bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Multi-Factor Authentication of Potential 5' Splice Sites by the *Saccharomyces cerevisiae* U1 snRNP

AUTHORS: Sarah R. Hansen[1,2], Mark Scalf[3], Ivan R. Corrêa Jr[4], Lloyd M. Smith[3], and Aaron A. Hoskins[1,2,3]*

CORRESPONDING AUTHOR:

*Aaron A. Hoskins, ahoskins@wisc.edu

AFFILIATIONS:

[1] Department of Biochemistry, U. Wisconsin-Madison, Madison, WI 53706

[2] Integrated Program in Biochemistry, U. Wisconsin-Madison, Madison, WI 53706

[3] Department of Chemistry, U. Wisconsin-Madison, Madison, WI 53706

[4] New England Biolabs, Ipswich, MA 01938

MAJOR SUBJECT AREAS: Biochemistry, Biophysics

KEYWORDS: Splicing, Single Molecule Fluorescence, CoSMoS, Spliceosome, snRNP, RNA

RESEARCH ORGANISMS: *S. cerevisiae* (Baker's yeast)

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

In eukaryotes, splice sites define the introns of pre-mRNAs and must be recognized and excised with nucleotide precision by the spliceosome to make the correct mRNA product. In one of the earliest steps of spliceosome assembly, the U1 small nuclear ribonucleoprotein (snRNP) recognizes the 5' splice site (5' SS) through a combination of base pairing, protein-RNA contacts, and interactions with other splicing factors. Previous studies investigating the mechanisms of 5' SS recognition have largely been done *in vivo* or in cellular extracts where the U1/5' SS interaction is difficult to deconvolute from the effects of *trans*-acting factors or RNA structure. In this work we used co-localization single-molecule spectroscopy (CoSMoS) to elucidate the pathway of 5' SS selection by purified yeast U1 snRNP. We determined that U1 reversibly selects 5' SS in a sequence-dependent, two-step mechanism. A kinetic selection scheme enforces pairing at particular positions rather than overall duplex stability to achieve long-lived U1 binding. Our results provide a kinetic basis for how U1 may rapidly surveil nascent transcripts for 5' SS and preferentially accumulate at these sequences rather than on close cognates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

IMPACT STATEMENT

The yeast U1 snRNP recognizes multiple features of target RNAs to reversibly identify splicing-competent 5' splice sites.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

In eukaryotes, the introns of precursor messenger RNA (pre-mRNA) must be identified and removed with nucleotide precision by the spliceosome to produce mRNA (Wahl et al., 2009). The junction between an intron and the upstream exon is marked by the 5' splice site (5' SS) sequence, a motif that is essential for assembly of the spliceosome and both catalytic steps of splicing. Though the 5' SS is marked by a conserved consensus sequence (5'-GUAUGU in yeast, 5'-GURAG in humans) only the first two nucleotides are nearly invariant (99% GU, <1% GC) as they are necessary for catalysis (Fouser and Friesen, 1986; Konarska, 1998; Parker and Siliciano, 1993; Roca et al., 2013; Vijayraghavan et al., 1989; Wilkinson et al., 2017). The other positions are degenerate, especially in the human genome where more than 9,000 variants of the -3 to +6 region of the 5' SS are utilized (Carmel et al., 2004; Roca et al., 2013). Despite this degeneracy, the precise determination of exon-intron boundaries is essential to healthy cellular function. An estimated 50% of all disease-related point mutations alter splicing in some way, with 14% of all disease-related point mutations occurring at splice sites (Soemedi et al., 2017).

U1 small nuclear ribonucleoprotein complex (snRNP) is responsible for 5' SS selection during the earliest steps of spliceosome assembly (Lacadie and Rosbash, 2005; Rosbash and Séraphin, 1991; Ruby and Abelson, 1988). The 5' SS consensus sequence is complementary to the 5' end of U1 small nuclear RNA (Lerner et al., 1980). Since the first ten nucleotides of U1 snRNA (splice site recognition sequence, SSRS) are perfectly conserved between yeast and humans, this implies a conserved mechanism of 5' SS selection determined, in part, by base pairing (Rosbash and Séraphin, 1991). However, the degeneracy of certain positions within the 5' SS consensus shows that SSRS/5' SS duplexes often form with less than complete complementarity, and a subset of SSRS/5' SS interactions can even occur with noncanonical registers (Roca and Krainer, 2009). Additionally, there are many sequences which have a high degree of complementarity to U1 but are not utilized as splice sites (pseudo 5' SS) or only used

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

when nearby canonical 5' SS are inactivated (cryptic 5' SS) (Roca et al., 2013). Together these observations show that base pairing strength with the U1 SSRS alone cannot predict 5' SS usage. Since the 5' SS must be transferred from the U1 to the U6 snRNA for splicing to occur, spliced mRNA formation is a convolution of multiple 5' SS recognition events (Brow, 2002). As a result, it is difficult to determine how U1 SSRS/5' SS interactions change between different sequences based on analysis of mRNAs.

Structural biology of both *Saccharomyces cerevisiae* (yeast) and human U1 snRNP has revealed how snRNP proteins could play key roles in 5' SS recognition in addition to base pairing with the SSRS. In crystal structures of human U1 snRNP bound to a 5' SS-containing RNA oligonucleotide (oligo), the conserved U1-C protein (Yhc1 in yeast) contacts the SSRS/5' SS duplex in the minor groove at the pairing site between the nearly invariant 5' SS G(+1) and U(+2) nucleotides with the snRNA (Kondo et al., 2015; Krummel et al., 2009). Similarly, in cryo-EM structures containing yeast U1 snRNP, Yhc1 contacts the SSRS/5' SS duplex, also near G(+1), while a second yeast splicing factor, Luc7, contacts the snRNA strand opposite (Bai et al., 2018; Li et al., 2019; Plaschka et al., 2018). The proximity of Yhc1 and Luc7 to the SSRS/5' SS duplex are also consistent with genetic data supporting roles for these proteins in 5' SS recognition (Chen et al., 2001; Fortes et al., 1999; Schwer and Shuman, 2015, 2014). Filter-binding competition assays using a reconstituted human U1 snRNP showed that U1-C contributes to the affinity and specificity of U1 for 5' SS RNA oligos (Kondo et al., 2015). However, these assays are difficult to interpret with respect to a mechanism of 5' SS discrimination since it is unclear if equilibrium was reached during the experiment (Jarmoskaite et al., 2020), the assay was limited in its ability to directly detect interactions with non-consensus 5' SS, and it provided no information on how or if the kinetics of U1 interactions differed between different 5' SS RNAs. Thus, it is unkonwn if recognition of 5' SS originates from U1’s failure to bind mismatched RNAs or due to a selection event occurring after association.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Colocalization single molecule spectroscopy (CoSMoS) has previously been used to study the kinetics of both yeast and human U1/RNA interactions in cell extracts (Braun et al., 2018; Hoskins et al., 2011; Larson and Hoskins, 2017; Shcherbakova et al., 2013). In all cases, short (<60 s) and long-lived (>60 s), 5' SS-dependent interactions were observed between U1 and immobilized pre-mRNAs. In previous work from our laboratory with yeast U1 in whole cell extract (WCE), we showed how the populations of short- and long-lived interactions as well as their lifetimes can vary depending on the presence of a consensus or weak (containing additional mismatches) 5' SS or due to mutation of Yhc1 (Larson and Hoskins, 2017). These interactions were also strongly influenced by the presence or absence of *trans*-acting factors that bind elsewhere on the pre-mRNA, including the nuclear cap-binding complex (CBC) or the branch site bridging protein (BBP)/Mud2 complex, that together with U1 form the yeast E complex spliceosome or commitment complex (Larson and Hoskins, 2017; Séraphin and Rosbash, 1991, 1989). Our results were consistent with a two-step mechanism for 5' SS recognition by U1 that involves reversible formation of an initial weakly bound complex with RNA that can transition to a more stably bound state, as was proposed previously by others (Du et al., 2004; McGrail and O'Keefe, 2008). Yet, neither our prior experiments nor those from other laboratories could exclude roles for other, non-U1 splicing factors present in the WCE in this process or potential influence of pre-mRNA structure on the observed kinetics.

In this study, we use CoSMoS to directly observe how individual yeast U1 snRNP molecules interact with short RNA oligos. The short- and long-lived interactions observed in cell extracts with large pre-mRNA substrates are also observed when purified U1 snRNP binds cognate RNAs providing direct evidence for these kinetic features being inherent to 5' SS recognition. By using RNA oligos with varying base-pairing strength to the snRNA as well as with different locations and types of mismatches, we show that 5' SS recognition occurs

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

subsequent to binding. RNAs with limited pairing to the SSRS are released quickly after association while those with extended complementarity and pairing at certain positions are more likely to be retained and form long-lived complexes. Significantly, formation of long-lived U1/RNA complexes does not always correlate with the predicted thermodynamic stabilities of the SSRS/5' SS RNA duplexes. We propose that U1 uses a multi-step kinetic pathway to discriminate between RNAs and that formation of long-lived complexes is dependent on multiple factors that together favor U1 accumulation on introns competent for splicing.

RESULTS

U1 forms Short- and Long-Lived Complexes with RNAs Containing a 5' SS Sequence

Since we wished to study U1/5' SS interactions in the absence of *trans*-acting factors, we first developed a protocol for purifying fluorophore-labeled U1 snRNP from yeast extract. We genetically encoded a tandem-affinity purification (TAP) tag on the U1 protein Snu71 and a SNAP-tag on the U1 protein Snp1 in a protease-deficient, haploid yeast strain (**Fig. 1A**). TAP-tagged Snu71 has previously been used to purify U1 snRNP (Feltz and Krummel, 2016; Rigaut et al., 1999), and SNAP-tagged Snp1 has been used to fluorescently label and visualize U1 binding events by single molecule fluorescence in WCE (Hoskins et al., 2011; Larson and Hoskins, 2017). Extracts were prepared from the dual-tagged strain, and U1 snRNP purified using published protocols (Feltz and Krummel, 2016). Fluorophore labeling was carried concertedly with TEV protease cleavage of the TAP tag, and excess fluorophore was removed during calmodulin affinity purification. In these experiments, a tri-functional SNAP-tag ligand containing a Dy649 fluorophore, biotin, and benzyl-guanine leaving group (Smith et al., 2013) was used to simultaneously fluorophore label and biotinylate U1 on the Snp1 protein.

Purified U1 was characterized by mass spectrometry, and samples contained all known U1 components (**Fig. 1-Supplement 1**). Only a small number of peptides from other yeast splicing factors were identified, and these were not present in all replicates. Dideoxy sequencing

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of the isolated U1 confirmed the presence of the snRNA and SSRS, and the purified U1 was able to restore the splicing activity of WCE in which the endogenous U1 was degraded by targeted RNase H cleavage of the snRNA (Du and Rosbash, 2001) (**Fig. 1-Supplement 2**). Together the data support purification of functional U1 particles.

For substrates, we designed a set of Cy3-labeled, 29 nucleotide (nt)-long RNA oligonucleotides with varying degrees of complementarity to U1 (**Fig. 1-Supplemental Table 1**). The RNAs are based on the RP51A pre-mRNA 5' SS sequence, a well-studied splicing substrate (Hoskins et al., 2011; Larson and Hoskins, 2017; Rymond and Rosbash, 1985) and are identical except for substitutions within the 5' SS region. Importantly, the RNAs contain the entire region known to cross-link with the U1 snRNA (McGrail and O'Keefe, 2008) and all U1-interacting nt that could be modeled into cryo-EM densities of the spliceosome E and A complexes (the ACT1 intron stem loop observed in E complex being an exception) (Li et al., 2019; Plaschka et al., 2018). The RNAs contain all of the sites shown to cross-link with U1 snRNP proteins with the exception of non-conserved poly-U tracts located downstream of the 5' SS (+27-46) that likely interact with the RRM domains of Nam8 (Plaschka et al., 2018; Puig et al., 1999; Zhang and Rosbash, 1999). We omitted this region to avoid potential interferences from 5' SS-independent RRM/RNA interactions and folding of larger RNA substrates into structures that could compete with U1 interactions. The RNAs are predicted to have minimal stable secondary structure by mFold (Zuker, 2003) and range from limited complementarity with U1 (no more than two predicted contiguous base pairs; **Fig. 1-Supplemental Table 1**, RNA-Control or RNA-C) to a maximum of 10 contiguous potential base pairs (RNA-10).

We immobilized the purified U1 snRNP with streptavidin on a passivated and biotinylated glass slide (Salomon et al., 2015) and readily observed single spots of fluorescence from the Dy649 fluorophore upon excitation at 633 nm (**Fig. 1B**). When a 29-nt RNA oligo containing a consensus 5' SS and Cy3 fluorophore (RNA-4+2) was introduced, spots of Cy3 fluorescence

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

168 began to transiently appear on the surface (**Fig. 1B, C**). The spots of Cy3 fluorescence co-
169 localized with the immobilized U1 molecules, and spots repeatedly appeared and disappeared
170 from the same U1 molecule. This is consistent with multiple rounds of binding and release of the
171 RNA-4+2 oligo during the experiment. As a control, we added a Cy3-labeled oligo which lacked
172 any significant complementarity to U1 (**Fig. 1-Supplemental Table 1** RNA-C). We observed few
173 Cy3 signals on the surface, and the event density (frequency of co-localized binding events) of
174 RNA-C was 40-fold less than that of RNA-4+2 (**Fig. 1-Supplemental Fig. 3**). While it is possible
175 that non-specific interactions between RNA-C and U1 occurred too rapidly for us to detect, the
176 large differences in event density between RNA-C and RNA-4+2 indicate that the vast majority
177 of the detected binding events represent sequence specific interactions.

178 We analyzed the dwell time distribution of the RNA-4+2 binding events and were able to
179 fit the data to a function containing two exponential terms (**Fig. 1D**). This is consistent with the
180 appearance of both short- and long-lived binding events observed in the time trajectories of
181 single U1 molecules (**Fig. 1C**). The short-lived kinetic parameter (τ_1) was ~12 s with an
182 amplitude of 0.9, while the long-lived kinetic parameter (τ_2) was much larger (~204 s) but with a
183 smaller amplitude (0.1). Previous analysis of U1 binding events on immobilized RP51A pre-
184 mRNAs in yeast WCE also resulted in multi-exponential dwell time distributions (Hoskins et al.,
185 2011; Larson and Hoskins, 2017). The exponential fits of dwell times for RNA-4+2 binding to
186 purified, immobilized U1 snRNP and for U1 snRNP binding to immobilized RP51A pre-mRNAs
187 containing the same 5' SS have similar parameters (Larson and Hoskins, 2017). In both cases,
188 the majority of binding events are short-lived and with lifetimes of ~12 s. In WCE without ATP,
189 the long-lived kinetic parameter was smaller (64 vs. 204 s) but with a larger amplitude (0.3 vs.
190 0.1). Longer binding events of ~200 s were observed in WCE with this 5' SS but only when
191 either CBC or BBP were also capable of binding the pre-mRNA.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Together, our data indicate that short- and long-lived interactions with RNA substrates are an inherent property of U1. Since we purified and immobilized U1 and studied its interactions with small RNAs, the diversity of binding events cannot solely originate from the influence of *trans*-acting factors present in a WCE or folding/unfolding of large RNA substrates. We do not exactly know how these factors influence U1 binding in complex environments, but they may be the origins of differences we observed between experiments carried out with purified U1 and with U1 present in WCE.

Base Pairing Potential Accelerates U1/RNA Complex Formation

We next systematically studied how the base-pairing potential of the RNA oligo influenced binding by U1 snRNP. We carried out single molecule binding assays with RNAs capable of forming between 4 and 10 contiguous base pairs with the snRNA (**Fig. 2A**). All of these substrates can form base pairs at the highly conserved G+1 and U+2 positions of the 5' SS, and we extended base pairing outwards from these positions towards the 5' and 3' ends of the SSRS. For several positions we also varied the duplex position with pairing extending away from or towards the 5' end of the U1 snRNA without altering the number of potential base pairs (RNA-6a vs. -6b, for example).

When the number of binding interactions to immobilized U1 and the apparent association rates were measured, the RNA oligos exhibited two distinct classes of behavior. In the first class, RNA oligos capable of forming <6 contiguous base pairs showed very few colocalized binding events with U1 (**Fig. 2B**). While these oligos may have been able to form sequence-specific interactions with U1, these interactions were either too rapid or infrequent for us to observe. The few measurable events were essentially indistinguishable in frequency to background binding of RNA-C. RNAs capable of forming ≥6 contiguous base pairs exhibited a second class of behavior. These RNAs had a 100-fold increase in detectable U1 binding event density compared to RNAs in the first class (**Fig. 2B**). The dependence of the event density on

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the number of potential base pairs with the snRNA supports that the interactions are not only sequence dependent (**Fig. 1**) but are also due to interactions with the U1 SSRS.

For RNAs with detectable U1 binding, we were able to calculate the observed association rate ($k_{association}$) to U1 under these conditions (**Fig. 2C, Fig. 2-Supplemental Table 1**). RNAs capable of forming more potential base pairs with U1 bound more quickly. The correlation of the association rate with extent of base pairing could be due to RNAs with greater complementarity also having a greater probability of nucleating duplex formation due to the increased number of possible toeholds or short stretches of pairing interactions. This hypothesis is consistent with previous single molecule fluorescence resonance energy transfer studies and ensemble measurements of DNA and RNA oligo hybridization that show nucleation of nucleic acid duplex formation by base pairing interactions of only 2-4 nt in length (Cisse et al., 2012; Craig et al., 1971; Marimuthu and Chakrabarti, 2014; Wetmur, 1991; Wetmur and Davidson, 1968).

Additionally, we observed that oligos capable of pairing towards the 3' end of the SSRS formed observable complexes more quickly than those where the pairing was shifted towards the 5' end (**Fig. 2C**, RNAs 6a, 7a, and 8a vs. 6b, 7b, 8b). This indicates that the 3' end of the SSRS might be either more accessible to the RNAs or can more easily facilitate nucleation of RNA interactions that lead to the observable binding events. This latter possibility may be related to the increased calculated thermodynamic stability of duplexes with pairing interactions closer to the 3' end of the SSRS due to the presence of a G/C pair in this region: RNAs 6a and 7a are predicted to form more stable duplexes than RNAs 6b and 7b (**Fig. 2A**).

The Abundance of Short- and Long-Lived U1/RNA Complexes Depends on Base Pairing

We next studied the dwell times with U1 for the same series of RNA oligos. By visually inspecting the individual fluorescence time trajectories, we were immediately struck by apparent differences in binding behaviors. We observed predominantly very long dwell times with RNAs

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

241 capable of forming a large number of base pairs with U1 (**Fig. 3A, RNA-10**), very short dwell
242 times with RNAs capable of only forming a small number of base pairs (**RNA-6a**), and a mixture
243 of short and long dwell times for RNAs capable of forming an intermediate number of base pairs
244 (**RNA8a**). When the individual dwell times from each experiment were combined and analyzed
245 using probability density histograms and fits to exponential equations, the resulting plots and
246 kinetic parameters confirmed these observations (**Fig. 3B**). RNA-10 is capable of forming 10
247 base pairs with U1 and the distribution of dwell times for RNA-10 on U1 is best fit using a single
248 exponential term ($\tau \approx 178$ s, **Figure 3-Supplemental Table 1**), consistent with long-lived binding.
249 RNA-6a is capable of only forming 6 base pairs with U1 and its distribution of dwell times could
250 also be fit using only a single exponential ($\tau \approx 12$ s), consistent with short lived binding. RNA-8a
251 can make up to 8 base pairs with U1 and its dwell times were best fit using an equation with two
252 kinetic parameters describing short- ($\tau \approx 41$ s) and long-lived binding events ($\tau \approx 133$ s).

253 When we examined all of the RNAs in this series (**Fig. 2A**), we observed a trend: RNAs
254 capable of forming few base pairs with U1 exhibited predominantly short-lived dwell times
255 (defined here as $\tau_S < 60$ s) with a small fraction of long-lived ($\tau_L > 60$ s) binding events and
256 correspondingly small amplitude for the long-lived kinetic parameter (**Fig. 3C**). As the number
257 of potential base pairs increased, so did the amplitude of τ_L. It is unlikely that these results arose
258 from presence of two subpopulations of U1 snRNPs in our experiments (one capable of only
259 making short-lived interactions and one capable of only making long-lived interactions) since we
260 would not expect these subpopulations to change in abundance between experiments carried
261 out with the same preparations of U1.

262 Instead, these data are most consistent with a simpler mechanism in which U1
263 association with RNAs involves two steps and formation of short-lived intermediate (**Fig. 3F**). All
264 RNAs that we can observe interacting with U1 (those capable of making $\geq$ 6 base pairs) are

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

capable of forming the short-lived complex. RNAs with a limited number of base pairs (i.e., RNAs-6a, 6b) rarely progress through the second step to form the long-lived complex and most often dissociate from the intermediate state. On the other hand, RNAs with a large number of base pairs (RNAs-9a, 10) transition to the long-lived complex much more frequently than they dissociate from the intermediate. RNAs with in-between numbers of base pairs (e.g., RNAs-8a, 8b) face a competition between dissociation from the intermediate and formation of the long-lived complex.

Finally, it is interesting to note that RNAs in which the base pairing extends to the 3' end of the SSRS (**Fig. 3C**, RNAs 8a and 9a) also had higher fitted amplitudes for the long-lived state than those capable of forming the same number of base pairs but not reaching the 3' end of the SSRS (RNAs 8b and 9b). This suggests that pairing within the 3'-most nt of the SSRS closest to the zinc finger of Yhc1 is not only important for increasing the rate of U1 binding but also contributes to formation of the longest-lived U1/RNA complexes. Combined, these results support formation of a short-lived, intermediate between U1 and RNAs that is dependent on base pairing for its formation. The RNA can then dissociate from this intermediate or the U1/RNA complex can transition to more tightly bound state.

Some U1/5' SS Duplexes are Destabilized in the U1 snRNP

In addition to the amplitudes of the short- and long-lived parameters, we also determined their values from the fits to the dwell time data (**Fig. 3D, E**). The short-lived dwell time parameter (τ_S) ranged from 12 – 48s for RNA oligos capable of forming 6 – 9 contiguous, potential base pairs. The long-lived dwell time parameter (τ_L) ranged from 133 – 288s for RNA oligos capable of forming 7 – 10 base pairs. For both parameters, RNAs capable of forming more base pairs also tended to have longer dwell times. However, this relationship did not hold true in all cases: RNA-9b and RNA-10 have similar values of τ_L despite the presence of an additional potential

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

289 base pair. Both the t_S and t_L parameters only varied within a range of 2- to 4-fold, which was
290 surprising since a previous single molecule fluorescence study of RNA oligo hybridization
291 reported a 10-fold decrease in off-rate due to presence of one additional base pair (Cisse et al.,
292 2012).

293 It is possible that the particular properties of the RNA oligos we used in these studies
294 contributed to the small range in τ_S and τ_L we determined. To test this, we constructed a RNA-
295 only mimic of the U1 SSRS (**Fig. 4A**). In this case, binding kinetics would only be influenced by
296 the nucleic acid complexes being formed and not be influenced snRNP proteins or structural
297 constraints imposed by U1. Unlike U1 snRNP, the surface-immobilized mimic did not efficiently
298 bind to the RNA oligos when they were present in solution at nM concentrations (the upper
299 concentration limit of our single-molecule assay). So, we instead pre-annealed each oligo to the
300 mimic and then measured its off-rate by monitoring disappearance of colocalized oligo
301 fluorescence signals over time (**Figs. 4A,B**).

302 For each of the RNAs, we were able to fit the dissociation data to equations containing a
303 single exponential term (**Fig. 4-Supplemental Table 1**). This signifies that the RNAs are
304 dissociating in a single observable step from the immobilized mimic and that dissociation was
305 occurring from only a single type of RNA/mimic complex. The single exponential kinetics are in
306 contrast with results obtained for many of the same RNAs with U1 snRNP, for which multi-
307 exponential kinetic equations were required to fit the dwell time data (RNA-7a,b; 8a,b; and 9a).
308 This was true for both a mimic that, like U1, contains pseudouridines in the SSRS as well as for
309 one with uridine substitutions at those positions. Thus, the dissociation pathways for a given
310 RNA oligo are not identical between the RNA-only mimic and the U1 snRNP under these two
311 experimental conditions.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In addition to differences in the dissociation pathways, the amount of time the oligos remained bound differed dramatically between the RNA mimic and U1. Dissociation rates from the mimic varied linearly with base pairing potential over 20-fold, a larger range than for binding of the same oligos to U1 snRNP (**Fig. 4C**). Surprisingly, the lifetimes of many of the RNAs bound to the mimic were also much longer than their lifetimes bound to U1. For example, RNA-10 had a dissociation rate of 5.5×10^{-4} s^{-1} when bound to the pseudouridine-containing mimic. This corresponds to an average lifetime of 1,818 s—more than 10-fold larger than the τ_L obtained for binding of the same RNA to U1. From these observations we conclude that some U1/5' SS duplexes can be destabilized in the context of the U1 snRNP. Thus, the lifetimes of U1/5' SS interactions in the snRNP cannot be predicted from base-pairing potential or studies of model RNA duplexes alone.

Long-Lived U1/RNA Interactions Are Sensitive to the Location and Type of Mismatches

Splice sites with perfect and uninterrupted complementarity to U1 are very rare in yeast. In fact, only 14 annotated 5' SS in yeast contain 6 contiguous base pairs (corresponding to RNA-6b) and only one (a cryptic 5' SS in RPL18A) may contain more than 7 contiguous base pairs (Grate and Ares, 2002). Most 5' SS are interrupted by one or more mismatches in complementarity to the U1 snRNA. We next tested how these mismatches impacted interactions of the RNA oligos with U1 snRNP. We analyzed and compared the binding interactions of RNAs capable of forming various numbers of contiguous base pairs between U1 SSRS nt +3 to +9. We incorporated mismatches systematically at each position resulting in RNAs that can form uninterrupted duplexes of 7 or 6 base pairs (RNAs-7a, 6a, 6b) or interrupted duplexes of a total length of 7 nucleotides (**Fig 5A**). One of the RNA oligos within this comparison group contains the U1 consensus 5' SS found within the well-spliced RP51A transcript (**RNA-4+2, Fig. 5A**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

335 Within this group, the mismatches result in a range of predicted duplex stabilities from -0.4 – 9.1
336 kcal/mol (**Fig. 5**).

337 We observed long-lived complexes for RNA-7a, which has a 7 bp predicted duplex length,
338 and only short-lived complexes for RNA-6a and -6b, which have only 6 bp predicted duplex
339 lengths (**Fig. 5B**; replotted from **Fig. 3D, E**). Whether or not RNAs containing mismatches that
340 disrupt the duplex with the SSRS showed long-lived interactions (like RNA-7a) or only short-
341 lived interactions (like RNA-6a, 6b) depended the position of the mismatch. Neither RNA oligos
342 containing a C/C mismatch at the +1 site nor an A/A mismatch at the +2 site were able form
343 long-lived complexes with U1. However, RNA oligos containing U/U mismatches at +3 or +4 or
344 a C/C mismatch at +5 could form long lived complexes (**Fig. 5B**). From these data we conclude
345 that long-lived complex formation is sensitive to the tested mismatches at some positions (+1,
346 +2) and not others (+3, +4) within a substrate of 7bp end-to-end length. In addition, the same
347 type of mismatch (C/C) could either prevent or permit long-lived complex formation depending
348 on its position within the 7bp duplex. As a consequence, formation of long-lived U1/5' SS
349 interactions does not correlate with predicted duplex stabilities (*cf.*, RNA-5+1 vs. RNA-2+4, 6a,
350 or 6b in **Fig. 5B**).

351 **Long-Lived U1/RNA Interactions Depend on Base Pairing at the G+1 Position of the 5' SS**

352 We next tested if a single mismatch could eliminate long-lived binding even if all other
353 positions within the 5' SS oligo could potentially pair with SSRS. We incorporated single
354 mismatches at the +1 position of RNA-10 (**Fig. 6A**). This results in a mismatch at the first
355 position of the highly conserved 5' SS GU. All RNAs containing a mismatch at +1 were able to
356 associate with U1 at rates ~100-fold greater than background binding by RNA-C (**Fig. 6B**).
357 However, none of them were able to form appreciable amounts of the long-lived complex (**Fig.**
358 **6C**). The observed distributions of dwell times for RNAs containing mismatches at +1 could still

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

359 be best fit to two exponential distributions containing short- and long-lived parameters (**Fig. 6-**
360 **Supplemental Table 1**). However, the amplitudes of the long-lived parameters were very small
361 as expected from the scarcity of the long-lived events. Consistent with data shown in **Fig. 5**, the
362 predicted thermodynamic stabilities again did not correlate with observation of the long-lived
363 complexes. For example, RNA-2+7 (A+1) containing a A/C mismatch at the +1 position is
364 predicted to form a more stable duplex than RNA 5+1 (ΔG^o -4.4 vs. -2.7 kcal/mol). Yet, the
365 amplitude of the long-lived parameter for RNA 5+1 is ~14x greater than that for RNA-2+7 (A+1).
366 These results show that long-lived complex formation between U1 and the RNA oligos is
367 intolerant of mismatches at the +1 position. Failure of U1 to accumulate on RNAs with
368 mismatches at the +1 site is not due to lack of association. Rather, recognition of a mismatch at
369 +1 involves a discrimination step occurring after association and mismatched RNAs are rapidly
370 released.

371 **Discussion**

372 By studying single molecules of yeast U1 snRNPs interacting with a diverse range of RNA
373 oligos, our experiments have revealed the dynamics associated with the earliest step of 5' SS
374 recognition. U1 can form both short and long-lived, sequence-dependent complexes with RNAs
375 (**Figs. 1, 3**). RNA binding is accelerated by increased numbers of potential base pairs as well as
376 by their positioning closer to the 3' end of the SSRS—the same region in which the Yhc1 and
377 Luc7 proteins contact the 5' SS/SSRS duplex (Kondo et al., 2015; Krummel et al., 2009; Li et
378 al., 2017; Plaschka et al., 2018) (**Fig. 2**). Sequence-dependent interactions with lifetimes of
379 several seconds are only observed with oligos capable of forming duplexes of at least 6 bp in
380 length (**Fig. 2**) while additional potential base-pairing increases the probability of forming long-
381 lived interactions lasting several minutes (**Fig. 3**). Relative to an RNA-only mimic, U1 snRNP
382 binds freely diffusing oligos more readily when they are present at nM concentrations and

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

accelerates the release of RNAs capable of forming the largest number of potential base pairs (**Figs. 3, 4**), indicating that snRNP proteins and/or U1 snRNA structure can destabilize the SSRS/5' SS duplex. Formation of long-lived interactions is dependent on the position of mismatches as well as pairing at the G+1 position rather than predicted thermodynamic stabilities of the RNA duplexes (**Fig. 5, 6**).

These results support a reversible multi-step binding process in which U1 first forms sequence-dependent short-lived interactions with RNAs via the SSRS and then transitions to a more stably bound complex if certain requirements are met (**Fig. 3F**). This agrees with previous single molecule experiments with U1 in WCE (Larson and Hoskins, 2017) and biochemical studies of the temperature dependence of yeast U1/RNA interactions (Du et al., 2004) while providing further details of the 5' SS recognition process in the absence of *trans*-acting factors or confounding effects due to RNA secondary structures. We have not yet identified what specific event or conformational change is associated with long-lived complex formation; however, comparison of cryo-EM densities for free yeast U1 and the U1/U2-containing yeast pre-spliceosome (in which U1 is bound to a 5' SS) reveals that Yhc1 and Luc7 are more ordered in the latter (Li et al., 2017; Plaschka et al., 2018). It is possible that a disorder-to-ordered conformational change for Yhc1 and Luc7 results in the short- to long-lived transition inferred from our single molecule assays. Mismatches at particular positions such as G+1 may inhibit this transition and/or prevent formation of stable contacts between the two proteins or with the RNA. In agreement with the importance of potential Yhc1/Luc7 contacts, mutants of Luc7 located at the Yhc1 interface exhibit numerous genetic interactions with U1 snRNP proteins and splicing factors involved in E complex formation (Agarwal et al., 2016; Li et al., 2019). We also note that interactions between U1 proteins and RNA sequences or structures not included in these

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

406 experiments, such as RNA hairpins or binding sites for Nam8 (Li et al., 2019; Plaschka et al.,
407 2018), may influence this process to further tune U1 binding on specific transcripts.

408 **Checkpoints during 5' SS Recognition**

409 Combined, our data also indicate that binding involves several checkpoints before long-
410 lived complexes are formed (**Fig. 7**). By analogy, 5' SS recognition is similar to “multi-factor
411 authentication” processes that are commonly used to verify identities on-line. The initial barrier
412 to forming short-lived complexes is low and requires limited complementarity to the SSRS.
413 Formation of this complex is readily reversible which permits rapid surveillance of transcripts by
414 U1 for 5' SS and prevents accumulation of U1 on RNAs lacking features necessary for splicing.

415 Passage to the long-lived complex is more stringent and dependent on a G nucleotide at
416 the +1 position as well as increased complementarity. It is likely that the need for pairing with
417 G+1 evolved in U1 due to the importance of this nucleotide in splicing chemistry since G+1 at
418 the 5' SS must form a non-Watson Crick pair with G-1 at the 3' SS during exon ligation (Parker
419 and Siliciano, 1993; Plaschka et al., 2019; Yan et al., 2019). Thus, U1 binding kinetics prioritize
420 identification of a nucleotide used for splicing chemistry even though U1 itself does not
421 participate in those steps.

422 Interestingly, we found that formation of the long-lived state often correlated with
423 predicted duplex length more so than predicted duplex stability (**Fig. 5**, RNA-5+1 vs. RNA 6a or
424 6b, for example). This suggests 5' SS recognition involves a “molecular ruler” similar to those
425 observed in other RNPs (Kwon et al., 2016; MacRae et al., 2006). From our data, we would
426 predict that U1 uses this ruler activity to preferentially form long-lived interactions with RNA
427 duplexes $\geq$ 7 bp in end-to-end length. In terms of U1 snRNP structure, duplexes of 6-7 bp in
428 length are needed to span the distance between the zinc finger of Yhc1 to the second zinc finger

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Luc7 and the Yhc1/Luc7 interface (Li et al., 2019)—suggesting a role for duplex length in conformational changes or intermolecular contacts associated with stable 5' SS binding.

A duplex length requirement of at least 7 bp is surprising since the consensus sequence of the 5' SS indicates only 6 highly conserved positions (**Fig. 7B**). How then would most natural 5' SS in yeast be able to stably bind U1? When predicted end-to-end duplex lengths between the SSRS and yeast 5' SS are analyzed, rather than their specific nucleotide sequences, it is apparent that the over-whelming majority of 5' SS are capable of forming extended duplexes with U1 with the caveat that these duplexes may contain one or more internal mismatches (**Fig. 7C**). Consequently, we would predict that most yeast 5' SS are able to meet duplex length requirements for long-lived complex formation.

It is likely that non-U1 splicing factors can permit bypass of the molecular ruler to facilitate accumulation of U1 on weak SS. As an example of bypass, we previously observed that pre-mRNAs containing a weak 5' SS with only a 5 bp end-to-end length can form long-lived complexes with U1 in WCE but only when either the CBC or BBP/Mud2 could also bind the pre-mRNA (Larson and Hoskins, 2017). One limitation of our model is that we do not yet know if these *trans*-acting factors can also bypass the need for pairing with G+1 in the assays used here. Regardless, it is notable that only the branch site-binding factors, BBP/Mud2, increased the probability of long-lived complex formation and not just its lifetime (Larson and Hoskins, 2017). This is consistent with the idea that retention of U1 on transcripts is governed through direct and indirect interactions with intronic sequences involved in splicing chemistry including both the 5' SS and branch site.

Base Pairing Potential Does Not Predict U1 Interaction Kinetics

Our data show that the lifetime of the U1/5' SS interaction cannot be predicted based on the thermodynamic stability of the base pairing interactions alone. Rather, strong positional

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

effects lead to prioritization of base pairing at particular positions for stable retention of U1 (**Figs. 5, 6**). How kinetic stability of U1 at a particular 5' SS correlates with its subsequent use by the spliceosome has not yet been determined. If 5' SS usage does correlate with U1 lifetimes, then our results would be in strongest agreement with computational models for 5' SS identification that take into account positional effects and interdependencies rather than simply the thermodynamic stabilities of the predicted base pairs (Roca et al., 2013; Yeo and Burge, 2004).

Since failure to release U1 can inhibit splicing (Staley and Guthrie, 1999), we would expect that optimal promotion of splicing by yeast U1 would occur by balancing its recruitment and retention with release. This “Goldilocks” model has previously been proposed for 5' SS recognition by human U1 (Chiou et al., 2013). Interestingly, we observed that the lifetimes of the longest-lived U1/RNA complexes were similar to one another (**Fig. 3, 5**) and lifetimes of the RNAs capable of making 9 or 10 base pairs were much shorter when bound to U1 than expected based on their off-rates from a RNA mimic of the SSRS (**Fig. 3, 4**). U1 may equilibrate its interactions with 5' SS by both stabilizing and destabilizing RNA duplexes. As a result, U1’s interactions with sequence-diverse substrates may all be “just right” for subsequent steps in splicing. This, in turn, may impact the ATP requirement for U1 release during activation (Staley and Guthrie, 1999).

Implications for Human 5' SS Recognition

While much of the catalytic machinery of the spliceosome is well-conserved between yeast and humans, we do not yet know if the mechanism of 5' SS recognition we propose also holds true for human U1. Chemical probing data of human U1 revealed allosteric modulation of the SSRS based on positioning of splicing regulatory elements (Shenasa et al., 2020). Differing conformations of the SSRS could lead to differences in binding behavior and give rise to short- or long-lived binding interactions similar to those we observe with yeast U1. Whether or not the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

477 SSRS of yeast U1 displays similar changes in conformation has not been determined. Ensemble
478 binding data for human U1/RNA interactions also revealed a strong preference for formation of
479 stable complexes on pairing at the G +1 site as well as position- and mismatch-dependent effects
480 at other locations (Kondo et al., 2015; Tatei et al., 1987). This is in agreement with our single
481 molecule data on yeast U1 interactions. While it is possible that yeast and human U1 may have
482 differing pathways for splice site recognition, the outcome of longer-lived binding on particular
483 RNA sequences may be the same.

484 Finally, the large number of non-obligate accessory factors that associate with human U1
485 may yield highly malleable pathways for RNA binding. Different factors may tune 5' SS
486 recognition by a holo-U1 complex to yield distinct kinetic mechanisms. This in turn could lead to
487 enhancement or repression of U1 accumulation at particular sites or functional differences
488 between U1 complexes involved in splicing or telescripting (Kaida et al., 2010; Oh et al., 2017).
489 The mechanism we propose for yeast U1 may be most relevant for the subset of human U1
490 snRNPs that associate with alternative splicing factors such as LUC7L or PRPF39 that are
491 homologs of obligate components of the yeast U1 snRNP (Li et al., 2017).

492 **General Features of Nucleic Acid Recognition by RNPs**

493 Other cellular RNPs face similar challenges as U1 in finding specific nucleotide
494 sequences and preventing accumulation on close cognates. While Cas9 (involved in bacterial
495 CRISPR-based immunity), Hfq (involved in bacterial small RNA regulation) and Argonaute
496 (AGO, involved in mRNA repression and silencing) RNPs are involved in very different biological
497 processes than RNA splicing, single molecule studies of each of these RNPs reveal striking
498 similarities with yeast U1 (Globyte et al., 2019; Małecka and Woodson, 2021; Salomon et al.,
499 2015; Sternberg et al., 2014). All of these RNPs exhibit kinetic behaviors that lead to prioritization
500 of certain sequences over others and are distinct from “all-or-nothing” models for hybridization

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of nucleic acids in the absence of proteins (Cisse et al., 2012; Wetmur, 1991; Wetmur and Davidson, 1968). In the cases of AGO and Cas9, correct base pairing with the micro-RNA seed sequence (AGO) or PAM (Cas9) is necessary for fast and stable binding. Rapid reversibility of this interaction ensures that these RNPs can dissociate and find other targets if mismatches are detected within the priority region. Additional base-pairing with the target then leads to the most stable binding interaction, consistent with binding occurring in multiple steps. These results are analogous to reversible interrogation of RNAs by U1 that prioritizes pairing at the G+1 site and formation of extended duplexes for stable interaction. Cas9 also accelerates target search by diffusion along DNA molecules (Globyte et al., 2019; Sternberg et al., 2014). While this has not been directly tested with U1, tethering of U1 to the pol II transcription machinery (Kotovic et al., 2003; Zhang et al., 2021), rather than RNAs themselves, may lead to similar acceleration in binding site identification.

Funding

This work was supported by the National Institutes of Health (R01 GM 122735 and R35 GM136261 to AAH and R35 GM126914 to LMS and MS). SRH was supported in part by the NIH Chemistry-Biology Interface Training Grant (T32 GM008505).

Competing Interests

AAH is conducting sponsored research with and a scientific advisor for Remix Therapeutics, Inc.

Acknowledgements

We thank Boehm, Gehring, et al. whose work on “two-factor authentication” of nonsense-mediated mRNA decay inspired our multi-factor authentication model (https://www.biorxiv.org/content/10.1101/2020.07.07.191437v1). We thank David Brow, Sam Butcher, Joshua Larson, Margaret Rodgers, and Tucker Carrocci for critical reading of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

manuscript. We thank Clarisse van der Feltz and Daniel Pomeranz Krummel for assistance in U1 snRNP purification.

Methods

Key resources and materials

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
RNasin® Ribonuclease Inhibitor	Promega	Cat No. N2611
Pierce Protease Inhibitor Tablet	ThermoFisher Scientific	Cat No. A32965
TEV Protease	Sigma Aldrich	Cat No. T4455
BG-649-PEG-biotin	(Smith et al., 2013)	
m7G(5')ppp(5')G RNA Cap Structure Analog	New England BioLabs	Cat No. S1404S
AMV Reverse Transcriptase	Promega	Cat No. M5101
RNaseH (2U/µL)	ThermoFisher Scientific	Cat No. 18021014
Vectabond	Fisher Scientific	Cat No. NC9280699
Biotin-PEG-SVA (MW 5,000)	Laysan Bio	Cat No. Biotin-PEG-SVA-5000-100mg
mPEG-SVA (MW 5,000)	Laysan Bio	Cat No. mPEG-SVA-5000-1G
Poly-L-lysine	Sigma Aldrich	Cat No. P7890
Glucose Oxidase from Aspergillus niger type VII	Sigma Aldrich	Cat No. G2133-50KU
Catalase from bovine liver	Sigma Aldrich	Cat No. C40-100MG
(±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox)	Sigma Aldrich	Cat No. 238813-1G
TransFluoSpheres® Streptavidin-Labeled Microspheres (488/645), 0.04µm, 0.5% solids	Life Technologies/ Invitrogen	Cat No. T-10711
Yeast tRNA (10 mg/mL)	ThermoFisher Scientific	Cat No. AM7119
Streptavidin, 10mg	Prozyme	Cat No. SA10-10mg
Heparin sodium salt from porcine intestinal mucosa	Sigma Aldrich	H4784-250MG
Commercial Materials and Kits		
Ultra-clear centrifuge tubes (14mL capacity)	Beckman Coulter	Cat. No. 344060
Criterion TGX Precast Gel (4-20%)	Bio-Rad	Cat No. 567-1093
Precision Plus Protein™ All Blue Prestained Protein Standards	Bio-Rad	Cat No. 161-0373
Silver Stain Plus™ Kit	Bio-Rad	Cat No. 161-0449
GE Healthcare IgG Sepharose 6 Fast Flow resin	VWR Scientific	Cat No. 95017-050
0.8x4cm Poly-Prep Chromatography Columns	Bio-Rad	Cat No. 731-1550
Calmodulin Affinity Resin	Agilent	Cat No. 214303

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

10kDa MWCO Slide-A-Lyzer dialysis cassette	ThermoFisher Scientific	Cat No. 66380
Amicon Ultra 100kDa MWCO centrifugal filters	Sigma Aldrich	Cat No. Z677906-24
Gold Seal Cover Slips (#1, 24x60mm)	Fisher Scientific	Cat No. 5031132
Gold Seal Cover Slips (#1, 25x25mm)	ThermoFisher Scientific	Cat No. 3307
Fisherbrand Five-Slide Mailer	Fisher Scientific	Cat No. HS15986
Experimental Models: Organisms/Strains		
BJ2168 *S. cerevisiae* (MATa prc1–407 prb1–1122 pep4–3 leu2 trp1 ura3–52 gal2)	Bruce Goode Lab (Crawford et al., 2008)	yAAH0001
U1-SNAP-TAP *S. cerevisiae* (BJ2168 + *SNP1*::*SNP1-fSNAP-Hyg* + *SNU71*::*SNU71-TAP-URA*)	This study	yAAH0393
Oligonucleotides		
U1 cOligo (DNA): 5' - CTT AAG GTA AGT AT	Integrated DNA Technologies	JL-U1 5' complement
U1 RT Oligo (DNA): 5' - TCA GTA GGA CTT CTT GAT	Integrated DNA Technologies	SRH15
U1 snRNA mimic (UU, RNA): 5' - AUA CUU ACC UUA AGA UAU CAG AGG AGA UCA AGA AG /3Cy5Sp/	Integrated DNA Technologies	SRH21
U1 snRNA mimic (ΨΨ, RNA): 5' - AUA CΨΨ ACC UUA AGA UAU CAG AGG AGA UCA AGA AG /3Cy5Sp/	Integrated DNA Technologies	SRH36
Handle for U1 mimic (DNA): 5' - /Biotin/ TCT CTT CTT GAT CTC CTC TGA TAT CTT A	Integrated DNA Technologies	SRH22
RNA-Cy3 oligomers are listed separately in ***Figure 1-Supplemental Table 1****. All oligonucleotides were purchased from Integrated DNA Technologies*		
Recombinant DNA		
Plasmid for *in vitro* transcription of RP51A (pBS117)	Michael Rosbash Lab (Seraphin and Rosbash, 1991)	pAAH0016
Equipment and Instruments		
Typhoon FLA 9000	GE Healthcare Life Sciences	https://www.gelifesciences.com
ImageQuant LAS 4000	GE Healthcare Life Sciences	https://www.gelifesciences.com
Retsch Mixer Mill MM 400	Retsch	https://www.retsch.com/
Software and Algorithms		
ImageQuant TL 8.1 software	GE Healthcare Life Sciences	https://www.gelifesciences.com
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
ChemDraw Prime 15.0	PerkinElmer	http://www.cambridgesoft.com/

TAP Tagging of Yeast U1 snRNP

C-terminal TAP and fSNAP tags were appended to the endogenous SNU71 and SNP1 proteins, respectively, by homologous recombination in the protease-deficient *S. cerevisiae*

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

strain BJ2168 and selection for growth in the absence of uracil (TAP) or in the presence of hygromycin (fSNAP) (Larson and Hoskins, 2017; Puig et al., 2001).

Purification of labeled U1 snRNP

A total of 10 L U1-SNAP-TAP yeast cultures were grown in 1 L batches of rich media (YPD) in a shaking incubator (30°C, 220 rpm) to late log stage. The cells were pelleted, washed, and resuspended in 3.5 mL (per 1 L culture) Lysis Buffer (10 mM Tris-Cl pH 8.0, 300 mM NaCl, 10 mM KCl, 0.2 mM EDTA, 5 mM imidazole, 10% v/v glycerol, 0.1% v/v NP40, 1 mM PMSF, 0.5 mM DTT). The resuspended cells were flash frozen in a drop-wise fashion in liquid nitrogen and stored at -80°C until lysed. The frozen pellets were lysed in batches using a Retsch Mixer Mill MM 400 (five rounds of 3 min at 10 Hz, with 2 min cooling in liquid nitrogen between rounds). The frozen lysate powder was stored at -80°C.

The total cell lysate from 10 L was thawed at 4°C. Lysis Buffer (10 mL) was used to dissolve one EDTA-free Protease Inhibitor Tablet (Pierce), and this solution was combined with the cell lysate. Insoluble material was removed by centrifugation (15,000 rpm, 4°C, 30 min). The supernatant was then cleared in an ultracentrifuge at 36,000 rpm, 4°C, for 75 min. The resulting middle layer was carefully removed and added to 300 µL GE Healthcare IgG Sepharose 6 Fast Flow resin that had been equilibrated with IgG150 Buffer (10 mM Tris-Cl pH 8.0, 150 mM NaCl, 10 mM KCl, 1 mM $MgCl_2$, 5 mM imidazole, 0.1% v/v NP40, no reducing agent) to incubate at 4°C with rotation for 2 hr.

The resin slurry was divided between two 0.8x4cm Poly-Prep Chromatography Columns. After the lysate had flowed through and without the resin running dry, each column was washed with 3x10mL IgG150 Buffer (plus 1 mM DTT) containing one dissolved Protease Inhibitor Tablet (Pierce) per 50 mL buffer. The flow was stopped by capping the columns with 1.0 mL of resin plus solution remaining. TEV protease (40U) and the SNAP-tag dye (2 µM) were then added. The columns were sealed with caps, and the resin was incubated for 45 min at room temperature

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

560 in the dark with mixing. Subsequent steps were carried out with as little exposure to light as
561 possible.

562 The labeled, TEV-cleaved eluent was added directly to calmodulin affinity resin (400 µL)
563 that had been equilibrated with Calmodulin Binding Buffer (10 mM Tris-Cl pH 8.0, 150 mM NaCl,
564 10 mM KCl, 1 mM $MgCl_2$, 5 mM imidazole, 2 mM $CaCl_2$, 0.1% v/v NP40, no reducing agent,
565 4°C). The IgG resin was washed with an additional 200-300 µL IgG150 Buffer (plus 1 mM DTT)
566 to ensure all sample was transferred to the calmodulin resin. To this slurry, three equivalent
567 volumes (with respect to the volume of the TEV eluate) of Calmodulin Binding Buffer containing
568 10mM β-mercaptoethanol were added. The slurry was then incubated at 4°C with rotation for
569 60 min.

570 The resin slurry was divided between two 0.8x4cm Poly-Prep Chromatography Columns.
571 After the flow through was eluted, each column was washed with 3x5mL Calmodulin Binding
572 Buffer containing 10 mM β-mercaptoethanol. Before elution, the columns were capped at the
573 bottom to control the timing of subsequent steps. Buffer exchange was performed by washing
574 the resin with 100 µL (approximate resin bed volume) Calmodulin Elution Buffer (CEB, 10 mM
575 Tris-Cl pH 8.0, 150 mM NaCl, 10 mM KCl, 1 mM $MgCl_2$, 5 mM imidazole, 4 mM EGTA, 0.08%
576 v/v NP40, 10 mM β-mercaptoethanol). Immediately afterwards, labeled U1 snRNP was eluted
577 in 4x150 µL fractions using incubation times of 0, 2.5, 5, and 10 min with the elution buffer.

578 Fractions were then analyzed by SDS-PAGE, and fractions E1-E3 typically had the
579 highest concentrations of U1. These fractions were pooled and dialyzed in a 10kDa MWCO
580 Slide-A-Lyzer dialysis cassette in 1 L Dialysis Buffer (10 mM Tris-Cl pH 8.0, 150 mM NaCl, 10
581 mM KCl, 1 mM $MgCl_2$, 5 mM imidazole, 10 mM β-mercaptoethanol) overnight at 4°C. In the
582 morning, the cassette was moved to fresh Dialysis Buffer (1 L) for 4 h. The dialyzed sample was
583 concentrated in an Amicon Ultra 100kDa MWCO centrifugal filter unit (14,000 rpm, 4°C, in 1 min
584 intervals). The sample was mixed by pipetting up and down between spins and by addition of

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

585 more dialyzed sample. The final sample volume (~100 µL) was divided into 5 µL aliquots, flash
586 frozen, and stored at -80°C.

587 **HPLC-ESI-MS/MS Analysis**

588 Samples were analyzed by HPLC-ESI-MS/MS using a system consisting of a high-
589 performance liquid chromatograph (nanoAcquity, Waters) connected to an electrospray
590 ionization (ESI) Orbitrap mass spectrometer (LTQ Velos, ThermoFisher Scientific). HPLC
591 separation employed a 100 x 365 mm fused silica capillary micro-column packed with 20 cm of
592 1.7µm-diameter, 130 Angstrom pore size, C18 beads (Waters BEH), with an emitter tip pulled
593 to approximately 1 µm using a laser puller (Sutter Instruments). Peptides were loaded on-
594 column at a flow-rate of 400nL/min for 30 minutes and then eluted over 120 min at a flow-rate of
595 300 nl/minute with a gradient of 2-30% acetonitrile in 0.1% formic acid. Full-mass profile scans
596 were performed in the orbitrap between 300-1500 m/z at a resolution of 60,000, followed by ten
597 MS/MS HCD scans of the ten highest intensity parent ions at 42% relative collision energy and
598 7,500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a
599 repeat count of two over the duration of 30 seconds and an exclusion window of 120 seconds.

600 **Activity Assays**

601 Splicing extracts (yWCE) were prepared from a BJ2168-derived strain of *S. cerevisiae* as
602 previously described (Ansari and Schwer, 1995). Aliquots were flash frozen in liquid nitrogen,
603 stored at -80°C, and thawed on ice once before use. Capped, [^{32}P] -labeled RP51A pre-mRNA
604 was prepared by *in vitro* transcription and gel purified and splicing conditions were adapted from
605 previously described protocols (Crawford et al., 2008). Splicing reactions contained 100 mM
606 potassium phosphate pH 7.3, 3% w/v PEG-8000, 2.5 mM $MgCl_2$, 1 mM DTT, 2 mM ATP, 0.4
607 U/µL RNasin, 40% v/v yWCE, 0.2nM [^{32}P]-labeled RP51A pre-mRNA, and 0.048 U/µL RNaseH.
608 To ablate the U1 snRNA, these reactions were first prepared without ATP, RNasin, [^{32}P]-labeled
609 RP51A, or U1 snRNP and with the inclusion of 0.016 µg/µL U1 cOligo (5'-CTTAAGGTAAGTAT-

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

310 3') so that RNaseH would digest the 5' end of endogenous U1 snRNA in the yWCE (Du and
311 Rosbash, 2001; Larson and Hoskins, 2017). After 30 min at 30°C, the remaining components of
312 the splicing reaction were added along with 0.04 µg/µL purified U1 snRNP. As controls,
313 reactions were prepared without purified U1 snRNP or without U1 cOligo in the ablation reaction.
314 After 60 min at room temperature the reactions were stopped, and RNA was extracted as
315 previously described (Crawford et al., 2008). The products were resolved on a 9% acrylamide
316 (19:1) gel (8M urea, 1X TBE buffer). The gel was dried and imaged using a Phosphor Screen
317 and a Typhoon FLA 9000. The bands were quantified using ImageQuant software.

318 **5' End Analysis by Dideoxynucleotide Sequencing**

319 RNA from purified U1 snRNP or 40 µL U1-SNAP-TAP yWCE was isolated by phenol-
320 chloroform extraction and ethanol precipitation. All of the RNA isolated from labeled, purified U1
321 snRNP was used for reverse transcription while only 10% of the isolated RNA from yWCE was
322 necessary. The isolated RNA was combined with 1 pmol [^{32}P]-labeled primer complementary to
323 nucleotides 27-44 of U1 snRNA (5'- TCAGTAGGACTTCTTGAT) in Annealing Buffer (250mM
324 KCl, 10mM Tris pH8.0) and the reaction was incubated at 90°C for 3 min, snap cooled on ice for
325 3 min, then pre-heated to 45°C for 5 min. A reverse transcriptase (2xRT) master mix was
326 prepared containing 1 U/µL AMV Reverse Transcriptase in 25mM Tris pH 8.0, 8µM DTT, and
327 0.4mM dNTPs.

328 For dideoxynulceotide sequencing, five parallel reactions were set up for each sample.
329 The reactions were made with 3.0 µL 2xRT master mix, 1.0 µL ddNTP/H_2O (1mM ddATP, 1mM
330 ddCTP, 1mM ddTTP, 0.3mM ddGTP, or RNase-free water), and 2.0µL annealing reaction. The
331 reverse transcription reaction was incubated at 45°C for 45min. To stop the reaction, 2 µL
332 formamide loading dye (95% v/v deionized formamide, 0.025% w/v bromophenol blue, 0.025%
333 w/v xylene cyanol FF, 5 mM EDTA pH 8.0) was added then the samples were cooled on ice for
334 3 min then heated to 90°C for 3 min. A portion (5 µL) of each reaction was loaded onto a 0.4

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mM thick 20% acrylamide (19:1) / 7.5M urea / 1xTBE gel. The gel was run until bromophenol blue neared the bottom. The gel was dried and imaged using a phosphorscreen and a Typhoon FLA 9000.

RNA Oligo Secondary Structure Prediction and Calculation of Free Energy of Unwinding

mFold was used to identify potential stable secondary structures formed by the RNA oligos (Zuker, 2003).

The approximate stability of the duplex between U1 snRNA or the U1 mimic RNAs and the Cy3-RNA oligomers was predicted by calculating the stability of hybridization of the uridine-substituted SSRS to the complementary sequence of the RNA oligo using the Hybridization function of DINAMelt (Markham and Zuker, 2005). We note that while base pairs with pseudouridine are predicted to be more stable than those to uridine [(Deb et al., 2019); also see **Fig. 4**], thermodynamic parameters for base pairing to consecutive pseudouridine bases, such as those found within the U1 SSRS (5'-AUACΨΨACCU-3'), have not to our knowledge been determined. Therefore, we were unable to use nearest-neighbor methods to calculate the thermodynamic stabilities for RNAs pairing to the U1 SSRS and instead approximated these stabilities using a uridine-substituted SSRS.

Microscope Slide Preparation

Microscope slides and coverslips were cleaned and assembled into flow as previously described (Crawford et al., 2008). Briefly, top and bottom coverslips were cleaned by sonication for 60 min at 40°C in successive washes of 2% v/v Micro-90 solution, absolute ethanol, 1 M KOH and water with intermittent rinsing with MilliQ water between each wash step. The cleaned coverslips were silanized using freshly prepared 1% v/v Vectabond in acetone (~30mL to cover) for 10 min at room temperature. After silanization, the slides were immediately and thoroughly rinsed with absolute ethanol. The coverslips were thoroughly dried again then assembled into flow cells using vacuum grease to demark lanes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Poly-L-lysine-graft-PEG copolymer (PLL-g-PEG) passivation was used to coat the slide surface and heparin was included in slide washing and imaging buffers to produce a negatively-charged surface (Salomon et al., 2015). Dry aliquots (2 mg) of PLL-g-PEG were dissolved to a final concentration of 4 mg/mL PLL-g-PEG in 100mM HEPES-KOH pH7.4 just before use. The silanized lanes of the flow cell were filled with the PLL-g-PEG solution (~30 µL each) and incubated at room temperature overnight in the dark. For experiments using the U1 snRNA oligo mimic, slides were coated with PEG as previously described (Crawford et al., 2008).

U1 snRNA Mimic Preparation

The U1 snRNA mimic, the biotinylated DNA handle, and an RNA oligomer, were annealed and the tripartite complex was immobilized on the slide surface. Annealing reactions consisted of 2 µM Cy5-labeled U1 mimic (UU or ΨΨ), 200 nM biotinylated DNA handle, and 10 µM Cy3-labeled RNA oligomer in 50mM Tris-HCl pH 7.4, 400mM NaCl. The reactions were heated to 95°C in a thermocycler and cooled by decreasing temperature in 5°C intervals every 2 min until the reaction reached 25°C. After heating, reactions were immediately stored on until use in single-molecule experiments.

Single-molecule Microscopy

CoSMoS experiments were performed on a custom-built, objective-based micromirror total internal reflection fluorescence microscope (Larson et al., 2014). The red laser (633nm) was set to 250µW, and the green laser (532nm) was set to 400µW for data collection. The fluorescence signal was imaged at 1 sec exposure at 5 sec intervals unless otherwise specified. For all experiments, the imaging buffer included glucose, glucose oxidase, and catalase, as oxygen scavengers (OSS), and trolox as a triplet state quencher (TSQ) (Crawford et al., 2008). Drift correction was performed, as necessary, by tracking the movement of individual immobilized spots for the duration of the experiment. Auto-focusing was carried out using a 785nm laser and was done every minute in between exposures. Mapping files were generated

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

each day using TransFluorSpheres (Thermo Fisher) fluorescent in both the <635nm and >635nm fields of view (FOV).

For experiments with the U1 snRNA mimic, prepared slides were first washed with 200 µL Annealing Buffer (50mM Tris-HCl pH7.4, 400mM NaCl) with 0.01 mg/mL yeast tRNA. Prior to imaging, each lane was washed with 70 µL 0.2 mg/mL streptavidin in Annealing Buffer (+tRNA) which was immediately washed away with 70 µL Annealing Buffer (+tRNA). The annealed mimic/handle/oligomer complex was diluted by a factor of 1:2,000-5,000 in Annealing Buffer (+OSS +TSQ +tRNA) and the lane was washed with 70 µL of this solution. The accumulation of the complex on the slide surface was monitored in real time in the >635nm FOV and when the desired density of spots was achieved, excess components were washed away using 70µL Annealing Buffer (+OSS +TSQ +tRNA). To initiate movies, the buffer in the lane was exchanged with 90 µL Mock Splicing Buffer (100mM potassium phosphate pH 7.3, 10mM HEPES-KOH pH 7.9, 20 mM KCl, 2.5 mM $MgCl_2$, 8% v/v glycerol, 5% w/v PEG-8000, 1.4 mM DTT, +OSS +TSQ +tRNA) and data recording was immediately started. For less stable complexes (*e.g.*, RNA-7a), 30 min of data collection was sufficient to observe the dissociation of most (>90%) RNA oligomers. For the most stable complexes (*e.g.*, ΨΨ mimic + RNA-10), 80 min of data collection was necessary and imaging intervals were reduced to 1 exposure/10s.

For experiments with U1 snRNP, prepared slides were first washed with 200 µL Mock Splicing Buffer with 0.05 mg/mL heparin and 0.01 mg/mL yeast tRNA. Subsequent steps were performed one lane at a time. The lane was washed with 70 µL 0.2-0.4 mg/mL streptavidin in Mock Splicing Buffer (+heparin +tRNA) which was allowed to incubate with the slide surface for 2-5 min. The lane was then washed with 70 µL 0.5 mg/mL (10x) heparin in Mock Splicing Buffer (+tRNA) and incubated for 10 min before U1 snRNP was added. U1 snRNP was diluted to a final concentration of 5-20nM in Mock Splicing Buffer (+heparin +tRNA +OSS +TSQ) and added to the lane. The accumulation of fluorescent spots was monitored periodically in the >635nm

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FOV until an optimal density was achieved (usually 2-5 min), then the excess complexes were washed away with 70 µL Mock Splicing Buffer (+heparin +tRNA +OSS +TSQ). Finally, the lane was washed with 70µL 10nM RNA-Cy3 in Mock Splicing Buffer (+heparin +tRNA +OSS +TSQ) and data recording was immediately started. In U1 snRNP experiments, the Cy3 signal was typically imaged at 1 sec exposure at 5 sec intervals for 360 frames (30 min). We determined that the lifetimes measured in these experiments were not being limited by photobleaching by performing control experiments where the power of the 532nm laser was varied from 200 to 600 µW or where the periodicity of the 1 sec exposure was increased to 10 sec.

Data Analysis

Data analysis was performed as previously described (Hoskins et al., 2011; Shcherbakova et al., 2013). In brief, the fluorescence signal detected in the >635 FOV was used to select areas of interest (AOIs). After drift correction, these locations were mapped to the <635 FOV and the pixel intensity was integrated for each AOI using custom MATLAB software. Each colocalization event was manually inspected to confirm the presence of a colocalized spot in the AOI.

For fitting dwell times of oligos binding to the U1 mimic, the distributions were analyzed using survival fraction plots and fit with single exponential decay functions which generated a 95% confidence interval (CI) for the calculated k_{off} and an R-square parameter for the fit. The reciprocal of k_{off} is the mean lifetime (µ).

For analysis of oligo binding to immobilized U1 snRNPs, the distribution of observed dwell times was used to construct a probability density plot in which the dwell times were binned and the probability of each bin was divided by the product of the bin width and the total number of events in the data set. Error bars for each bin were calculated as previously described based on the error of binomial distributions (Hoskins et al., 2011). These plots were fit using a maximum likelihood function for a single- or double-exponential distribution as described by Hansen, et al.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Equations 1 and **2**, respectively (Hoskins et al., 2011). In these equations, t_m is the time between consecutive frames and t_{max} is the duration of the experiment; A_1 and A_2 are the fitted amplitudes for a bi-exponential distribution; and the "taus" are the fitted dwell time parameters for single- (τ_0) or bi-exponential (τ_1 and τ_2) distributions. Errors in the fit were determined by bootstrapping 1000 random samples of the data and determining the standard deviation of the resulting normal distribution.

$$\left[\left(A_0 \cdot \left(e^{-\frac{t_m}{\tau_0}} - e^{-\frac{t_{max}}{\tau_0}}\right)\right)\right]^{-1} \cdot \left[\frac{A_0}{\tau_0} \cdot e^{-\frac{t}{\tau_0}}\right] \quad \textbf{(1)}$$

$$\left[\left(A_1 \cdot \left(e^{-\frac{t_m}{\tau_1} - e^{-\frac{t_{max}}{\tau_1}}}\right)\right) + \left((A_2) \cdot \left(e^{-\frac{t_m}{\tau_1}} - e^{-\frac{t_{max}}{\tau_1}}\right)\right)\right]^{-1} \cdot \left[\frac{A_1}{\tau_1} \cdot e^{-\frac{t}{\tau_1}} + \frac{A_1}{\tau_2} \cdot e^{-\frac{t}{\tau_2}}\right] \quad \textbf{(2)}$$

where $A_1 + A_2 = 1$

To judge the goodness of the fits, the log likelihood ratio test was used to determine if the simplest kinetic model was sufficient to describe the data (Kaur et al., 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures and Figure Legends

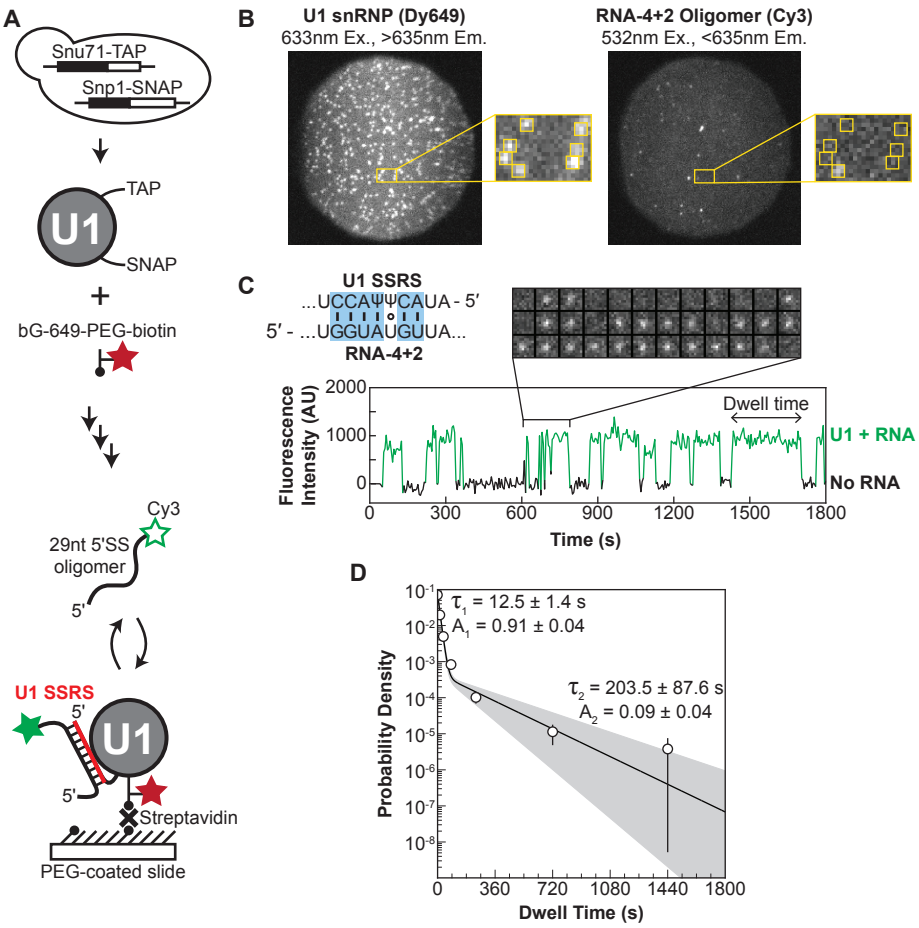

Figure 1. Immobilized Yeast U1 snRNP forms Reversible Short- and Long-Lived Interactions with a 5' SS Oligo. (**A**) Preparation of purified, fluorescently-labeled yeast U1 snRNP using SNAP- and TAP-tags. In single-molecule experiments U1 snRNP is immobilized to the slide surface and its interactions with Cy3-labeled RNA oligomers are observed using CoSMoS. The U1 SSRS that binds to the oligo is shown in red. (**B**) Images showing individual U1 snRNP molecules tethered to the slide surface (left field of view, FOV) and colocalized Cy3-labeled RNA-4+2 molecules (right FOV). (**C**) Representative fluorescence trajectory of changes in Cy3 intensity (green) due to oligo binding to a single immobilized U1 molecule. RNA binding events appear as spots of fluorescence in the recorded images (see inset). Also shown is the predicted pairing interactions (blue) between the RNA-4+2 oligo and the U1 SSRS. (**D**) Probability density histogram of dwell times for the RNA-4+2 oligo and the fitted parameters of the data to an equation containing two exponential terms; the shaded region represents the uncertainty associated with the parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

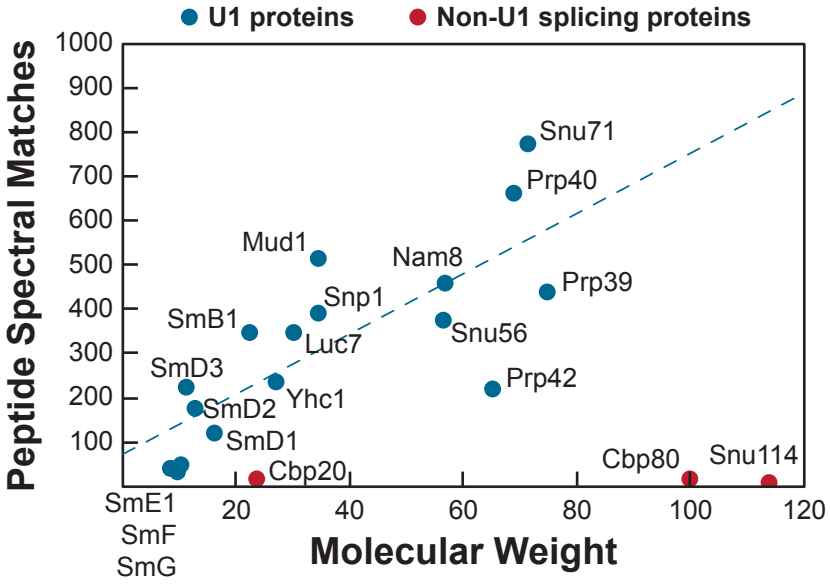

Figure 1-Supplement 1. Mass Spectrometry Analysis of Purified U1 Samples. Plotted are the number of peptide spectral matches observed for the indicated U1 snRNP proteins (blue) vs. the predicted molecular weight of the protein in kDa. The indicated U1 proteins were observed in all preparations of U1 analyzed by mass spectrometry. In some preparations, peptides corresponding to Cbp20 (2 out of 3 preparations), Cbp80 (1 out of 3 preparations), and Snu114 (1 out of 3 preparations) were also observed (red). These were the only known non-U1 splicing factors observed in the samples and these factors were likely present at very low levels since few peptides were observed given the molecular weights of the proteins.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

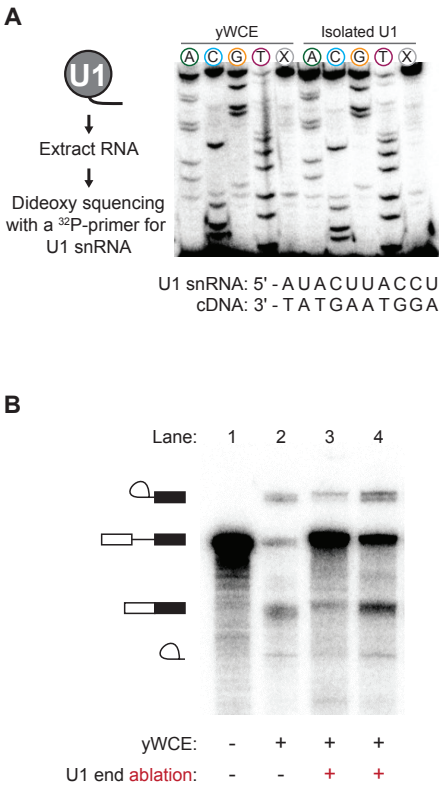

Figure 1-Supplement 2. Dideoxy Sequencing of the Purified U1 snRNA and Activity Assay. (**A**) The presence of the U1 SSRS in the purified U1 snRNP was confirmed by dideoxy sequencing of the SSRS and comparison with sequencing of the snRNA present in total RNA isolated from yeast whole cell extract (yWCE). The dideoxynucleotide present in each reaction is noted above the corresponding lane. Lanes marked X did not contain any dideoxynucleotides. Similar patterns are obtained for the U1 snRNA present in the yWCE as in the isolated U1 and confirm presence of the SSRS. (**B**) Purified U1 can restore splicing activity of yWCE in which the endogenous U1 was ablated by addition of a complementary DNA oligo and RNase H cleavage. Relative splicing efficiencies shown were calculated as the amounts of mRNA products formed compared to the total of the observed RNA species. The bar graph represents the average of 3 replicate experiments ±SD.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

788 **Figure 1-Supplemental Table 1**

Name	Sequence[a]	ΔG°$_{unw}$ [b] (kcal/mol)
RNA-10	5' - ACU GAA AA**A GGU AAG UAU** AUA UGG ACU GA - **Cy3**	12.5
RNA-9a	5' - ACU GAA AA**A GGU AAG UA**A AUA UGG ACU GA - **Cy3**	12.0
RNA-9b	5' - ACU GAA AAU **GGU AAG UAU** AUA UGG ACU GA - **Cy3**	12.4
RNA-8a	5' - ACU GAA AA**A GGU AAG U**UA AUA UGG ACU GA - **Cy3**	9.3
RNA-8b	5' - ACU GAA AAU **GGU AAG UA**A AUA UGG ACU GA - **Cy3**	11.8
RNA-7a	5' - ACU GAA AAU **GGU AAG U**UA AUA UGG ACU GA - **Cy3**	9.1
RNA-7b	5' - ACU GAA AAU C**GU AAG UA**A AUA UGG ACU GA - **Cy3**	6.7
RNA-6a	5' - ACU GAA AAU **GGU AAG** AUA AUA UGG ACU GA - **Cy3**	8.7
RNA-6b	5' - ACU GAA AAU C**GU AAG U**UA AUA UGG ACU GA - **Cy3**	3.9
RNA-5	5' - ACU GAA AAU **GGU AA**C AUA AUA UGG ACU GA - **Cy3**	2.7
RNA-4	5' - ACU GAA AAU **GGU A**UC AUA AUA UGG ACU GA - **Cy3**	3.2
RNA-C	5' - ACU GAA AAU CCA UUC AUA AUA UGG ACU GA - **Cy3**	ND
RNA-1+5	5' - ACU GAA AAU **GcU AAG U**UA AUA UGG ACU GA - **Cy3**	0.4
RNA-2+4	5' - ACU GAA AAU **GGa AAG U**UA AUA UGG ACU GA - **Cy3**	3.3
RNA-3+3	5' - ACU GAA AAU **GGU uAG U**UA AUA UGG ACU GA - **Cy3**	4.0
RNA-4+2	5' - ACU GAA AAU **GGU AuG U**UA AUA UGG ACU GA - **Cy3**	4.7
RNA-5+1	5' - ACU GAA AAU **GGU AAc U**UA AUA UGG ACU GA - **Cy3**	2.7
RNA-2+7 A(+1)	5' - ACU GAA AA**A GaU AAG UAU** AUA UGG ACU GA - **Cy3**	4.4
RNA-2+7 C(+1)	5' - ACU GAA AA**A GcU AAG UAU** AUA UGG ACU GA - **Cy3**	4.4
RNA-2+7 U(+1)	5' - ACU GAA AA**A GuU AAG UAU** AUA UGG ACU GA - **Cy3**	4.7

789 [a] The regions of predicted complementarity to the U1 SSRS are in bold and highlighted. The
790 G(+1) position of the 5' SS (based on RP51A) is underlined.
791 [b] The predicted standard free energy change for duplex unwinding from U1 using DINAMelt at
792 37°C in 1 M NaCl. Note that these free energies were calculated using a uridine-substituted
793 SSRS, see **Methods**.
794

795
796

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

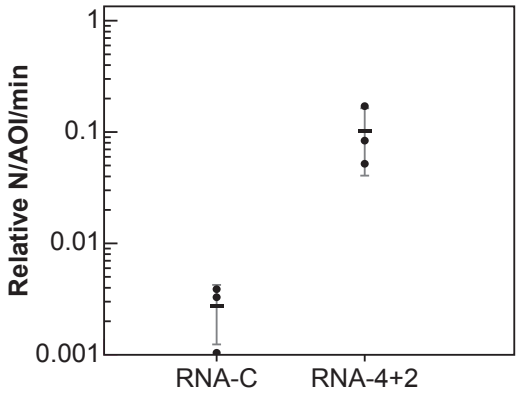

Figure 1-Supplement 3. Observed U1 Binding Events are Sequence-Dependent. Relative event densities of oligo binding to immobilized U1 molecules for RNA-C (little to no pairing with the SSRS) and RNA-4+2 (the WT RP51A 5' SS with 6 predicted base pairs). Plotted are the results from 3 replicate experiments (dots) along with the average ±SD (horizontal bars and vertical lines).

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

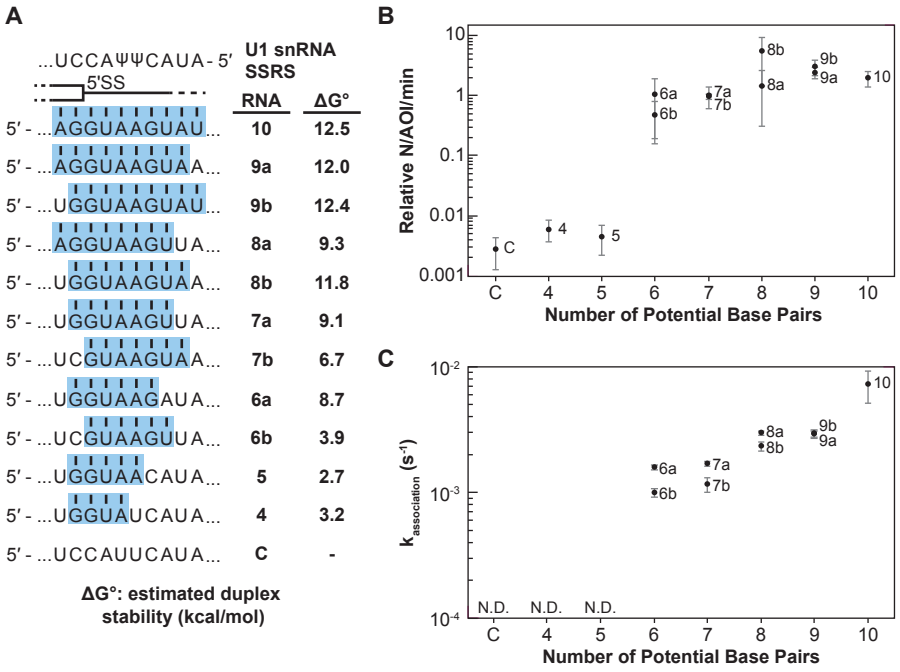

Figure 2. Impact of Base Pairing Potential on RNA Oligo Binding to U1. (**A**) RNA oligos tested for interaction with U1 containing 4-10 predicted base pairs and the calculated free energy changes for duplex unwinding/formation based on nearest neighbor analysis. The regions shaded in blue are predicted to pair with the SSRS (**B**) Relative event densities of oligo binding to immobilized U1 molecules as a function of potential base pairs. (**C**) Measured association rates of the oligos to U1 as a function of potential base pairs. For (B), the plotted points represent the average results from at least 3 replicate experiments ±SD. For (C), the plotted points represent the fitted parameters ± the uncertainties of the fits.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

320 **Figure 2-Supplemental Table 1**

RNA	*N*[a]	$k_{association}$[b]
RNA-10	91	7.18 ± 2.04
RNA-9a	95	2.90 ± 0.16
RNA-9b	124	2.93 ± 0.23
RNA-8a	153	2.96 ± 0.12
RNA-8b	77	2.32 ± 0.19
RNA-7a	87	1.68 ± 0.07
RNA-6a [c]	44	1.57 ± 0.07
RNA-6b [c]	41	0.99 ± 0.07
RNA-4+2	85	2.95 ± 0.16
		(x 10^{-3} sec^{-1})

321 [a] *N* represents the number of measured time intervals
322 [b] $k_{association}$ determined for experiments where [RNA] = 10 nM
323

324

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

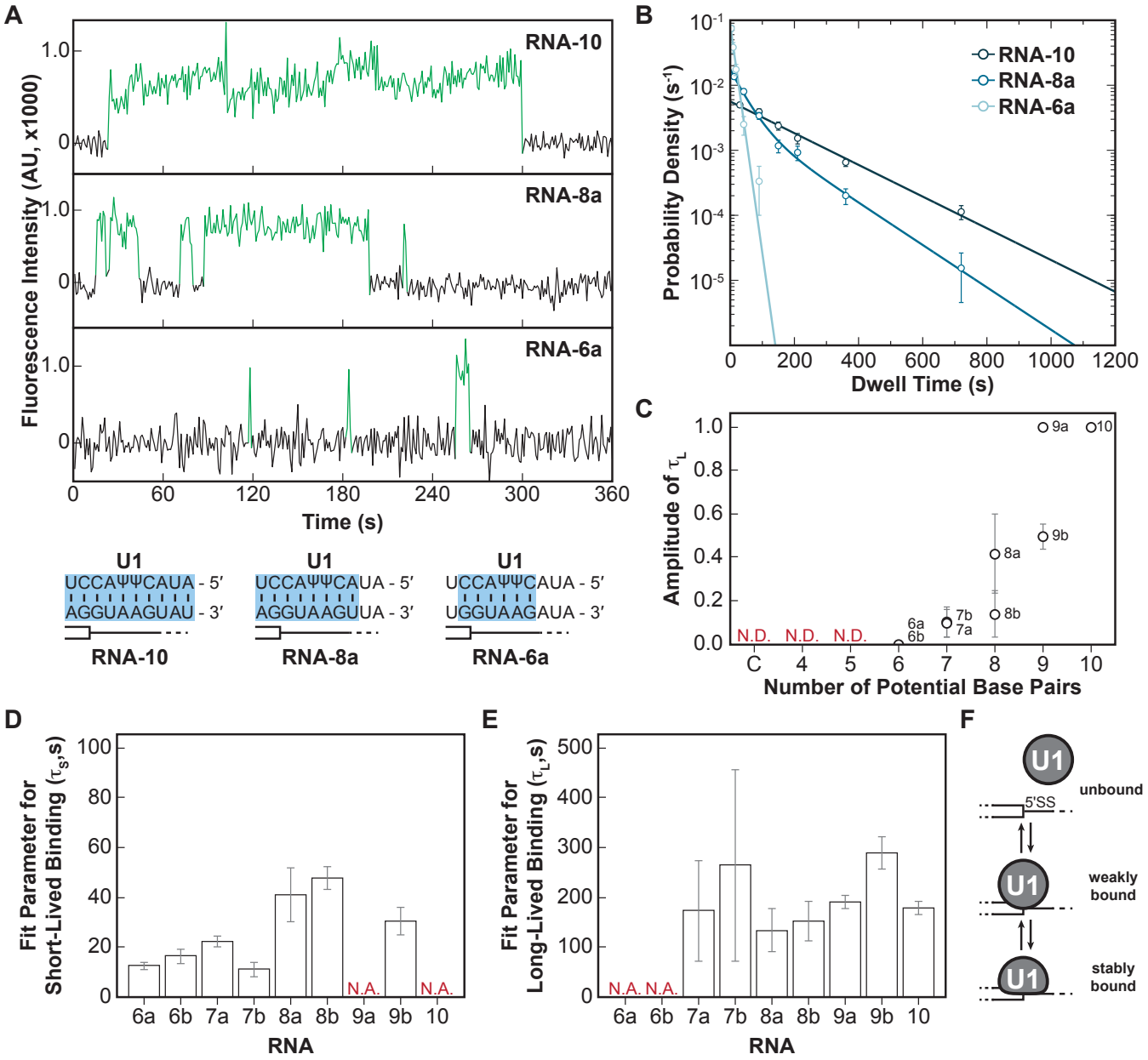

Figure 3. The long-lived state is dependent on the length of the snRNA-RNA duplex. (**A**) Representative fluorescence trajectories of changes in Cy3 intensity (green) due to oligo binding to a single immobilized U1 molecules for RNAs 6a, 8a, and 10. Also shown are the predicted pairing interactions (blue) between the oligos and the U1 SSRS. (**B**) Probability density histograms for dwell times for RNAs 6a, 8a, and 10 binding to U1. Lines represent the single- or double-exponential distribution obtained for the fitted parameter from each data set. (**C**) The amplitude of the long-lived population (τ_L > 60 s) is shown for each RNA oligomer in **Fig. 2A**. The amplitude is not determined (N.D.) for oligomers for which little binding was observed. (**D-E**) The fit parameters for short-lived binding (panel D, τ_S < 60 s) and long-lived binding (panel E, τ_L > 60 s) shown for each RNA oligomer in **Fig. 2A**. If there is only one fit parameter, then the other is not applicable (N.A.). (**F**) A two-step mechanism for U1/RNA association that can account for both short and long-lived binding events.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

338

339 **Figure 3-Supplemental Table 1**

RNA	N[a]	Tau	Tau1 (τ_S)	Tau2 (τ_L)	Amp. of Tau2
RNA-10	295	178.1 ± 13.3 sec			
RNA-9a	232	190.1 ± 13.8 sec			
RNA-9b	273		30.4 ± 5.7 sec	287.7 ± 32.7 sec	0.50 ± 0.06
RNA-8a	270		40.9 ± 10.8 sec	133.3 ± 43.8 sec	0.42 ± 0.18
RNA-8b	351		47.7 ± 4.6 sec	151.2 ± 39.7 sec	0.14 ± 0.11
RNA-7a	132		22.2 ± 2.4 sec	172.8 ± 100.9 sec	0.10 ± 0.07
RNA-7b	67		10.3 ± 2.8 sec	236.0 ± 158.2 sec	0.13 ± 0.07
RNA-6a	100	12.4 ± 1.4 sec			
RNA-6b	59	16.2 ± 2.8 sec			

340 [a] Number of dwell times combined from the replicates and fit to obtain the corresponding
341 parameters.
342

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

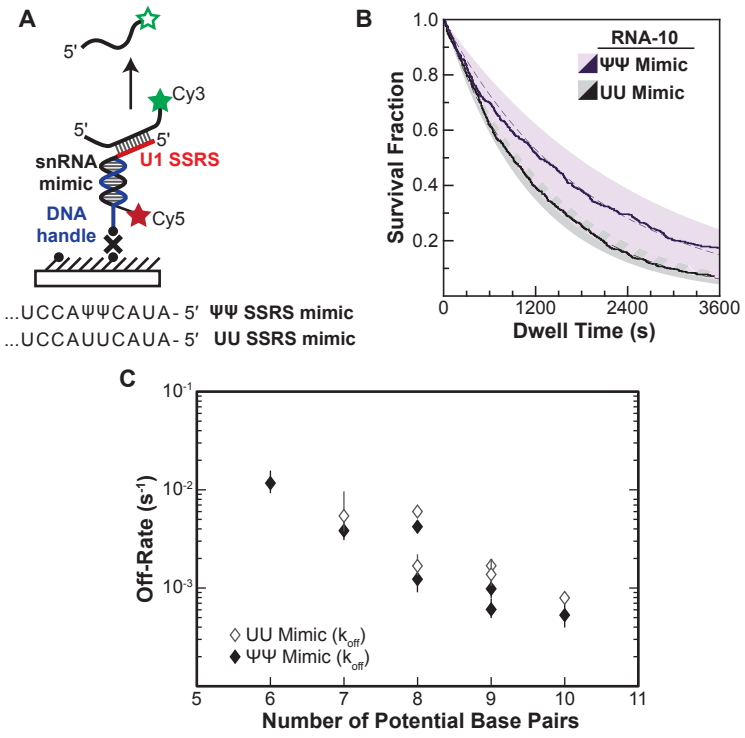

Figure 4. Lifetimes of 5' SS Oligo/RNA Interactions are Dependent on Base-Pairing Potential in a RNA-Only Mimic of the U1 SSRS. (**A**). Schematic of a single molecule assay for monitoring dissociation of RNA oligos from the RNA-only mimic of the U1 SSRS. Two mimics were used that contain pseudouridine (Ψ) or uridine (U) at two positions in the SSRS that have Ψ in the native U1 snRNA. (**B**) The fraction of colocalized RNA oligos remaining was plotted over time to yield survival fraction curves for determining RNA oligo off-rates (black lines). The curves were then fit to exponential decay functions to yield off-rates as well as 95% confidence intervals for the fits (dashed lines and shaded regions, respectively). Shown are the survival fraction curves for RNA-10 dissociation (see **Fig. 2A**). (**C**) Measured off-rates for RNA Oligos to the SSRS mimics (see **Figure 4-Supplemental Table 1**) plotted as a function of potential base pairs.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4—Supplemental Table 1

Mimic	RNA	N[a]	k_{off} (95% CI) [b]
UU mimic	RNA-10	580	**7.71** (7.69 7.73) **x 10^{-4} sec^{-1}**
UU mimic	RNA-9a	235	**1.40** (1.37 1.43) **x 10^{-3} sec^{-1}**
UU mimic	RNA-9b	409	**1.67** (1.64 1.70) **x 10^{-3} sec^{-1}**
UU mimic	RNA-8a	170	**1.65** (1.62 1.67) **x 10^{-3} sec^{-1}**
UU mimic	RNA-8b	126	**5.83** (5.71 5.95) **x 10^{-3} sec^{-1}**
UU mimic	RNA-7a	48	**5.23** (4.86 5.66) **x 10^{-3} sec^{-1}**
ΨΨ mimic	RNA-10	550	**5.50** (5.46 5.54) **x 10^{-4} sec^{-1}**
ΨΨ mimic	RNA-9a	287	**10.10** (9.97 10.20) **x 10^{-4} sec^{-1}**
ΨΨ mimic	RNA-9b	276	**6.06** (5.97 6.14) **x 10^{-4} sec^{-1}**
ΨΨ mimic	RNA-8a	172	**1.33** (1.29 1.36) **x 10^{-3} sec^{-1}**
ΨΨ mimic	RNA-8b	212	**4.15** (4.10 4.21) **x 10^{-3} sec^{-1}**
ΨΨ mimic	RNA-7a	157	**3.73** (3.57 3.88) **x 10^{-3} sec^{-1}**
ΨΨ mimic	RNA-6a	67	**1.28** (1.17 1.39) **x 10^{-2} sec^{-1}**

[a] Number of (dwell times) combined from multiple replicates.

[b] The calculated k_{off} (and 95% confidence interval for this value) that results from fitting the combined dwell times.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

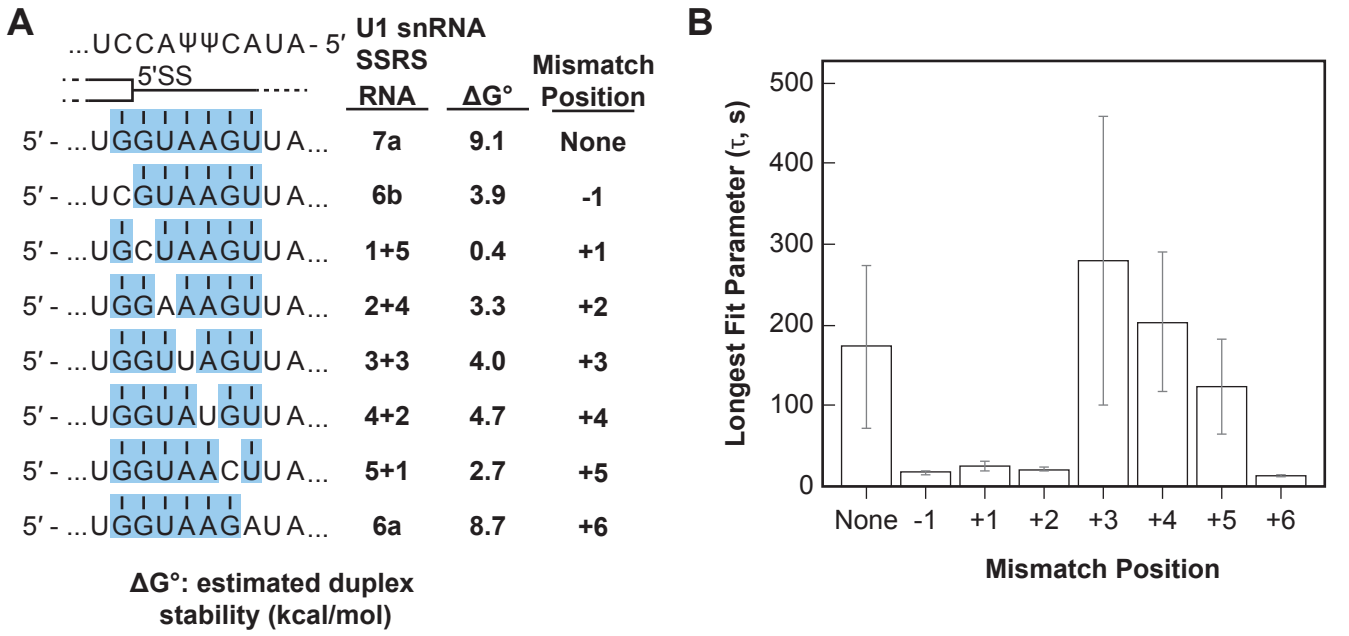

Figure 5. Long-lived U1/RNA Interactions are Dependent on Mismatch Position. (**A**) RNA oligos tested for interaction with U1 containing mismatches at the -1 to +6 positions and the calculated free energy changes for duplex unwinding/formation based on nearest neighbor analysis. The regions shaded in blue are predicted to pair with the SSRS. (**B**) The value of the longest-lived parameter (τ_0 or τ_2, **Figure 5-Supplemental Table 1**) obtained by fits of the distributions of dwell times to U1 for each RNA oligomer in panel (A). The plotted bars represent the fitted parameters ± the uncertainties of the fits. Note that data for RNA oligos 7a, 6a, and 6b were replotted from Fig. 3D, E for comparision.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

375 **Figure 5-Supplemental Table 1**

RNA (Mismatch Position)	**N[a]**	**Tau (τ_0)**	**Tau1 (τ_1)**	**Tau2 (τ_2)**	**A_2**
RNA-7a (None)	132		22.2 ± 2.4 sec	172.8 ± 100.9 sec	0.10 ± 0.07
RNA-6a (+6)	100	12.4 ± 1.4 sec			
RNA-6b (-1)	59	16.2 ± 2.8 sec			
RNA-1+5 (+1)	68	23.8 ± 6.3 sec			
RNA-2+4 (+2)	45	20.3 ± 3.1 sec			
RNA-3+3 (+3)	59		11.4 ± 1.9 sec	279.2 ± 178.6 sec	0.21 ± 0.07
RNA-4+2 (+4)	367		12.5 ± 1.4 sec	203.5 ± 87.6 sec	0.09 ± 0.04
RNA-5+1 (+5)	113		12.6 ± 3.3 sec	122.8 ± 59.8 sec	0.14 ± 0.06

376 [a] Number of dwell times combined from the replicates and fit to obtain the corresponding
377 parameters.
378
379

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

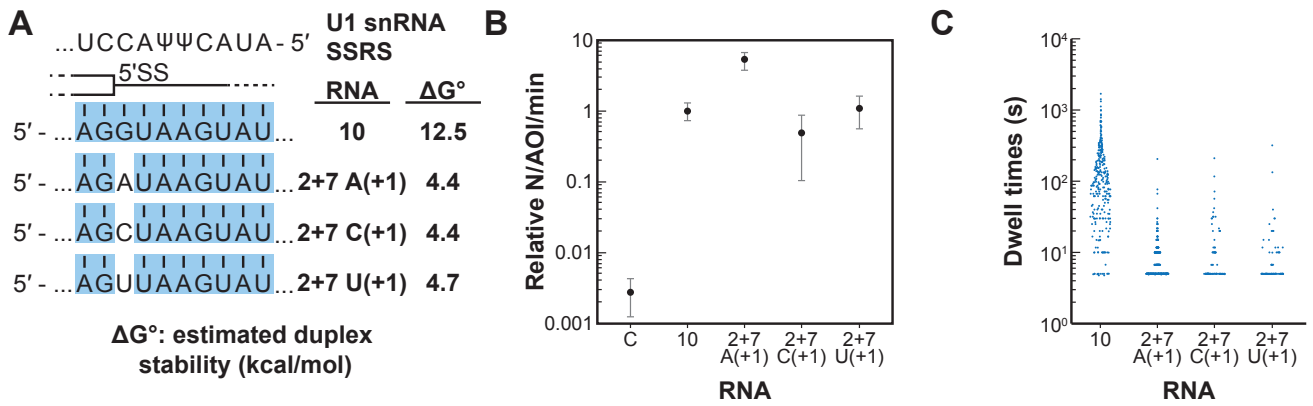

Figure 6. Long-lived Interactions are Greatly Stimulated by G at the 5' SS +1 Position. (**A**) RNA oligos tested for interaction with U1 containing mismatches only at the +1 positions and the calculated free energy changes for duplex unwinding/formation based on nearest neighbor analysis. The regions shaded in blue are predicted to pair with the SSRS. (**B**) Relative event densities of oligo binding to immobilized U1 molecules for RNAs shown in panel (A). Plotted are averages from replicate experiments ±SD (dots and vertical lines). (**C**) Distribution of observed dwell times for U1 interactions with oligos from panel (A). Each dot corresponds to a single dwell time.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 6-Supplemental Table 1

RNA	Tau (τ)[a]	Log-likelihood (1 exponential term) [b]	Tau1 (τ_S) [c]	Tau2 (τ_L)[c]	A_1[c]	A_2[c]	Log-likelihood (2 exponential terms)[d]
RNA-2+7 - A(+1)	7.5 ± 0.5	-1563.0	6.9 ± 0.2	92.0 ± 42.4	0.99 ±0.01	0.01	-1543.3
RNA-2+7 - C(+1)	13.8 ± 2.8	-322.3	8.6 ± 1.1	72.2 ± 33.2	0.90 ± 0.07	0.10	-308.9
RNA-2+7 - U(+1)	12.1 ± 3.7	-1563.0	7.1 ± 0.6	188.1 ± 68.1	0.97 ± 0.02	0.03	-284.3

[a] Fit parameter to an equation containing a single exponential term.
[b] Log-likelihood output for the fit to an equation with a single exponential term.
[c] Fit parameters to an equation containing two exponential terms. A_1 is the amplitude of the first exponential term, τ_S. A_2 is the amplitude of the second exponential term, τ_L.
[d] Log-likelihood output for the fit to an equation with two exponential terms. The more positive log-likelihood indicates these data are better fit to equations containing two terms.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

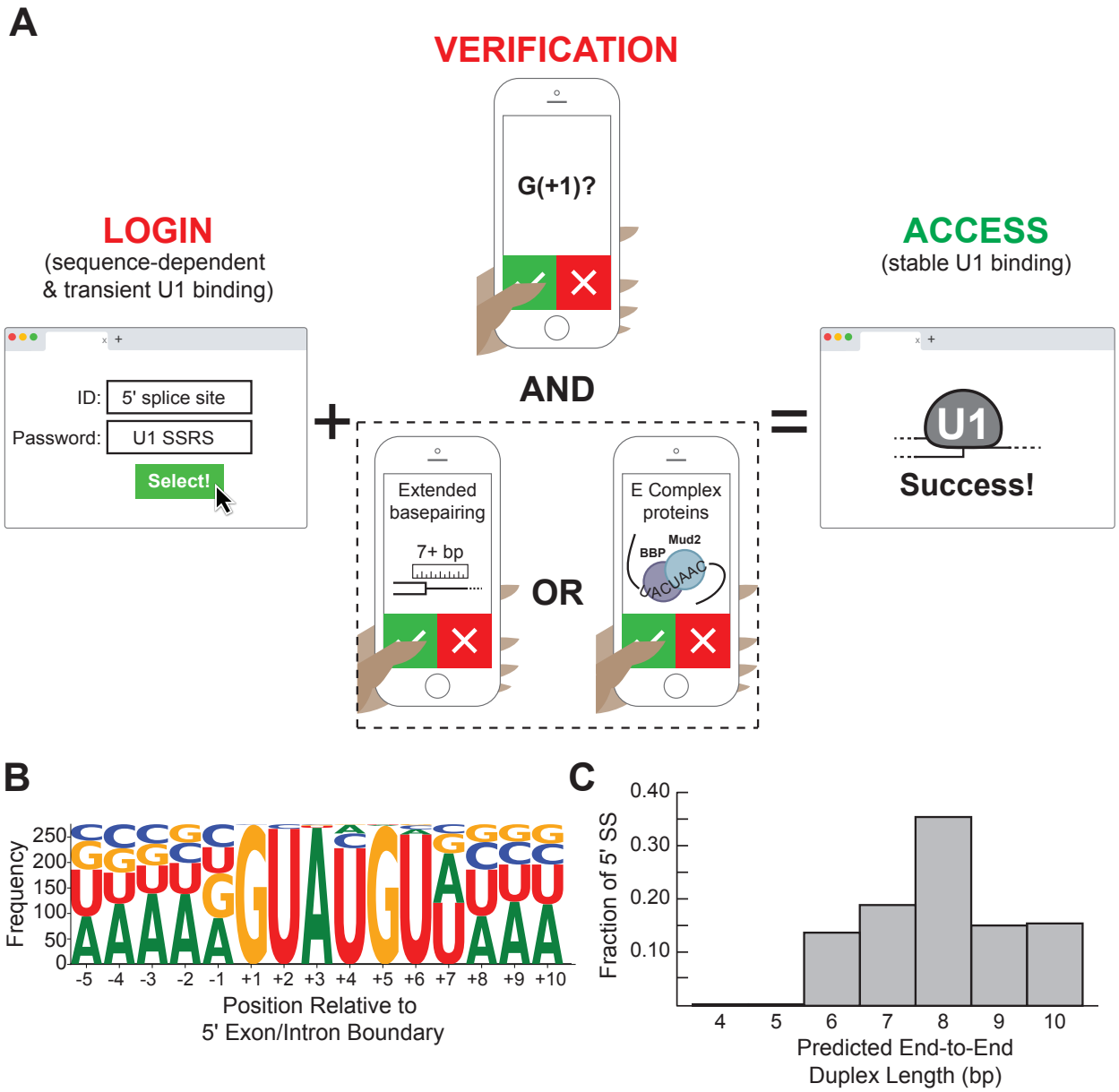

Figure 7. Multi-Factor Authentication Model for 5' SS Recognition. (**A**) U1 binding initially occurs by formation of a weakly interacting complex that is dependent on base pairing potential between the RNA and U1 SSRS. Stable binding is dependent on presence of G+1 at the 5' SS and formation of an extended duplex with an end-to-end length of at least 7 bp or the presence of *trans*-acting splicing factors such as E complex proteins (Larson and Hoskins, 2017). (**B**) Sequence LOGO for annotated yeast 5' SS (Lim and Burge, 2001). (**C**) Histogram of end-to-end duplex lengths based on base pairing potential between annotated yeast 5' SS and the U1 SSRS. Most of these duplexes contain one or more mismatches with the SSRS.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

Agarwal R, Schwer B, Shuman S. 2016. Structure–function analysis and genetic interactions of the Luc7 subunit of the Saccharomyces cerevisiae U1 snRNP. RNA 22:1302–1310. doi:10.1261/rna.056911.116

Ansari A, Schwer B. 1995. SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. The EMBO journal 14:4001–4009.

Bai R, Wan R, Yan C, Lei J, Shi Y. 2018. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science (New York, NY) 360:1423–1429. doi:10.1126/science.aau0325

Braun JE, Friedman LJ, Gelles J, Moore MJ. 2018. Synergistic assembly of human pre-spliceosomes across introns and exons. eLife 7:20221. doi:10.7554/elife.37751

Brow DA. 2002. Allosteric cascade of spliceosome activation. Annual review of genetics 36:333–360. doi:10.1146/annurev.genet.36.043002.091635

Carmel I, Tal S, Vig I, Ast G. 2004. Comparative analysis detects dependencies among the 5′ splice-site positions. Rna 10:828–840. doi:10.1261/rna.5196404

Chen JY, Stands L, Staley JP, Jackups RR, Latus LJ, Chang TH. 2001. Specific alterations of U1-C protein or U1 small nuclear RNA can eliminate the requirement of Prp28p, an essential DEAD box splicing factor. Molecular Cell 7:227–232.

Chiou N-T, Shankarling G, Lynch KW. 2013. HnRNP L and HnRNP A1 Induce Extended U1 snRNA Interactions with an Exon to Repress Spliceosome Assembly. Molecular Cell. doi:10.1016/j.molcel.2012.12.025

Cisse II, Kim H, Ha T. 2012. A rule of seven in Watson-Crick base-pairing of mismatched sequences. Nature Structural & Molecular Biology. doi:10.1038/nsmb.2294

Craig ME, Crothers DM, Doty P. 1971. Relaxation kinetics of dimer formation by self complementary oligonucleotides. J Mol Biol 62:383–401. doi:10.1016/0022-2836(71)90434-7

Crawford DJ, Hoskins AA, Friedman LJ, Gelles J, Moore MJ. 2008. Visualizing the splicing of single pre-mRNA molecules in whole cell extract. RNA (New York, NY) 14:170–179. doi:10.1261/rna.794808

Deb I, Popenda Ł, Sarzyńska J, Małgowska M, Lahiri A, Gdaniec Z, Kierzek R. 2019. Computational and NMR studies of RNA duplexes with an internal pseudouridine-adenosine base pair. Sci Rep-uk 9:16278. doi:10.1038/s41598-019-52637-0

Du H, Rosbash M. 2001. Yeast U1 snRNP-pre-mRNA complex formation without U1snRNA-pre-mRNA base pairing. RNA (New York, NY) 7:133–142.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Du H, Tardiff DF, Moore MJ, Rosbash M. 2004. Effects of the U1C L13 mutation and temperature regulation of yeast commitment complex formation. Proceedings of the National Academy of Sciences of the United States of America 101:14841–14846. doi:10.1073/pnas.0406319101

Feltz C van der, Krummel DP. 2016. Purification of Native Complexes for Structural Study Using a Tandem Affinity Tag Method. Journal of visualized experiments : JoVE e54389. doi:10.3791/54389

Fortes P, Bilbao-Cortés D, Fornerod M, Rigaut G, Raymond W, Séraphin B, Mattaj IW. 1999. Luc7p, a novel yeast U1 snRNP protein with a role in 5′ splice site recognition. Gene Dev 13:2425–2438. doi:10.1101/gad.13.18.2425

Fouser LA, Friesen JD. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. Cell 45:81–93.

Globyte V, Lee SH, Bae T, Kim J, Joo C. 2019. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. Embo J 38. doi:10.15252/embj.201899466

Grate L, Ares M. 2002. Searching yeast intron data at ares lab web site. Methods Enzymol 350:380–392. doi:10.1016/s0076-6879(02)50975-7

Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ. 2011. Ordered and dynamic assembly of single spliceosomes. Science (New York, NY) 331:1289–1295. doi:10.1126/science.1198830

Jarmoskaite I, AlSadhan I, Vaidyanathan PP, Herschlag D. 2020. How to measure and evaluate binding affinities. Elife 9:e57264. doi:10.7554/elife.57264

Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, Dreyfuss G. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature 468:664–668. doi:10.1038/nature09479

Kaur H, Jamalidinan F, Condon SGF, Senes A, Hoskins AA. 2019. Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions. Methods (San Diego, Calif) 153:13–21. doi:10.1016/j.ymeth.2018.11.014

Konarska MM. 1998. Recognition of the 5’ splice site by the spliceosome. Acta Biochim Pol 45:869–881. doi:10.18388/abp.1998_4346

Kondo Y, Oubridge C, Roon A-MM van, Nagai K. 2015. Crystal structure of human U1 snRNP, a small nuclear ribonucleoprotein particle, reveals the mechanism of 5’ splice site recognition. eLife 4. doi:10.7554/elife.04986

Kotovic KM, Lockshon D, Boric L, Neugebauer KM. 2003. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. Molecular and cellular biology 23:5768–5779.

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Krummel DAP, Oubridge C, Leung AKW, Li J, Nagai K. 2009. Crystal structure of human spliceosomal U1 snRNP at 5.5 A resolution. Nature 458:475–480. doi:10.1038/nature07851

Kwon SC, Nguyen TA, Choi Y-G, Jo MH, Hohng S, Kim VN, Woo J-S. 2016. Structure of Human DROSHA. Cell 164:81–90. doi:10.1016/j.cell.2015.12.019

Lacadie SA, Rosbash M. 2005. Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. Molecular Cell 19:65–75. doi:10.1016/j.molcel.2005.05.006

Larson J, Kirk M, Drier EA, O'Brien W, MacKay JF, Friedman LJ, Hoskins AA. 2014. Design and construction of a multiwavelength, micromirror total internal reflectance fluorescence microscope. Nature protocols 9:2317–2328. doi:10.1038/nprot.2014.155

Larson JD, Hoskins AA. 2017. Dynamics and consequences of spliceosome E complex formation. eLife 6:e27592. doi:10.7554/elife.27592

Lerner MR, Boyle JA, Mount SM, Wolin SL, Steitz JA. 1980. Are snRNPs involved in splicing? Nature 283:220–224.

Li X, Liu S, Jiang J, Zhang L, Espinosa S, Hill RC, Hansen KC, Zhou ZH, Zhao R. 2017. CryoEM structure of Saccharomyces cerevisiae U1 snRNP offers insight into alternative splicing. Nature communications 8:1035. doi:10.1038/s41467-017-01241-9

Li X, Liu S, Zhang L, Issaian A, Hill RC, Espinosa S, Shi S, Cui Y, Kappel K, Das R, Hansen KC, Zhou ZH, Zhao R. 2019. A unified mechanism for intron and exon definition and back-splicing. Nature 573:375–380. doi:10.1038/s41586-019-1523-6

Lim LP, Burge CB. 2001. A computational analysis of sequence features involved in recognition of short introns. Proceedings of the National Academy of Sciences of the United States of America 98:11193–11198. doi:10.1073/pnas.201407298

MacRae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA. 2006. Structural Basis for Double-Stranded RNA Processing by Dicer. Science 311:195–198. doi:10.1126/science.1121638

Małecka EM, Woodson SA. 2021. Stepwise sRNA targeting of structured bacterial mRNAs leads to abortive annealing. Mol Cell 81:1988-1999.e4. doi:10.1016/j.molcel.2021.02.019

Marimuthu K, Chakrabarti R. 2014. Sequence-dependent theory of oligonucleotide hybridization kinetics. J Chem Phys 140:175104. doi:10.1063/1.4873585

Markham NR, Zuker M. 2005. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res 33:W577–W581. doi:10.1093/nar/gki591

McGrail JC, O'Keefe RT. 2008. The U1, U2 and U5 snRNAs crosslink to the 5' exon during yeast pre-mRNA splicing. Nucleic Acids Research 36:814–825. doi:10.1093/nar/gkm1098

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Oh J-M, Di C, Venters CC, Guo J, Arai C, So BR, Pinto AM, Zhang Z, Wan L, Younis I, Dreyfuss G. 2017. U1 snRNP telescripting regulates a size-function-stratified human genome. Nature Structural & Molecular Biology 24:nsmb.3473-999. doi:10.1038/nsmb.3473

Parker R, Siliciano PG. 1993. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. Nature 361:660–662. doi:10.1038/361660a0

Plaschka C, Lin P-C, Charenton C, Nagai K. 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature 559:419–422. doi:10.1038/s41586-018-0323-8

Plaschka C, Newman AJ, Nagai K. 2019. Structural Basis of Nuclear pre-mRNA Splicing: Lessons from Yeast. Cold Spring Harbor Perspectives in Biology a032391. doi:10.1101/cshperspect.a032391

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Séraphin B. 2001. The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. Methods (San Diego, Calif) 24:218–229. doi:10.1006/meth.2001.1183

Puig O, Gottschalk A, Fabrizio P, Seraphin B. 1999. Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. Gene Dev 13:569–580. doi:10.1101/gad.13.5.569

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. 1999. A generic protein purification method for protein complex characterization and proteome exploration. Nature Biotechnology 17:1030–1032. doi:10.1038/13732

Roca X, Krainer AR. 2009. Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA. Nature Structural & Molecular Biology 16:176–182. doi:10.1038/nsmb.1546

Roca X, Krainer AR, Eperon IC. 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. 27:129–144. doi:10.1101/gad.209759.112

Rosbash M, Séraphin B. 1991. Who's on first? The U1 snRNP-5 splice site interaction and splicing. Trends in Biochemical Sciences 16:187–190. doi:10.1016/0968-0004(91)90073-5

Ruby SW, Abelson J. 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science (New York, NY) 242:1028–1035.

Rymond BC, Rosbash M. 1985. Cleavage of 5′ splice site and lariat formation are independent of 3′ splice site in yeast mRNA splicing. Nature 317:735–737. doi:10.1038/317735a0

Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. 2015. Single-Molecule Imaging Reveals that Argonaute Reshapes the Binding Properties of Its Nucleic Acid Guides. Cell 162:84–95. doi:10.1016/j.cell.2015.06.029

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Schwer B, Shuman S. 2015. Structure-function analysis and genetic interactions of the Yhc1, SmD3, SmB, and Snp1 subunits of yeast U1 snRNP and genetic interactions of SmD3 with U2 snRNP subunit Lea1. RNA (New York, NY) 21:1173–1186. doi:10.1261/rna.050583.115

Schwer B, Shuman S. 2014. Structure-function analysis of the Yhc1 subunit of yeast U1 snRNP and genetic interactions of Yhc1 with Mud2, Nam8, Mud1, Tgs1, U1 snRNA, SmD3 and Prp28. Nucleic Acids Research 42:4697–4711. doi:10.1093/nar/gku097

Séraphin B, Rosbash M. 1991. The yeast branchpoint sequence is not required for the formation of a stable U1 snRNA-pre-mRNA complex and is recognized in the absence of U2 snRNA. The EMBO journal 10:1209–1216.

Séraphin B, Rosbash M. 1989. Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell 59:349–358.

Shcherbakova I, Hoskins AA, Friedman LJ, Serebrov V, Corrêa IR, Xu M-Q, Gelles J, Moore MJ. 2013. Alternative Spliceosome Assembly Pathways Revealed by Single-Molecule Fluorescence Microscopy. Cell Reports 5:151–165. doi:10.1016/j.celrep.2013.08.026

Shenasa H, Movassat M, Forouzmand E, Hertel KJ. 2020. Allosteric regulation of U1 snRNP by splicing regulatory proteins controls spliceosomal assembly. Rna 26:1389–1399. doi:10.1261/rna.075135.120

Smith BA, Padrick SB, Doolittle LK, Daugherty-Clarke K, Corrêa JIR, Xu M-Q, Goode BL, Rosen MK, Gelles J, Sundquist W. 2013. Three-color single molecule imaging shows WASP detachment from Arp2/3 complex triggers actin filament branch formation. eLife 2. doi:10.7554/elife.01008

Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, Bayrak-Toydemir P, McDonald J, Fairbrother WG. 2017. Pathogenic variants that alter protein code often disrupt splicing. Nature Publishing Group 49:848–855. doi:10.1038/ng.3837

Staley JP, Guthrie C. 1999. An RNA switch at the 5’ splice site requires ATP and the DEAD box protein Prp28p. Molecular Cell 3:55–64.

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature 507:62–67. doi:10.1038/nature13011

Tatei K, Takemura K, Tanaka H, Masaki T, Ohshima Y. 1987. Recognition of 5’ and 3’ splice site sequences in pre-mRNA studied with a filter binding technique. J Biological Chem 262:11667–74.

Vijayraghavan U, Company M, Abelson J. 1989. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. 3:1206–1216.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. Cell 136:701–718. doi:10.1016/j.cell.2009.02.009

bioRxiv preprint doi: https://doi.org/10.1101/2021.05.18.443434; this version posted May 18, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Wetmur JG. 1991. DNA probes: applications of the principles of nucleic acid hybridization. Critical reviews in biochemistry and molecular biology 26:227–259. doi:10.3109/10409239109114069

Wetmur JG, Davidson N. 1968. Kinetics of renaturation of DNA. Journal of molecular biology 31:349–370.

Wilkinson ME, Fica SM, Galej WP, Norman CM, Newman AJ, Nagai K. 2017. Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection. Science (New York, NY) 358:1283–1288. doi:10.1126/science.aar3729

Yan C, Wan R, Shi Y. 2019. Molecular Mechanisms of pre-mRNA Splicing through Structural Biology of the Spliceosome. Cold Spring Harbor Perspectives in Biology 11:a032409. doi:10.1101/cshperspect.a032409

Yeo G, Burge CB. 2004. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. J Comput Biol 11:377–394. doi:10.1089/1066527041410418

Zhang D, Rosbash M. 1999. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. Gene Dev 13:581–592. doi:10.1101/gad.13.5.581

Zhang S, Aibara S, Vos SM, Agafonov DE, Lührmann R, Cramer P. 2021. Structure of a transcribing RNA polymerase II–U1 snRNP complex. Science 371:305–309. doi:10.1126/science.abf1870

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. doi:10.1093/nar/gkg595