Alternative Modes of Binding by U2AF65 at the Polypyrimidine Tract†

Kristy L. Henscheid, Rodger B. Voelker, and J. Andrew Berglund*

Institute of Molecular Biology, 1229 University of Oregon, Eugene, Oregon 97403

Received June 23, 2007; Revised Manuscript Received October 1, 2007

ABSTRACT: During initial recognition of an intron in pre-mRNA, the 3′ end of the intron is bound by essential splicing factors. Notably, the consensus RNA sequences bound by these proteins are highly degenerate in humans. This raises the question of 3′ splicing factor function in introns lacking canonical binding sites. Investigating the introns of the model organism *Neurospora crassa* revealed a different organization at the 3′ end of the intron compared to most eukaryotic organisms. The predicted branch point sequences of *Neurospora* introns are much closer to the 3′ splice site compared to those in human introns. In addition, *Neurospora* introns lack the canonical polypyrimidine tract found at the end of introns in most eukaryotic organisms. The large subunit of the U2 snRNP associated factor (U2AF65), which is essential for splicing of human introns and specifically recognizes the polypyrimidine tract, is also present in *Neurospora*. We show that *Neurospora* U2AF65 binds RNA with low affinity and specificity, apparently evolving with its disappearing binding site. The arginine/serine rich domain at the N-terminus of *Neurospora* U2AF65 regulates its RNA binding. We find that this regulated binding can be recapitulated in human U2AF65 which has been mutated to decrease both affinity and overall charge. Finally, we show that the addition of the small U2AF subunit (U2AF35) to U2AF65 with weakened RNA binding affinity significantly enhances the affinity of the resulting U2AF heterodimer.

Removal of intervening sequences (introns) from messenger RNAs is an essential part of gene expression in eukaryotes. Splicing, the process of intron removal, is carried out by a large conserved protein−RNA complex, the spliceosome, which consists of 5 snRNPs[1] (small nuclear ribonucleoproteins) containing short RNAs and several proteins, as well as over 100 additional protein factors (*1*). Accurate recognition of introns by the spliceosome is crucial to generating the correct mRNAs. Intron recognition is also the likely point of regulation for alternative splicing, a process by which multiple mRNAs (and thus RNA or protein products) can be generated from a single gene (*2*). Generally conserved consensus sequences have been identified (*1*, *3*) which occur at the 5′ and 3′ ends of the intron as well as the branch point (where the chemistry of splicing is initiated). Though the factors that bind these sequences are largely essential, the canonical sequences are too degenerate in vertebrates to provide all the information required for defining most introns (*1*). Accordingly, an outstanding challenge is understanding how splicing factors define the intron using the canonical splicing motifs, often regulated by additional intronic and exonic sequences (*2*).

In the current model of pre-mRNA splicing, each end of an intron is recognized by separate RNA−RNA interactions: the 5′ splice site (5′ss) is bound by the U1 snRNP, and the branch point sequence (BPS) near the 3′ splice site (3′ss) is bound by the U2 snRNP (*1*, *4*). Targeting of the U2 snRNP to the intron requires the essential heterodimeric splicing factor U2AF (U2 snRNP auxiliary factor) (*5*). The large subunit, U2AF65, contains three C-terminal RNA recognition motifs (RRMs), and an N-terminal RS domain (*6*). Only two of the RRMs are involved in RNA binding, making sequence-specific contacts with the polypyrimidine tract (PyT) between BPS and 3′ss. The atypical C-terminal third RRM is a protein interaction domain (*7*−*9*). RRM3 interacts with the splicing factor SF1/BBP to recognize the PyT and BPS in a cooperative fashion (*8*). U2 snRNP then replaces SF1/BBP at the BPS by interacting with RRM3 through the U2 snRNP-specific splicing factor SF3b155 (*10*, *11*). Deletion studies have shown that RRM1 and RRM2 are required for *in vitro* splicing; RRM3 is dispensable for some substrates *in vitro* but is necessary *in vivo* (*6*, *7*). The RS domain, common to many splicing factors, is a basic region enriched in arginine-serine repeats and is commonly involved in protein−protein interactions (*12*). The RS domain of U2AF65, however, contacts the BPS and may assist in annealing it with the U2 snRNA (*13*, *14*). Truncation studies have shown that the RS domain is important for U2AF65's splicing function, but it appears to have no role in binding the PyT (*6*, *13*).

U2AF35, the small subunit of U2AF, consists of an atypical (pseudo-)RRM flanked on each end by zinc fingers, and C-terminal glycine-rich and RS domains (*15*). The two

† K.L.H. was supported by a National Science Foundation Graduate Research Fellowship. This work was supported by an NSF grant (0616264-MCB) to J.A.B.

* Author to whom correspondence should be addressed. E-mail: aberglund@molbio.uoregon.edu. Phone: 541-346-5097. Fax: 541-346-5891.

[1] Abbreviations: snRNP, small nuclear ribonucleoprotein; 5′ss, 5′ splice site; BPS, branch point sequence; 3′ss, 3′ splice site; U2AF, U2 snRNP auxiliary factor; U2AF65, U2AF large subunit; RRM, RNA recognition motif; PyT, polypyrimidine tract; U2AF35, U2AF small subunit; NcU2AF65, *Neurospora* U2AF65; HsU2AF65, human U2AF65; QM, U2AF65 quadruple mutant (R150A, K195A, K260A, K300A); RNPmut, U2AF65 RNP mutant (Y152A, F199A, F262A, F304A); chU2AF65, chimeric human/*Neurospora* U2AF65; Nc35, *Neurospora* U2AF35; Hs35, human U2AF35; K_d, dissociation constant; p*I*, isoelectric point.

10.1021/bi701240t CCC: $40.75
Published on Web 12/08/2007

RS domains in the U2AF heterodimer are functionally redundant; deletion of the RS domain from a single subunit of the *Drosophila* U2AF does not affect binding affinity or viability (*16*, *17*). U2AF35 interacts with U2AF65 through U2AF35's pseudo-RRM; the other face of this domain may bind RNA (*15*). U2AF35 binds RNA weakly on its own, increases the RNA affinity of U2AF over that of U2AF65 alone, and has been shown to specifically recognize the AG dinucleotide at the 3′ss (*16*, *18*−*20*). The specific domains with which it recognizes RNA are still unclear. In mammals, though U2AF35 is considered an essential splicing factor, it does not appear to be important for most *in vitro* splicing. A splicing-deficient nuclear extract specifically depleted of U2AF can be rescued by the addition of U2AF65 alone (*6*, *21*). The requirement for U2AF35 appears to be substrate-specific and may also depend upon experimental conditions (*15*, *22*).

We examined intron sequences in the filamentous fungus *Neurospora crassa*, a well-established model organism for genetic studies (*23*). Our results indicate that *Neurospora* is an excellent model for PyT-poor introns, as it contains a significant portion of introns lacking PyTs. The *Neurospora* orthologue of U2AF65 binds RNA weakly, and its binding is dependent upon the RS domain, most likely because its RNA binding domains are negatively charged. We corroborate this conclusion with site-directed mutagenesis of human U2AF65. We also show that U2AF35 can significantly enhance RNA binding in a heterodimer with a weakly binding U2AF65.

MATERIALS AND METHODS

Human and Neurospora Intron Comparison. Intron databases were constructed using human release NCBI build 36.1 and *Neurospora* Broad Institute Assembly 7 (annotation version 2), from which introns were extracted using custom software. Redundant sequences and annotated introns that did not begin with GT and end with AG were removed. The final *Neurospora* dataset consisted of 17,049 intronic sequences. The human dataset was described previously (*24*). The occurrence of all pentamers within the last 40 bases of all annotated introns was determined using a sliding window enumeration. Occurrences for putative BPSs (represented as pentamers containing CTRAC, R = A or G) were normalized by dividing the occurrences in each position by the total number of occurrences of that *n*-mer within the entire region. The distribution therefore reflects the fraction of occurrences at each position. For the PyT, the occurrences in each bin were divided by the total number of introns examined to yield the fraction of introns that have the *n*-mer in each position.

Cloning. All proteins were cloned into pGEX-6P-1 (Amersham) except for HsU2AF35 and NcU2AF35 (below). HsU2AF65 was previously cloned (*25*). The quadruple mutant (QM), RRM1 mutant, RRM2 mutant, and RNP mutant (RNPmut) constructs were created using sequential Quikchange (Stratagene) reactions using HsU2AF65 as the first template. Chimeric human/*Neurospora* U2AF65 (chU2AF65) was assembled by PCR from fragments containing bases 1−435 and 1027−1428 of HsU2AF65 and bases 751−1404 of NcU2AF65. HsU2AF65ΔRS, QMΔRS, RNPmutΔRS, and chU2AF65ΔRS were cloned by PCR of the full-length constructs with primers which removed the first 255 bases (corresponding to amino acids 1−85). Chimeric Nc(SF1)-U2AF65 was assembled by PCR from fragments containing bases 1−204 of NcSF1 (RS domain), and bases 541−1755 of NcU2AF65 (Nc65ΔRS). $(RS)_7$NcU2AF65 was created by amplification of Nc65ΔRS with a forward primer containing the coding sequence for seven RS repeats. Primer sequences are given in Table 1.

NcU2AF65 was cloned using gene-specific RT-PCR from strain N150 (Oak Ridge wild type, *matA*) RNA. RNA was prepared from frozen tissue (grown in Vogel's minimal medium at 32 °C for 2 days) pulverized and extracted 4 times with acid phenol:chloroform 5:1 pH 4.5 (Ambion) in 100 mM sodium acetate pH 5.2, 1 mM EDTA, 4% SDS followed by a single chloroform extraction and ethanol precipitation. NcU2AF65ΔRS was cloned by PCR from a pooled lambda cDNA library from all stages of *Neurospora* (kindly provided by G. Kothe and S. Honda).

HsU2AF35 was amplified by PCR from the coexpression plasmid of Rudner et al. (*26*) and inserted into pAC28, which is compatible for co-transformation with pGEX and designed for two-plasmid coexpression (*27*). NcU2AF35 was amplified from the *Neurospora* lambda cDNA library and inserted into pAC28.

All plasmids were verified by sequencing and transformed into *E. coli* strain BL21 STAR (Invitrogen) for protein expression. For heterodimer coexpression, the BL21 strain carrying the U2AF65 construct was made chemically competent and transformed with the U2AF35 construct.

Protein Expression and Purification. HsU2AF65 was purified as previously described (*25*). All U2AF65 constructs, as well as NcU2AF65ΔRS+Nc35, were also purified by this method, except that ΔRS constructs were purified on a SOURCE 15Q (Amersham) column with a linear elution gradient from buffer A (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA) to buffer B (10 mM Tris pH 7.5, 1 M NaCl, 1 mM EDTA). Cell growth, protein induction, and lysis for all constructs followed our previously described procedure (*25*) with exceptions noted below. All purified proteins were dialyzed overnight into U2AF65 storage buffer (25 mM Tris pH 7.5, 300 mM NaCl, 1 mM DTT, 15% glycerol) and stored at −80 °C.

RS-domain-deleted U2AF heterodimers (HsU2AF65ΔRS+HsU2AF35, QMΔRS+HsU2AF35, and chU2AF65ΔRS+Hs35) were purified similarly except that the chromatography was omitted. These were bound to glutathione Sepharose (Amersham), directly cleaved from bound GST with Precision Protease (Amersham), and concentrated, after filtration through a 0.45 μm membrane, on a Centricon YM-30 centrifugal filtration device (Amicon). The His_6 tag was retained on HsU2AF35 and NcU2AF35, and the GST tag was retained on NcU2AF65ΔRS in the NcU2AF65ΔRS+Nc35 heterodimer.

Protein−RNA Binding. Various concentrations of protein were incubated for 10 min at room temperature with 0.1−0.2 nM 5′ ^{32}P-labeled RNA oligonucleotide, 50 mM Tris pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 0.01% Triton X-100, 1 mM DTT, 0.05% bromophenol blue, 4% glycerol, and 0.25 mg/mL heparin. The reactions were separated on native polyacrylamide gels (5% 37.5:1 mono:bis acrylamide, 4% glycerol, 0.5X TB) at 150 V for 2 h at 4 °C. Gels were dried, exposed to phosphor screens overnight, and quantitated using

Table 1: Primer Sequences Used for Cloning

construct	sequence
QM: R150A	5′ GAC AAG CCC GGG CCC TCT ACG TGG GCA AC 3′
	5′ GTT GCC CAC GTA GAG GGC CCG GGC TTG TC 3′
QM: K195A	5′ CAG ATT AAC CAG GAC GCG AAT TTT GCC TTT TTG G 3′
	5′ CCA AAA AGG CAA AAT TCG CGT CCT GGT TAA TCT G 3′
QM: K260A	5′ CCG ACT CTG CCC ACG CGC TGT TCA TCG GGG 3′
	5′ CCC CGA TGA ACA GCG CGT GGG CAG AGT CGG 3′
QM: K300A	5′ CCA CGG GGC TCT CCG CGG GCT ACG CCT TCT G 3′
	5′ CAG AAG GCG TAG CCC GCG GAG AGC CCC GTG G 3′
RNPmut: Y152A	5′ GCC CGG CGC CTC GCC GTG GGC AAC ATC CCC 3′
	5′ GGG GAT GTT GCC CAC GGC GAG GCG CCG GGC 3′
RNPmut: F199A	5′ GGA CAA GAA TTT TGC CGC TTT GGA GTT CCG C 3′
	5′ GCG GAA CTC CAA AGC GGC AAA ATT CTT GTC C 3′
RNPmut: F262A	5′ GCC CAC AAG CTG GCC ATC GGG GGC TTA CCC 3′
	5′ GGG TAA GCC CCC GAT GGC CAG CTT GTG GGC 3′
RNPmut: F304A	5′ CCA AGG GCT ACG CCG CCT GTG AGT ACG TGG 3′
	5′ CCA CGT ACT CAC AGG CGG CGT AGC CCT TGG 3′
ch65: Nc65 751−1404 w/Hs65 overlap	5′ TGA CGG TCT GGC TGT GAC CCC AAC GCC GGT GCC CGT GGT CCT TAA GCC TAC GAA CTC CCG CCA GA 3′
	5′ AGC CCG GCA CTT GCA GGG TCA CAG GCG TCT GAT TGA TGG TGT TCA CAC TCA GTT CTC CAG CTA CC 3′
ch65: Hs65 1−435	5′ GGG GGA TCC ATG TCG GAC TTC GAC GAG TTC GAG CGG CAG 3′
	5′ GGT CAT CTG GCT CCC GAC CAC GGG CAC CGG 3′
ch65: Hs65 1027−1428	5′ GGG GAA TTC TCA CCA GAA GTC CCG GCG GTG ATA AGA GTC GGG 3′
	5′ ACG CTG GTG AGC CCC CCG AGC ACC ATC AAT 3′
Hs65ΔRS	5′ GGG GGA TCC AAG AAG GTC CGT AAA TAC TGG 3′
NcU2AF65	5′ GGG GGA TCC ATG AAC GGG GAC ACC TAT TCC 3′
	5′ CCC GCG GCC GCT TAC CAG GCG TTG ACA TC 3′
Nc65ΔRS	5′ GGG GGA TCC GAC TTG ACC GAT GTC GTT 3′
	5′ CCC GTC GAC TTA CCA GGC CTT GAC ATC 3′
HsU2AF35	5′ GGG GGA TCC ATG GCG GAG TAT CTG GCC TCC 3′
	5′ GGG GAA TTC TCA GAA TCG CCC AGA TCT TTC ACG 3′
Nc(SF1)65: SF1 w/Nc65 overlap	5′ CGC CTC TTG CGC TCT ATA ACA GGA ACG ACA TCG GTC AAG TCG CTG CGG CCG CGC TTG ACA TCG CG 3′
$Nc(RS)_7$65	5′ GGG GGA TCC CGT AGC CGT AGC CGT AGC CGT AGC CGT AGC CGT AGC CGT AGC GAC TTG ACC GAT GTC GTT CC 3′
NcU2AF35	5′ GGG GGA TCC ATG GCC AAC TTT CTC GCC 3′
	5′ CCC GAA TTC CTA ATA CCG TCT CCT AGT 3′

Table 2: RNA−Protein Affinities (μM)[a]

	RNA	
protein	AV3	AV3pp
HsU2AF65	0.19 ± 0.05	16[b]
HsU2AF65ΔRS	0.66 ± 0.14	33.8 ± 3.2
HsU2AF65ΔRS+Hs35	0.22 ± 0.04	3.9 ± 0.4
NcU2AF65	37.4 ± 14.8[b]	74[b]
NcU2AF65ΔRS	≫100[c]	≫100[c]
NcU2AF65ΔRS+Nc35	38.1 ± 18.3[b]	80[b]
$Nc(RS)_7$U2AF65	7.6 ± 1.9	16.3 ± 2.6
Nc(SF1)U2AF65	0.32 ± 0.11	0.86 ± 0.24
QM	30[b]	160[b]
QMΔRS	≫60[c]	≫60[c]
QMΔRS+Hs35	0.70 ± 0.42	6.9 ± 1.1
RNPmut	80[b]	210[b]
RNPmutΔRS	110[b]	270[b]
RRM1mut	11.1 ± 2.4	140[b]
RRM2mut	53[b]	440[b]
chU2AF65	22[b]	28[b]
chU2AF65ΔRS	≫140[c]	≫140[c]
chU2AF65ΔRS+Hs35	1.6 ± 0.9	8.7 ± 4.9

[a] Calculated average K_d (μM) (±one standard deviation) for *in vitro* binding to RNA indicated. All values represent the average of at least three independent experiments. [b] Extrapolated K_d (beyond highest available protein concentration). [c] K_d too high to extrapolate.

ImageQuant software (Molecular Dynamics). Dissociation constants (K_d) (Table 2) were calculated as previously described (*10*) with all fraction bound values corrected for background by subtracting the fraction bound with no protein added.

RESULTS

Neurospora as a Model for Studying Introns Lacking the Canonical Polypyrimidine Tract. We examined the makeup of *Neurospora* PyTs in comparison to human using two approaches. We first calculated the frequency of predicted BPSs as a function of their distance from the 3′ end of the intron. We defined the BPS sequence to be CURAC (R = A or G), which corresponds with both the BPS identified by Kupfer et al. (*28*) and the high affinity site for the branch point binding protein SF1/BBP (*29*). Sequences matching this motif were found in the last 50 bases of 58% of *Neurospora* introns. Both human and *Neurospora* introns show a spatial distribution of BPS occurrences with a notable peak (Figure 1A). The peak is significantly closer to the 3′ss in *Neurospora* (13 nucleotides) than in human (21 nucleotides), suggesting that a large portion of *Neurospora* introns have short PyTs.

Enrichment of pyrimidines, especially uridines, at the 3′ end is also characteristic of many introns. In human introns the last 35 nucleotides of most introns are 35−50% uridine (Figure 1B). However, this enrichment is largely absent from *Neurospora* introns, which are only 25−30% uridine at the 3′ end (Figure 1B, thin lines). This value for *Neurospora* represents a virtually random base composition. *Neurospora* introns are also deficient in polypyrimidine tracts. The occurrence of four uridines in a row, representative of a strong PyT, is much less common in *Neurospora* than in

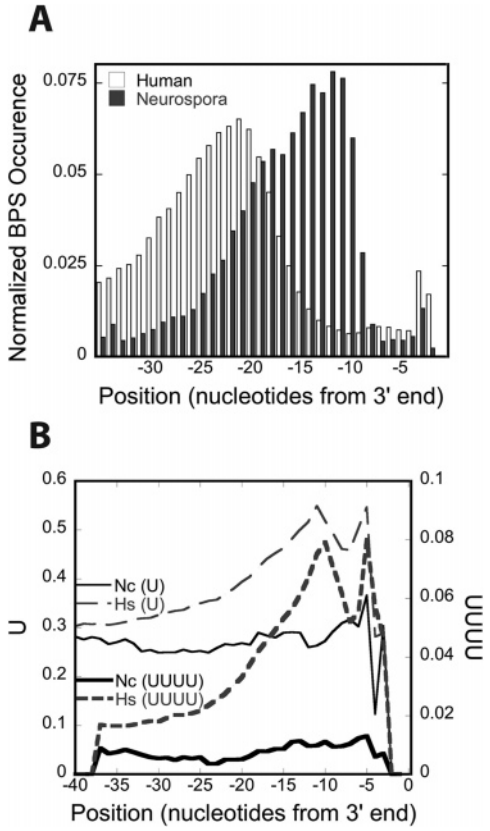


FIGURE 1: *Neurospora* introns contain degenerate polypyrimidine tracts. (A) BPS−3′ss distance is shortened in *Neurospora* introns. The normalized positional occurrence of the predicted branch point sequence CURAC is plotted relative to the 3′ splice site at 0. The total number of occurrences at each position was divided by the total number of occurrences throughout the region. Open bars, human; filled bars, *Neurospora*. (B) The 3′ ends of *Neurospora* introns are not enriched in uridine. The frequency of occurrence of U (thin lines, left axis) or poly-U (defined to be a contiguous run of four Us; thick lines, right axis) is plotted by position relative to the 3′ss (0). The frequency of occurrence (*Y*-axis) indicates the fraction of introns containing these *n*-mers at each position. Note that the two *Y*-axes have different scales. Solid lines, *Neurospora* (Nc); broken lines, human (Hs).

human introns (Figure 1B, thick lines). Taken together, these data indicate that many *Neurospora* introns contain significantly shortened and uridine-poor polypyrimidine tracts.

RNA Binding of Neurospora U2AF65 Requires the RS Domain. If *Neurospora* introns generally lack polypyrimidine tracts, and if U2AF65 binds to the PyT in order to accurately target the U2 snRNP to the 3′ss of the intron, then what are the RNA binding properties of *Neurospora* U2AF65 (NcU2AF65)? To study RNA−protein binding of all U2AF65 constructs, we used model PyT RNA oligonucleotides (Figure 2A, bottom) derived from the 3′ end of the Adenovirus major late (AdML) intron. AV3 is the wild-type sequence and contains BPS, PyT, and 3′ss motifs. AV3pp is a weak PyT model where purine mutations have interrupted the PyT.

To avoid problems with expression and solubility, we first cloned NcU2AF65 without its RS domain, which is extended (180 amino acids, compared to 85 for HsU2AF65). Surprisingly, binding of NcU2AF65ΔRS to AV3 as well as AV3pp RNA is nearly undetectable (Figure 2A, lanes 25−32). Previous studies have shown that human U2AF65 (HsU2AF65) does not require its RS domain for high affinity binding (*6*, *13*). Similarly, in our hands the affinity of HsU2AF65ΔRS is equivalent to full-length (Figure 2A, lanes 1−8 versus 9−16; Table 2).

The RNA-binding domains of NcU2AF65 are much more negatively charged than HsU2AF65 (Table 3). RRM1 of NcU2AF65 has a predicted isoelectric point (p*I*) of 4.47 (HsU2AF65 RRM1, 7.54), and the p*I* of RRM2 is 5.66 (HsU2AF65 RRM2, 7.35). These low p*I*s are similar to that of RRM3 (NcU2AF65, 4.23; HsU2AF65, 4.33), which does not bind RNA and is instead a protein−protein interaction domain (*30*). It is possible that the RS domain of NcU2AF65 is extended because its positive charge contributes to RNA binding. Indeed, full-length NcU2AF65 does bind to PyT RNA (Figure 2A, lanes 17−24). The affinity of NcU2AF65 is weak (K_d for AV3: 37 μM) compared to that for HsU2AF65 (190 nM). Combined with the lack of binding by NcU2AF65ΔRS, this suggests that the RNA-binding domains of NcU2AF65 are weak.

The addition of an unrelated RS domain to NcU2AF65ΔRS also restores RNA binding (Figure 2C). We fused either the RS domain of *Neurospora* SF1 or seven RS dipeptides [$(RS)_7$] to the N-terminus of NcU2AF65ΔRS. The resulting chimeric proteins are Nc(SF1)U2AF65 and Nc$(RS)_7$U2AF65, respectively. The RS domain from SF1 dramatically increased affinity (Figure 2C, lanes 1−12). Nc(SF1)U2AF65 binds to the AV3 RNA with a K_d of 320 nM, in the same range as HsU2AF65 (Table 2). Addition of the artificial domain $(RS)_7$ also resulted in affinity greater than that of full-length NcU2AF65 (Figure 2C, lanes 13−24). The K_d of Nc(RS_7)U2AF65 for AV3 is 7.6 μM (Table 2). It is unclear why these three different RS domains have such varying effects on NcU2AF65 RNA binding.

NcU2AF65 exhibits much less sequence selectivity than HsU2AF65. Whereas the difference in affinity between strong (AV3) and weak (AV3pp) PyT RNAs by HsU2AF65 is 50−80-fold, NcU2AF65, as well as Nc(SF1)U2AF65 and Nc$(RS)_7$U2AF65, has only 2-fold selectivity (K_ds, Table 2). These results indicate that *Neurospora* U2AF65 is much less specific in RNA binding than the human protein. If the RS domain is the primary contributor to RNA binding in NcU2AF65, low sequence selectivity would be expected. The positively charged RS domain is not expected to have any structure and is probably interacting nonspecifically with the negatively charged phosphate backbone.

The RS Domain Is Required for RNA Binding by Human U2AF65 When Its RRMs Are Mutated To Be More Acidic. The decrease in affinity and selectivity of NcU2AF65 could be a property of its RNA binding domain (RRM1 and RRM2), or of another portion of the protein. We mutated RRM1 and RRM2 of HsU2AF65 to decrease their isoelectric point. If low p*I* is an indicator of a weak RRM, this mutagenesis should decrease the RNA binding activity.

HsU2AF65 residues R150, K195, K260, and K300 were chosen for mutation to alanine (Figure 3A). These basic amino acids (or their equivalents) contact RNA in both the HsU2AF65−RNA (*31*) and Sxl−RNA (*32*) crystal structures through hydrogen bonding and ionic interactions, largely as members of the highly conserved RNP motifs, through which most RRMs bind RNA. The resulting quadruply mutated U2AF65 (R150A, K195A, K260A, K300A) will be referred to in this paper as QM. RRM1 and RRM2

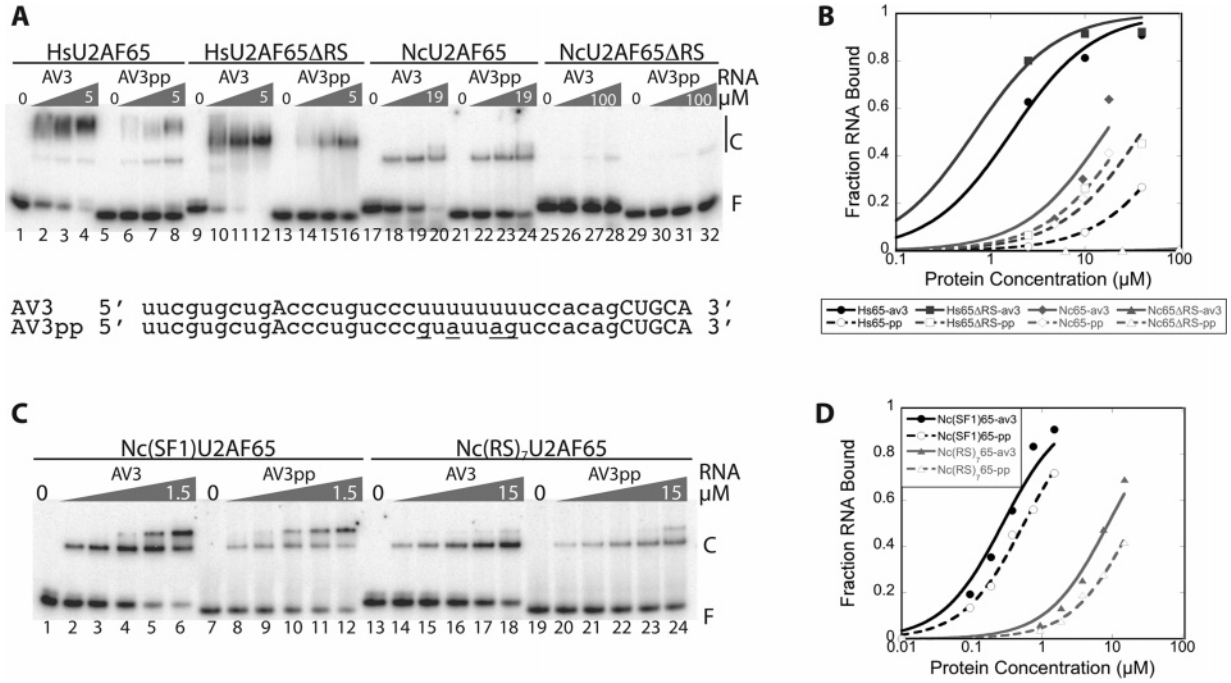


FIGURE 2: RNA binding by *Neurospora* U2AF65 is dependent upon its RS domain. (A) Electrophoretic mobility shift assay of human and *Neurospora* U2AF65. Final protein concentrations: HsU2AF65 (lanes 1−8) and HsU2AF65ΔRS (lanes 9−16), 0, 0.3125, 1.25, 5 μM; NcU2AF65, 0, 4.75, 9.5, 19 μM (lanes 17−24); NcU2AF65ΔRS, 0, 6.25, 25, 100 μM (lanes 25−32). C, RNA−protein complex; F, free RNA. RNA sequences are given below the gel. Intron sequence is lowercase, and exon sequence and the branch point A are in uppercase. Mutations in AV3pp are underlined. (B) Quantitation of (A) with best-fit line. Circles, HsU2AF65; squares, HsU2AF65ΔRS; diamonds, NcU2AF65; triangles, NcU2AF65ΔRS. Closed markers and solid lines indicate AV3 RNA; open markers and broken lines indicate AV3pp RNA. (C) Electrophoretic mobility shift assay of *Neurospora* U2AF65−artificial RS constructs. Labels are as in (B). Final protein concentrations: Nc(SF1)U2AF65, 0, 0.09375, 0.1875, 0.375, 0.75, 1.5 μM (lanes 1−12); Nc(RS)$_7$U2AF65, 0, 0.9375, 1.875, 3.75, 7.5, 15 μM (lanes 13−24). (D) Quantitation of (C) with best-fit line. Circles, Nc(SF1)U2AF65; triangles, Nc(RS)$_7$U2AF65. Closed markers and solid lines indicate AV3 RNA; open markers and broken lines indicate AV3pp RNA.

Table 3: Predicted Isoelectric Points (p*I*) of Proteins and Domains[a]

protein	full length	ΔRS	RRM1[b]	RRM2[c]	RRM3[d]
HsU2AF65	9.49	4.98	7.54	7.35	4.33
NcU2AF65	6.43	4.51	4.47	5.66	4.23
QM	8.73	4.71	4.91	4.87	4.33
RNPmut	9.38	4.99	7.55	5.77[e]	4.33
chU2AF65	7.52	4.54	4.47	5.66	4.33

[a] p*I*s calculated with EMBOSS (*46*). [b] Amino acids used for RRM1: HsU2AF65, 145−323; NcU2AF65, 257−351; chU2AF65, 147−241. [c] Amino acids used for RRM2: HsU2AF65, 255−340; NcU2AF65, 374−457; chU2AF65, 264−349. [d] Amino acids used for RRM3: HsU2AF65, 370−475; NcU2AF65, 479−584; chU2AF65, 379−484. [e] Includes G338E mutation; see text.

of QM have significantly lowered p*I*s (4.91 and 4.87, respectively; Table 3).

The QM construct binds a PyT RNA (AV3) with measurable affinity, though much more weakly than HsU2AF65, similar to NcU2AF65 (~30 μM (QM) vs 190 nM (HsU2AF65); Figure 3B, Table 2). Its sequence selectivity is also partially reduced; QM shows 5−8-fold selectivity, whereas U2AF65 displays 50-fold selectivity between AV3 and AV3pp RNAs (Table 2). It should be noted that this is an estimate; the K_d for each protein binding to AV3pp RNA (as well as QM for AV3) is extrapolated past the highest protein concentration we were able to test. Comparisons of the fraction of each RNA bound at equivalent protein concentrations, however, yield a similar fold selectivity. Basic amino acids in RRM1 and RRM2 of U2AF65 thus appear to play a role in RNA affinity as well as selectivity for pyrimidine-rich sequences.

Deletion of the RS domain (amino acids 1−85) from QM (forming QMΔRS) causes it to lose all appreciable RNA binding (Figure 3B, lanes 13−18; Table 2). In contrast, the RNA binding of wild-type HsU2AF65 is unaffected when the RS domain is removed (Figure 2A). This result indicates that the RS domain is able to compensate for the weakened interactions of the RRMs to RNA in the context of the QM mutations, though it is not required for wild-type U2AF65 to bind RNA (*6*, *13*).

U2AF35 Aids Binding by U2AF. Although U2AF65 is capable of recognizing and binding to a strong PyT on its own, during intron recognition it interacts with other splicing factors like U2AF35 and SF1/BBP. We predicted that these protein−protein interactions would be sufficient to bring U2AF65 to RNA lacking a high-affinity binding site. Alternatively, SF1/BBP and/or U2AF35 could assist a weak U2AF65 to bind RNA (regardless of sequence) between BPS and 3′ss. To test this model, we measured the binding affinity of a QMΔRS+HsU2AF35 heterodimer. Because recombinant U2AF35 is insoluble, we modified the coexpression strategy of Rudner et al. (*16*) to copurify U2AF heterodimers containing full-length U2AF35 and our constructs of U2AF65 (see Materials and Methods).

Addition of HsU2AF35 dramatically restored RNA binding to QMΔRS as a heterodimer (QMΔRS+Hs35). The affinity of QMΔRS+Hs35 is only 3.5-fold less than that of wild-type HsU2AF65 (Figure 3B, lanes 1−6 versus 19−24; Table 2). HsU2AF35 appears to be supplying a significant amount of RNA binding affinity in the QMΔRS heterodimer−RNA complex, perhaps by recruiting the large subunit

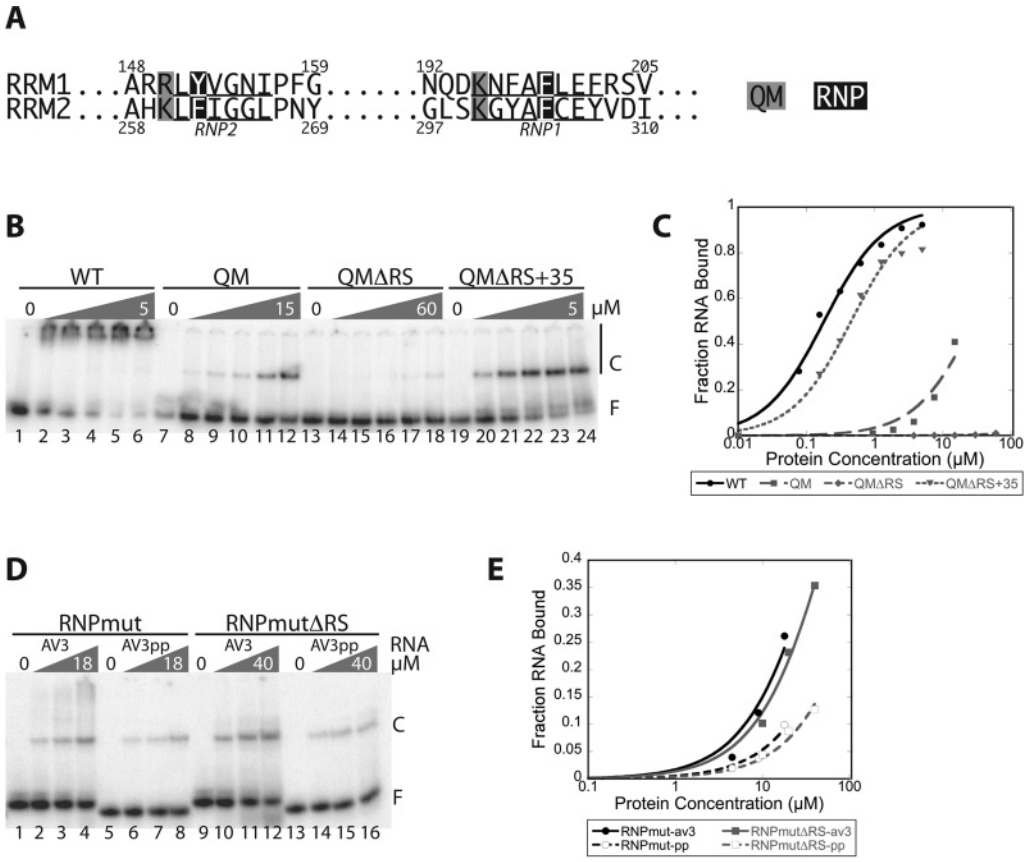


FIGURE 3: The RS domain is required for binding of U2AF65 with weakly binding acidic RRMs. (A) Schematic of mutations made in the first two RRMs of HsU2AF65. The primary sequences encompassing the RNP motifs of RRMs 1 and 2 are shown, with the RNP motifs underlined. Amino acids mutated to alanine in the quadruple mutant (QM) are in gray boxes (R150, K195, K260, K300). Amino acids mutated to alanine in the RNP mutant (RNP) are in black boxes (Y152, F199, F262, F304). (B) Electrophoretic mobility shift of QM constructs on a model 3′ splice site (AV3 RNA). Final protein concentrations: HsU2AF65 (WT) 0, 0.3125, 0.625, 1.25, 2.5, 5 μM (lanes 1−6); QM 0, 0.94, 1.88, 3.75, 7.5, 15 μM (lanes 7−12); QMΔRS 0, 3.75, 7.5, 15, 30, 60 μM (lanes 13−18); QMΔRS+Hs35 0, 0.3125, 0.625, 1.25, 5 μM (lanes 19−24). C, RNA−protein complex; F, free RNA. (C) Quantitation of (B) with best-fit line, average of at least three gels. Circles, WT; squares, QM; diamonds, QMΔRS; triangles, QMΔRS+Hs35. (D) Electrophoretic mobility shift assay of RNPmut constructs with AV3, AV3pp RNA. Labels as in (B). Final protein concentrations: RNPmut 0, 4.5, 9, 18 μM (lanes 1−8); RNPmutΔRS 0, 10, 20, 40 μM (lanes 9−16). (E) Quantitation of (D). Circles, RNPmut; squares, RNPmutΔRS. Closed markers and solid lines indicate AV3 RNA; open markers and broken lines indicate AV3pp RNA.

to the RNA, though its own RNA affinity is weak (*16*). Alternatively, U2AF35′s role may be supplying an RS domain to the U2AF complex. However, its RNA-binding contribution to QMΔRS does not appear to be equivalent to the contribution of the RS domain in full-length QM. The affinity of QM is 40-fold less than that of QMΔRS+Hs35 (Figure 3B, lanes 7−12 versus 19−24; Table 2). Therefore, a combination of these two mechanisms may be at work.

The combination of *Neurospora* U2AF35 (NcU2AF35) with NcU2AF65ΔRS also restored binding (data not shown). Unlike with QMΔRS, where addition of HsU2AF35 resulted in a heterodimer with affinity much greater than that of full-length QM, NcU2AF65ΔRS+Nc35 has similar affinity to NcU2AF65 (K_d for AV3: 38 μM and 37 μM, respectively; Table 2). This may reflect lower activity *in vitro* for NcU2AF35 compared to HsU2AF35, or could be due to much weaker RNA binding of NcU2AF65ΔRS compared to QMΔRS (although both were unmeasurable). Whereas a shorter construct of HsU2AF35 (amino acids 43−146) is soluble alone (*33*), we have had no success expressing truncated versions of NcU2AF35. However, the activity of NcU2AF35 is enough to lend measurable affinity to a U2AF heterodimer (NcU2AF65ΔRS+Nc35) in which the large subunit (NcU2AF65ΔRS) has very low binding activity.

Whereas both single-subunit HsU2AF65ΔRS and the HsU2AF65ΔRS+Hs35 heterodimer show sequence selectivity between AV3 and AV3pp RNAs, HsU2AF65ΔRS alone has 50-fold selectivity while that for HsU2AF65+Hs35 is approximately 18-fold (Table 2 and data not shown). A similar 5- to 10-fold change in selectivity is likely true for QMΔRS and its heterodimer, though this cannot be directly measured, because QMΔRS alone does not bind RNA. Coupled with the increased affinities of the human heterodimers for a mutated PyT (AV3pp) compared to that of the large subunit alone, these results indicate that U2AF35 aids binding of the U2AF heterodimer to RNA with a poor PyT, regardless of the binding affinity of U2AF65.

The RS Domain Is Not Required for Weakened U2AF65 with Unchanged pI. We made another construct, the RNP mutant (RNPmut), containing a set of mutations (Y152A, F199A, F262A, F304A) in HsU2AF65 to decrease RNA affinity without affecting charge (Figure 3A). These conserved aromatic amino acids in the RNP motifs stack with bases in the RNA and are the most important residues in

RRM−RNA interactions (*31*). Consistent with this role for the targeted residues, the RNPmut has affinity for PyT RNA even lower than that of QM (AV3, 80 μM, AV3pp, 210 μM; Figure 3D, Table 2). Mutating individual RRMs (RRM1mut, Y152A+F199A; RRM2mut, F262A+F304A) also decreased affinity to a lesser extent (Table 2 and data not shown). Mutating RRM2 had a larger effect than RRM1 (K_d for AV3: 53 μM versus 11 μM, respectively; Table 2), consistent with previous studies which showed that RRM2 contributes more to affinity than RRM1 (*31*, *34*).

The RRMs of RNPmut have nearly identical charge properties to the wild-type HsU2AF65 (Table 3). It should be noted that the core of RRM2 in RNPmut has a very similar p*I* to wild-type; our calculations for Table 3 included approximately 5 amino acids on either side of the RRMs. An additional mutation just outside the RRM (G338E) introduced an acidic amino acid during RNPmut mutagenesis that does not have a significant effect on RNA binding. RNP mutations in RRM2 alone (Table 2, RRM2mut) cause a significant decrease in binding in the absence of the G338E mutation.

Removing the RS domain from RNPmut has minimal effect on affinity (Figure 3D, lanes 9−16; Table 2). This corroborates previous ΔRS results: those constructs with acidic RRMs (NcU2AF65, QM) lose all RNA binding when the RS domain is deleted, and the basic RRMs of wild-type HsU2AF65 are unaffected by deletion of the RS domain. Interestingly, overall protein p*I* is not a reliable indicator of RNA binding activity; all the ΔRS constructs have p*I*s between 4.5 and 5 (Table 3), but display very different affinities (Table 2).

An RS Domain or U2AF35 Compensates for Lack of RNA Binding by Neurospora RRMs. In order to directly compare the RNA binding of NcU2AF65 with the mutant versions of HsU2AF65, we constructed a chimeric protein (chU2AF65; Figure 4A) where RRMs 1 and 2 of HsU2AF65 (amino acids 140−350) were replaced with RRMs 1 and 2 from NcU2AF65 (amino acids 250−468). The protein context surrounding the RNA-binding domain is thus identical and the activity of RRMs 1 and 2 can be directly measured between constructs.

The chU2AF65 construct binds RNA very similarly to NcU2AF65 (Figure 4B, lanes 1−8). This result confirms that RRMs 1 and 2 are major contributors to RNA binding for NcU2AF65, and not other portions of the protein, which have been completely replaced in chU2AF65 with segments of HsU2AF65. Sequence selectivity between AV3 and AV3pp is approximately 2-fold or less at the protein concentrations tested (Table 2 and data not shown). This suggests that, as seen in the wild-type protein, NcU2AF65's RRMs are not very selective for "strong" PyT sequences.

The RS domain is, once again, an important contributor to affinity; chU2AF65ΔRS (Δaa1−85) does not bind RNA (Figure 4B, lanes 9−16). Length of the RS domain does not appear to be as important as the presence of one. While NcU2AF65 has a 180-residue RS domain, chU2AF65 has the 85-residue RS of HsU2AF65.

Adding HsU2AF35 also restores affinity to chU2AF65ΔRS (Figure 4B, lanes 17−24). As with QMΔRS+Hs35, the affinity of chU2AF65ΔRS+Hs35 is much greater than that of full-length chU2AF65 alone (K_d for AV3: 1.6 μM and 22 μM, respectively; see Table 2). The heterodimer has a 5-fold selectivity of between strong and weak PyTs. Because

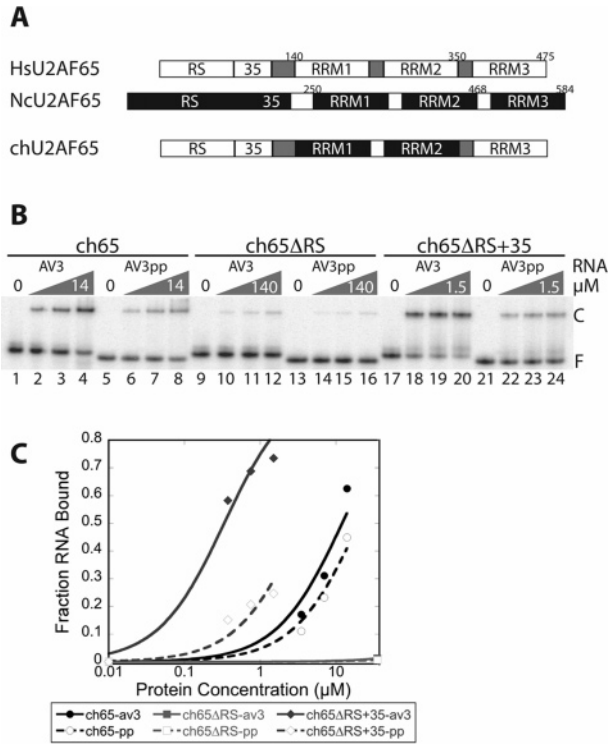
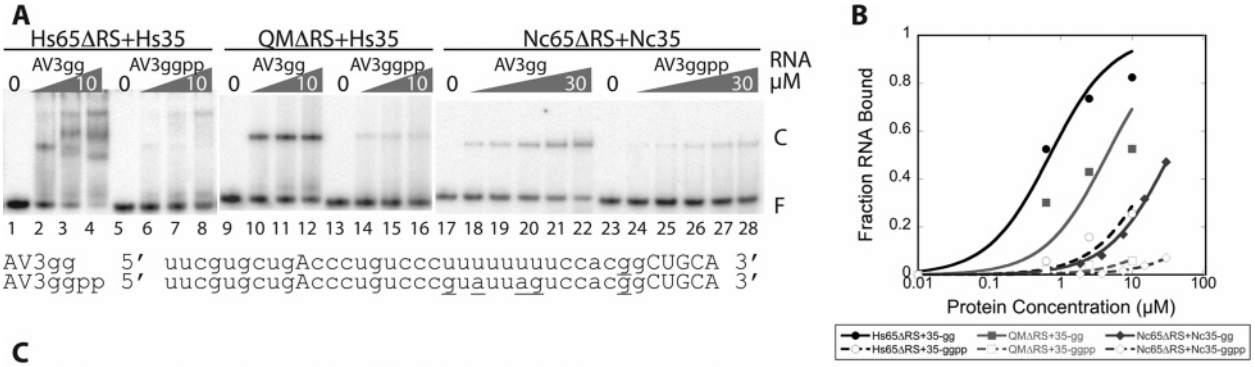


FIGURE 4: A chimera of HsU2AF65 with NcU2AF65 RRMs 1+2 binds similarly to NcU2AF65. (A) Schematic of chimera (chU2AF65) construction. Amino acids 140−350 (RRM1+2) of HsU2AF65 were replaced with amino acids 250−468 (RRM1+2) of NcU2AF65. (B) Electrophoretic mobility shift assay of chimera constructs with AV3, AV3pp RNA. Final protein concentrations: chU2AF65 (ch65), 0, 3.5, 7, 14 μM (lanes 1−8); ch65ΔRS 0, 35, 70, 140 μM (lanes 9−16); ch65ΔRS+Hs35 0, 0.375, 0.75, 1.5 μM (lanes 17−24). C, RNA−protein complex; F, free RNA. (C) Quantitation of (B). Circles, chU2AF65; squares, chU2AF65ΔRS; diamonds, ch65ΔRS+Hs35. Closed markers and solid lines indicate AV3 RNA; open markers and broken lines indicate AV3pp RNA.

U2AF35 recognizes the 3′ss and exon bases downstream (*18*), which are identical between the two RNAs, this selectivity is likely due to the NcRRMs present in chU2AF65. Greater selectivity may be seen in the heterodimer than in full-length chU2AF65 alone due to the increase in affinity, which suggests that U2AF35 contributes more than the RS domain to helping the weak NcRRMs bind RNA.

Effect of 3′ss AG on U2AF35 Affinity of U2AF Heterodimer. We further explored the way that U2AF35 was contributing to the RNA binding activity of U2AF with additional mutations of the RNA. U2AF35 has been shown to recognize the AG dinucleotide at the 3′ss, and mutation to GG or UG abolishes cross-linking (*18*−*20*). Interestingly, mutation of the 3′ss AG to GG in the AV3 RNA binding substrate (AV3gg; Figure 5A, bottom) does not significantly affect U2AF affinity in the heterodimers tested (Figure 5). The K_d of HsU2AF65ΔRS+Hs35 is 0.22 μM for AV3 and 0.33 μM for AV3gg (Figure 5A, lanes 1−4; Figure 5C), equivalent within experimental error. Similarly, NcU2AF65ΔRS+Nc35 has a K_d of 38 μM for AV3 and 30 μM for AV3gg (Figure 5A, lanes 17−22; Figure 5C). An effect is seen with QMΔRS+Hs35, which binds to AV3 and AV3gg with K_ds of 0.7 μM and 3 μM, respectively (Figure 5A, lanes 9−12; Figure 5C).

A greater effect on binding is apparent when the 3′ss AG is mutated in an RNA which also has a weakened



C

Protein	RNA			
	AV3	AV3pp	AV3gg	AV3ggpp
Hs65ΔRS+Hs35	0.22±0.04	3.9±0.4	0.33±0.09	15.0±7.1
QMΔRS+Hs35	0.70±0.42	6.9±1.1	3	60*
Nc65ΔRS+Nc35	38.1±18.3	80*	29.9±3.6	320*

FIGURE 5: U2AF affinity is partially dependent on the 3′ss AG dinucleotide. (A) Electrophoretic mobility shift assay of U2AF constructs with 3′ss mutated RNAs AV3gg and AV3ggpp. Final protein concentrations: HsU2AF65ΔRS+Hs35 (lanes 1−8) and QMΔRS+Hs35 (lanes 9−16), 0, 0.625, 2.5, 10 μM; NcU2AF65ΔRS+Nc35, 0, 1.875, 3.75, 7.5, 15, 30 μM. C, RNA−protein complex; F, free RNA. RNA sequences are given below the gel, with intron and exon sequences, branch point, and mutations indicated as in Figure 2A. (B) Quantitation of (A). Circles, HsU2AF65ΔRS+Hs35; squares, QMΔRS+Hs35; diamonds, NcU2AF65ΔRS+Nc35. Closed markers and solid lines indicate AV3gg RNA; open markers and broken lines indicate AV3ggpp RNA. (C) K_d values (μM) for U2AF constructs. The values for AV3 and AV3pp are reiterated from Table 2 for comparison. Asterisk (*) denotes an extrapolated K_d (beyond highest available protein concentration).

PyT (AV3ggpp; Figure 5A, bottom). The affinity of HsU2AF35ΔRS+Hs35 for AV3ggpp is 2−4-fold lower than for AV3pp (Figure 5C; Figure 5A, lanes 5−8). The *Neurospora* heterodimer (NcU2AF65ΔRS+Nc35) also exhibits a 4-fold decrease in affinity (Figure 5C; Figure 5A, lanes 23−28). For QMΔRS+Hs35, mutation of the AG to GG with a poor PyT decreases binding as much as 10-fold (Figure 5C; Figure 5A, lanes 13−16). Similar results were observed with chU2AF65ΔRS+Hs35 (data not shown). These effects on U2AF heterodimer binding of mutating the 3′ss AG to GG are much less dramatic than expected. However, the greater effects of the AG-GG mutation in the poor PyT RNA are logical in light of the role of U2AF35 in recruiting U2AF65 to an intron with a weak PyT. Taken together, these results complement earlier studies implicating the 3′ss AG dinucleotide as a key recognition element for U2AF35.

DISCUSSION

The pyrimidine enrichment and length of many vertebrate introns is highly variable, and a number of introns appear to lack polypyrimidine tracts altogether. We have determined that *Neurospora crassa* is an excellent model organism for studying introns without PyTs. Our comparison of human and *Neurospora* PyTs (Figure 1) confirms a previous study (*28*) which used a different PyT scoring method and also found a deficiency of PyTs in *Neurospora* introns. The notable lack of uridine enrichment in the last 40 bases of *Neurospora* introns (Figure 1B) is indicative of a scarcity of strong PyTs.

The short distance between BPS and 3′ss we observed may even be too short for U2AF65 to bind. Experimental evidence has shown that, in human splicing, moving the branch point A closer than 17 bases to the 3′ss makes the intron require an AG at the 3′ss for the first step of splicing (*35*). The requirement for an AG in these introns suggests it is possible that U2AF65 bridges these PyTs solely through protein−protein interactions with other splicing factors bound to surrounding sequences, especially U2AF35 at the 3′ss, a model also supported by our *in vitro* binding studies. The peak of BPS distribution in *Neurospora* is 13 nucleotides from the 3′ss, well within this limit, showing that many *Neurospora* PyTs are unusual in their short length and probably dependent upon U2AF35. All of the introns analyzed for Figure 1 end in AG, by definition during the building of the intron databases. Notably, in *Schizosaccharomyces pombe*, the median branch point−3′ss distance is 9 nucleotides, 3′ enrichment of pyrimidines is absent, and all introns appear to be AG-dependent (*36*−*38*).

The PyT is bound by U2AF65 as part of the early recognition of the 3′ end of the intron. In *Neurospora*, U2AF65 has apparently adapted to the general lack of PyTs by virtue of low sequence selectivity. Additionally, it relies on U2AF35 (Figures 4 and 5) and probably other factors like DEK (*39*) and SF1/BBP, with which it interacts *in vitro* (data not shown), to target it to the correct 3′ss. It is possible that *Neurospora* contains more than one U2AF65-like protein, and another splicing factor is present which does bind PyT RNA with higher selectivity and fulfills the canonical role of U2AF65. However, no obvious paralogue exists in the genome, and we have been unable to resolve a heterokaryotic NcU2AF65 knockout strain made by the *Neurospora* Knockout Project (*40*), which is ascospore lethal (data not shown). This indicates that NcU2AF65 is an essential gene, likely required to splice at least a subset of essential genes.

It is also possible that the highest affinity sequence for NcU2AF65 is not a polypyrimidine tract. We consider this unlikely, as the *in vitro* activity of NcU2AF65 is consistent with that of U2AF65 from other organisms (*6*, *16*, *25*, *41*). We observe that polyU is the best competitor for NcU2AF65

in a polynucleotide competition assay (data not shown), and NcU2AF65 is selective for an uninterrupted PyT (AV3, AV3gg) over a mutated PyT (AV3pp, AV3ggpp) (Figures 2, 5; Table 2). While identification of additional, perhaps compensating, motifs at the 3′end of *Neurospora* introns is beyond the scope of this study, we note that the first base of the 3′ exon in *Neurospora*, as in other organisms, is often a G (data not shown). A previous study found apparent (though unspecified) information content in the 3′ exon in *Neurospora* (*28*). This may represent an extended NcU2AF35 binding site for some introns.

It was surprising to observe any requirement for the RS domain in RNA binding by U2AF65, as we and others have shown this to be unnecessary for HsU2AF65 (*6*, *13*). Our results suggest that the RS domain does have a role in RNA binding, which is minor in comparison to the affinity of the wild-type HsU2AF65 RRMs. When the RRMs are weakened, as in NcU2AF65 (Figure 2) and QM (Figure 3), the contribution of the RS domain to RNA affinity becomes more apparent. We do not believe that the RNA binding activity of these proteins is solely through the intact RS domain, as they do display sequence selectivity, which the RS domain is unlikely to confer. The role of the RS domain is likely nonspecific, particularly in light of our ability to restore affinity to NcU2AF65ΔRS with the addition of unrelated RS domains (Figure 2C, 2D). The role of the RS domain may be to lend enough positive charge for negatively charged RRMs to approach the RNA and function through their usual mode of binding. Thus, the RNA binding of RRM-containing proteins may be regulated by RS domains and other basic domains. This model is substantiated by our mutagenesis of HsU2AF65. The QM, which was designed to decrease the p*I* of RRMs 1 and 2 (Table 3), requires an RS domain in order to bind RNA (Figure 3B, 3C). However, the affinity of the RNPmut, where p*I* was unaffected, remains unchanged when the RS domain is deleted (Figure 3D, 3E).

Recent studies have shown that the RS domains of multiple splicing factors contact the pre-mRNA to facilitate spliceosome assembly by promoting interactions between snRNA and pre-mRNA (*13*, *14*, *42*). This function is largely dependent upon positive charge *in vitro*. While our study looked at interactions between protein and single-stranded, rather than double-stranded, RNA, we also observed that positively charged amino acids were important. It is likely that nonspecific contact of the negatively charged RNA backbone is a common function of RS domains. This contact facilitates both RNA−RNA and RNA−protein interactions. Although charge in the RS domain is a key factor, other aspects are important as well, as demonstrated by our finding that different RS domains affected NcU2AF65 binding differently (Figure 2). The RS domain of NcSF1 (pI 8.34) greatly enhances binding, while a more basic RS domain, that of NcU2AF65 (pI of 11.5), only modestly increases binding, and an even more basic artificial RS domain, $(RS)_7$ (p*I* 12.8), has an intermediate effect. Interestingly, the SF1 RS domain is conserved in *Drosophila* and *Caenorhabditis elegans* (*43*) but absent from human SF1 and yeast BBP, and might play an important role in SF1 RNA binding or RNA−RNA annealing in flies, worms, and fungi.

The remarkable increase in affinity of NcU2AF65, QM, and chU2AF65 as heterodimers with U2AF35 shows that the small subunit is a major contributor to RNA binding by the U2AF heterodimer. Because U2AF35 binds to the 3′ss (*18*−*20*), probably without regard to PyT strength, it is probably instrumental in the recognition of introns with weak PyTs, which may not be easily recognized by U2AF65. Indeed, the sequence selectivity of HsU2AF35ΔRS+Hs35 (Table 2) as well as full-length U2AF (data not shown) is significantly less than that of the large subunit alone. It may be more accurate to refer to recruitment of U2AF rather than its subunits; the heterodimer as a whole is required to bind at the 3′ end of the intron, but this binding can come through U2AF65 or U2AF35 (or both).

It is likely that U2AF35 binds nonspecifically to RNA, especially when not positioned by U2AF65. While we observed a decrease in affinity for all heterodimers when the AG was changed to GG (Figure 5), some binding remained. Binding to the GG mutant RNAs was most surprising for the quadruple mutant (QM) heterodimers, as QMΔRS alone has no measurable RNA-binding activity. The RNA affinity of the QM heterodimers was thus expected to be dependent upon U2AF35, which, based on previous studies (*18*−*20*), was expected to be dependent on an AG dinucleotide. The affinity of QMΔRS+Hs35 for AV3gg and AV3ggpp suggests that the 3′ss AG, while important, is not the sole sequence target of U2AF35.

The cross-linking study described by Wu et al. (*18*) implicating the AG was done with U2AF from nuclear extracts and an adenovirus RNA with a strong PyT. While they saw a loss of cross-linking at the 3′ss when they mutated the AG, cross-linking at other nucleotides was not examined with this mutation, so it is possible that U2AF35 contacts nearby RNA nonspecifically, such as in the downstream exon. U2AF from HeLa extracts showed a clear preference for AG following a PyT in SELEX experiments, and the 3′ss AG is known to be important for splicing, especially when the PyT is weak (*15*, *18*, *22*, *35*, *44*). Our binding results suggest that the selection of AG may be only partially due to U2AF35 binding. DEK (which may have been copurified with the U2AF used by Wu et al.) and other splicing factors may have significant roles in modulation of 3′ss selection (*39*).

It is becoming apparent that the RNA-binding activity of splicing factors is only a part of their function. Equally or even more important is their binding to each other: cooperatively, to tether each other when recognition sequences are not optimal (*8*), as well as simply to establish cross-talk across intervening sequences. Our data suggest a model of U2AF65 binding where additional domains (such as the RS) or splicing factors (like U2AF35) may compensate for weak RNA binding by RRMs 1+2 of U2AF65. In addition, interactions with other splicing factors bound to neighboring RNA sequences could bring U2AF65 to the 3′ss without it binding RNA, especially in those introns with BPS-3′ss spacing too short for U2AF65. Such a model has been proposed for the splicing of introns that require splicing enhancers (*15*, *18*, *22*, *45*).

While we have looked primarily at how mutations in the large subunit affect RNA binding activity, our results may also be predictive for the role of U2AF in identifying introns which lack strong PyTs. U2AF35, SF1/BBP, and other splicing factors are likely to play important roles in bringing U2AF65 to a weak polypyrimidine tract to which it might not bind on its own.

ACKNOWLEDGMENT

Sincere thanks are due to A. Mahady for consummate technical assistance, A. Barkan for comments on the manuscript, G. Kothe and S. Honda for *Neurospora* cDNA, and Z. Lewis, K. Adhvaryu, and E. Selker for assistance and discussions on *Neurospora*.

REFERENCES

1. Burge, C. B., Tuschl, T., and Sharp, P. A. (1999) Splicing of precursors to mRNAs by the spliceosomes, in *The RNA World* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 525−560, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Black, D. L. (2003) Mechanisms of alternative pre-messenger RNA splicing, *Annu. Rev. Biochem. 72*, 291−336.
3. Lim, L. P., and Burge, C. B. (2001) A computational analysis of sequence features involved in recognition of short introns, *Proc. Natl. Acad. Sci. U.S.A. 98*, 11193−11198.
4. Nilsen, T. W. (1998) RNA-RNA interactions in nuclear pre-mRNA splicing, in *RNA Structure and Function* (Simons, R. W., and Grunberg-Manago, M., Eds.) 279−307, Cold Spring Harbor Laboratory Press, Plainview, NY.
5. Ruskin, B., Zamore, P. D., and Green, M. R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly, *Cell 52*, 207−219.
6. Zamore, P. D., Patton, J. G., and Green, M. R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF, *Nature 355*, 609−614.
7. Banerjee, H., Rahn, A., Gawande, B., Guth, S., Valcárcel, J., and Singh, R. (2004) The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF 65) is essential in vivo but dispensable for activity in vitro, *RNA 10*, 240−253.
8. Berglund, J. A., Abovich, N., and Rosbash, M. (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition, *Genes Dev., 12*, 858−867.
9. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Krämer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP, *Mol. Cell 11*, 965−976.
10. Cass, D. M., and Berglund, J. A. (2006) The SF3b155 N-Terminal Domain Is a Scaffold Important for Splicing, *Biochemistry 45*, 10092−10101.
11. Gozani, O., Potashkin, J., and Reed, R. (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site, *Mol. Cell. Biol. 18*, 4752−4760.
12. Singh, R., and Valcárcel, J. (2005) Building specificity with nonspecific RNA-binding proteins, *Nat. Struct. Mol. Biol. 12*, 645−653.
13. Valcárcel, J., Gaur, R. K., Singh, R., and Green, M. R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA, *Science 273*, 1706−1709.
14. Shen, H., and Green, M. R. (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly, *Mol. Cell 16*, 363−373.
15. Guth, S., Tange, T. Ø., Kellenberger, E., and Valcárcel, J. (2001) Dual function for U2AF35 in AG-dependent pre-mRNA splicing, *Mol. Cell. Biol. 21*, 7673−7681.
16. Rudner, D. Z., Breger, K. S., Kanaar, R., Adams, M. D., and Rio, D. C. (1998) RNA binding activity of heterodimeric splicing factor U2AF: at least one RS domain is required for high-affinity binding, *Mol. Cell. Biol. 18*, 4004−4011.
17. Rudner, D. Z., Breger, K. S., and Rio, D. C. (1998) Molecular genetic analysis of the heterodimeric splicing factor U2AF: the RS domain on either the large or small *Drosophila* subunit is dispensable in vivo, *Genes Dev. 12*, 1010−1021.
18. Wu, S., Romfo, C. M., Nilsen, T. W., and Green, M. R. (1999) Functional recognition of the 3′ splice site AG by the splicing factor U2AF35, *Nature 402*, 832−835.
19. Merendino, L., Guth, S., Bilbao, D., Martínez, C., and Valcárcel, J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3′ splice site AG, *Nature 402*, 838−841.
20. Zorio, D. A. R., and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3′ splice site in *Caenorhabditis elegans*, *Nature 402*, 835−838.
21. Zamore, P. D., and Green, M. R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution, *EMBO J. 10*, 207−214.
22. Guth, S., Martínez, C., Gaur, R. K., and Valcárcel, J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF35 and for its function after polypyrimidine tract recognition by U2AF65, *Mol. Cell. Biol. 19*, 8263−8271.
23. Davis, R. H. (2000) *Neurospora: Contributions of a Model Organism*, Oxford University Press, New York.
24. Voelker, R. B., and Berglund, J. A. (2007) A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing, *Genome Res. 17*, 1023−1033.
25. Henscheid, K. L., Shin, D. S., Cary, S. C., and Berglund, J. A. (2005) The splicing factor U2AF65 is functionally conserved in the thermotolerant deep-sea worm *Alvinella pompejana*, *Biochim. Biophys. Acta 1727*, 197−207.
26. Rudner, D. Z., Kanaar, R., Breger, K. S., and Rio, D. C. (1998) Interaction between subunits of heterodimeric splicing factor U2AF is essential in vivo, *Mol. Cell. Biol. 18*, 1765−1773.
27. Kholod, N., and Mustelin, T. (2001) Novel vectors for co-expression of two proteins in *E. coli*, *Biotechniques 31*, 322−328.
28. Kupfer, D. M., Drabenstot, S. D., Buchanan, K. L., Lai, H., Zhu, H., Dyer, D. W., Roe, B. A., and Murphy, J. W. (2004) Introns and splicing elements of five diverse fungi, *Eukaryotic Cell 3*, 1088−1100.
29. Garrey, S. M., Voelker, R., and Berglund, J. A. (2006) An extended RNA binding site for the yeast branch point-binding protein and the role of its zinc knuckle domains in RNA binding, *J. Biol. Chem. 281*, 27443−27453.
30. Kielkopf, C. L., Lucke, S., and Green, M. R. (2004) U2AF homology motifs: protein recognition in the RRM world, *Genes Dev. 18*, 1513−1526.
31. Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., and Kielkopf, C. L. (2006) Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65, *Mol. Cell 23*, 49−59.
32. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein, *Nature 398*, 579−585.
33. Kielkopf, C. L., Rodionova, N. A., Green, M. R., and Burley, S. K. (2001) A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer, *Cell 106*, 595−605.
34. Ito, T., Muto, Y., Green, M. R., and Yokoyama, S. (1999) Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF65), *EMBO J. 18*, 4523−4534.
35. Reed, R. (1989) The organization of 3′ splice-site sequences in mammalian introns, *Genes Dev. 3*, 2113−2123.
36. Csank, C., Taylor, F. M., and Martindale, D. W. (1990) Nuclear pre-mRNA introns: analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes, *Nucleic Acids Res. 18*, 5133−5141.
37. Romfo, C. M., and Wise, J. A. (1997) Both the polypyrimidine tract and the 3′ splice site function prior to the first step of splicing in fission yeast, *Nucleic Acids Res. 25*, 4658−4665.
38. Zhang, M. Q., and Marr, T. G. (1994) Fission yeast gene structure and recognition, *Nucleic Acids Res. 22*, 1750−1759.
39. Soares, L. M., Zanier, K., Mackereth, C., Sattler, M., and Valcárcel, J. (2006) Intron removal requires proofreading of U2AF/3′ splice site recognition by DEK, *Science 312*, 1961−1965.
40. Colot, H. V., Park, G., Turner, G. E., Ringelberg, C., Crew, C. M., Litvinkova, L., Weiss, R. L., Borkovich, K. A., and Dunlap, J. C. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors, *Proc. Natl. Acad. Sci. U.S.A. 103*, 10352−10357.

41. Singh, R., Valcárcel, J., and Green, M. R. (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins, *Science 268*, 1173−1176.
42. Shen, H., Kan, J. L., and Green, M. R. (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly, *Mol. Cell 13*, 367−376.
43. Mazroui, R., Puoti, A., and Krämer, A. (1999) Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability, *RNA 5*, 1615−1631.
44. Zuo, P., and Maniatis, T. (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing, *Genes Dev. 10*, 1356−1368.
45. Graveley, B. R., Hertel, K. J., and Maniatis, T. (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing, *RNA 7*, 806−818.
46. Rice, P., Longden, I., and Bleasby, A. (2000) EMBOSS: The European Molecular Biology Open Software Suite, *Trends Genet. 16*, 276−277.

BI701240T