Structure-guided U2AF^{65} variant improves recognition and splicing of a defective pre-mRNA

Anant A. Agrawal, Krystle J. McLaughlin[1], Jermaine L. Jenkins, and Clara L. Kielkopf[2]

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

Edited by Daniel A. Pomeranz Krummel, Brandeis University, Waltham, MA, and accepted by the Editorial Board October 29, 2014 (received for review July 6, 2014)

Purine interruptions of polypyrimidine (Py) tract splice site signals contribute to human genetic diseases. The essential splicing factor U2AF^{65} normally recognizes a Py tract consensus sequence preceding the major class of 3′ splice sites. We found that neurofibromatosis- or retinitis pigmentosa-causing mutations in the 5′ regions of Py tracts severely reduce U2AF^{65} affinity. Conversely, we identified a preferred binding site of U2AF^{65} for purine substitutions in the 3′ regions of Py tracts. Based on a comparison of new U2AF^{65} structures bound to either A- or G-containing Py tracts with previously identified pyrimidine-containing structures, we expected to find that a D231V amino acid change in U2AF^{65} would specify U over other nucleotides. We found that the crystal structure of the U2AF^{65}-D231V variant confirms favorable packing between the engineered valine and a target uracil base. The D231V amino acid change restores U2AF^{65} affinity for two mutated splice sites that cause human genetic diseases and successfully promotes splicing of a defective retinitis pigmentosa-causing transcript. We conclude that reduced U2AF^{65} binding is a molecular consequence of disease-relevant mutations, and that a structure-guided U2AF^{65} variant is capable of manipulating gene expression in eukaryotic cells.

pre-mRNA splicing | protein–RNA complex | protein engineering | crystal structure | RRM

Approximately 15% of the documented disease-causing point mutations disrupt consensus splice site elements in pre-mRNAs, including a polypyrimidine (Py) tract between a branch point sequence (BPS) and an AG dinucleotide at the junction of the 3′ splice site (1) (Fig. 1*A*). For example, disease-causing mutations in Py tracts have been documented in ∼3,000 genes in the Human Gene Mutation Database (2), and an estimated 20% of these mutations affect regulatory splice site signals (3, 4). One of the earliest reports of a splice site mutation as a major cause of inherited human disease was for β-thalassemia (reviewed in ref. 5), for which splice site mutations in the human β-globin gene (*HBB*) are found in ∼14% of patients, causing symptoms of mild to severe anemia (reviewed in ref. 6).

With the emergence of high-throughput sequencing technologies, splice site mutations in specific transcripts have been identified as common contributors to neuromuscular disorders, metabolic disorders, cancers, leukemias, deafness, and blindness, among other disorders (reviewed in ref. 4). Retinitis pigmentosa, the most prevalent form of inherited blindness in adults, represents one such disease that is primarily the consequence of mutations in splice sites of vision-relevant transcripts or splicing factors responsible for their recognition (reviewed in ref. 7). Neurofibromatosis type I, a disease characterized by tumors of nerve tissue, is an inherited disorder in which nearly 30% of the documented mutations disrupt neurofibromin 1 (*NF1*) splice sites (reviewed in ref. 8). Despite etiologic progress, the relationships between disease-causing pre-mRNA splice site mutations and downstream inhibition of pre-mRNA splicing factors remain largely unclear at the molecular level.

The Py tract splice site signals of the major class of introns are recognized by the U2 small nuclear ribonucleoprotein (snRNP) auxiliary factor, 65 kDa (U2AF^{65}) (Fig. 1*A*), which acts in a complex with Splicing Factor 1 (SF1) (9) and small (35 kDa) U2AF (U2AF^{35}) subunits (10) that recognize the upstream BPS (11) and consensus AG dinucleotide at the 3′ splice site junction (12–14), respectively. The U2AF^{65}-SF1-U2AF^{35} complex in turn stabilizes the association of core spliceosome components with the pre-mRNA. U2AF^{65} has been shown to bind the SF3b155 subunit of the U2 snRNP (15), which ultimately displaces SF1 (16), whereas SF1 interacts with the U1 snRNP at the 5′ splice site (9, 17) and appears to be dispensable for the splicing of most human transcripts (18, 19). The U2AF^{35} small subunit is an accessory factor to U2AF^{65}, required for splicing a subset of introns with degenerate Py tracts and conserved AG consensus (12, 20). The central U2AF^{65} subunit is required for splicing of most of the major U2 class of introns (21).

The two central U2AF^{65} RNA recognition motifs of U2AF^{65}, RRM1 and RRM2, recognize the Py tract splice site signals (Fig. 1 *A* and *B*) (22). We have contributed milestone structures of the U2AF^{65} RRM1 and RRM2 connected by a shortened inter-RRM linker (dU2AF^{65}) that visualize the nucleotide interactions at a subset of U2AF^{65}-binding sites (23, 24). An NMR structure comprising U2AF^{65} RRM1 and RRM2 (U2AF^{65}1,2) shows side-by-side binding of the tandem RRMs to polyU RNA (25). Nevertheless, the structural basis for the ability of U2AF^{65} to adapt to degenerate purine-containing Py tracts and, conversely, the consequences for U2AF^{65} association of disease-causing purine mutations in human splice sites remain unknown.

Considering its central role in spliceosome recruitment (15, 21, 26), engineering U2AF^{65} variants for improved affinity at specific Py tracts offers a potential approach to increase the use of an adjacent 3′ splice site. In the case of Pumilio homology (Puf) domains, designer RNA-binding proteins that improve splicing have been successfully constructed by fusion with an RS-rich splicing domain [(27); reviewed in ref. 28]; however, these Puf-RS fusions cannot readily substitute for U2AF^{65}, which is a central hub for

Significance

The essential U2AF^{65} protein recognizes a splice site signal that is frequently mutated in inherited human diseases. Herein we show that reduced U2AF^{65} binding is a molecular consequence of splice site mutations that commonly underlie human genetic disease. We demonstrate for a proof-of-principle case that structure-guided U2AF^{65} variants are a feasible tool to evoke disease-relevant changes in pre-mRNA splicing.

Author contributions: A.A.A. and C.L.K. designed research; A.A.A., K.J.M., and J.L.J. performed research; A.A.A., K.J.M., J.L.J., and C.L.K. analyzed data; and A.A.A., K.J.M., and C.L.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. D.A.P.K. is a guest editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4TU7, 4TU8, and 4TU9).

[1]Present address: Department of Biological Sciences, Lehigh University, Bethlehem, PA 18015.

[2]To whom correspondence should be addressed. Email: clara_kielkopf@urmc.rochester.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412743111/-/DCSupplemental.

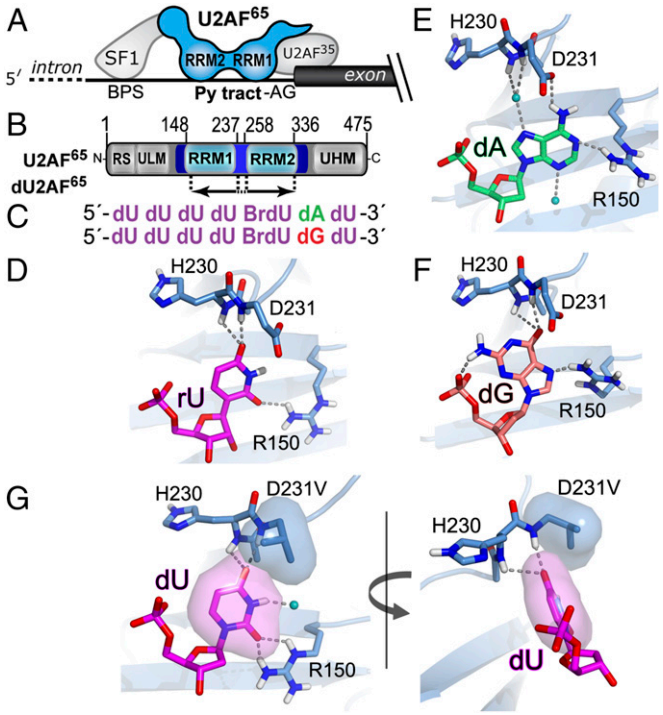

Fig. 1. Schematic diagrams of the U2AF^{65}, SF1, and U2AF^{35} splicing factors recognizing the 3′ splice site *(A)* and U2AF^{65} domains *(B)*. Boundaries of the dU2AF^{65} construct for crystallization (residues 148–336 except an internal linker deletion of residues 238–257) are indicated below. (*C*) Cocrystallized oligonucleotide sequences. (*D–F*) Views of the penultimate dU2AF^{65}-bound nucleotide including rU interactions (PDB ID code 2G4B) (*D*), which are indistinguishable from dU at this site, dA interactions (*E*), and dG interactions (*F*). *(G)* dU2AF^{65}-D231V bound to the dUdUdUdU(5-Br-dU)**dU**dU.

the recruitment of SF1, SF3b155, U2AF^{35}, and UAP56 splicing proteins to the pre-mRNA site. Alternatively, so-called "splice-site switching" oligonucleotides (SSOs) are well-developed substances used to block aberrant splice sites that are currently in clinical trials for the treatment of several major diseases (for examples, see www.isispharm.com; reviewed in ref. 5); however, SSO strategies are limited to a steric-blocking mechanism that is distinct from splice site activation, such as the function of U2AF^{65}. "Tailored" U2AF^{65} variants have the potential to improve or synergize with SSOs for the manipulation of pre-mRNA splicing.

Here we determined the penalties incurred by representative disease-causing Py tract mutations for recognition by U2AF^{65}. We leveraged this analysis along with structural information to engineer a U2AF^{65} variant that can relieve the consequences of a representative disease-causing splice site mutation. The results point to inhibition of U2AF^{65} as a contributing factor in human genetic disease and set a precedent for structure-guided modification of U2AF^{65} as a way to alter the splicing of therapeutically relevant pre-mRNAs.

Results

Disease-Causing Mutations in the 5′ Regions of Py Tracts Penalize U2AF^{65} Association. Although numerous examples of disease-causing mutations in Py tracts have been identified, few studies are available on the consequences of these mutations for the splice site affinities of U2AF^{65}. To investigate the involvement of U2AF^{65} inhibition, we examined the effects on U2AF^{65} association of mutations in two representative Py tracts (Table 1): (*i*) a U→A transversion in the Py tract of the neurofibromin 1 gene [*NF1(U3 > A)*], which creates a new 3′ splice site junction (consensus AG) and leads to familial neurofibromatosis type I (29), and (*ii*) a U→A transversion in the Py tract of the retinitis pigmentosa 2 (*RP2*) gene [*RP2(U4 > A)*] that also introduces an AG dinucleotide and induces exon skipping and, consequently, X-linked retinitis pigmentosa (30). We titrated the Py tract recognition domain of U2AF^{65} comprising RRM1 and RRM2 and bordering residues into either the WT or mutated fluorescein-labeled Py tract RNAs, and then fit the apparent equilibrium dissociation constant (K_D) values to anisotropy changes as described previously (31) (Fig. S1).

To enhance affinity and hence reduce protein consumption, we initially used 13-mer sequences directly preceding the 3′ splice site junctions in a series of *NF1* oligonucleotides. Subsequently, we focused on the core 9-mer Py tracts for *RP2* and for comparison tested that of *NF1*, which although reducing the avidity due to the high local concentration of sites (32, 33), represented the key Py tract interactions with U2AF^{65} and omitted flanking RNA sequences that bind other subunits of the assembling spliceosome. For both *NF1* and *RP2*, the disease-causing splice site mutations substantially reduced U2AF^{65} affinity [by a factor of three for *NF1 (U3 > A)* and a factor of four for *RP2(U4 > A)* relative to WT Py tracts; Table 1 and Fig. S1]. We conclude that these Py tract mutations disrupt splicing not only by introducing an aberrant AG consensus that normally dictates the junction of the 3′ splice site, but also by penalizing U2AF^{65} association.

Purine Substitutions in the 3′ Regions of Py Tracts Have Little Impact on U2AF^{65} Binding. Serendipitously, both of the disease-causing U→A mutations in *NF1* [*NF1(U3 > A)*] and *RP2* [*RP2(U4 > A)*] were located in the 5′ regions of the affected Py tracts. Although the region-dependence of disease-causing Py tract mutations has yet to be surveyed comprehensively, we noted qualitatively that many disease-causing Py tract mutations, such as the *NF1(U3 > A)* and *RP2(U4 > A)* investigated here, are located in the 5′ region of this splice site signal. Accordingly, we previously found that the C-terminal U2AF^{65} RRM2 has a strict preference for U nucleotides in the 5′ regions of Py tracts, whereas the N-terminal RRM1 is promiscuous for U or C nucleotides in the 3′ regions (23).

To compare the impact of purine substitutions in the 3′ regions and 5′ regions of the Py tract, we introduced artificial purine substitutions at the penultimate nucleotide of the *NF1* and *RP2* Py tracts [*NF1(U8 > A)*, *NF1(U8 > G)*, and *RP2(U8 > A)*] and determined the affinities of U2AF^{65} for these RNA oligonucleotides (Table 1 and Fig. S1). We also compared an analogous A substitution in a well-characterized hemoglobin β splice site [*HBB(U8 > A)*] that is disrupted in cases of β-thalassemia (34, 35). In all cases, these purine mutations in the 3′ region of the Py tract had little detectable effect on U2AF^{65} binding. Observation of crystal and NMR structures confirmed that U2AF^{65} directly binds both the 5′ and 3′ regions of the Py tracts with equivalent UV cross-linking efficiencies with photoactivated 4-thio-U nucleotides in the 5′ and 3′ regions of the *RP2* Py tract (*RP2* U3 and A9, respectively, as enumerated in Table 1) (Fig. S2). Overall, these results are in agreement with a sequence-stringent U2AF^{65} RRM2 and promiscuous RRM1 bound to the 5′ and 3′ regions, respectively, of the Py tract (23).

A Binding Pocket of U2AF^{65} RRM1 Tolerates Purines. We previously determined that sites on U2AF^{65} RRM1 locally adjust to U→C transitions through hydrogen bond rearrangements (23). Although valuable, the resulting structures left unanswered the question of how U2AF^{65} can adapt to purine substitutions in the 3′ regions of degenerate Py tracts. To address this question, we examined structures of U2AF^{65} bound to U tracts containing A or G substitutions at the penultimate nucleotide of an otherwise all-U Py tract (Table S1). We used a comparable crystallization approach as described previously (23, 36), composed of a U2AF^{65} variant (dU2AF^{65}, lacking residues 238–257 of the inter-RRM linker; Fig. 1*B*) in complex with a deoxy-ribose (d) oligonucleotide backbone. We focused our structure determinations on oligonucleotides containing dA or dG at the penultimate nucleotide and included 5-Br-dU as a marker for the sequence registers (Fig. 1*C* and Fig. S3 *A* and *B*). The polypeptide and oligonucleotide conformations

Table 1. Apparent equilibrium dissociation constants

Parent transcript	RNA sequence: 123456789	D231V:WT affinity ratio
HBB intron1 (13-mer)	UU CCCACCC*UU* AG	2.2
HBB(U8 > A)	UU CCCACCC***A***U AG	0.4
NF1 intron10a (13-mer)	GU UUUGUUU*UU* AG	2.3
NF1(U3 > A)	GU UU***A***GUUU*UU* AG	2.2
NF1(U8 > A)	GU UUUGUUU***A***U AG	0.3
NF1(U8 > G)	GU UUUGUUU***G***U AG	0.5
NF1 intron10a (9-mer)	UUUGUUU*UU*	2.0
RP2 intron3	GUUUGCU*UA*	2.5
RP2(U4 > A)	GUU***A***GCU*UA*	2.6
RP2(U8 > A)	GUUUGCU***A***A	1.0
RP2 intron 4	UAUUUAAAA	0.7

Four nucleotides in the 5′ region and four nucleotides in the 3′ region are expected to contact RRM2 and RRM1, respectively, of U2AF^{65}. The anticipated D231V-bound nucleotide is in italic type. Nucleotide mutations relative to the WT (Wt) RNA sequence are highlighted in bold.

of the two copies in the asymmetric unit closely match one another and the previous, baseline dU-bound dU2AF^{65} structure (23) (rmsd, 0.4–0.6 for matching Cα and C1′ atoms) (Fig. S3*D*). With the exception of one alternative dG conformation that engages in crystallographic contacts, the two complexes in each asymmetric unit share similar interactions with the bound purines.

As observed previously for uracil and cytosine (23, 24), the bound adenine or guanine bases stack on the consensus ribonucleoprotein motif (RNP1) of U2AF^{65} RRM1 and are engaged by hydrogen bonds with the protein backbone, as well as the D231 and R150 side chains (Fig. 1 *E* and *F* and Movies S1 and S2). For any type of bound nucleotide base, R150 consistently donates hydrogen bonds to the pyrimidine-O2, dA-N1, or dG-N7 acceptors. In previously identified ribose-(r)U– or dU-bound structures, the U2AF^{65} H230/D231 backbone amides donate hydrogen bonds to the lone pairs of uracil-O4 (Fig. 1*D*). In contrast, when bound to dA (Fig. 1*E* and Movie S1), an ordered water molecule mediates these hydrogen bonds with the protein backbone, which is relatively distant from the adenine (heavy atom distances, 6.5 Å for D231-NH–dA-N7 and 3.2 Å for D231-NH–dU-O4). However, the carboxylate side chain of U2AF^{65} D231 is newly positioned to accept a direct hydrogen bond from the adenine exocyclic amine. These dA contacts are reminiscent of the water-mediated and D231 interactions of U2AF^{65} and the exocyclic amine of a bound cytosine (23). In contrast, the U2AF^{65}-bound dG flips to the *syn* conformer (Fig. 1*F* and Movie S2), which differs from the *anti* glycosidic bonds of other types of nucleotides in the U2AF^{65} structures. In this conformation, the guanosine-O6 accepts hydrogen bonds from the backbone amides with only slightly less optimal geometry than a uracil-O4, whereas in the *anti* conformer, the exocyclic amine of the guanosine would be expected to sterically interfere with the R150 side chain. Considered together, the structures reveal a binding site on U2AF^{65} RRM1 that can accommodate diverse nucleotides in the 3′ region of the Py tract.

Design and Structure of a U2AF^{65}-D231V Variant. The relatively weak RNA affinity and promiscuity of U2AF^{65} RRM1 led us to hypothesize that these characteristics could be ameliorated in synthetic U2AF^{65} variants. We focused on optimizing the binding site on U2AF^{65} RRM1 that accommodates diverse nucleotides at the penultimate position of the Py tract. Based on comparisons among the structures of dU2AF^{65} bound to rU-, dU-, dC-, dA-, or dG-containing Py tracts, we reasoned that replacement of the negatively charged D231 carboxylate group with a hydrophobic valine side chain (D231V) would specifically increase U2AF^{65} affinity for a U at the corresponding nucleotide position of the Py tract splice site signal. To characterize the modified interactions between the D231V mutant and uracil-containing oligonucleotide, we determined the 2.1-Å resolution structure of dU2AF^{65}-D231V bound to a poly-dU oligonucleotide (Table S1 and Fig. S3 *C* and *D*). The dU2AF^{65}-D231V structure demonstrates that the engineered valine side chain packs against the uracil base while maintaining hydrogen bonds with the protein backbone (Fig. 1*G*).

U2AF^{65}-D231V Variant Prefers Uridine. We proceeded to test the structural hypothesis that the D231V substitution would preferentially increase U2AF^{65} affinity for U over other nucleotides at its binding site. We compared the affinities of unmodified U2AF^{65} or the D231V variant for the *NF1*, *HBB*, or *RP2* Py tracts, all of which share a U at the expected D231V-binding site (U8) (Table 1 and Fig. S1). Accordingly, the U-specifying D231V substitution increased U2AF^{65} affinities for these RNAs by more than twofold. The net free energy gain of approximately −0.5 kcal mol^{-1} after the D231V mutation agrees with the ∼100-Å^2 increase in buried hydrophobic surface area (37) observed in the structure (Fig. 1*G*), which corresponds to the burial of approximately one methyl group (135 Å^2).

We next confirmed that the U2AF^{65}-D231V variant specifies Us over purine nucleotides at the penultimate position of the Py tract. Based on fluorescence anisotropy changes during titration of fluorescein-labeled RNAs, we determined the affinities of the unmodified U2AF^{65} or the D231V variant for either the A or the G mutations in the 3′ region of the *NF1* Py tract [*NF1(U8 > A)* or *NF1(U8 > G)*], as well as the corresponding A mutations of the *RP2* [*RP2(U8 > A)*] and *HBB* [*HBB(U8 > A)*] Py tracts (Table 1 and Fig. S1). The affinities of unmodified U2AF^{65} for these Py tracts were the same within the margin of error as the WT U8 regardless of substitution by a purine, consistent with the sequence promiscuity of the expected binding site in the unmodified U2AF^{65} RRM1 (Fig. 1). After the D231V modification, the U2AF^{65}-D231V variant discriminated against the U→A or U→G transversions in the *NF1* Py tract. The U2AF^{65}-D231V affinities were more than sixfold and threefold greater for the WT *NF1* U8 counterpart than for the respective *NF1(U8 > A)* and *NF1(U8 > G)* mutants. Likewise, the U2AF^{65}-D231V variant preferred the WT, U8-containing *HBB* by nearly fivefold over its U8→A transversion in *HBB(U8 > A)*. We conclude that in these sequence contexts, the D231V substitution selectively increases U2AF^{65}'s affinity for U nucleotides and discriminates against binding of purine nucleotides.

The discrimination of U2AF^{65}-D231V against the U8→A mutation in the *RP2(U8 > A)* relative to the WT *RP2* Py tract was more subtle, possibly owing to the introduction of tandem adenosines at the 3′ terminus. For comparison, we tested the Py tract signal of the downstream *RP2* splice site (in intron 4 as opposed to intron 3) that naturally comprises four consecutive adenosines in the penultimate nucleotide sites (Table 1 and Fig. S1). This "spliced-into" Py tract has a twofold higher U2AF^{65} affinity than that of the "skipped" *RP2(U4 > A)* Py tract, supporting U2AF^{65} inhibition as a contributing mechanism to the splicing defect. We found that the D231V-modified U2AF^{65} discriminated against the tandem adenosines of the Py tract in *RP2* intron 4, yet, as for *RP2(U8 > A)*, this difference was slightly less than that observed for the *NF1* or *HBB* Py tracts. Nevertheless, the abilities of the D231V-modified U2AF^{65} to selectively increase affinity for the disease-relevant *RP2* splice site yet bind the downstream *RP2* splice site demonstrates its possible utility in selectively targeting and improving splicing of the *RP2(U4 > A)* defective transcript.

U2AF^{65}-D231V Corrects a Representative Splicing Defect in Human Cell Culture. We hypothesized that the increased U affinity of the U2AF^{65}-D231V variant could indirectly overcome inhibition of U2AF^{65} binding by mutations at other sites of the Py tract. We first tested this hypothesis by determining the affinity of U2AF^{65}-D231V for the *NF1* and *RP2* Py tracts, which naturally present U nucleotides at the expected D231V- binding site, with disease-causing mutations in distinct 5′ regions of the Py tracts [*NF1(U3 > A)* and *RP2(U4 > A)*] (Table 1 and Fig. S1). Compared with unmodified U2AF^{65}, the U2AF^{65}-D231V affinities for the defective

NF1(U3 > A) and *RP2(U4 > A)* Py tracts increased by more than twofold. We conclude that when bound to a U nucleotide at the penultimate position of a Py tract, the U2AF^{65}-D231V substitution has the capacity to compensate for the binding penalties incurred by purine mutations at other sites.

We next tested whether the synthetic U2AF^{65}-D231V variant could correct defective splicing of the retinitis pigmentosa-causing *RP2* mutation [*RP2(U4 > A)*] using a splicing reporter minigene in a human cell line (30) (Fig. 2 and Fig. S4). The WT and D231V-modified U2AF^{65} showed similar levels of expression after transient transfection of their expression constructs into HEK293T cells (Fig. S4*A*). As reported previously (30), nearly all of the detected *RP2* mRNAs composing the WT Py tract included the central exon (exon 4) (Fig. S4*B*). As such, cotransfection with either the WT or D231V-modified U2AF^{65} increased exon inclusion only slightly (by 4% or 7% of the total spliced transcript, respectively). Conversely, nearly all of the *RP2 (U4 > A)* transcript harboring the retinitis pigmentosa-causing mutation skipped exon 4, instead joining exons 3 and 5 (30) (Fig. 2*B* and Fig. S4*C*). Exon incorporation [i.e., correct splicing of the *RP2(U4 > A)* minigene transcript] increased proportionately after cotransfection of the minigene with increasing amounts of the U2AF^{65}-D231V expression construct (Fig. S4*C*). In contrast, cotransfection with unmodified U2AF^{65} had little detectable effect.

By including both the splicing minigene and U2AF^{65} coding region on a bicistronic vector to ensure coexpression in the same cell (Fig. 2*A*), the U2AF^{65}-D231V variant improved exon inclusion in the mutant *RP2(U4 > A)* transcript to levels approaching those of the WT control transcript (Fig. 2 *B* and *C* and Fig. S4*D*). Importantly, this result establishes proof-of-principle of the ability of a structure-guided U2AF^{65} variant (D231V) to improve splicing of a defective Py tract.

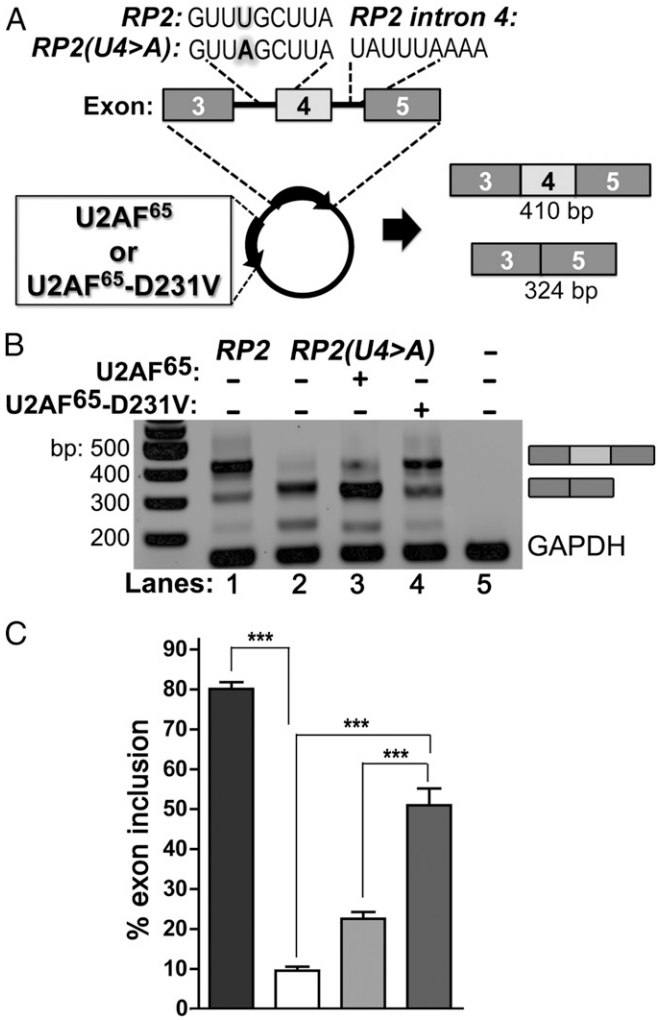

Fig. 2. The U2AF^{65}-D231V variant improves splicing of the X-linked retinitis pigmentosa-causing *RP2(U4 > A)* mutant splice site in human cells. (*A*) Experimental scheme. A bicistronic vector comprising either WT *RP2* or mutated *RP2(U4 > A)* minigenes (Py tract sequences; *Inset*) and either WT U2AF^{65} or the U2AF^{65}-D231V variant is transfected in HEK293T cells. (*B*) Representative RT-PCR of mRNA isolated from transfected cells. (*C*) The bar graph plots percent of the exon-included band relative to the total amplified *RP2* product. The average percentages and SDs of five independent biological replicates are given. ***$P < 0.0003$ with 95% CI.

Discussion

U2AF^{65} Inhibition Contributes to Disease Outcomes in Py Tract Mutations. Disease-causing mutations often introduce new, harmful AG splice site signals within Py tracts (2). In most cases, these mutations are assumed to interfere with splicing by falsely matching the AG consensus sequence of a 3′ splice site junctions. In the present study, we show that certain purine mutations that introduce AG dinucleotides within Py tracts also penalize association of the essential pre-mRNA splicing factor U2AF^{65}; for example, the reduced binding affinity of U2AF^{65} for the mutated 3′ splice site can explain the distinct effects of the AG disruptions on *NF1* and *RP2* splicing. In the former *NF1* case, the introduced AG is preceded by pyrimidine nucleotides that offer alternative recognition sites for U2AF^{65}, such that the *NF1* mutation is "spliced into," thereby adding several nucleotides beyond the bona fide junction (29). In the latter case, *RP2* lacks a detectable Py tract preceding the introduced AG, and as such, the mutated exon is entirely skipped in favor of the natural 3′ splice site in the downstream intron (30).

Beyond the U2AF^{65}–Py tract interaction focused on herein, it remains to be determined whether pre-mRNA motifs with sequence similarity to the BPS or 3′ junction-like sequences, respectively, can dictate the position of a cryptic splice site via favorable interactions with the respective SF1 or U2AF^{35} subunits. Studies in fission yeast suggest that introns with degenerate Py tracts depend on recognition by other subunits of this splicing factor complex (38). Likewise, humans have conditional requirements for U2AF^{35} in the binding and splicing of poorly conserved splice sites (12, 39, 40) and for SF1 for alternative splicing (18). Together with the documented disruption of splicing, our findings suggest that a mutation that introduces an AG dinucleotide within a Py tract not only generates a false splice acceptor, but also disrupts a balanced competition among pre-mRNA sequences for binding the ternary U2AF^{65}-SF1-U2AF^{35} complex.

Human U2AF^{65} Evolved a Proofreading RRM2 and an Enabling RRM1. For the mutations studied here, U2AF^{65} association is strongly inhibited by transversions in the 5′ regions of Py tracts yet tolerates them in the 3′ region. This finding extends previous evidence indicating that U2AF^{65} is selectively inhibited by C tracts in the 5′ region of the Py tract, whereas those in the 3′ region have little effect (23). Furthermore, structure-based amino acid changes in RRM2 confer tolerance for C tracts in the 5′ region of the Py tract (23). The region-dependent sequence tolerances of U2AF^{65} indicate that the respective human U2AF^{65} RRM2 and RRM1 have evolved distinct functions of specifying U-rich Py tracts and adjusting to alternative splice site sequences. These sequence preferences of the human factor may differ from other homologs; for example, *Caenorhabditis elegans* 3′ splice sites are preceded by four strict Us, but also require the U2AF^{35} small subunit for accurate recognition (41). Considering that the human U2AF^{65} RRM2 and RRM1 each recognize approximately four nucleotides in the respective 5′ and 3′ halves of the Py tract (23–25), the region-dependent penalties of mutations in Py tracts indicate that U2AF^{65} RRM2 is stringent for Us, whereas RRM1 is promiscuous for other nucleotides.

The differential sensitivity of human U2AF^{65} RRM2 and RRM1 to purine interruptions of Py tracts also provides a

molecular explanation for previously reported biochemical results. For example, a modified adenovirus major late promoter (AdML) substrate is inefficiently spliced when an A is substituted for the penultimate U of the Py tract (42). In comparison, splicing substrates with A substitutions at either the fourth U of the AdML Py tract or the second U of a *sex-lethal* Py tract in *Drosophila* nearly abolish detectable splicing. Likewise, purine mutations of a β-globin Py tract are more detrimental in the 5′ region than in the 3′ region (43). The sequence-specific U2AF^{65} RRM2, which recognizes the 5′ region of the Py tract, is likely to proofread so-called "AG-exclusion zones" in the emerging transcript, identifying regions entirely devoid of AG dinucleotides that precede normal 3′ splice sites (44). In contrast, the promiscuous RRM1 offers versatility for adapting to human Py tracts, which are often degenerate, owing in part to the evolution of alternative splice site signals (45, 46).

Although comprehensive surveys mapping the positions of disease-causing mutations within splice site signals remain to be determined, this region-dependent inhibition of U2AF^{65} further predicts that the consequences for splicing, and hence progression to disease, would be more severe for purine mutations in the 5′ nucleotides as opposed to the 3′ nucleotides of a Py tract.

Common RRM-Binding Site for *syn* Purine Nucleotides. At the penultimate nucleotide-binding site, we observe that human U2AF^{65} RRM1 adapts to the Hoogsteen face of a *syn* G interruption, which offers a similar hydrogen-bonding pattern as a uracil after a 180° rotation about the *N*-glycosidic bond. Traditionally, the Watson–Crick faces of nucleotides in the *anti* conformation are considered available for recognition by protein. This structure adds to a growing category of RRMs known to recognize *syn* G-nucleotides, including hnRNP A1 (47), hnRNP D (48), and SRSF2 (49). These published structures and the U2AF^{65} structure consistently share a *syn* G-binding site on a conserved aromatic residue (F199 of U2AF^{65}) in the ribonucleoprotein consensus motif (RNP1) (50). Several cases of RRMs recognizing the *syn* conformer of adenosine at this RNP1-binding site, including SRp20 (51), RBMY (52), and Tra2-β1 (53), have been detected as well. The *syn* G conformation is stabilized a priori by a distinctive intramolecular hydrogen bond between the exocyclic amine and the 5′ phosphate. Accordingly, the equilibrium between the *syn* and *anti* conformers of guanosine-5′-phosphate reaches an ∼50:50 mixture in solution (54), indicating that RRMs such as U2AF^{65} RRM1 engage in local conformational selection of the *syn* over *anti* G conformers. As such, preference for *syn* purines at a specific RNP1 site is emerging as a predictable theme of RRM/RNA recognition.

Predicting Nucleotide Recognition by RRMs: Emerging Themes and Remaining Challenges. Whether the RRM can offer a scaffold for the design of RNA-binding proteins of a chosen specificity remains a topic of ongoing debate (28, 50). Although RNP1 and RNP2 consensus motifs of the canonical RRM generally share similar interactions with the nucleotide bases, the diverse roles of the surrounding residues and loop regions have posed a considerable challenge to designing the sequence-specificity of synthetic RRMs. A valine-to-alanine amino acid change that increases the RNA affinity of a U1A RRM was discovered more than 20 y ago by phage-display selection (55); however, to our knowledge, this serendipitous U1A mutation remains the singular successful example of RRM improvement until our present work.

Here we demonstrate the feasibility of designed RRM-RNA recognition in two ways. As described above, we add to a growing body of structural evidence for a predictable preference for *syn* over *anti* purine nucleotides at a specific RRM-binding site. Most importantly, we establish a successful precedent for rational RRM improvement through a structure-based D231V amino acid change that specifically improves the U affinity of U2AF^{65} RRM1. This U2AF^{65}-D231V variant is sufficient to indirectly compensate for the penalty of a disease-causing Py tract mutation and restore splicing of a representative transcript in human cells.

Outstanding challenges for designer U2AF^{65} proteins were identified during the development of our milestone U2AF^{65}-D231V variant. First, affinity analyses imply that the identity of the flanking nucleotides influences the exact sequence preferences of U2AF^{65} and its engineered derivatives. Specifically, the ability of U2AF^{65}-D231V to discriminate against an adenine base was slightly decreased for AA (in *RP2* intron 3 and *RP2* intron 4) compared with AU (in *HBB* and *NF1*) dinucleotide steps (Table 1). Nevertheless, in support of the feasibility of modular U2AF^{65} engineering, the U2AF^{65} RRMs have not yet been seen to extrude nucleotides or otherwise skip sequence registers, as has been detected for Puf proteins (56, 57).

Second, whether the addition of U2AF^{65} or engineered U2AF^{65} variants will have the capacity to improve splicing that has been composed indirectly by mutations outside of the Py tract (e.g., in exonic splicing enhancers) remains unknown. The minimal effect of U2AF^{65}-D231V on splicing of the WT *RP2* minigene transcript, for which nearly all of the spliced mRNAs (more than 80%) included exon 4 a priori, indicated that engineered U2AF^{65} variants would have little utility to alter splicing of "strong" (i.e., U-rich) Py tracts. In contrast, the large U2AF^{65}-D231V–dependent increase in splicing of the defective *RP2(4U > A)* splice site (by fivefold for the bicistronic vector) demonstrates that engineered U2AF^{65} variants can increase splicing of disrupted target Py tracts, and by analogy, are likely to affect "weak" (i.e., degenerate) Py tracts.

Third, it remains to be determined whether future U2AF^{65} alterations can be identified that specifically target each of the approximately nine nucleotides comprising natural Py tracts. Fortuitously, the extended binding site size of the ternary U2AF^{65}-SF1-U2AF^{35} complex inherently decreases the likelihood of random off-target sequence matches amid the transcriptome.

Here we have successfully demonstrated that the optimization of U2AF^{65} binding to RNA is feasible for a proof-of-principle case of a D231V variant. Numerous groups are intensely invested in the development of RNA-based therapeutics, which often are targeted at the level of pre-mRNA splicing. Considering the broad potential benefits of U2AF^{65} variants as tools for biochemical investigation or gene therapies, it will be well worthwhile to build from the proof-of-principle U2AF^{65}-D231V in future generations of optimization to achieve highly sequence specific proteins that alter pre-mRNA splicing.

Methods

The experimental procedures are described in detail in *SI Methods*.

RNA-Binding Experiments. Human U2AF^{65} (residues 141–342) was titrated into 5′-fluorescein–labeled RNA (sequences presented in Table 1), and the fluorescence anisotropy changes were fit to obtain the K_D values as described previously (31) (Fig. S1).

Crystallization and Structure Determination. Human dU2AF^{65} protein (residues 148–237 and 258–336; Fig. 1*B*) was cocrystallized with 5′-dUdUdUdU(5-Br-dU)dAdU and 5′-dUdUdUdU(5-Br-dU)dGdU. The dU2AF^{65}-D231V variant was cocrystallized with 5′-dUdUdUdU(5-Br-dU)dUdU. Structures were determined by difference Fourier using PDB ID code 3VAK as a starting model (Table S1 and Fig. S3).

Transfection and RT-PCR Analyses. RT-PCR and qRT-PCR procedures are described in *SI Methods* and illustrated in Fig. 2 and Fig. S4.

ACKNOWLEDGMENTS. We are grateful to M. Salim, J. Ashton, and J. Wedekind for their helpful advice. This work was supported by National Institutes of Health (NIH) Grant R01 GM070503. Crystallographic data were collected with the support of NIH National Center for Research Resources (NCRR) Grant S10 RR026501 for in-house equipment; the US Department of Energy, NIH Grant P41 RR001209, and the National Institute of General Medical Sciences for Stanford Synchrotron Radiation Lightsource; and National Science Foundation Grant DMR-0936384 and NIH NCRR Grant RR-01646 for Cornell High Energy Synchrotron Source.

1. Burge CB, Tuschl T, Burge CB, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosomes. *The RNA World*, eds Gesteland RF, Cech TR, Atkins JF (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY), pp 525–560.
2. Stenson PD, et al. (2014) The Human Gene Mutation Database: Building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 133(1):1–9.
3. Krawczak M, et al. (2007) Single base-pair substitutions in exon-intron junctions of human genes: Nature, distribution, and consequences for mRNA splicing. *Hum Mutat* 28(2):150–158.
4. Singh RK, Cooper TA (2012) Pre-mRNA splicing in disease and therapeutics. *Trends Mol Med* 18(8):472–482.
5. Kole R, Krainer AR, Altman S (2012) RNA therapeutics: Beyond RNA interference and antisense oligonucleotides. *Nat Rev Drug Discov* 11(2):125–140.
6. Galanello R, Origa R (2010) β-thalassemia. *Orphanet J Rare Dis* 5:11.
7. Mordes D, et al. (2006) Pre-mRNA splicing and retinitis pigmentosa. *Mol Vis* 12: 1259–1271.
8. Barron VA, Lou H (2012) Alternative splicing of the neurofibromatosis type I pre-mRNA. *Biosci Rep* 32(2):131–138.
9. Abovich N, Rosbash M (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89(3):403–412.
10. Zamore PD, Green MR (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci USA* 86(23):9243–9247.
11. Berglund JA, Chua K, Abovich N, Reed R, Rosbash M (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89(5): 781–787.
12. Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF^{35}. *Nature* 402(6763):832–835.
13. Merendino L, Guth S, Bilbao D, Martínez C, Valcárcel J (1999) Inhibition of *msl-2* splicing by Sex-lethal reveals interaction between U2AF^{35} and the 3′ splice site AG. *Nature* 402(6763):838–841.
14. Zorio DA, Blumenthal T (1999) Both subunits of U2AF recognize the 3′ splice site in *Caenorhabditis elegans*. *Nature* 402(6763):835–838.
15. Gozani O, Potashkin J, Reed R (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 18(8):4752–4760.
16. Rutz B, Séraphin B (1999) Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* 5(6):819–831.
17. Bedford MT, Reed R, Leder P (1998) WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: The proline glycine and methionine-rich motif. *Proc Natl Acad Sci USA* 95(18):10602–10607.
18. Corioni M, Antih N, Tanackovic G, Zavolan M, Krämer A (2011) Analysis of in situ pre-mRNA targets of human splicing factor SF1 reveals a function in alternative splicing. *Nucleic Acids Res* 39(5):1868–1879.
19. Tanackovic G, Krämer A (2005) Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing *in vivo*. *Mol Biol Cell* 16(3):1366–1377.
20. Guth S, Tange TO, Kellenberger E, Valcárcel J (2001) Dual function for U2AF(^{35}) in AG-dependent pre-mRNA splicing. *Mol Cell Biol* 21(22):7673–7681.
21. Ruskin B, Zamore PD, Green MR (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* 52(2):207–219.
22. Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355(6361):609–614.
23. Jenkins JL, Agrawal AA, Gupta A, Green MR, Kielkopf CL (2013) U2AF^{65} adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res* 41(6):3859–3873.
24. Sickmier EA, et al. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF^{65}. *Mol Cell* 23(1):49–59.
25. Mackereth CD, et al. (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475(7356):408–411.
26. Shen H, Green MR (2004) A pathway of sequential arginine-serine–rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell* 16(3): 363–373.
27. Wang Y, Cheong CG, Hall TM, Wang Z (2009) Engineering splicing factors with designed specificities. *Nat Methods* 6(11):825–830.
28. Mackay JP, Font J, Segal DJ (2011) The prospects for designer single-stranded RNA-binding proteins. *Nat Struct Mol Biol* 18(3):256–261.
29. Ars E, et al. (2000) Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum Mol Genet* 9(2):237–247.
30. Pomares E, et al. (2009) Identification of an intronic single-point mutation in RP2 as the cause of semidominant X-linked retinitis pigmentosa. *Invest Ophthalmol Vis Sci* 50(11):5107–5114.
31. Jenkins JL, Shen H, Green MR, Kielkopf CL (2008) Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF^{65}. *J Biol Chem* 283(48):33641–33649.
32. Bienz M (2014) Signalosome assembly by domains undergoing dynamic head-to-tail polymerization. *Trends Biochem Sci* 39(10):487–495.
33. Rudnick SI, Adams GP (2009) Affinity and avidity in antibody-based tumor targeting. *Cancer Biother Radiopharm* 24(2):155–161.
34. Spritz RA, et al. (1981) Base substitution in an intervening sequence of a beta$^+$-thalassemic human globin gene. *Proc Natl Acad Sci USA* 78(4):2455–2459.
35. Metherall JE, Collins FS, Pan J, Weissman SM, Forget BG (1986) Beta zero thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron. *EMBO J* 5(10):2551–2557.
36. Sickmier EA, Frato KE, Kielkopf CL (2006) Crystallization and preliminary X-ray analysis of a U2AF^{65} variant in complex with a polypyrimidine-tract analogue by use of protein engineering. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 62(Pt 5): 457–459.
37. Pace CN, et al. (2011) Contribution of hydrophobic interactions to protein stability. *J Mol Biol* 408(3):514–528.
38. Sridharan V, Singh R (2007) A conditional role of U2AF in splicing of introns with unconventional polypyrimidine tracts. *Mol Cell Biol* 27(20):7334–7344.
39. Henscheid KL, Voelker RB, Berglund JA (2008) Alternative modes of binding by U2AF^{65} at the polypyrimidine tract. *Biochemistry* 47(1):449–459.
40. Guth S, Martínez C, Gaur RK, Valcárcel J (1999) Evidence for substrate-specific requirement of the splicing factor U2AF^{35} and for its function after polypyrimidine tract recognition by U2AF^{65}. *Mol Cell Biol* 19(12):8263–8271.
41. Hollins C, Zorio DA, MacMorris M, Blumenthal T (2005) U2AF binding selects for the high conservation of the *C. elegans* 3′ splice site. *RNA* 11(3):248–253.
42. Roscigno RF, Weiner M, Garcia-Blanco MA (1993) A mutational analysis of the polypyrimidine tract of introns: Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem* 268(15):11222–11229.
43. Reed R (1989) The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev* 3(12B):2113–2123.
44. Gooding C, et al. (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol* 7(1):R1.
45. Senapathy P, Shapiro MB, Harris NL (1990) Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project. *Methods Enzymol* 183:252–278.
46. Irimia M, Roy SW (2008) Evolutionary convergence on highly-conserved 3′ intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *PLoS Genet* 4(8):e1000148.
47. Ding J, et al. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13(9):1102–1115.
48. Enokizono Y, et al. (2005) Structure of hnRNP D complexed with single-stranded telomere DNA and unfolding of the quadruplex by heterogeneous nuclear ribonucleoprotein D. *J Biol Chem* 280(19):18862–18870.
49. Daubner GM, Cléry A, Jayne S, Stevenin J, Allain FH (2012) A syn-anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. *EMBO J* 31(1):162–174.
50. Auweter SD, Oberstrass FC, Allain FH (2006) Sequence-specific binding of single-stranded RNA: Is there a code for recognition? *Nucleic Acids Res* 34(17):4943–4959.
51. Hargous Y, et al. (2006) Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. *EMBO J* 25(21):5126–5137.
52. Skrisovska L, et al. (2007) The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction. *EMBO Rep* 8(4):372–379.
53. Cléry A, et al. (2011) Molecular basis of purine-rich RNA recognition by the human SR-like protein Tra2-β1. *Nat Struct Mol Biol* 18(4):443–450.
54. Son TD, Guschlbauer W, Guéron M (1972) Flexibility and conformations of guanosine monophosphates by the Overhauser effect. *J Am Chem Soc* 94(22):7903–7911.
55. Laird-Offringa IA, Belasco JG (1995) Analysis of RNA-binding proteins by in vitro genetic selection: Identification of an amino acid residue important for locking U1A onto its RNA target. *Proc Natl Acad Sci USA* 92(25):11859–11863.
56. Lu G, Hall TM (2011) Alternate modes of cognate RNA recognition by human PUMILIO proteins. *Structure* 19(3):361–367.
57. Miller MT, Higgin JJ, Hall TM (2008) Basis of altered RNA-binding specificity by PUF proteins revealed by crystal structures of yeast Puf4p. *Nat Struct Mol Biol* 15(4): 397–402.