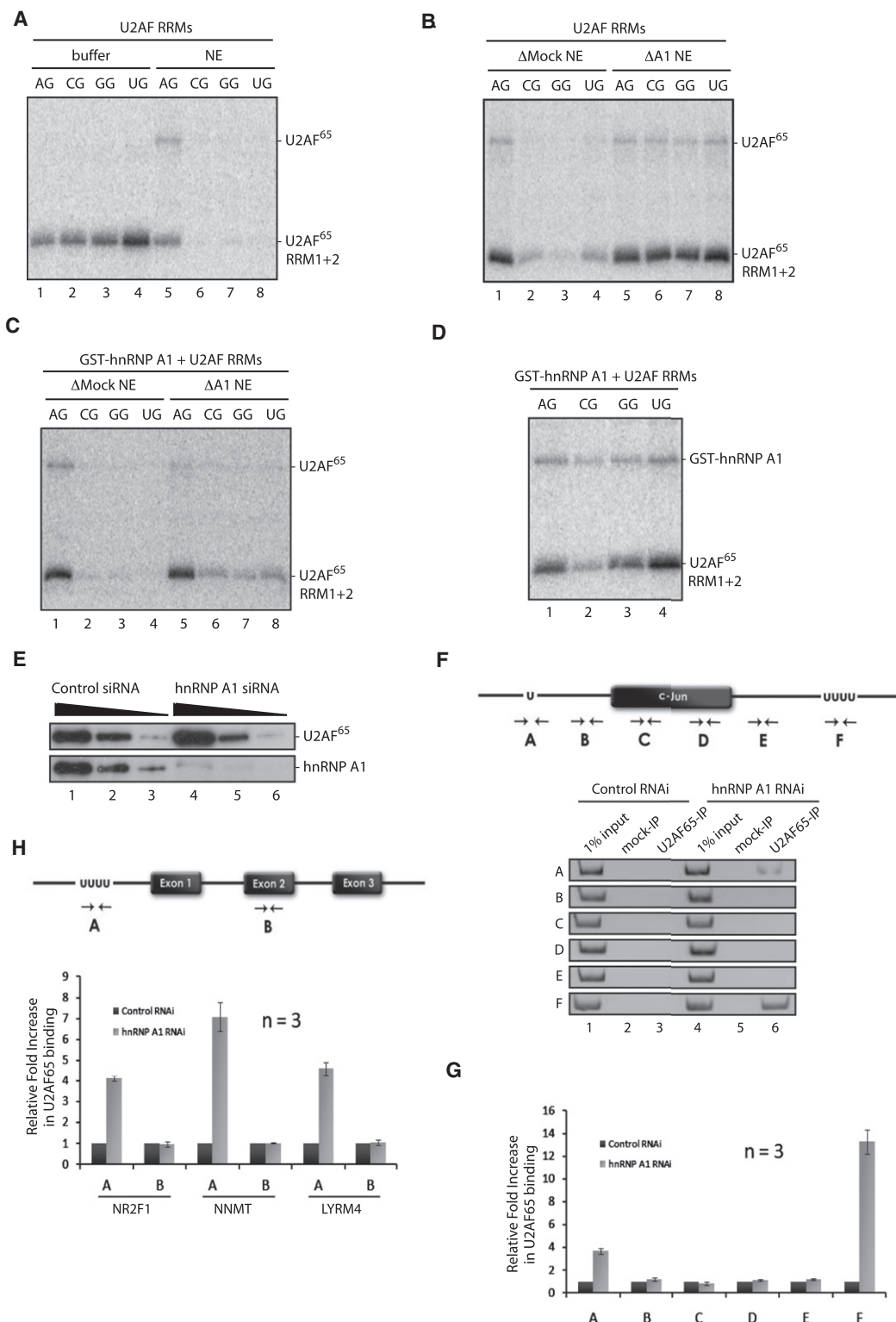


Figure 3. In Vitro Proofreading of Other 3' Splice Site Mutations and In Vivo Proofreading of U2AF Binding to an Intronsless Transcript Require hnRNP A1

(A) UV crosslinking and immunoprecipitation of endogenous U2AF65 and U2AF65 RRM 1+2 in the presence of the indicated RNAs (32 P-uridine-labeled RNAs 5' GGG(U)₁₃AC-AG/-CG/-GG/ or -UG).

(B) Discrimination assay as in (A) in the presence of mock-depleted and hnRNP A1-depleted nuclear extracts.

procedure (Figure S4A). Second, incubation with unlabeled RNA corresponding to the hnRNP A1 SELEX winner sequence (Burd and Dreyfuss, 1994) inhibited binding of hnRNP A1 to the radioactively labeled transcripts and also inhibited ternary complex formation on AG-RNA, allowing binding of the U2AF heterodimer only (Figure S4B). Taken together, the results of Figures 4C–4F and Figure S4 indicate that both U2AF subunits and hnRNP A1 simultaneously assemble on AG-containing pyrimidine-rich RNAs, while U2AF is displaced by hnRNP A1 from CG-RNAs.

To further document the formation of a ternary complex on AG-RNAs, the proteins precipitated by pull-down of GST-hnRNP A1 with glutathione beads were resolved by SDS-polyacrylamide gel electrophoresis. Coomassie blue staining detected the presence of the two subunits of the minimal U2AF heterodimer (Figure 5A, lane 3). Coprecipitation was RNA dependent (compare lanes 2 and 3 in Figure 5A) and was significantly stronger (2.3-fold on average) using AG- than using CG-RNAs (compare lanes 3 and 4 in Figure 5A). Consistent with the requirement for the glycine-rich domain of hnRNP A1 for AG versus CG discrimination by U2AF, GST-UP1 precipitates contained equal levels of U2AF associated to AG-RNAs and CG-RNAs (Figure 5A, lanes 7 and 8). Consistent with the requirement for U2AF35, no U2AF65 RRM were coprecipitated with either GST-hnRNP A1 or GST-UP1 in the absence of the 35 kDa subunit (Figure 5B).

Collectively, the results of the RNA-protein interaction assays of Figures 4 and 5 indicate that hnRNP A1 and U2AF heterodimer form a ternary complex with AG-containing pyrimidine-rich RNAs, while hnRNP A1 displaces U2AF from CG-RNAs. The glycine-rich region of hnRNP A1 is necessary for AG- versus CG- discrimination by U2AF but is not required for formation of a ternary complex.

NMR Analysis of the U2AF-hnRNP A1-RNA Interactions

We used NMR spectroscopy to characterize the molecular interactions and binding interfaces involving the minimal U2AF heterodimer, hnRNP A1, and the AG- and CG-containing pyrimidine-rich RNAs. Protein-specific isotope-labeled complexes were prepared and chemical shifts were analyzed using ^1H , ^{15}N correlation spectra (Figure 6A). Large chemical shift changes were observed when comparing NMR spectra of isotope-labeled U2AF65 (assembled in the minimal heterodimer) alone and bound to AG-RNA or CG-RNA (Figures 6B and 6C). The

RNA-bound states (black spectra in Figures 6B and 6C, superimposed in Figure S5) show identical NMR spectra, indicating that—as expected—U2AF65 contacts the Py tract region and not the AG or CG nucleotides of these RNAs. Upon addition of hnRNP A1 and in the presence of CG-RNA, the NMR spectrum of U2AF65-labeled heterodimer is very similar to the free heterodimer (Figure 6B, magenta spectrum; notice the more extensive signal overlap between the green [free U2AF65] and the magenta [U2AF65 in the presence of CG-RNA and hnRNP A1] than with the black [U2AF65 bound to CG-RNA] spectra within the circled residues). This is consistent with the notion that A1 blocks the U2AF heterodimer from binding to CG-RNAs, as supported by the results of crosslinking and electrophoretic mobility shift experiments (Figure 1, Figure 2, Figure 3, Figure 4, and Figure 5). In contrast, a higher-order complex is formed with AG-RNA, as indicated by line-broadening and disappearance of particular signals in the U2AF65-RNA binding interface (Figure 6C, magenta spectrum; notice, for example, the reduction of magenta signals for F202, G154, G301, or T296). The overall increased line widths are consistent with the formation of a higher molecular weight complex involving U2AF, hnRNP A1, and the AG-RNA. The chemical shifts of observable amide signals for a number of residues surrounding the RNA binding interface resemble the RNA-bound state of U2AF65, as expected (I160, L270 in Figure 6C).

These conclusions are further supported by analysis of NMR spectra of hnRNP A1 (Figures 6D and 6E). Addition of CG-RNA to hnRNP A1 shows chemical shift perturbations consistent with formation of hnRNP A1 contacts with the RNA (Figure 6D, black spectrum). NMR spectra of isotope-labeled hnRNP A1 with CG-RNA are virtually identical in the presence or absence of U2AF (Figure 6D, compare black and blue spectra), confirming that hnRNP A1 binds to the CG-RNA independently of U2AF. The line widths and overall good quality of the NMR spectra are consistent with the formation of only a binary complex involving hnRNP A1 and CG-RNA, without the involvement of U2AF. NMR spectra of hnRNP A1 upon addition of AG RNA show large chemical shift perturbations, demonstrating the formation of a protein-RNA complex (Figure 6E, compare black and orange spectra). As observed with the NMR data of the U2AF65-labeled heterodimer, the spectra of hnRNP A1 bound to the AG-RNA in the presence of U2AF show line broadening and partial disappearance of NMR signals (Figure 6E, compare

(C) Discrimination assay as in (B), complementing the indicated depleted nuclear extracts with 2 ng/ μl of recombinant GST-hnRNP A1.

(D) UV crosslinking and immunoprecipitation of U2AF65 RRM 1+2 in a minimal recombinant system containing purified minimal heterodimer and hnRNP A1 as in Figure 2E, using the RNAs described in (A).

(E) Depletion of hnRNP A1 from HeLa cells. The levels of hnRNP A1 and U2AF65 in extracts of cells (10 μl and serial 2-fold dilutions) treated with control siRNAs or siRNAs targeting hnRNP A1 were analyzed by western blot using specific antibodies.

(F) (Upper panel) Schematic representation of *c-jun* transcripts (thin line corresponds to 5' and 3'UTRs, box represents ORF) and of the regions amplified by RT-PCR with the indicated primer pairs. U indicates location of uridine-rich sequences (defined as stretches of at least four consecutive uridines within a context of at least ten pyrimidines). (Lower panel) Increased binding of U2AF to uridine-rich regions of intronless *c-jun* transcripts upon hnRNP A1 depletion. RNA extracted from control or U2AF65 immunoprecipitates of sonicated extracts of cells transfected with control siRNAs or siRNAs targeting hnRNP A1 was reverse transcribed and PCR amplified with primers specific for the transcript regions indicated in the upper panel and analyzed by acrylamide gel electrophoresis.

(G) Quantitative RT-PCR analysis of the RNA samples in (F) for the regions indicated in the upper panel. The graph indicates the fold increase and standard deviations ($n = 3$) in amplification signals of RNAs present in U2AF65 immunoprecipitates over input, from extracts of cells transfected with siRNAs targeting hnRNP A1; the amplification values from extracts of cells transfected with control siRNAs have been normalized to 1.

(H) Quantitative RT-PCR assays as in (G) for transcripts from the indicated genes. Two regions of the RNAs were analyzed: uridine-rich stretches in the 5'UTR, A, and exonic sequences of balanced nucleotide composition, B. Increased association of U2AF65 to uridine-rich sequences upon hnRNP A1 depletion is observed for the three transcripts. Quantitative analyses are as in (G).

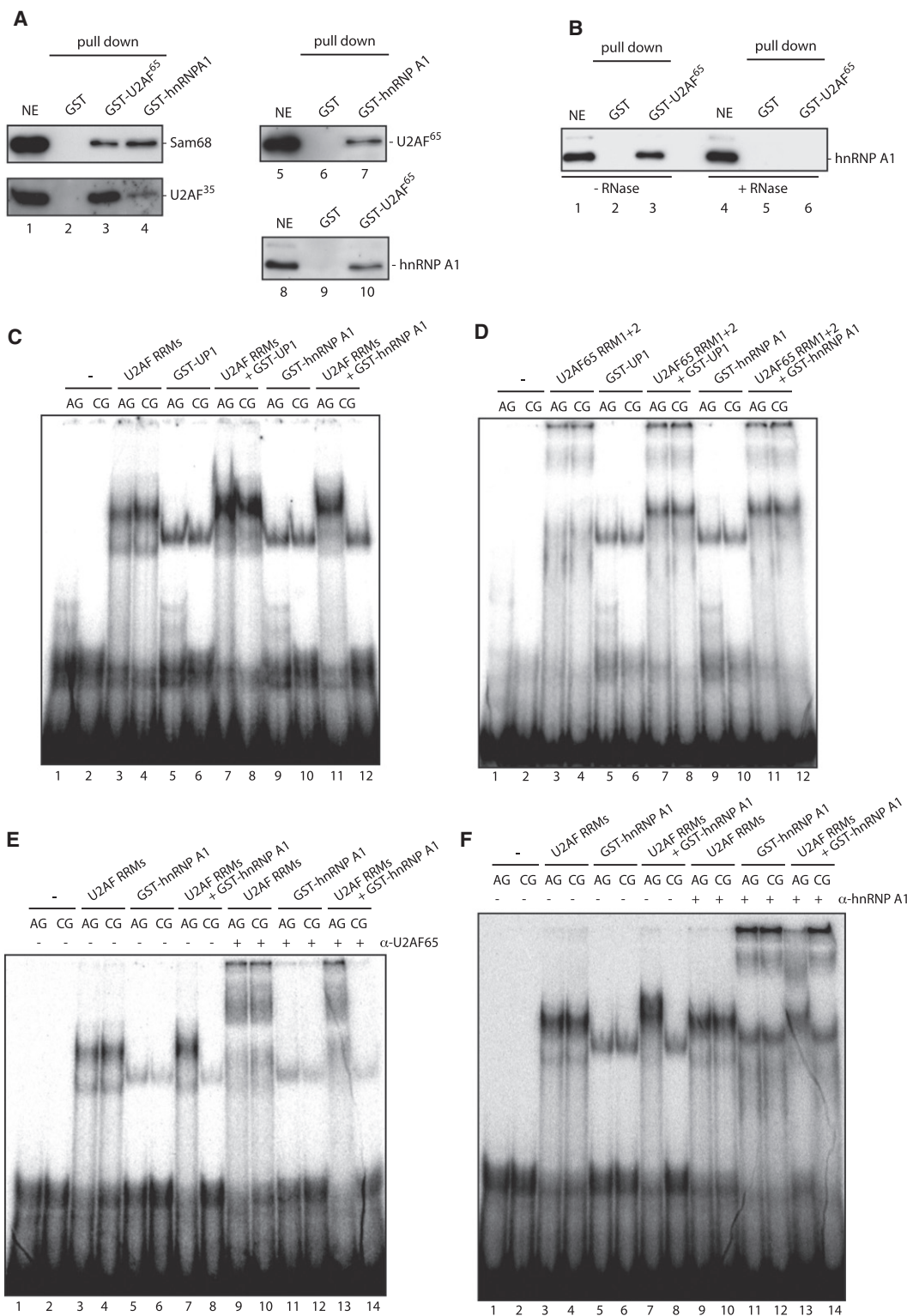


Figure 4. RNA- and U2AF35-Dependent Ternary Complexes Mediate hnRNP A1-Induced AG Discrimination by U2AF65

(A) Interaction between U2AF and hnRNP A1 detected by GST-pull-down assays. The indicated GST fusion proteins were incubated with HeLa nuclear extracts and precipitated with glutathione agarose beads, and after extensive washes the precipitates were fractionated by SDS-polyacrylamide gel electrophoresis and analyzed by western blot using antibodies against the indicated proteins.

(B) The U2AF65/hnRNP A1 interaction is RNA dependent. Pull-down assays were carried out as in (A) in the absence or presence of RNase treatment.

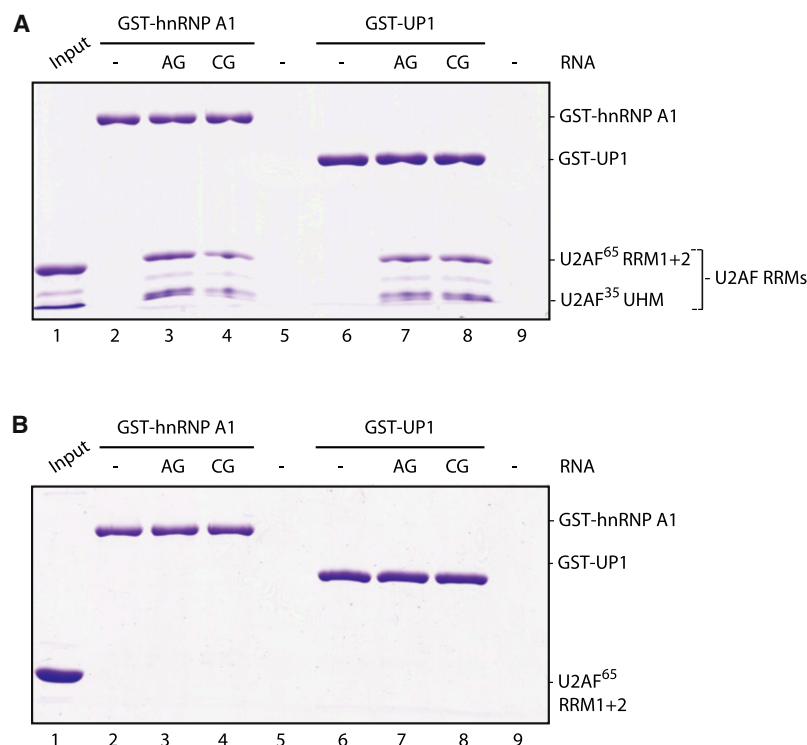


Figure 5. U2AF and hnRNP A1 Form Ternary Complexes on AG-RNAs

(A) Protein analysis of complexes formed on uridine-rich AG- and CG-containing RNAs. Pull-down of the indicated GST fusion proteins was carried out in the absence or presence of the indicated RNAs, and the proteins in the precipitates were analyzed by SDS-polyacrylamide gel electrophoresis and Coomassie staining.

(B) U2AF35 is required for interaction in the ternary complex. Pull-down assays were carried out as in (A) except that U2AF35 was omitted from the mixes.

Functional Relevance of hnRNP A1-Mediated Proofreading for Spliceosome Assembly

To evaluate the impact of the activity of hnRNP A1 reported above on spliceosome assembly, functional proofreading assays were carried out. In these assays, schematically represented in Figure 7A, U2 snRNP recruitment to a 3' splice site-containing radioactively labeled RNA (AdML) is competed by addition of an excess of unlabeled RNAs containing a uridine tract followed by AG or CG at the 3' end. When AG versus CG discrimination is active, an excess of AG-RNAs titrate U2AF and prevent U2 snRNP assembly (a 3' complex formation)

black and blue spectra), consistent with the formation of a higher-order complex involving hnRNP A1, U2AF, and the AG-RNA.

NMR spectra also provide evidence for the requirement of full-length hnRNP A1 for discriminating CG-RNAs versus AG-RNAs. Both hnRNP A1 and UP1 (which lacks the C-terminal glycine-rich tail) show significant and comparable chemical shift changes upon binding to AG or CG RNAs (Figure S6). However, some additional NMR signals are observed only in the complex with AG-RNA and full-length hnRNP A1 (red arrows/circles in Figures 6D and 6E, also Figure S6). Based on their characteristic chemical shift, these signals can be tentatively assigned to glycines in the C-terminal tail of hnRNP A1. This suggests that the glycine-rich tail in hnRNP A1 mediates additional contacts with the RNA or protein components of the ternary complex, which are required for CG versus AG discrimination.

In summary, the NMR data are consistent with the biochemical assays and confirm the displacement of U2AF from CG RNAs by hnRNP A1, and the formation of a ternary complex containing hnRNP A1 and the U2AF heterodimer on AG RNAs.

on the AdML substrate. The presence of proofreading activities prevents titration of U2AF by the CG-RNAs, which therefore do not compromise U2 snRNP recruitment on AdML. In the absence of proofreading activities, however, CG-RNAs become competitors as effective as the AG-RNAs, titrating U2AF and inhibiting complex A formation on AdML RNA (Soares et al., 2006).

When nuclear extracts were preincubated with mutant-telomeric DNA oligos that do not provide a binding site for hnRNP A1, complex A formation in the 3' splice site of AdML was competed by the AG-RNA but not by the CG-RNA (Figure 7B, lanes 1–7). In contrast, both RNAs were effective competitors when nuclear extracts were preincubated with telomeric DNA oligos containing hnRNP A1 binding sites (Figure 7B, lanes 8–14), indicating that these extracts are deficient in an activity that normally discriminates between AG-RNAs and CG-RNAs and allows only the former to proceed through spliceosome assembly under conditions of competition.

The same experiment was carried out in two types of hnRNP A1-depleted extracts. In affinity-depleted nuclear extracts (as in Figure 2), complex A formation was competed only by addition

(C) AG versus CG discrimination by recombinant proteins using mobility shift assays. The proteins indicated on the top were incubated with AG- or CG-containing uridine-rich RNAs and fractionated by electrophoresis on native polyacrylamide gels.

(D) U2AF35 is required for AG versus CG discrimination. Mobility shift assays were carried out as in (C) in the absence of U2AF35.

(E) U2AF65 is present in discrimination complexes. Mobility shift assays as in (C) were carried out in the absence or presence of anti-U2AF65 antibodies added after initial incubation of the proteins with RNA. Note the change in mobility of U2AF complexes and of the complex formed in the presence of U2AF and hnRNP A1 upon addition of the antibody, indicative of the presence of U2AF65.

(F) hnRNP A1 is present in discrimination complexes. Mobility shift assays as in (C) were carried out in the absence or presence of anti-hnRNP A1 antibodies added after initial incubation of the proteins with RNA. Note the partial change in mobility of a hnRNP A1 complexes and of the complex formed in the presence of U2AF and hnRNP A1 upon addition of the antibody, indicative of the presence of hnRNP A1. See also Figure S4.

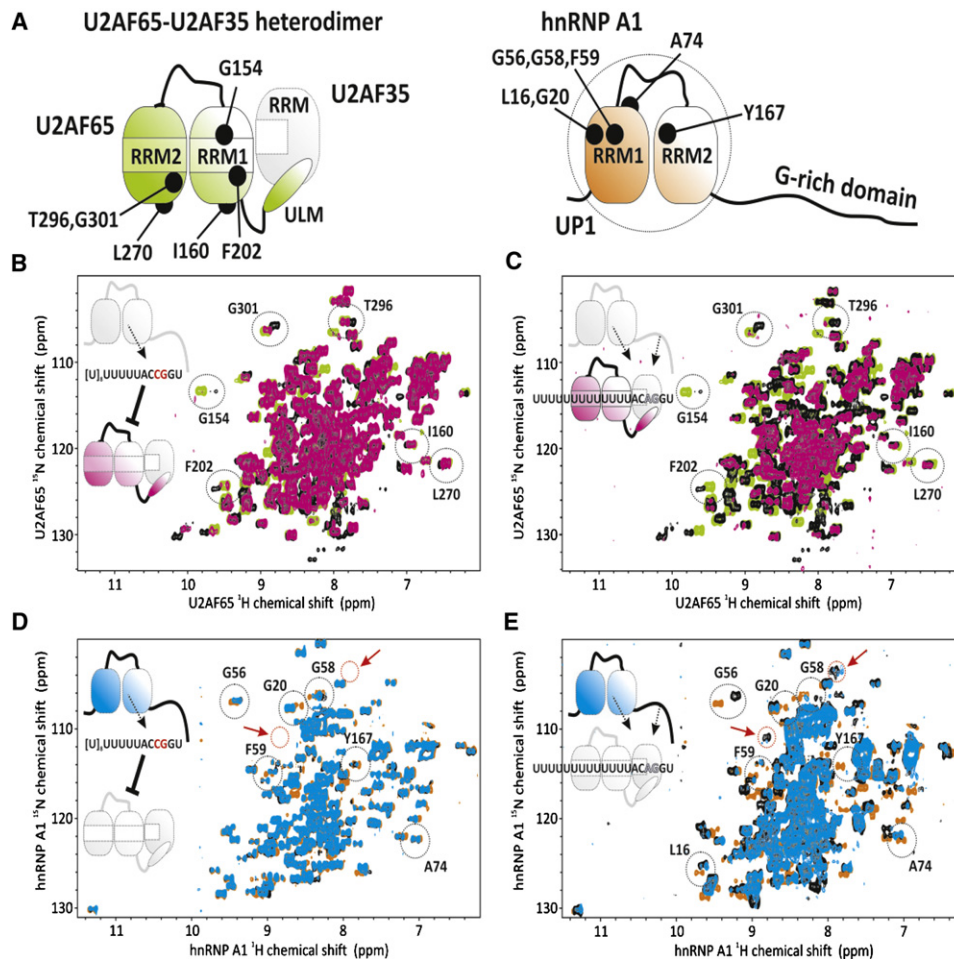


Figure 6. NMR Analysis of the U2AF/hnRNP A1 Complexes with CG and AG RNA

(A) Diagram of the U2AF heterodimer representing domains of U2AF65 and U2AF35 (left) and of hnRNP A1 (right). Characteristic domains (RRM, RNA recognition motif; ULM, U2AF homology domain [UHM] ligand motif) and selected residues highlighted in the spectra are labeled. U2AF65 and hnRNP A1 are colored green and orange, respectively, with the same color code for the RNA-unbound proteins being used in the NMR spectra in (B)–(E). (B) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of U2AF65-labeled U2AF heterodimer free (green), when bound to a 19 nucleotide CG-containing RNA (black) or in the presence of RNA and unlabeled hnRNP A1 (magenta). (C) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of U2AF65-labeled U2AF heterodimer free (green), when bound to a 19 nucleotide AG-containing RNA (black) or in the presence of RNA and unlabeled hnRNP A1 (magenta). (D) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of isotope-labeled hnRNP A1 free (orange), bound to the CG-containing RNA (black) and in the presence of RNA and unlabeled U2AF heterodimer (blue). (E) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of isotope-labeled hnRNP A1 free (orange) bound to the AG RNA (black) or in the presence of RNA and unlabeled U2AF heterodimer (blue). In all spectra, some characteristic residues affected by the binding are encircled and labeled. Two additional NMR signals which can be tentatively assigned to the glycine-rich tail of hnRNP A1 are encircled in red and marked by red arrows. Protein concentrations used for the NMR experiments were 50 μM (green, black, orange spectra) or 25 μM (magenta spectra), respectively. The CG- and AG-containing RNAs were added in a 1:1 heterodimer:RNA ratio. See also Figures S5 and S6.

of the AG-RNA under conditions of mock depletion (Figure 7C, lanes 1–7), while both AG-RNAs and CG-RNAs were effective competitors in hnRNP A1-depleted extracts (Figure 7B, lanes 8–14). When these depleted extracts were complemented with full-length recombinant GST-hnRNP A1—but not with GST alone—CG-RNAs failed to compete (Figure 7B, compare lanes 18–20 with lanes 15–17). These results support the notion that the activity of hnRNP A1 that provides U2AF binding discrimination is relevant for U2 snRNP recruitment

under conditions of competition with an excess of pyrimidine-rich RNAs. Also consistent with this idea, both AG-RNAs and CG-RNAs were effective competitors when complex A formation was analyzed in extracts depleted of both hnRNP A1 and U2AF by chromatography in oligo-dT cellulose (as in Figure 1), complemented with full-length recombinant GST-U2AF65 (Figure S7A). As expected, binding of the purified U2AF heterodimer is competed with the same efficiency by AG-RNAs and CG-RNAs (Figure S7B).

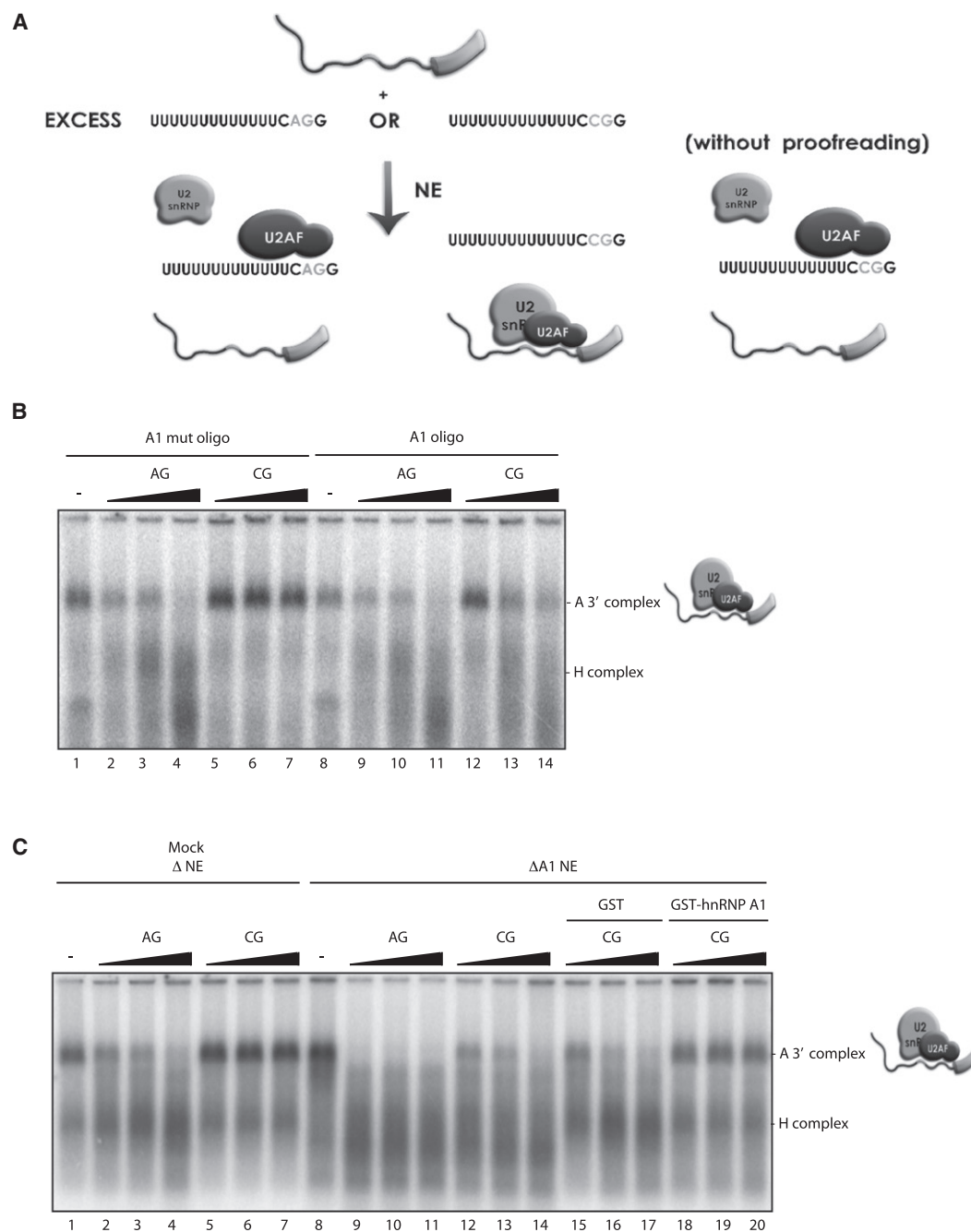


Figure 7. Proofreading Activity of hnRNP A1 Is Important for Spliceosome Assembly

(A) Schematic representation of the proofreading assay. A radioactively labeled 3' splice site-containing RNA is incubated in HeLa nuclear extracts in the presence of an excess of uridine-rich RNAs followed by either an AG or a CG dinucleotide. In the presence of proofreading activities, only the AG-containing pyrimidine-rich RNA prevents spliceosome assembly. In the absence of proofreading activities—i.e., in hnRNP A1-depleted extracts—both AG- and CG-containing RNAs prevent U2AF association and spliceosome assembly on the pre-mRNA.

(B) Spliceosome assembly assays using AdML 3'RNA containing 3' half of the intron and exon 2. Nuclear extracts were preincubated with WT and mutant telomeric DNA oligos as in Figure 2D, and assembly of spliceosomal complexes was competed by the addition of a 10, 30, and 90 molar excess of AG- and CG-containing uridine-rich RNAs.

(C) Spliceosome assembly assays as in (A) using mock- and hnRNP A1-depleted nuclear extracts. hnRNP A1-depleted nuclear extracts were complemented with 10 ng/μl recombinant GST and GST-hnRNP A1 as indicated. See also Figure S7.

We conclude that hnRNP A1-mediated proofreading of U2AF binding is important for complex A formation on bona fide 3' splice site regions.

DISCUSSION

Degeneracy of splice site sequences is characteristic of organisms harboring extensive alternative splicing and may play a key role in facilitating regulation of splice site selection. One mechanism to reconcile sequence diversity with efficient splice site recognition entails cooperative binding of interacting factors to neighboring sequences. For example, identification of the three loosely defined signals at 3' splice sites involves interactions between BBP/SF1 and U2AF65 and between U2AF65 and U2AF35 that provide a platform to recognize arrangements of branch site, Py tract, and AG sequences even if the affinity of each individual factor to its cognate sequence is limited (Zamore et al., 1992; Berglund et al., 1998; Wu et al., 1999; Zorio and Blumenthal, 1999; Merendino et al., 1999; Selenko et al., 2003; Mackereth et al., 2011). This strategy, however, opens the question of how binding of individual factors to high-affinity individual cognate sequences out of the context of splice sites is prevented. In this manuscript, we report that hnRNP A1 facilitates discrimination by the U2AF heterodimer between pyrimidine-rich RNAs followed or not by the 3' splice site AG dinucleotide, thus preventing spurious association of U2AF outside of splice sites.

Recruitment of U2AF to sequences outside of 3' splice sites has been detected by RNA immunoprecipitation assays (Gama-Carvalho et al., 2006). While the functional significance of these binding events is still not well understood, the balance between the activity of hnRNP A1 and other factors is likely to allow recruitment of U2AF to particular pyrimidine-rich regions outside of 3' splice sites. Additional effects of hnRNP A1 on various aspects of RNA metabolism, including alternative splicing, may also be facilitated by its ability to proofread binding of U2AF or other RNA binding proteins.

Proofreading of 3' Splice Site Recognition

Our initial motivation to identify factors that proofread 3' splice site recognition by U2AF was the observation that the protein product of the oncogene DEK is necessary, but not sufficient, for discrimination between Py tracts followed by AG or CG dinucleotides by U2AF (Soares et al., 2006). The results of this manuscript indicate that hnRNP A1 is necessary and sufficient for 3' splice site discrimination between AG- and CG-containing pyrimidine-rich RNAs by the U2AF heterodimer. DEK (Soares et al., 2006) is therefore required for this intrinsic activity of hnRNP A1 to be deployed in the context of the complex composition of hnRNP particles existing in nuclear extracts/cell nucleus. DEK may facilitate the association of hnRNP A1 with pyrimidine-rich regions otherwise occupied by other factors. DEK interacts—dependent upon phosphorylation of two serine residues—with U2AF35 (Soares et al., 2006), and hnRNP A1 requires U2AF35 for association with U2AF65 and with the RNA, as well as to facilitate 3' splice site discrimination (Figure 4, Figure 5, Figure 6). One possible model is that U2AF35-mediated recruitment of hnRNP A1 to the ternary complex is blocked by

other factors unless phosphorylated DEK displaces these hnRNP A1 repressors.

Proofreading of U2AF binding is only one of the steps at which fidelity mechanisms ensure proper 3' splice site recognition. For example, a stem-loop structure in U2 snRNA helps to establish initial interactions between the branchpoint recognition site in U2 (mainly located in the loop) and the pre-mRNA branch site. Upon initial recognition, the stem-loop structure is remodeled to allow stable U2 snRNP recruitment (Perriman and Ares, 2010). This remodeling step is likely to be guided by the DEAD box helicase/ATPase Prp5, which kinetically proofreads U2 snRNP assembly by monitoring U2 snRNA/branch site base pairing (Xu and Query, 2007). Kinetic proofreading appears as a general mechanism to ensure fidelity in splice site recognition at multiple steps in the assembly and catalysis by the spliceosome, and more generally in the function of other ribonucleoprotein complex machineries (Smith et al., 2008; Staley and Woolford, 2009; Egecioglu and Chanfreau, 2011).

A Ternary Complex for Initial 3' Splice Site Proofreading

Displacement of the U2AF heterodimer from non-AG-containing pyrimidine-rich RNAs by hnRNP A1 could be viewed as the result of competition between proteins with overlapping RNA binding specificities. However, the prediction from SELEX and other RNA binding data (Swanson and Dreyfuss, 1988; Buoli et al., 1990; Burd and Dreyfuss, 1994) is that hnRNP A1 binds better to AG than CG-containing pyrimidine-rich RNAs. In a simple competition model, hnRNP A1 would therefore more readily outcompete U2AF from AG-RNAs than from CG-RNAs, which is in contrast with our observations. In addition, our mobility shift assays did not reveal a significant difference in hnRNP A1 apparent binding affinity for AG-RNAs versus CG-RNAs (Figure 4).

In an alternative model consistent with the results of our experiments and the NMR data (Figure 4, Figure 5, Figure 6), the presence of an AG dinucleotide on pyrimidine-rich RNAs allows formation of a ternary ribonucleoprotein complex where hnRNP A1 is arranged in a conformation compatible with U2AF binding. Because U2AF35 is required both for formation of the ternary complex and for proofreading, recognition of the AG dinucleotide by U2AF35 (Wu et al., 1999; Zorio and Blumenthal, 1999; Merendino et al., 1999) may play a key role in accommodating hnRNP A1 and U2AF65 together on the RNA. The glycine-rich domain of hnRNP A1 is critical for displacement of U2AF from CG-containing, uridine-rich RNAs, but the domain is dispensable for binding to these RNAs or assembly of ternary complexes (Figure 2F, Figure 4, Figure 5, Figure 6). The domain may establish contacts with the Py tract (or with U2AF, or intramolecular contacts with the RRM of hnRNP A1) such that the interaction of hnRNP A1 with CG-containing RNAs prevents U2AF binding by steric occlusion. On AG-containing RNAs, the glycine-rich domain would adopt a conformation compatible with—or even enhance—U2AF65 interactions with the Py tract. hnRNP A1-induced discrimination between AG- and CG-containing pyrimidine-rich RNAs by U2AF is quantitatively more striking in mobility shift assays (Figure 4) than in pull-down assays (Figure 5), possibly reflecting less stable U2AF

interactions with CG-RNAs under conditions of prolonged electrophoresis on native gels.

Collectively, the biochemical and structural analyses reported in this manuscript reveal that initial recognition of 3' splice sites by U2AF is subject to proofreading by hnRNP A1 such that U2AF binding to Py tracts not followed by a bona fide AG 3' splice site is destabilized and displaced. The presence of an AG accommodates both subunits of U2AF and hnRNP A1 in a ternary complex that is compatible with U2AF recognition of the Py tract and with subsequent events in spliceosome assembly. More generally, the results illustrate how the conformation of ribonucleoprotein complexes can be exquisitely sensitive to minimal sequence variations, leading to outcomes that determine further assembly and processing steps on the RNA.

EXPERIMENTAL PROCEDURES

Oligo dT and hnRNP A1 Depletion of Nuclear Extracts

HeLa nuclear extract was depleted of U2AF as previously described (Valcárcel, 1997), passing the extract over a 1 ml oligo dT-cellulose column at 1 M KCl. Column was subsequently extensively washed with 1 M KCl, followed by elution of 200 μ l fractions with 2.5 M KCl and subsequently with 2 M guanidine-HCl. Biotinylated DNA oligos (A1-oligo 5'-CTAGTATGATAGGGACT TAGGGTG-3', A1mut-oligo 5'-CTAGTATGAGATGGACTGATGGTG-3') were used to deplete hnRNP A1 from nuclear extracts as described (Bonnal et al., 2005). Oligonucleotide (2 nmol) was incubated with 50 μ l of streptavidin beads (Sigma-Aldrich, St. Louis, MO) in 150 μ l of buffer D at 4°C, for 1 hr, with constant rotation. Beads were washed three times for 10 min and incubated with 100 μ l of NE adjusted to 0.6 M KCl for 90 min at 4°C with rotation. Samples were then centrifuged 5 min at 3,000 rpm, and the supernatants (depleted nuclear extracts) were collected and subjected to four additional rounds of depletion following the same procedure.

UV Crosslinking and Immunoprecipitation

³²P-labeled RNAs were incubated under in vitro splicing conditions in a total volume of 27 μ l. After incubation for 30 min at 30°C (or 0°C), mixtures were irradiated with UV light (254 nm; 0.4 J). Samples were subsequently incubated with 1 mg/ml RNase A at 37°C for 30 min, and U2AF65 was immunoprecipitated by addition of 25 μ l of MC3 antibody and incubated for 1.5 hr at 4°C. Protein A/G Sepharose bead slurry (25 μ l) was subsequently added, and samples were incubated for 1 hr at 4°C. Sedimented beads were washed four times with 500 μ l high-salt buffer (500 mM NaCl, 50 mM Tris-HCl [pH 8.0], 1% NP-40) and twice with low-salt buffer (150 mM NaCl, 50 mM Tris-HCl [pH 8.0], 1% NP-40). Sedimented beads were resuspended in 4 $\times$ SDS loading dye, boiled, centrifuged, and the supernatant loaded on 10% SDS-polyacrylamide gels. Dried gels were exposed to a PhosphorImager screen.

Cell Culture, RNAi, and RNA Immunoprecipitation

HeLa cells were grown in Dulbecco's modified Eagle's medium (GIBCO-BRL) supplemented with 10% fetal bovine serum (FBS) (GIBCO-BRL). For RNAi experiments, HeLa cells were plated the day before in 100 mm plates and transfected at 40%–50% confluence with 20 nM of both scrambled and hnRNP A1 siRNAs (CAACUUCGUGUGGAGGA) oligonucleotides (Thermo Scientific Dharmacon) using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Two sequential rounds of transfection were performed and cells were collected 60 hr after the initial transfection. RNAi efficiency was monitored by western blot.

For RNA immunoprecipitation, control and hnRNP A1 interfered cells were trypsinized, resuspended in fresh medium, and crosslinked with a final concentration of 1% (v/v) formaldehyde. Pellets were resuspended in RIPA buffer (50 mM Tris-HCl [pH 7.5], 1% NP-40, 0.5% sodium deoxycholate, 0.05% SDS, 1 mM EDTA, 150 mM NaCl) supplemented with protease inhibitors and RNasin (Promega). Extracts were sonicated with eight 30 s bursts using a microprobe at 30% amplitude yielding nucleic acid fragments with

a bulk size of 300–200 bp. After sonication, insoluble material was pelleted, and homogenates were further diluted 1:1 in RIPA buffer, precleared with A/G Sepharose beads for 1 hr, and incubated overnight with 20 μ l of anti-U2AF65 antibody or nonimmune IgG.

Immunocomplexes were recovered by the addition of 50 μ l of A/G Sepharose beads for 2 hr, washed five times in RIPA buffer, washed two times in RIPA buffer containing 500 mM NaCl, washed once in LiCl buffer, and washed once in TE (pH 7.5). Protein-nucleic acid complexes were eluted from the beads with elution buffer (TE with 1% SDS) for 30 min at 75°C and treated with 0.5 mg/mL of Proteinase K. Crosslinks were reversed at 65°C for 5 hr, nucleic acids were extracted with phenol-chloroform and treated with RQ1 DNase (Promega), and RNA was ethanol precipitated with glycogen as a carrier.

NMR Spectroscopy

Samples for NMR measurements typically contained 0.05–0.8 mM protein in 20 mM sodium phosphate (pH 6.5), 50 mM (300 mM for hnRNP A1) NaCl, and 1 mM DTT with 10% ²H₂O added for the lock signal. NMR spectra were recorded at 298 K on Avance 900 and Avance III 600 Bruker NMR spectrometers equipped with cryogenic triple resonance gradient probes. All titration spectra were recorded at 900 MHz on a cryo TXI probehead with a recycle delay of 1.0 s and spectral widths of 15/34 ppm centered at 4.7/117 ppm in ¹H/¹⁵N, with 1024 and 128 points, respectively, using 256 scans per increment. To improve sensitivity ¹H, ¹⁵N-CRINEPT-HMQC experiments (Riek et al., 1999) were employed for all spectra except those involving UP1. Spectra were processed with NMRPipe/Draw and analyzed with Sparky 3 (T.D. Goddard and D.G. Kneller, University of California, San Francisco, USA). Protein backbone assignments were obtained from standard HNCACB, HN(CO)CACB, HN(CO)CA, and HNCA spectra (Sattler et al., 1999), or by comparison to related ¹H, ¹⁵N-HSQC, and -TROSY spectra and previously published data (Soares et al., 2006).

Additional methods, including detailed protein purification procedures, can be found in the Supplemental Information.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, Supplemental Experimental Procedures, and Supplemental References and can be found with this article online at doi:10.1016/j.molcel.2011.11.033.

ACKNOWLEDGMENTS

We thank numerous colleagues in our institutes who have provided valuable input throughout this project and useful comments on the manuscript, particularly P. Papasaikas for bioinformatic predictions, A. Corriero and R. Tejedor for help with figures preparation, and the CRG Proteomics Facility for mass-spectrometry determination. J.P.T. was supported by an EMBO postdoctoral fellowship. T.M. thanks the Austrian Science Fund (FWF) and EMBO for postdoctoral fellowships. H.K. acknowledges support by the TUM Graduate School. Work in J.V.'s laboratory is supported by Fundación Marcelino Botín, Fundación Alicia Koplowicz, AICR, Consolider RNAREG, EURASNET, Ministerio de Ciencia e Innovación, and AGAUR. Work in M.S.'s laboratory is supported by the European Commission, grant NIM3 number 226507, and the Deutsche Forschungsgemeinschaft.

Received: March 21, 2011

Revised: August 1, 2011

Accepted: November 23, 2011

Published online: February 9, 2012

REFERENCES

Banerjee, H., Rahn, A., Davis, W., and Singh, R. (2003). Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. RNA 9, 88–99.

- Banerjee, H., Rahn, A., Gawande, B., Guth, S., Valcarcel, J., and Singh, R. (2004). The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF 65) is essential in vivo but dispensable for activity in vitro. *RNA* 10, 240–253.
- Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89, 781–787.
- Berglund, J.A., Abovich, N., and Rosbash, M. (1998). A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* 12, 858–867.
- Blanchette, M., and Chabot, B. (1999). Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* 18, 1939–1952.
- Bonnal, S., Pileur, F., Orsini, C., Parker, F., Pujol, F., Prats, A.C., and Vagner, S. (2005). Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. *J. Biol. Chem.* 280, 4144–4153.
- Burd, C.G., and Dreyfuss, G. (1994). RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 13, 1197–1204.
- Buvoli, M., Cobianchi, F., Biamonti, G., and Riva, S. (1990). Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res.* 18, 6595–6600.
- Buvoli, M., Cobianchi, F., and Riva, S. (1992). Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly. *Nucleic Acids Res.* 20, 5017–5025.
- Caceres, J.F., Stamm, S., Helfman, D.M., and Krainer, A.R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* 265, 1706–1709.
- Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E., and Krainer, A.R. (2006). Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.* 78, 63–77.
- Dallaire, F., Dupuis, S., Fiset, S., and Chabot, B. (2000). Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J. Biol. Chem.* 275, 14509–14516.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* 13, 1102–1115.
- Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu. Rev. Biochem.* 62, 289–321.
- Egecioglu, D.E., and Chanfreau, G. (2011). Proofreading and spellchecking: a two-tier strategy for pre-mRNA splicing quality control. *RNA* 17, 383–389.
- Gama-Carvalho, M., Barbosa-Morais, N.L., Brodsky, A.S., Silver, P.A., and Carmo-Fonseca, M. (2006). Genome-wide identification of functionally distinct subsets of cellular mRNAs associated with two nucleocytoplasmic-shuttling mammalian splicing factors. *Genome Biol.* 7, R113.
- Gozani, O., Potashkin, J., and Reed, R. (1998). A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol. Cell. Biol.* 18, 4752–4760.
- Guth, S., Tange, T.O., Kellenberger, E., and Valcarcel, J. (2001). Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol. Cell. Biol.* 21, 7673–7681.
- Ishikawa, F., Matunis, M.J., Dreyfuss, G., and Cech, T.R. (1993). Nuclear proteins that bind the pre-mRNA 3' splice site sequence r(UUAG/G) and the human telomeric DNA sequence d(TTAGGG)n. *Mol. Cell. Biol.* 13, 4301–4310.
- Kashima, T., and Manley, J.L. (2003). A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* 34, 460–463.
- Kent, O.A., Reay, A., Foong, L., Chilibeck, K.A., and MacMillan, A.M. (2003). Structuring of the 3' splice site by U2AF65. *J. Biol. Chem.* 278, 50572–50577.
- Keren, H., Lev-Maor, G., and Ast, G. (2010). Alternative splicing and evolution: diversification, exon definition and function. *Nat. Rev. Genet.* 11, 345–355.
- Krecic, A.M., and Swanson, M.S. (1999). hnRNP complexes: composition, structure, and function. *Curr. Opin. Cell Biol.* 11, 363–371.
- LaBranche, H., Dupuis, S., Ben-David, Y., Bani, M.R., Wellinger, R.J., and Chabot, B. (1998). Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat. Genet.* 19, 199–202.
- Lee, C.G., Zamore, P.D., Green, M.R., and Hurwitz, J. (1993). RNA annealing activity is intrinsically associated with U2AF. *J. Biol. Chem.* 268, 13472–13478.
- Liu, Z., Luyten, I., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A., and Sattler, M. (2001). Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 294, 1098–1102.
- Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcarcel, J., and Sattler, M. (2011). Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475, 408–411.
- Mayeda, A., and Krainer, A.R. (1992). Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68, 365–375.
- Mayeda, A., Munroe, S.H., Caceres, J.F., and Krainer, A.R. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* 13, 5483–5495.
- McAfee, J.G., Soltaninassab, S.R., Lindsay, M.E., and LeSturgeon, W.M. (1996). Proteins C1 and C2 of heterogeneous nuclear ribonucleoprotein complexes bind RNA in a highly cooperative fashion: support for their contiguous deposition on pre-mRNA during transcription. *Biochemistry* 35, 1212–1222.
- McKay, S.J., and Cooke, H. (1992). A protein which specifically binds to single stranded TTAGGGn repeats. *Nucleic Acids Res.* 20, 1387–1391.
- Merendino, L., Guth, S., Bilbao, D., Martinez, C., and Valcarcel, J. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 402, 838–841.
- Nelson, K.K., and Green, M.R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev.* 3, 1562–1571.
- Nilsen, T.W., and Graveley, B.R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463.
- 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.
- Pandolfo, M., Valentini, O., Biamonti, G., Morandi, C., and Riva, S. (1985). Single stranded DNA binding proteins derive from hnRNP proteins by proteolysis in mammalian cells. *Nucleic Acids Res.* 13, 6577–6590.
- Parker, R., Siliciano, P.G., and Guthrie, C. (1987). Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49, 229–239.
- Paronetto, M.P., Achsel, T., Massiello, A., Chalfant, C.E., and Sette, C. (2007). The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* 176, 929–939.
- Perriman, R., and Ares, M., Jr. (2010). Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing. *Mol. Cell* 38, 416–427.
- Reed, R., and Maniatis, T. (1986). A role for exon sequences and splice-site proximity in splice-site selection. *Cell* 46, 681–690.
- Riek, R., Wider, G., Pervushin, K., and Wuthrich, K. (1999). Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc. Natl. Acad. Sci. USA* 96, 4918–4923.
- Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 34, 93–158.
- Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003). Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol. Cell* 11, 965–976.

- Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* 12, 645–653.
- Singh, R., Valcarcel, J., and Green, M.R. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* 268, 1173–1176.
- Smith, C., Song, H., and You, L. (2008). Signal discrimination by differential regulation of protein stability in quorum sensing. *J. Mol. Biol.* 382, 1290–1297.
- Soares, L.M., Zanier, K., Mackereth, C., Sattler, M., and Valcarcel, J. (2006). Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* 312, 1961–1965.
- Sridharan, V., and Singh, R. (2007). A conditional role of U2AF in splicing of introns with unconventional polypyrimidine tracts. *Mol. Cell. Biol.* 27, 7334–7344.
- Sridharan, V., Heimiller, J., and Singh, R. (2010). Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF. *Mol. Cell. Biol.* 31, 652–661.
- Staley, J.P., and Woolford, J.L., Jr. (2009). Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr. Opin. Cell Biol.* 21, 109–118.
- Swanson, M.S., and Dreyfuss, G. (1988). RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J.* 7, 3519–3529.
- Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J., and Kjems, J. (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J.* 20, 5748–5758.
- Valcárcel, J. (1997). *Functional Analysis of Splicing Factors and Regulators* (New York: Academic Press).
- Valcárcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* 273, 1706–1709.
- Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* 136, 701–718.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, 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.
- Wu, J., and Manley, J.L. (1989). Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.* 3, 1553–1561.
- Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402, 832–835.
- Xu, Y.Z., and Query, C.C. (2007). Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly. *Mol. Cell* 28, 838–849.
- Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. USA* 86, 9243–9247.
- Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355, 609–614.
- Zhang, M.Q. (1998). Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* 7, 919–932.
- Zhang, M., Zamore, P.D., Carmo-Fonseca, M., Lamond, A.I., and Green, M.R. (1992). Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc. Natl. Acad. Sci. USA* 89, 8769–8773.
- Zhang, Q.S., Manche, L., Xu, R.M., and Krainer, A.R. (2006). hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* 12, 1116–1128.
- Zhuang, Y., and Weiner, A.M. (1989). A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev.* 3, 1545–1552.
- Zorio, D.A., and Blumenthal, T. (1999). Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 402, 835–838.