Pre-mRNA Splicing Factor U2AF2 Recognizes Distinct Conformations of Nucleotide Variants at the Center of the pre-mRNA Splice Site Signal

Eliezra Glasser^{1,†}, Debanjana Maji^{1,†}, Giulia Biancon², Anees Mohammed Keedakkatt Puthenpeedikakkal¹, Chapin E. Cavender¹, Toma Tebaldi², Mary J. Pulvino¹, Jermaine L. Jenkins¹, David H. Mathews¹, Stephanie Halene², and Clara L. Kielkopf^{1,*}

¹ Department of Biochemistry and Biophysics, Center for RNA Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA

² Section of Hematology, Department of Internal Medicine and Yale Cancer Center, Yale University School of Medicine, New Haven, CT 06520, USA

**Correspondence:*
clara_kielkopf@urmc.rochester.edu

[†]These authors contributed equally to this work.

ABSTRACT

The essential pre-mRNA splicing factor U2AF2 (also called U2AF⁶⁵) identifies polypyrimidine (Py) tract signals of nascent transcripts, despite length and sequence variations. Previous studies have shown that the U2AF2 RNA recognition motifs (RRM1 and RRM2) preferentially bind uridine-rich RNAs. Nonetheless, the specificity of the RRM1/RRM2 interface for the central Py tract nucleotide has yet to be investigated. We addressed this question by determining crystal structures of U2AF2 bound to a cytidine, guanosine, or adenosine at the central position of the Py tract, and compared U2AF2-bound uridine structures. Local movements of the RNA site accommodated the different nucleotides, whereas the polypeptide backbone remained similar among the structures. Accordingly, molecular dynamics simulations revealed flexible conformations of the central, U2AF2-bound nucleotide. The RNA binding affinities and splicing efficiencies of structure-guided mutants demonstrated that U2AF2 tolerates nucleotide substitutions at the central position of the Py tract. Moreover, enhanced crosslinking and immunoprecipitation of endogenous U2AF2 in human erythroleukemia cells showed uridine-sensitive binding sites with lower sequence conservation at the central nucleotide positions of otherwise uridine-rich, U2AF2-bound splice sites. Altogether, these results highlight the importance of RNA flexibility for protein recognition and take a step towards relating splice site motifs to pre-mRNA splicing efficiencies.

INTRODUCTION

The vast majority of human genes contain intervening introns that need to be spliced from the nascent transcript and the exons joined to form the mRNA before translation into a protein. Alternative splicing to join different subsets of exons expands the diversity of proteins encoded by a limited number of genes (1). The pre-mRNA splice sites are marked by relatively short, consensus motifs that can vary in length and sequence. Uridine (U)-rich polypyrimidine (Py) signals precede the major class of 3' splice sites. Yet, purines often interrupt Py tract signals and can regulate alternative 3' splice site selection in multicellular eukaryotes (2).

The essential pre-mRNA splicing factor U2AF2 (also called U2AF⁶⁵) recognizes the Py tract signal to promote the earliest stage of pre-mRNA splicing. The U2AF2 protein forms a ternary complex with SF1 and U2AF1 (also called U2AF³⁵), which ensures 3' splice site fidelity by identifying the branchpoint and AG consensus sequences flanking the Py tract. In a series of ATP-dependent steps, the 5' and 3' splice sites ultimately are positioned for catalysis in the active spliceosome. Breakthrough cryo-electron microscopy structures have revealed the later stages of spliceosome assembly (reviewed in (3)), whereas piecewise X-ray crystallography and NMR structures provide snapshots of splicing factor domains during the transient, early stages of 3' splice site recognition. The U2AF2 protein recognizes the Py tract via two tandem RNA recognition motifs (RRM1 and RRM2). In the absence of RNA or in the presence of degenerate Py tracts comprising less than four consecutive uridines, U2AF2 adopts a “closed” conformation in which RRM1 is masked and only RRM2 is available for RNA binding (4-6). When bound to a longer uridine tract such as the 3' splice site consensus, the U2AF2 RRMs have an “open”, side-by-side conformation with both RRM1 and RRM2 contacting the RNA (4,7). Both RRMs prefer uridines(8,9), although the N-terminal RRM1 is more tolerant of cytidine and purine substitutions in the Py tract than is RRM2 (10,11). In particular, the uridine-specificity of a promiscuous RRM1 site can be enhanced by a structure-guided mutation (10). Yet, unlike the well-characterized RRM1 and RRM2 of U2AF2, the sequence specificity of the RRM1/RRM2 interface for the central nucleotide of the Py tract is unknown.

U2AF2 defects have been associated with a variety of human diseases. Acquired U2AF2 mutations recur among certain cancers (12-14), although with lower frequency than in the U2AF1 subunit (15). De novo mutations of U2AF2 are significantly associated with developmental delay and malformation (16). U2AF2 binding to *RP2* and *NF1* Py tracts is reduced by purine substitutions associated with retinitis pigmentosa and neurofibromatosis (10). U2AF2 has been shown to regulate splicing of an *IL7R* exon that is dysregulated in autoimmune disorders including multiple sclerosis (17). Moreover, disrupted association between U2AF2 and PTEN correlates with autism spectrum disorder (18). Structure/function studies of these disease-associated U2AF2 mutations highlight key interfaces for the normal functions of the protein and provide insight into mechanisms of disease progression. However, understanding the normal sequence specificity and adaptability of the protein is an important baseline for comparison with disease-associated mutants.

Here, we investigate the interactions and nucleotide sequence specificity of the U2AF2 RRM1/RRM2 interface. By X-ray crystallography and complementary molecular dynamics simulations, we find that a protein scaffold accommodates bulky purines at the RRM1/RRM2 interface by repositioning the central nucleotides of the bound Py tract. The specificity of the pre-mRNA splicing and RNA binding activities of U2AF2 and its structure-guided variants was relaxed at the central Py tract position. In human cells, we found that the nucleotide consensus was more variable at the central positions of sequence logos for U2AF2 binding to sites that were otherwise uridine-rich. These results reveal that U2AF2, a key factor for early spliceosome assembly, adapts to natural splice site variations by offering alternative binding sites for different RNA conformations.

MATERIALS AND METHODS

Preparation of U2AF2^{12L} proteins and oligonucleotides

The wild-type and mutant U2AF2^{12L} proteins (residues 141–342 of NCBI RefSeq NP_009210) were expressed and purified as described (7,12). The final protein buffer was 100 mM NaCl, 15 mM HEPES pH 6.8, 0.2 mM TCEP following size exclusion chromatography. Purified, deprotected RNA oligonucleotides were purchased from Horizon Discovery Ltd.

Fluorescence anisotropy RNA binding assays

The RNA-binding experiments followed protocols described in (12,19). The 5'-fluorescein-labeled, RNA oligonucleotides were diluted >100-fold to 30 nM final concentration in a binding buffer comprising 100 mM NaCl, 15 mM HEPES at pH 6.8, 0.2 mM TCEP, 0.1 U ml⁻¹ Suprase-InTM (InvitrogenTM). The changes in total volume following addition of the protein were <10% to minimize dilution effects. The fluorescence anisotropy changes during titration were measured using a FluoroMax-3 spectrophotometer, temperature-controlled at 23 °C by a circulating water bath. Samples were excited at 490 nm and emission intensities recorded at 520 nm with slit widths of 5 nm. The fluorescence emission spectra also were monitored for similarity throughout the experiment. Each titration was fit with a nonlinear equation (12,19) to obtain the apparent equilibrium dissociation constant (K_D). These fits and the *p*-values of a two-tailed unpaired t-test with Welch's correction were calculated using Prism v6.0 (GraphPad Software Inc.). The apparent equilibrium affinities (K_A) are the reciprocals of each K_D . The average K_D or K_A values and standard deviations are given for three replicates of each experiment.

Crystallization, data collection, and structure determination

Crystallization conditions were similar to those described (12). Following concentration to 20 mg mL⁻¹, U2AF2^{12L} protein was mixed with 1.2-fold molar excess purified oligonucleotide variant (5'-phosphoryl-UU(dU)NU(5BrdU)CC-3', where N is cytosine (C5), adenosine (A5), or guanosine (G5)). Crystals were obtained by hanging drop vapor diffusion experiments with precipitants composed either of 0.60 M succinic acid, 0.10 M HEPES pH 7.0, 2% PEG monomethyl ether 2000 (C5) or 0.24 M Na malonate, 26% PEG 3350 (G5, A5). Addition of 0.1 μL of 5% w/v LDAO detergent (Hampton Research) to the G5 or A5 drops and 10% sucrose to the A5 drops prior to incubation improved crystal quality. Crystals were flash-cooled in liquid nitrogen after coating with a mixture of 1:1 (v/v) paratone-N and silicone oil (G5), or sequential transfers to precipitant solutions containing either 21% glycerol (C5) or 28% sucrose/8% PEG 200 (A5). Crystallographic data sets were collected at 100 K by remotely using the Stanford Synchrotron Radiation Light (SSRL) source Beamline 12-2 (20) and processed using the SSRL AUTOXDS script (A. Gonzalez and Y. Tsai) implementation of XDS (21) and CCP4 packages (22). The structures were determined using the Fourier synthesis method starting from PDB code 6XLW. The models were adjusted using COOT (23) and refined using PHENIX (24). The crystallographic data and refinement statistics are given in **Supplementary Table S1** and reduced-bias electron density maps (25) are shown in **Figure S1**.

Molecular dynamics simulations and analysis

Molecular dynamics (MD) simulations were run using Amber 18 (26). The U2AF2-A5, U2AF2-G5 and U2AF2-U5 crystal structures were solvated in a truncated octahedron of OPC water (27) with a 12 Å margin of the solute using Leap. The system was neutralized using 8 Na⁺ atoms, and 20 Na⁺ and Cl⁻ ions were added to model NaCl at a bulk concentration of 150 mM (28). The starting structures were energy-minimized using the steepest descent and then conjugate gradient methods, each for 500 steps. Subsequently, the systems were heated to 298K in 200 ps with a timestep of 2 fs. Then, dynamics were

performed for 500 ps with structures saved at every 100 ps. These five structures were used to run the final production dynamics for 1 μ s using Amber ff14SB (29) + RNA.OL3 (30-32) forcefields with periodic boundary conditions, using a 2 fs timestep and a direct space cutoff of 10 Å for non-bonded interactions. The structures were written to a trajectory file every 100 ps. Pressure was maintained at 1 atm using a Monte Carlo barostat and the temperature was maintained at 298K using Langevin thermostat with a collision frequency of 1.0 ps⁻¹. For the oligonucleotide-only simulations, the U2AF2 protein coordinates were removed to generate the starting structures, then the same steps used for the protein–RNA complex were followed.

For analysis of MD simulations, all the trajectories were merged, and the water and ions were removed using Ambertools 18 (33). The trajectories were aligned using the C α of RRM1s with the starting structures for U2AF2-RNA simulations and six-membered base rings for the simulations of the isolated oligonucleotide, using *aligner* in LOOS (34). Root mean square fluctuations were calculated for six-membered rings of RNA residues using *rmsf* in LOOS (34). Root mean squared deviations (RMSD) of the C α were calculated using *rmsd2ref* tool in LOOS. Pairwise RMSD was calculated using custom python script, *rmsds-align.py*.

Tissue culture, immunoblotting and RT-PCR for minigene analysis

Human cell culture, reverse transcriptase polymerase chain reactions, and immunoblotting were completed essentially as described (12). Human embryonic kidney epithelial cells (HEK 293T, ATCC[®] CRL-3216TM) maintained at 37 °C in a humidified atmosphere containing 5% CO₂, were cultured in Dulbecco's modified Eagle's medium (ThermoFisher Sci., Cat. No. 11995-065), 10% fetal calf serum (R&D Systems Inc.), and penicillin-streptomycin (GibcoTM). The unmodified or *pyPY*-expressing cells were co-transfected in 12-well plates at 80% confluency with 0.5 μ g *pyPY* plasmid (original construction described in (35)) with 0.5 μ g empty pCMV vector, or 0.25 μ g pCMV6-XL5 encoding wild type or mutant human U2AF2 (NCBI RefSeq NP_009210) plus 0.25 μ g empty vector, using jetPRIME[®] (Polyplus-transfection[®]) as instructed by the manufacturer. The cells were harvested 24 hours after transfection for RNA and protein analysis.

Whole cell lysates for immunoblots were prepared in a buffer containing 50 mM Tris, pH 8.0, 10 mM EDTA, 1% w/v SDS, 1 mM DTT, and protease inhibitors and the protein content quantified using a detergent-compatible protein assay (Bio-Rad). Lysates (10 μ g) were separated by electrophoresis on 10% SDS-PAGE gels, then transferred to polyvinylidene difluoride membranes. Primary antibodies probed for U2AF2 (MilliporeSigma, Cat. No. U4758, 1:1000 dilution), H2B (Cell Signaling, Cat. No. 12364, 1:1000 dilution), or GAPDH (Cell Signaling, Cat. No. 2118S, 1:1000). Secondary antibodies included anti-mouse IgG horseradish peroxidase or anti-rabbit IgG horseradish peroxidase (Cytiva Inc., Cat. Nos. NA931 or NA934, 1:5000 dilution). The chemiluminescent developing reagents were Amersham ECLTM (Cytiva Inc., for U2AF2 or GAPDH) or Clarity blotting substrate (Bio-Rad, for H2B), detected on a ChemidocTM Touch Imaging System (Bio-Rad Lab. Inc.).

Total RNA was extracted using the RNeasy Plus Mini kit (Qiagen) from cells homogenized on QIAshredder columns (Qiagen). RNA extracts were treated with RNase-free DNase (Qiagen). Reverse transcription was performed with MMLV-RT (InvitrogenTM) and random primers according to manufacturer's protocol. The RT-PCR reactions of the *pyPY* minigene comprised an initial denaturation for 3 min then 25 cycles of 94 °C for 30 s, 55 °C for 10 s, 72 °C for 20 s. Products were analyzed by agarose gel electrophoresis with ethidium bromide staining. The quantitative real-time RT-PCR reactions of cDNAs were run with SYBRTM Green in triplicate using a Bio-Rad CFX thermal cycler and quantified by the standard curve method. The primer sequences were identical to those reported in (12).

Enhanced crosslinking and immunoprecipitation

U2AF2 eCLIP-seq was performed in duplicate following the protocol in (36) with modifications reported in (37). In summary, human erythroleukemia (HEL) cells (ATCC, Cat #TIB-180) were subjected to UV-crosslinking and U2AF2–RNA complexes were immunoprecipitated with 8 µg anti-U2AF2 antibody (Sigma-Aldrich, Cat #U4758) and Dynabeads Protein G (ThermoFisher Sci., Cat #10004D). RNA was partially digested with RNase I (ThermoFisher Sci., Cat #AM2295) and P³²-labeled (PerkinElmer, Cat #BLU002Z250UC), followed by RNA linker ligation. After SDS-PAGE and transferring to nitrocellulose membrane, a region between 65–110 kD was excised to obtain U2AF2-bound RNA complexes (**Figure S4**). RNA was isolated after treatment with proteinase K and subjected to library preparation. Libraries were sequenced on Illumina NovaSeq 6000 system at the Yale Center for Genome Analysis (YCGA). The U2AF2 eCLIP-seq reads were processed according to the pipeline reported in (37). After duplicate removal (FastUniq (38)) and adapter trimming (Cutadapt (39)), reads were aligned to the human genome (GRCh38.p10) with STAR (version 2.7.0f, GENCODE Release 27 for transcript annotation). Crosslinked nucleotides were extracted from BAM files considering the genomic position right after the end of each sequenced read. Bound junctions were confidently identified considering a nucleotide region from -40 to +10 around the 3' splice site in all the annotated splice junctions in the human genome and using a coverage threshold of at least 10 reads. Binding metaprofiles were built after trimming outlier signals at each nucleotide position from -20 to +5 around the 3' splice site.

RESULTS

U2AF2 has little sequence preference for the central Py tract nucleotide

To fill a missing gap in previous studies of U2AF2–RNA sequence specificity (10,11), we set out to determine the nucleotide-binding preferences of the RRM1/RRM2 interface and integrated inter-RRM linker (**Figure 1**). Since the central (fifth) position of a nine-nucleotide binding site is located at the junction of the RRMs, we addressed this question by comparing the binding affinities of U2AF2 for nine-nucleotide RNAs substituted with U, C, G, or A at the fifth nucleotide. We focused on the Py tract recognition domain of U2AF2, comprising RRM1, RRM2, and N-/C-terminal extensions (U2AF2^{12L}) (**Figure 1A, inset**). The U2AF2^{12L} region binds RNA with similar specificity and affinity as the full length U2AF2 (7). The changes in fluorescence anisotropy during titration of 5'-fluorescein-labeled oligonucleotides with protein were fit to obtain apparent equilibrium dissociation constants (K_D) as described (e.g., (19)).

The U2AF2^{12L} protein bound to the nine-nucleotide RNA corresponding to the sequence of a prototypical, strong Py tract from the Adenovirus Major Late Promoter (*AdML*) (**Figure 1A-B**) with approximately three-fold lower affinity than to a previously studied, 13-nucleotide Py tract from the same intron (K_D 100 nM vs. 30 nM) (7). Either avidity or the lack of a terminal phosphate, which we include in the co-crystallized oligonucleotide, could account for this difference. Substitution of a cytidine (C5) for the central, fifth uridine (U5), located between the RRM1 and RRM2 of the U2AF2^{12L} structure (4,7), does not significantly change the binding affinity. For purine substitutions, a guanosine (G5) incurred a subtle, less than two-fold penalty, whereas an adenosine (A5) produced a more substantial, four-fold decrease (1 kcal mol⁻¹). This discrimination against adenosine could contribute to defining the AG-exclusion zone, a region devoid of AG-dinucleotides between the branchpoint and bona-fide AG at the 3' splice site (40).

We next introduced the G5-substitution in the context of a consensus uridine tract (**Figure 1C-D**). The U2AF2^{12L} protein bound the uridine-tract with similar affinity as the *AdML* Py tract, consistent with a sequence difference of only two terminal cytidines. Likewise, a G5 substitution made little

difference in the binding affinity of U2AF2^{12L} (~0.3 kcal mol⁻¹). We then evaluated the consequence of a guanosine-substitution at the neighboring site, G6. Although the G6-induced change in U2AF2^{12L} binding affinity was moderate, the three-fold decrease was comparable to the penalties for U2AF2 binding of disease-associated mutations in the *RP2* and *NFI* Py tracts (10). Addition of G5 to the G6 substitution (G5G6) had no additional effect, again reflecting the promiscuity of the inter-RRM binding site.

Local shifts of the central nucleotides adapt to the U2AF2^{12L} structure

To view how U2AF2 adapts to different nucleotides at the RRM1/RRM2 interface, we determined three crystal structures of U2AF2^{12L} bound to U tracts containing C, G, or A substituted at the central position (**Figure 2, Supplementary Table S1**). To facilitate crystallization, we used an eight-nucleotide site that omits the 5' uridine while maintaining the original numbering of the nine-nucleotide site (here starting at nucleotide #2). The overall conformations of the protein backbones remained unchanged (0.1–0.3 Å pairwise RMSD between matching Cα atoms of C5, A5, or G5-containing structures compared to the U5 structure) (**Figure 2E**). Despite differences in the central, inter-RRM region, the nucleotides bound to the respective RRM2 and RRM1 also share similar positions (0.2–0.4 Å pairwise RMSD between all atoms of nucleotides 2–4/7–9 of C5, A5, or G5-containing structures compared to the U5 structure). However, substitutions for the central nucleotide dramatically shifted the local positions of the U2AF2-bound RNA (**Figure 2F, Figure 3, Supplementary Movies S1-S3**). A cytidine or adenosine (C5 or A5), for which the hydrogen bond groups differ from uridine, rotated approximately 25° away from the U2AF2 inter-RRM linker relative to the U5 position. Notably, networks of ordered water molecules filled the resulting gaps and mediated contacts between the extruded cytosine or adenine bases and the protein backbone (**Figure 3B, D and Supplementary Figure S1**). A guanosine at the central position (G5), on the other hand, shares identical hydrogen bonding with the protein as the base edge of uracil. Accordingly, the G5 substitution remains in a similar position as the U5 parent (**Figure 3A, C**).

Interestingly, the adjacent uridine on the 3' side (U6) also shifts position when purine nucleotides are substituted at the fifth site (**Figure 2F, Figure 3**). In the U2AF2-bound, all-uridine oligonucleotide, loops of the RRM2 and RRM1 folds sandwich the U6 base. In the presence of bulky A5 or G5 purines, the downstream U6 rotated approximately 25° away from the inter-RRM linker to settle in an alternative binding site, which also is located between the RRM1 and RRM2 loops. To achieve a comparable position of U6 despite the different conformations of A5 and G5, the A5-linked U6 phosphate rotated into an unusual conformer over the ribose group (**Figure 3D, Supplementary Movie S3**). Unlike the original position, no direct or water-mediated contacts with the protein were detected at the alternative U6 binding site in either purine-containing structure. Instead, several ordered water molecules that mediated U6 contacts with U2AF2 in the U5/C5 structures appeared absent in the presence of the purine substitutions (**Figure 3, Supplementary Figure S1**). The purine-induced perturbations of the adjacent U6 site, coupled with the shifted position of A5, can account for the subtle differences in U2AF2 binding affinity (U5/C5 > G5 > A5) for the oligonucleotides (**Figure 1B**).

U2AF2^{12L}-bound Py tract RNA is dynamic at the central nucleotides

To explore the conformations of the U2AF2–Py tract complex beyond the environment of the crystal structures, we performed all-atom molecular dynamics simulations using Amber (26). First, we ran 2 μs simulations of the A5, U5, and G5 crystal structures, repeated five times each. Each protein-RNA structure was stable (**Supplementary Figure S2**), and pairwise RMSD plots (**Supplementary Figure S3**) demonstrated convergence. The RRMs were found to be relatively static (**Figure 4A**). Most of the linker region connecting the RRMs was flexible in the simulations (**Figure 4B**). However, residues

250–255, the linker region bound to the central nucleotide of the Py tract, was static. The nucleotides bound to the RRM1s were found to fluctuate less compared to the central nucleotide (**Figure 4C**).

We also tested whether the conformation of the central nucleotide is related to an intrinsic property of the oligonucleotide. We ran five, 1 μ s all-atom simulations of oligonucleotides (A5, U5, and G5) in the absence of the protein. These simulations of the oligonucleotides exhibited substantial conformational fluctuations compared to the oligonucleotides bound to U2AF2 (**Figure 4D**). Specifically, the pairwise RMSD plots (**Supplementary Figure S3**) demonstrated no innate preferred conformation for the RNA. This suggests that the RNA is flexible in nature, allowing the central nucleotide to adopt a conformation that accommodates protein binding.

Structure-guided mutations enhance the specificity of U2AF2^{12L} for a central uridine

We examined the structures of U2AF2 bound to the different nucleotides, and considered whether the stringency for binding uridine at the central position of the Py tract could be increased by replacements of key interacting amino acids (**Figure 5**). First, we predicted that substituting the shorter side chain of asparagine for R227 would disrupt interactions with the guanosine exocyclic amine. Likewise, we thought that a K225N replacement would reduce U2AF2 binding to G5- more than to U5-containing RNAs, since K225 mediates one of the few contacts with the guanosine-containing site. Accordingly, the R227N and K225N variants of the U2AF2^{12L} protein showed six- and three-fold preferences for binding U5- over G5-containing oligonucleotides (**Figure 5A, C, and F**). Substituting either of these positively-charged residues with negatively-charged glutamates (R227E and K225E) significantly reduced the RNA binding affinity (by 80- and 20-fold compared to the wild-type protein) (**Figure 5B, D, and F**), most likely by electrostatic repulsion of the phosphodiester backbone. Next, we tested whether introducing an aspartate instead of a glycine near the purine-associated conformation of U6 (G297D) would selectively disfavor the G5-associated conformation. The G297D replacement increased the U5:G5 specificity of U2AF2^{12L} to approximately four-fold by reducing binding to the G5-containing oligonucleotide (**Figure 5E-F**). The results of these structure-guided mutants demonstrated that the specificity of U2AF2 for a uridine at the center of the Py tract could be increased by amino acid variations.

A model Py tract with a central guanosine efficiently supports splicing in cells

We next investigated the abilities of sequence variations of a model Py tract to support splicing in human cells (HEK 293T) (**Figure 6A-C, Supplementary Figure S4A-B**). We hypothesized that purine substitutions at the center of a uridine-rich Py tract would have a lower penalty for splicing than when located at surrounding positions. To test this hypothesis, we leveraged a well-characterized *pyPY* minigene (28), in which the *AdML* intron (“*PY*”) and exon are fused to *IgM* M1 and M2 exons (“*py*”) (**Figure 6A**). We prioritized guanosine over adenosine substitutions to avoid introducing cryptic splice site signals (40). We substituted single guanosines for each of five core uridines in the eight-uridine *PY* tract (-8 to -12 nucleotides from the intron-exon junction), which shares the sequence of the co-crystallized oligonucleotide.

Following expression of the minigene in human cells (HEK 293T), the spliced products were evaluated by RT-PCR and quantitative real-time RT-PCR (**Figure 6B-C**). In the absence of exogenous factors or substitutions, the *PY*-associated splice site was preferentially spliced in the HEK 293T cells. In most cases, guanosine mutations of the *PY* tract shifted the equilibrium towards splicing of the competing *py* site. Co-expression of U2AF2 with the minigene transcript generally increased splicing of the weak *py* site, consistent with previous observations (7,12). An exception to this trend resulted from the -10U>G mutation at the central *PY* nucleotide, which did not penalize *PY* splicing and instead slightly enhanced the *PY:py* product ratio. In agreement with the U2AF2 structures and binding affinities, these results indicated that a prototypical Py tract with guanosine interruptions was capable

of supporting splicing in cells and that a guanosine at the central position favored splicing more than when placed at other positions.

To further relate altered splicing of the guanosine-substituted PY tracts and the action of U2AF2, we next tested the effect of overexpressing U2AF2 variants with structure-guided R227E (**Figure 6D-E, Supplementary Figure S4C-D**). Consistent with its decreased RNA binding affinity, the R227E mutant decreased splicing of the U2AF2-sensitive *AdML* PY tract and its guanosine-containing variants. This mutation of RRM1 had little impact on the *IgM py* site, possibly because a closed U2AF2 conformation that binds degenerate Py tracts (such as *py*) relies on RRM2 (4). Consequently, the R227E mutant U2AF2 reduced the apparent ratios of *PY:py* splicing slightly more than the wild-type protein. Taken together with the diminished RNA binding affinity of the R227E mutant, these results support a conclusion that U2AF2 contributes to the different guanosine-sensitivities of central vs. flanking sites in a model Py tract.

U2AF2 interaction sites in human cells agree with U2AF2^{12L}-RNA binding specificity

While informative, these experiments left open the relationship between the central nucleotide site of the U2AF2^{12L} crystal structures and the natural binding sites for U2AF2 in cells. To further understand the organization of U2AF2 and the 3' splice site, we used the enhanced crosslinking and immunoprecipitation (eCLIP) assay (36,37,41) to map the RNA interactome of U2AF2 in human erythroleukemia (HEL) cells. Following U2AF2 immunoprecipitation and ³²P labeling of the crosslinked RNA, the immunoprecipitated complexes were separated by denaturing gel electrophoresis (**Supplementary Figure S5**). We focused on analyzing the region with a molecular weight between 65-110 kD, corresponding to the expected size of U2AF2-RNA complexes. We aligned and mapped an average of 86.2% usable reads from two biological replicates to identify specific U2AF2-binding locations in 78,720 splice junctions across the human transcriptome.

As expected, significant peaks for U2AF2 eCLIP-seq occurred in uridine-rich regions upstream of 3' splice site junctions (**Figure 7**). To investigate the relationship between U2AF2 binding and the sequence-content of 3' splice site signals, we divided the splice site junctions into three classes of uridine enrichment. These included splice sites with low (0-2), medium (3-5), or high (6-8) numbers of uridines in the zone from -4 to -11 nucleotides upstream of the intron 3' end (**Figure 7A**). The U2AF2 contacts with endogenous splice sites shifted position depending on the local uridine content. In particular, the interaction peak was narrowest, strongest, and closest to the intron 3' end for the high uridine class. Conversely, the peak was broader and more distant from the intron 3' end for the splice site junctions with few uridines (**Figure 7B**). Motif analysis of the high uridine-containing class showed two clusters of approximately two highly conserved uridines surrounding a core of less conserved uridines at the central positions. These results demonstrated that the U2AF2 binding sites are responsive to the uridine contents and locations within the pre-mRNA splice site signals. Additionally, the sequence logos of the U2AF2 interactions in human cells agreed with the RNA-sequence preferences of the U2AF2^{12L} protein.

DISCUSSION

Here we show that local rearrangements of the bound RNA, rather than protein backbone, enable U2AF2 to accommodate different nucleotides at the center of the Py tract. At the RRM1/RRM2 interface, U2AF2 selects an RNA conformation with the noncognate base (particularly C or A) rotated away from the RRM1/RRM2 interface, and instead engages intermediary water molecules to “glue” the RNA and protein surfaces. To fit bulky purines at this site, U2AF2 facilitates readjustment of the RNA backbone by offering an alternative binding site for the adjacent uridine. Otherwise, U2AF2

maintains unperturbed contacts between the RRM1s and the surrounding nucleotides of the bound Py tract, which is underscored by predictable, structure-guided increases in RRM2-RNA specificity (here and (10)). Already, flexible RNA conformations have been shown to help U2AF2 accommodate different splice site sequences at other nucleotide binding sites. In particular, we previously observed a noncognate *syn*-guanosine bound to a promiscuous site of U2AF2 RRM1 (10). A cytosine at this site likewise shifts its position to optimize hydrogen bonds with the U2AF2 backbone and side chains (11). The expanding views of U2AF2 bound to diverse RNA sites reinforce an emerging theme among ribonucleoprotein structures, in which the RNA conformation adapts to the protein surface as often as the protein adapts to the RNA. For example, *syn/anti* base flipping enables SRSF2 to recognize either tandem cytosines or guanosines with similar affinities (42). In the structures of Csr/Rsm with various noncoding RNA substrates, rearrangements of bound nucleotides facilitate recognition of the different RNA sequences (43). In another well-studied example, one mechanism for PUF family repeat proteins to bind a large set of degenerate RNA sequences is to eject noncognate nucleotides from the modular RNA binding surface (44). Altogether, these findings highlight the importance of RNA flexibility for proteins to associate with appropriate sites amidst the milieu of cellular RNAs.

Starting from the U2AF2–RNA crystal structures, molecular dynamics simulations reveal that the oligonucleotide is inherently flexible in the absence of protein, and that the central nucleotide remains flexible in the U2AF2-bound complex. Although more studies of ribonucleoproteins focused on the dynamics of the protein than on the RNA components, RNA flexibility clearly is an important contributor to versatile RNA–protein recognition. Several proteins have been shown to select an RNA structure with optimal intermolecular contacts among multiple conformations sampled by the protein-free RNA site (45–47). Indeed, a survey of RNA-binding proteins in the bound and free states implies that nucleic acid movements are a key aspect of protein-RNA recognition (48). In some cases, nucleotides making important contacts increase (rather than diminish) dynamics in the protein complex compared to the free state (46,47). Here, molecular dynamics simulations demonstrate that the Py tract RNAs likewise possesses a conformational repertoire in the absence of protein cofactors. Accordingly, polyuridine lacks a uniform structure in solution and shows the least base-stacking among the nucleotide polymers (49,50). From the ensemble of Py tract RNA conformations, we propose that U2AF2 selects a particular RNA conformation, thereby optimizing the intermolecular contacts with the altered central nucleotide and adjacent uridines. The molecular dynamics simulations further suggest that the central nucleotides remain flexible in the U2AF2–RNA complex, such as observed for other RRM-bound RNAs (46,47).

We view the ability of metazoan U2AF2 to structurally adapt to diverse splice sites as a key characteristic for its function. The transcriptome of human cells presents a vast number of potential splice sites for selection by U2AF2 during the initial stages of spliceosome assembly. Indeed, transcriptome-wide mapping of U2AF2 binding sites in cells (**Figure 7** and (51)) demonstrates broad association of U2AF2 with a plethora of RNA sites with varying sequences. Yet, accurately identifying 3′ splice site signals is critical for gene expression. As such, U2AF2 transcends a traditional classification of either a “specific” or “nonspecific” RNA binding protein. One means for U2AF2 to fulfill its multifaceted role is to rely on a modular architecture of tandem RRM1s, which differ in uridine-specificity and switch between “open” and “closed” conformations in response to the RNA sequence (4,6,11). Prior studies of U2AF2 RRM1/RRM2 bound to noncognate RNAs reveal a variety of changes, ranging from subtle shifts of the side chains and protein backbone to nucleotide rotations and *syn/anti*-conformer flips (10,11). Here, we have added flexibility of the central Py tract nucleotides to the repertoire of adaptive U2AF2–RNA recognition. The ability of structure-guided mutations (G297D here and D231V in (10)) to increase the uridine-specificity of U2AF2 implies that this splicing factor has evolved a subtle RNA binding specificity that serves its general function of identifying the major class of 3′ splice sites.

Often the specificity of modular RNA binding proteins is assisted in *trans* by auxiliary subunits. Multiple partners work to enhance and regulate U2AF2, including U2AF1, SF1, SF3B1, and PUF60/RBM39, among others. Cancer-associated mutations, most commonly in U2AF1 and in some cases U2AF2, also influence 3' splice site recognition (52-54) and the distribution of U2AF2 conformations (6). Resolving the influence of additional subunits and cancer-associated mutations on the U2AF2–RNA conformation remains an important direction for future research.

DATA AVAILABILITY

Data deposition: The coordinates for the U2AF structures have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 7S3A, 7S3B, 7S3C for C5, G5, and A5 structures). U2AF2 eCLIP-seq files are in the process of deposition in the GEO database, <https://www.ncbi.nlm.nih.gov/geo/> (GSEXXXXX). In the interim, the U2AF2 eCLIP-seq files are available at the following URL: https://drive.google.com/drive/folders/1I64zmdwpbqLt_c6ucohdp85qNTggK55p?usp=sharing

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank S. Henderson for preliminary refinement of the C5 and G5 structures.

FUNDING

This work was supported by National Institutes of Health grants R01 GM070503 to C.L.K., R01 GM132185 to D.H.M., and R01 DK102792 to S.H. The Edward P. Evans Foundation supported work in the labs of C.L.K. and S.H.

REFERENCES

1. Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature*, **456**, 470-476.
2. Nguyen, H. and Xie, J. (2018) Widespread Separation of the Polypyrimidine Tract From 3' AG by G Tracts in Association With Alternative Exons in Metazoa and Plants. *Front Genet*, **9**, 741.
3. Wan, R., Bai, R. and Shi, Y. (2019) Molecular choreography of pre-mRNA splicing by the spliceosome. *Curr Opin Struct Biol*, **59**, 124-133.
4. Mackereth, C.D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcarcel, J. and Sattler, M. (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature*, **475**, 408-411.
5. Voith von Voithenberg, L., Sanchez-Rico, C., Kang, H.S., Madl, T., Zanier, K., Barth, A., Warner, L.R., Sattler, M. and Lamb, D.C. (2016) Recognition of the 3' splice site RNA by the U2AF heterodimer involves a dynamic population shift. *Proc Natl Acad Sci U S A*, **113**, E7169-E7175.
6. Warnasooriya, C., Feeney, C.F., Laird, K.M., Ermolenko, D.N. and Kielkopf, C.L. (2020) A splice site-sensing conformational switch in U2AF2 is modulated by U2AF1 and its recurrent myelodysplasia-associated mutation. *Nucleic Acids Res*, **48**, 5695-5709.
7. Agrawal, A.A., Salsi, E., Chatrikhi, R., Henderson, S., Jenkins, J.L., Green, M.R., Ermolenko, D.N. and Kielkopf, C.L. (2016) An extended U2AF⁶⁵-RNA-binding domain recognizes the 3' splice site signal. *Nat Commun*, **7**, 10950.
8. Singh, R., Valcarcel, J. and Green, M.R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*, **268**, 1173-1176.
9. Singh, R., Banerjee, H. and Green, M.R. (2000) Differential recognition of the polypyrimidine-tract by the general splicing factor U2AF⁶⁵ and the splicing repressor Sex-lethal. *RNA*, **6**, 901-911.
10. Agrawal, A.A., McLaughlin, K.J., Jenkins, J.L. and Kielkopf, C.L. (2014) Structure-guided U2AF⁶⁵ variant improves recognition and splicing of a defective pre-mRNA. *Proc Natl Acad Sci U S A*, **111**, 17420-17425.
11. Jenkins, J.L., Agrawal, A.A., Gupta, A., Green, M.R. and Kielkopf, C.L. (2013) U2AF⁶⁵ adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res*, **41**, 3859-3873.
12. Maji, D., Glasser, E., Henderson, S., Galardi, J., Pulvino, M.J., Jenkins, J.L. and Kielkopf, C.L. (2020) Representative cancer-associated U2AF2 mutations alter RNA interactions and splicing. *J Biol Chem*, **295**, 17148-17157.

13. Glasser, E., Agrawal, A.A., Jenkins, J.L. and Kielkopf, C.L. (2017) Cancer-associated mutations mapped on high-resolution structures of the U2AF2 RNA recognition motifs. *Biochemistry*, **56**, 4757-4761.
14. Kralovicova, J., Borovska, I., Kubickova, M., Lukavsky, P.J. and Vorechovsky, I. (2020) Cancer-Associated Substitutions in RNA Recognition Motifs of PUF60 and U2AF65 Reveal Residues Required for Correct Folding and 3' Splice-Site Selection. *Cancers (Basel)*, **12**, 1865.
15. Kielkopf, C.L. (2017) Insights from structures of cancer-relevant pre-mRNA splicing factors. *Curr Opin Genet Dev*, **48**, 57-66.
16. Hiraide, T., Tanaka, T., Masunaga, Y., Ohkubo, Y., Nakashima, M., Fukuda, T., Ogata, T. and Saito, H. (2021) Global developmental delay, systemic dysmorphism and epilepsy in a patient with a de novo U2AF2 variant. *J Hum Genet*.
17. Schott, G., Galarza-Munoz, G., Trevino, N., Chen, X., Weirauch, M., Gregory, S.G., Bradrick, S.S. and Garcia-Blanco, M.A. (2021) U2AF2 binds IL7R exon 6 ectopically and represses its inclusion. *RNA*.
18. Thacker, S., Sefyi, M. and Eng, C. (2020) Alternative splicing landscape of the neural transcriptome in a cytoplasmic-predominant Pten expression murine model of autism-like Behavior. *Transl Psychiatry*, **10**, 380.
19. Jenkins, J.L., Shen, H., Green, M.R. and Kielkopf, C.L. (2008) Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF⁶⁵. *J Biol Chem*, **283**, 33641-33649.
20. Soltis, S.M., Cohen, A.E., Deacon, A., Eriksson, T., Gonzalez, A., McPhillips, S., Chui, H., Dunten, P., Hollenbeck, M., Mathews, I. *et al.* (2008) New paradigm for macromolecular crystallography experiments at SSRL: automated crystal screening and remote data collection. *Acta Crystallogr D Biol Crystallogr*, **64**, 1210-1221.
21. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr*, **66**, 133-144.
22. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr*, **67**, 235-242.
23. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.
24. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr*, **D66**, 213-221.

25. Afonine, P.V., Moriarty, N.W., Mustyakimov, M., Sobolev, O.V., Terwilliger, T.C., Turk, D., Urzhumtsev, A. and Adams, P.D. (2015) FEM: feature-enhanced map. *Acta Crystallogr.*, **D71**, 646-666.
26. Gotz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S. and Walker, R.C. (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput*, **8**, 1542-1555.
27. Izadi, S., Anandakrishnan, R. and Onufriev, A.V. (2014) Building Water Models: A Different Approach. *J Phys Chem Lett*, **5**, 3863-3871.
28. Schmit, J.D., Kariyawasam, N.L., Needham, V. and Smith, P.E. (2018) SLTCAP: A Simple Method for Calculating the Number of Ions Needed for MD Simulation. *J Chem Theory Comput*, **14**, 1823-1827.
29. Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*, **11**, 3696-3713.
30. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham, T.E., 3rd, Laughton, C.A. and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J*, **92**, 3817-3829.
31. Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E., 3rd and Jurecka, P. (2011) Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J Chem Theory Comput*, **7**, 2886-2902.
32. Wang, J., Cieplak, P. and Kollman, P.A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, **21**, 1049-1074.
33. Case, D.A., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Ghoreishi, D., Gilson, M.K. *et al.* (2018), University of California, San Francisco.
34. Romo, T.D., Leioatts, N. and Grossfield, A. (2014) Lightweight object oriented structure analysis: tools for building tools to analyze molecular dynamics simulations. *J Comput Chem*, **35**, 2305-2318.
35. Pacheco, T.R., Coelho, M.B., Desterro, J.M., Mollet, I. and Carmo-Fonseca, M. (2006) In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol Cell Biol*, **26**, 8183-8190.
36. Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat Methods*, **13**, 508-514.

37. Biancon, G., Joshi, P., Zimmer, J.T., Hunck, T., Gao, Y., Lessard, M.D., Courchaine, E., Barentine, A.E., Machyna, M., Botti, V. *et al.* (2021) Multi-omics profiling of U2AF1 mutants dissects pathogenic mechanisms affecting RNA granules in myeloid malignancies. *bioRxiv*, 2021.2004.2022.441020.
38. Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J. and Chen, S. (2012) FastUniq: a fast de novo duplicates removal tool for paired short reads. *PLoS One*, **7**, e52249.
39. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*, **17**, 3.
40. Gooding, C., Clark, F., Wollerton, M.C., Grellscheid, S.N., Groom, H. and Smith, C.W. (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol*, **7**, R1.
41. Van Nostrand, E.L., Pratt, G.A., Yee, B.A., Wheeler, E.C., Blue, S.M., Mueller, J., Park, S.S., Garcia, K.E., Gelboin-Burkhart, C., Nguyen, T.B. *et al.* (2020) Principles of RNA processing from analysis of enhanced CLIP maps for 150 RNA binding proteins. *Genome Biol*, **21**, 90.
42. Daubner, G.M., Clery, A., Jayne, S., Stevenin, J. and Allain, F.H. (2011) A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J*, **31**, 162-174.
43. Duss, O., Michel, E., Diarra dit Konte, N., Schubert, M. and Allain, F.H. (2014) Molecular basis for the wide range of affinity found in Csr/Rsm protein-RNA recognition. *Nucleic Acids Res*, **42**, 5332-5346.
44. Hall, T.M. (2016) De-coding and re-coding RNA recognition by PUF and PPR repeat proteins. *Curr Opin Struct Biol*, **36**, 116-121.
45. Leulliot, N. and Varani, G. (2001) Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry*, **40**, 7947-7956.
46. Shajani, Z., Drobny, G. and Varani, G. (2007) Binding of U1A protein changes RNA dynamics as observed by ¹³C NMR relaxation studies. *Biochemistry*, **46**, 5875-5883.
47. Oberstrass, F.C., Allain, F.H. and Ravindranathan, S. (2008) Changes in dynamics of SRE-RNA on binding to the VTS1p-SAM domain studied by ¹³C NMR relaxation. *J Am Chem Soc*, **130**, 12007-12020.
48. Ellis, J.J. and Jones, S. (2008) Evaluating conformational changes in protein structures binding RNA. *Proteins*, **70**, 1518-1526.
49. Inners, L.D. and Felsenfeld, G. (1970) Conformation of polyribouridylic acid in solution. *J Mol Biol*, **50**, 373-389.

50. Norberg, J. and Nilsson, L. (1995) Stacking free energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous solution. *J Am Chem Soc*, **117**, 10832-10840.
51. Shao, C., Yang, B., Wu, T., Huang, J., Tang, P., Zhou, Y., Zhou, J., Qiu, J., Jiang, L., Li, H. *et al.* (2014) Mechanisms for U2AF to define 3' splice sites and regulate alternative splicing in the human genome. *Nat Struct Mol Biol*, **21**, 997-1005.
52. Okeyo-Owuor, T., White, B.S., Chatrikhi, R., Mohan, D.R., Kim, S., Griffith, M., Ding, L., Ketkar-Kulkarni, S., Hundal, J., Laird, K.M. *et al.* (2015) U2AF1 mutations alter sequence specificity of pre-mRNA binding and splicing. *Leukemia*, **29**, 909-917.
53. Ilagan, J.O., Ramakrishnan, A., Hayes, B., Murphy, M.E., Zebari, A.S., Bradley, P. and Bradley, R.K. (2015) U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome Res*, **25**, 14-26.
54. Fei, D.L., Motowski, H., Chatrikhi, R., Prasad, S., Yu, J., Gao, S., Kielkopf, C.L., Bradley, R.K. and Varmus, H. (2016) Wild-type U2AF1 antagonizes the splicing program characteristic of U2AF1-mutant tumors and is required for cell survival. *PLoS Genet*, **12**, e1006384.

FIGURES

Figure 1. The specificity of U2AF2 for centrally-substituted Py tract oligonucleotides. The boundary of the U2AF2^{1,2L} construct (blue) used for RNA binding and structure determination is inset in panel (A). Fluorescence anisotropy measurements of U2AF2^{1,2L} titrated into the given RNAs, including (A-B) an *AdML* Py tract (blue) and its central cytosine (mustard), guanosine (salmon), or adenosine (green) substitutions, or (C-D) a nine-uridine tract (blue) and substituted with guanosines at G5 (salmon), G6 (grey), or G5/G6 (orange-striped). The average data points and standard deviations of three experiments are overlaid with the fitted binding curves in A and C. The sequences of the 5'-fluorescein-labeled RNA oligonucleotides are inset, alongside average apparent equilibrium dissociation constants (K_D) with standard deviations of three replicates. Bar graphs of the average apparent RNA binding affinities and standard deviations are shown in B and D. Two-tailed unpaired t-tests with Welch's correction were calculated using GraphPad Prism from comparisons with respective uridine counterparts. *p*-values: n.s., not significant, *, <0.05; **, <0.005; ***, <0.0005; ****, <0.0001.

Figure 2. Crystal structures of U2AF2^{12L} recognizing *AdML* Py tract variants that differ in the identities of the central nucleotide. The amino (N)/carboxy (C)-termini of the polypeptide and 5'/3' termini of the oligonucleotide are labeled. (A-D) Overall ribbon diagrams of the protein (blue) bound to oligonucleotides (grey) substituted with (A) uridine (U5, magenta, PDB ID 6XLW), (B) cytidine (C5, yellow), (C) guanosine (G5, salmon), or (D) adenosine (A5, green). (E) Superposition by matching C α atoms in the four structures shows conformational changes at the fifth and sixth nucleotide positions (numbered according to the nine nucleotide-binding sites of PDB ID 5EV4). (F) Closer view of the fifth and sixth nucleotides from (E), rotated 90° about the y-axis.

Figure 3. U2AF2^{12L} interactions with the fifth and sixth nucleotides of bound Py tracts (numbered with the convention of PDB ID 5EV4). Variants of the fifth nucleotide include (A) uridine (U5, magenta), (B) cytidine (C5, yellow), (C) guanosine (G5, salmon), or (D) adenosine (A5, green). Nitrogens (blue) and oxygens (red) are colored; interacting water molecules are spheres. Hydrogen bonds (<3.1 Å) are indicated by dashed lines. Mutated residues are bold and marked by an asterisk.

Figure 4: Molecular dynamics studies of conformational flexibility. (A) Root mean squared fluctuations (RMSF) of U2AF2 Cα by residue numbers. The RRM1 and RRM2 regions are relatively static. (B) Inset showing RMSF of the inter-RRM linker Cαs (residues 230-260). (C) RMSF values of six-membered rings of the RNA in the U2AF2-RNA complex simulation (2 μs). (D) RMSF values of six-membered rings of the RNA in the oligonucleotide simulations (1 μs). The oligonucleotide simulations have higher fluctuations than the U2AF2-RNA simulations.

Figure 5. Key residues at the U2AF2 interface with the central nucleotide influence RNA binding. (A-E) Average fluorescence anisotropy data points and standard deviations from three replicates of the indicated U2AF2^{1,2L} mutants titrated into 5'-fluorescein-labeled RNA oligonucleotides, overlaid with the fitted curves. The RNA sequences comprise either nine-uridines (blue) or its G5-variant (salmon), and are inset alongside the apparent equilibrium dissociation constants (K_D) and standard deviations. (F) Bar graph of average apparent affinities and standard deviations of U2AF2^{1,2L} mutants binding uridine tracts substituted at the central nucleotide. Two-tailed, unpaired t-tests with Welch's correction are calculated relative to wild-type protein binding the U5 oligonucleotide (asterisks directly above bars), or for the respective protein binding the U5 vs. G5 oligonucleotides (above horizontal lines): p -values: n.s., not significant, *, <0.05; **, <0.005.

Figure 6. G-to-U substitutions at the central position of the *AdML* Py tract have a lesser impact on splicing of the variant site in cells. **(A)** A minigene reporter (*pyPY*) comprises two exons and an intron of an *IgM* gene (“*py*”) fused with the uridine-rich *AdML* splice site (“*PY*”) and downstream exon. The sequences of G-substituted *PY* tract variants are inset and are numbered backwards from the spliced junction. **(B)** Representative RT-PCR of unmodified (“0”) or G-substituted *pyPY* variants transfected into HEK293T cells, together with either a control, “empty” vector (left) or one expressing wild-type (WT) U2AF2 protein (right). **(C)** Quantitative real-time reverse transcription PCR (qRT-PCR) quantifies the relative levels of *PY* compared to *py* splicing for the indicated variants. **(D-E)** Comparison of the structure-guided R227E- mutant with WT U2AF2 protein affecting splicing of G-substituted *AdML* Py tract variants by **(D)** RT-PCR or **(E)** qRT-PCR. Error bars result from three replicates. Two-tailed, unpaired t-tests with Welch’s correction either compare the parent *pyPY* splicing to each of the G-substituted variants with the matching empty or U2AF2-expression vector in panel C or WT to R227E U2AF2 for each of the corresponding *pyPY* variants in panel E. *p*-values: n.s., not significant; *, <0.05; **, <0.005; ***, <0.0005; ****, <0.0001. A 100 bp DNA ladder (Invitrogen) marks the size of the RT-PCR products (Lane M). **Supplementary Figure S4** shows the corresponding immunoblots.

Figure 7. U2AF2–RNA binding adapts to the local uridine content of the 3' splice site *in vivo*. **(A)** Classification of U2AF2-bound splice site junctions according to the number of uridines (Us) in the -11 to -4 region of the 3' splice site. Shaded areas distinguish the three classes of junctions: blue, poor uridine content (0-2 Us); grey, medium uridine content (3-5 Us); red, high uridine content (6-8 Us). This analysis is focused on internal exons of protein coding genes in U2AF2 eCLIP-seq data (replicates, n=2). **(B)** Binding metaprofiles (y-axis, mean $\pm$ SEM of the percentage of crosslinking events) for each class of uridine-containing junctions. Yellow area: -11 to -4 region; N, total number of analyzed splice site junctions. **(C)** 3' splice site sequence logos for each class. N, number of junctions per class. **Supplementary Figure S5** shows steps and results of the U2AF2 eCLIP-seq sample preparation.