Published in final edited form as:

J Mol Biol. 2006 February 24; 356(3): 664–683.

Multiple U2AF⁶⁵ binding sites within SF3b155: Thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors

Karen R. Thickman¹, Matthew Swenson¹, Joseph M. Kabogo^{1,2}, Zygmunt Gryczynski³, and Clara L. Kielkopf^{1,*}

¹ Department of Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA

³ University of Maryland, School of Medicine, Center for Fluorescence Spectroscopy, Department of Biochemistry and Molecular Biology, Baltimore, MD 21201, USA

Abstract

Essential, protein-protein complexes between the large subunit of the U2 small nuclear RNA Auxiliary Factor (U2AF⁶⁵) with the Splicing Factor 1 (SF1) or the spliceosomal component SF3b155 are exchanged during a critical, ATP-dependent step of pre-mRNA splicing. Both SF1 and the N-terminal domain of SF3b155 interact with a U2AF homology motif (UHM) of U2AF⁶⁵. SF3b155 contains seven tryptophan-containing sites with sequence similarity to the previously characterized U2AF⁶⁵-binding domain of SF1. We show that the SF3b155 domain lacks detectable secondary structure using circular dichroism spectroscopy, and demonstrate that five of the tryptophan-containing SF3b155 sites are recognized by the U2AF⁶⁵-UHM using intrinsic tryptophan fluorescence experiments with SF3b155 variants. When compared with SF1, similar spectral shifts and sequence requirements indicate that U2AF⁶⁵ interactions with each of the SF3b155 sites are similar to the minimal SF1 site. However, thermodynamic comparison of SF1 or SF3b155 proteins with minimal peptides demonstrates that formation the SF1/U2AF⁶⁵ complex is likely to affect regions of SF1 beyond the previously identified, linear interaction site, in a remarkably distinct manner from the local U2AF⁶⁵ binding mode of SF3b155. Furthermore, the complex of the SF1/U2AF⁶⁵ interacting domains is stabilized by 3.3 kcal mol⁻¹ relative to the complex of the SF3b155/U2AF⁶⁵ interacting domains, consistent with the need for ATP hydrolysis to drive exchange of these partners during pre-mRNA splicing. We propose that the multiple U2AF⁶⁵ binding sites within SF3b155 regulate conformational rearrangements during spliceosome assembly. Comparison of the SF3b155 sites defines an (R/K)_nXRW(DE) consensus sequence for predicting U2AF⁶⁵-UHM ligands from genomic sequences, where parentheses denote residues that contribute to, but are not required for binding.

Keywords

calorimetry; fluorescence; RNA splicing; spliceosome assembly; U2AF homology motif

Introduction

Proteins responsible for the initial steps of pre-mRNA splicing, Splicing Factor 1 (SF1)¹ and the heterodimeric splicing factor U2 small nuclear (sn)RNA Auxiliary Factor (U2AF)^{2; 3} are

* Corresponding author Phone: 443-287-4546, Fax: 410-955-2926, E-mail: ckielkop@jhsph.edu

²Current address: Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

essential for viability in higher eukaryotes. During pre-mRNA splicing, non-coding sequences called introns are precisely excised from pre-mRNA transcripts before translation of the processed mRNAs into proteins. Alternative splicing of different protein coding regions (exons) within a pre-mRNA provides one of the major sources of molecular diversity for cellular growth and development⁴. One of the most complicated molecular machines in the cell, an assembly of ribonucleoprotein particles (RNPs) called the spliceosome⁵, accomplishes the task of pre-mRNA splicing through a series of conformational changes among RNA-RNA, protein-RNA, and protein-protein complexes driven by ATP-dependent, RNP unwindases⁶. The components of the spliceosome are directed to appropriate splice sites by short, conserved pre-mRNA sequences, including a branch point sequence (BPS) that provides the proposed nucleophile of the splicing reaction and poly-pyrimidine tract (Py-tract) near the 3' splice site⁷.

The first ATP-dependent step of pre-mRNA splicing is a critical regulated step that commits a specific pair of splice sites to be joined⁸. To begin this initial step of pre-mRNA splicing, a protein-protein complex between SF1 and the U2AF heterodimer cooperatively associates with consensus sequences near the 3' splice site of the pre-mRNA⁹. The protein complex is arranged so that SF1 recognizes the BPS¹⁰ and the U2AF large subunit (U2AF⁶⁵) recognizes the Py-tract¹¹, thereby ensuring that adjacent pre-mRNA sequences are simultaneously recognized by the respective protein subunits. The N-terminal, 137-residues of human SF1 are necessary and sufficient to mediate interactions with U2AF⁶⁵ in nuclear extracts or yeast two-hybrid screens¹². The U2AF⁶⁵-interaction domain of SF1 has been further refined to a fifteen-residue peptide (SF1p, where p denotes peptide) by NMR titration experiments¹³. This SF1p peptide interacts with the C-terminal domain of U2AF⁶⁵^{9, 12}, which is called a U2AF homology motif (UHM) based upon structural and sequence homologies with the U2AF small subunit (U2AF³⁵)¹⁴. However, weak interactions remain between U2AF⁶⁵ and SF1 fragments lacking the N-terminal 28-residues^{12, 15}, raising the possibility that additional SF1 regions contribute to, but do not dominate, recognition of U2AF⁶⁵.

The complex between SF1 and U2AF⁶⁵ is transient^{16, 17}; the U2 snRNP replaces SF1 in a process that is driven by the ATP-dependent RNP unwindase UAP56, which facilitates protein and RNA rearrangements¹⁸. Following displacement of SF1, a U2 snRNP protein subunit called SF3b155 interacts with U2AF⁶⁵¹⁹. SF3b155 crosslinks with the pre-mRNA immediately adjacent to the Py-tract, and a region within the N-terminal half of SF3b155 specifically interacts with the U2AF⁶⁵-UHM¹⁹. A strong correlation with entry of the U2 snRNP during the first ATP-dependent step of pre-mRNA splicing supports the functional importance of the SF3b155/U2AF⁶⁵ interaction¹⁹. Additional U2 snRNP protein subunits, including p14 and the C-terminal domain of SF3b155, respectively crosslink with the BPS and nearby pre-mRNA sequences^{19, 20}. The complex of the U2 snRNP, U2AF⁶⁵, and pre-mRNA is also stabilized by a base-paired duplex between the U2 snRNA and the BPS.

Recent electron cryomicroscopy structures reveal that SF3b155 undergoes a dramatic conformational change to a more open form during assembly of a minor class of spliceosomes (U12-type spliceosome), which catalyze splicing of a rare class of pre-mRNA introns^{21, 22}. Based upon conservation of all SF3b subunits including SF3b155 and p14 in the U12-type spliceosome^{23, 24}, the SF3b155 conformational changes required for BPS recognition by the major (U2-type) spliceosome are expected to be similar. Nonetheless, the specific mechanisms and rearrangements guiding the exchange of SF1 for SF3b155 in the U2AF⁶⁵/pre-mRNA complex are not yet understood in molecular detail. The primary sequence organization and binding partners of the interacting domains of human SF3b155, SF1 and U2AF⁶⁵ are shown schematically in Figure 1.

The solution structure of the minimal complex between SF1 and U2AF⁶⁵ (SF1p/U2AF⁶⁵-UHM)¹³ reveals that the topology of the U2AF⁶⁵-UHM superficially resembles an RNA recognition motif (RRM), one of the most common types of single-stranded RNA binding domains²⁵. Nevertheless, a 10-residue, α -helical extension of the core RRM fold masks the putative RNA-binding surface (Figure 2), and the U2AF⁶⁵-UHM lacks detectable RNA affinity in NMR titration experiments¹³. Instead, the U2AF⁶⁵-UHM displays two distinctive features for recognition of tryptophan-containing protein ligands¹⁴. These include 1) a hydrophobic pocket that binds a key tryptophan of its protein partner (SF1-Trp22), and 2) acidic glutamates and aspartates in the first α -helix that interact with basic arginine or lysine residues preceding the tryptophan (residues 15–19 and 21 of SF1). The tryptophan and preceding basic residues of SF1 are required for association with U2AF⁶⁵ when monitored by pull-down assays with site-directed mutant proteins¹³. The key interactions of the SF1p/U2AF⁶⁵-UHM complex are shared by the minimal U2AF heterodimer (U2AF⁶⁵p/U2AF³⁵-UHM)²⁶, and similar sequence features have been noted in a variety of other putative UHM-containing proteins, including kinases and alternative splicing factors¹⁴.

The ability to predict specific UHM ligands based on primary sequence analysis would provide a powerful tool for extracting protein function from genomic sequences. The two well-characterized UHM ligands (SF1p and U2AF⁶⁵p) share the sequence features of tryptophan residues preceded by basic residues. However, without the framework of a given fold to serve as a preliminary filter during database searches, de novo prediction of UHM ligands is difficult due to the occurrence of many random sequence matches. Given the goal of developing a set of guidelines for predicting UHM ligands, we noted that an N-terminal region of human SF3b155 contains seven tryptophan residues, many of which are preceded by basic residues in a similar manner to the U2AF⁶⁵-interaction site of SF1 (Figure 3). Accordingly, a region within this SF3b155 domain (residues 267–369) is sufficient to mediate interactions with U2AF⁶⁵ in nuclear extract and yeast two-hybrid assays¹⁹. The observation of seven potential U2AF⁶⁵-interaction sites in SF3b155 raised the question of whether any or all of these sequences interact with the U2AF⁶⁵-UHM in a manner similar to the minimal SF1p. Such a model would predict that a short linear SF3b155 peptide containing a conserved tryptophan preceded by basic residues would be necessary and sufficient for association with U2AF⁶⁵. However, U2AF⁶⁵ interactions are strengthened by use of a larger SF3b155 fragment (residues 1–480), thus it is unclear whether the smaller SF3b155 subdomains are necessary as well as sufficient to mediate the SF3b155/U2AF⁶⁵ interaction.

We use spectroscopic techniques to establish that five of these seven SF3b155 tryptophan-containing sites contribute to interactions with the U2AF⁶⁵-UHM. Basic residues precede all five of the U2AF⁶⁵-‘binding’ tryptophans, in contrast with the neutral residues near the ‘nonbinding’ SF3b155 sites. Thermodynamic characterization using isothermal titration calorimetry (ITC) reveals that U2AF⁶⁵ has a lower affinity for SF3b155 than SF1, which may contribute to the need for an RNP unwindase to actively displace SF1 during entry of the U2 snRNP. Significant thermodynamic differences between U2AF⁶⁵ complexes with SF1 compared with variants of SF3b155, indicate that SF1 regions beyond the previously defined, linear binding site, are likely to be involved in complex formation. Rather than simply enhancing the affinity of the SF3b155/U2AF⁶⁵ complex, we suggest that SF3b155 provides alternative U2AF⁶⁵ binding sites to regulate or accommodate conformational rearrangements of the splicing process. Altogether, these results reinforce the importance of the positively charged residues preceding the tryptophan in the signature motif of UHM ligands, and establish that U2AF⁶⁵-UHM ligands can be successfully predicted given a functional protein partner.

Results

Homology models of the SF3b155 tryptophan-containing sites bound to U2AF⁶⁵

We noted seven tryptophan-containing sequence motifs near the N-terminus of SF3b155 (within residues 190–344 of human SF3b155, called SF3b155r). These sequences surround residues Trp200, Trp218, Trp232, Trp254, Trp293, Trp310, and Trp338 of human SF3b155, denoted for simplicity W1, W2, W3, W4, W5, W6, and W7 respectively. Four of these SF3b155 tryptophans (W1, W2, W5, and W7) are preceded by several basic arginines or lysines in a similar manner to the minimal U2AF⁶⁵-interaction site of SF1, SF1p (Figure 3(b)). Two of the SF3b155 sites with significant similarity to SF1p (W5 and W7) are located within the previously mapped, minimal U2AF⁶⁵-interaction site of SF3b155 (residues 267–369). The tryptophan-containing sequences within the SF3b155r domain are phylogenetically conserved (>40% pairwise sequence similarity among homologues represented in Figure 3(a)), although the number of the tryptophan-containing sites varies among different organisms. Moreover, the SF3b155/U2AF⁶⁵ interaction is conserved in *Schizosaccharomyces pombe*¹⁹, despite the presence of only two *S. pombe* SF3b155r tryptophans compared with seven in the human homologue.

To investigate whether any or all of the SF3b155r tryptophan-containing sites form plausible energetic contacts with the U2AF⁶⁵-UHM in a manner similar to SF1p, each of the sites were modeled with the U2AF⁶⁵-UHM based upon the structure of the SF1p/U2AF⁶⁵-UHM complex (Figure 2). Like SF1, each of the SF3b155r tryptophans could potentially insert into the hydrophobic pocket between two α -helices of the U2AF⁶⁵-UHM. In the SF3b155-W1, W2, W5, and W7 models, this tryptophan-mediated interaction positions the preceding arginine and lysine residues to interact favorably with the acidic α -helix of the U2AF⁶⁵-UHM, as observed for SF1p. Unlike SF1, phylogenetically conserved aspartate residues are located on the C-terminal side of each of the SF3b155r tryptophans, with the exception of the SF3b155-W6 site. In several cases (W3, W5, W6, W7), a second, acidic residue, glutamate, is also observed following the aspartate. Although these residues correspond to neutral side chains of SF1p (Asn23 and Gln24), the SF3b155r aspartates and glutamates respectively form salt bridges with U2AF⁶⁵-Lys453 or Arg452 in the structural models, suggesting that they may contribute to SF3b155/U2AF⁶⁵ interactions.

Several of the putative interaction sites (W3, W4, W6) lack extensive lysine or arginine residues and fail to suggest many favorable contacts beyond those with the conserved tryptophan. In the SF3b155-W6 model, a single histidine with a partial positive charge at physiological pH is oriented towards solvent rather than the U2AF⁶⁵-UHM. The W3 and W4 tryptophans of SF3b155 each are preceded by a single arginine or lysine, respectively. The single W3 arginine is conserved in SF1 and makes extensive contacts with both the acidic UHM helix and hydrophobic side chains bordering the tryptophan-binding pocket in the structural model. In contrast, the SF3b155-W4 lysine corresponds to a serine of SF1 (Ser20), which points away from the U2AF⁶⁵-UHM¹³. In addition to the lack of conserved basic residues, a steric clash between the UHM α -helix and an isoleucine preceding the tryptophan reduces the likelihood of U2AF⁶⁵-UHM interactions with the SF3b155-W4.

Phosphorylation of SF1p-Ser20 by cGMP-dependent kinase imparts a negative charge that inhibits association of SF1 with U2AF⁶⁵ in response to extracellular signals²⁷. Interestingly, a serine is also conserved at this position in three of the seven SF3b155 tryptophan-containing sites (W2, W6, and W7). Although the phosphorylation state and function of these SF3b155 serines has not been investigated, threonine-proline (TP) sequences within the SF3b155r domain are phosphorylated by cyclin-dependent kinases²⁸. These 'TP' sequences are repeated a remarkable 29 times between residues 142–437 of human SF3b155 (probability <10⁻²⁵ for a random human protein of equal size). Known phosphorylated SF3b155 threonines include

Thr248 preceding SF3b155-W4, and Thr313 following SF3b155-W6. A threonine corresponding to Thr313 is conserved in all of the SF3b155 sites except for SF3b155-W2, indicating that many of the tryptophan-containing sites are potentially regulated by phosphorylation.

Based upon structural similarity with the SF1p/U2AF⁶⁵-UHM structure, at least four of the seven tryptophan-containing SF3b155 sites were predicted to interact with U2AF⁶⁵, with possible weak or regulated interactions with the other three sites. Although U2AF⁶⁵ has been generally assumed to associate with the pre-mRNA and spliceosome as a monomer, we next examined the homology model to determine whether any of the tryptophan-containing sites of SF3b155 physically overlap, thereby preventing simultaneous association with multiple U2AF⁶⁵ molecules (Figure 2). With SF3b155 in a fully extended conformation, seven U2AF⁶⁵-UHMs bound to the SF3b155r tryptophan-containing sites can be concurrently modeled without interdomain contacts. However, three of the modeled U2AF⁶⁵-UHM cluster together near the N-terminus of the SF3b155r domain, where they would be likely to directly influence each other's association state given the natural flexibility of the polypeptide chains.

Unstructured nature of the SF3b155r domain

Encouraged by these homology models, we prepared a series of site-directed mutant SF3b155 proteins with the tryptophans replaced by alanines to investigate the ability of U2AF⁶⁵ to associate with each of the SF3b155r tryptophan-containing sites. However, tryptophans are often buried in the hydrophobic core of folded proteins, where secondary effects of a mutation on stability of the overall protein fold would be difficult to distinguish from direct effects on protein-protein interactions. Certain sequence characteristics suggested SF3b155r might belong to an emerging class of intrinsically unstructured proteins (IUP), which lack well-structured three-dimensional folds in the absence of appropriate binding partners²⁹. This would allow the key SF3b155r tryptophans to be altered without disrupting the overall protein fold. The SF3b155r domain does not match any known protein folds, has low sequence complexity, and lacks bulky hydrophobic residues in general, aside from the seven tryptophan residues. Consistently, the SF3b155r domain was identified as a highly probable IUP by several independent methods available for predicting unstructured regions from amino acid sequences^{30; 31; 32} (Figure 4(a)).

Circular dichroism (CD) was used to investigate the secondary structural organization of the SF3b155r domain (Figure 4(b)). The pronounced negative band just below 200 nm of the SF3b155r CD spectrum is characteristic of a protein that is in a largely unstructured conformation³³. Despite the large number of prolines in the SF3b155r domain (14% amino acid composition compared with 7% average proline content of human proteins), the CD spectrum lacks notable characteristics of left-handed poly-proline(II) helices, such as a negative band near 205 nm or a slight positive band in the vicinity of 220 nm³⁴. To further investigate whether the tryptophan residues contribute to the secondary structure of SF3b155r, a mutant SF3b155r protein with all seven tryptophan residues substituted with alanine was prepared (SF3b155-dW). The CD spectrum of SF3b155-dW is nearly identical to wild-type SF3b155r (Figure 4(b)). Since SF3b155r and SF3b155-dW both lack detectable secondary structure by CD, the SF3b155r tryptophan residues are unlikely to contribute to a folded protein core. These observations reduced concerns over unintended structural effects when site-directed mutant SF3b155 proteins were used to analyze SF3b155/U2AF⁶⁵ interactions.

Design of mutations to investigate U2AF⁶⁵ association with SF3b155

Site-directed mutant SF3b155 proteins were prepared to systematically investigate the ability of each of the tryptophan-containing sites to associate with U2AF⁶⁵ using intrinsic tryptophan fluorescence and isothermal titration calorimetry. Changes in the polarity of the local

environment of tryptophan side chains upon binding of a protein partner are reflected in the intensity and wavelength of maximum fluorescence emission³⁵. Since even slight, indirectly induced changes in the tryptophan environment would contribute to spectral changes, each single tryptophan-containing site must be studied in isolation. Starting with SF3b155-dW, in which all seven tryptophans were substituted with alanine, each of the SF3b155 tryptophans was separately reintroduced to create a series of variants containing individual tryptophans (SF3b155-W1 through W7, Figure 1(b)). Based upon our prediction that each SF3b155 site binds the U2AF⁶⁵-UHM in a similar manner to SF1p/U2AF⁶⁵-UHM, replacing the tryptophans with alanines was expected to nearly eliminate U2AF⁶⁵-binding; substitution of the SF1 tryptophan (Trp22) with alanine abolishes detectable association of SF1 with U2AF⁶⁵ in pull-down assays¹³, and alanine substitution of the corresponding U2AF⁶⁵ tryptophan (Trp92) decreases affinity for the U2AF³⁵-UHM by three orders of magnitude when measured by calorimetry²⁶. Given that the interaction of U2AF⁶⁵ with the alanine-substituted SF3b155 sites is likewise negligible, the SF3b155 variants with all but a single tryptophan replaced with alanine can be used to analyze U2AF⁶⁵-binding to individual tryptophan-containing sites using both intrinsic tryptophan fluorescence and isothermal titration calorimetry.

The U2AF⁶⁵-UHM contains a single tryptophan residue (Trp475), which would contribute to and complicate spectral changes during fluorescence experiments with SF3b155. Fortunately, this single U2AF⁶⁵-UHM tryptophan is the C-terminal residue of U2AF⁶⁵, and is located on the opposite surface from the SF1-binding site in the SF1p/U2AF⁶⁵-UHM structure (Figure 2). This tryptophan and the penultimate phenylalanine (Phe474) tether the α -helical extension of the core RRM fold to the putative RNA binding surface. Since phenylalanines are also weakly fluorescent³⁵, both Trp475 and Phe474 were removed from the UHM domain to create a shortened version (U2AF⁶⁵-dFW) that lacks significant fluorescence and can be used in titration experiments with the SF3b155 variants.

Thermodynamic comparison of SF1/U2AF⁶⁵-UHM and SF3b155/U2AF⁶⁵-UHM complexes

First, the thermodynamic contributions to the free energy of U2AF⁶⁵-UHM association with SF1 or SF3b155r were examined using isothermal titration calorimetry (ITC). The enthalpies of binding (ΔH°) were measured during titration of the U2AF⁶⁵-UHM into a solution of SF1 or SF3b155r (Figure 5). A shortened variant of SF1 (residues 1-255, hereafter called SF1) (Figure 1(c)) was used for all of the experiments, since it contains all of the essential domains for BPS recognition and U2AF⁶⁵-binding¹² and could be produced and concentrated to levels required for calorimetry and fluorescence experiments. The binding isotherms were fit to evaluate the Gibbs free energy (ΔG° , related to the apparent equilibrium dissociation constant K_d) and stoichiometry (n), from which the entropy (ΔS°) of binding was calculated (Table 1). Assuming identical and independent (non-cooperative) binding sites, the best fit of the wild-type SF3b155r binding isotherm was obtained with a stoichiometry of three binding sites for U2AF⁶⁵ within SF3b155r (Figure 5(a)). Although this stoichiometry suggests that U2AF⁶⁵ can simultaneously associate with a subset rather than all possible SF3b155 tryptophan-containing sites, deviation from the assumption of identical and independent binding sites may affect the apparent number of sites obtained from curve fitting. The fitted isotherms for U2AF⁶⁵-UHM binding with other partners, including the SF1/U2AF⁶⁵-UHM complex, were consistent with the formation of one-to-one complexes.

The values in Table 1 for SF3b155r represent an average, single site (assuming three identical and independent sites), and thus can be directly compared with the values for the complexes that were best fit with binding stoichiometries close to one. Consistent with the need for energy in the form of ATP-hydrolysis for the U2 snRNP to disrupt the SF1/U2AF⁶⁵-UHM complex, U2AF⁶⁵-UHM displayed 250-fold higher affinity for SF1 than for the average SF3b155r site. The relatively low affinity of the SF3b155r/U2AF⁶⁵-UHM complex may facilitate

disassociation of U2AF⁶⁵ from the pre-mRNA prior to the catalytic steps of splicing. On the other hand, the stability of the SF1/U2AF⁶⁵-UHM complex is remarkable ($K_d \sim 11$ nM) given the dynamic nature of the complex during the splicing process^{16; 17}. Nonetheless, the transient SF1/U2AF⁶⁵-UHM complex is less stable than the constitutive U2AF heterodimer ($K_d < 1.7$ nM)²⁶. The higher affinity of the U2AF heterodimer may be attributed to burial of an additional tryptophan, donated to the heterodimer interface by the U2AF³⁵-UHM. Comparison of the different UHM complexes illustrates a versatile repertoire of affinities attained through slight variations of a common structural scaffold.

Prior to the availability of human genomic sequences for validation, a point mutation of arginine to glycine (Arg19Gly) was introduced in an original SF1 cDNA by the polymerase chain reaction (PCR)³⁶. By ITC, the site-directed mutation reduced the U2AF⁶⁵ affinity of the SF1(Arg19Gly) variant by 3.4-fold compared with the wild-type SF1. A large unfavorable change in the entropic term (by $2.5 \text{ kcal mol}^{-1}$) was responsible for the affinity decrease, as expected based upon the torsional flexibility of glycine residues. The SF1(Arg19Gly) variant was inadvertently used instead of wild-type SF1 to quantitatively measure cooperative U2AF⁶⁵ interactions that facilitate pre-mRNA recognition⁹. Because the U2AF⁶⁵ affinity of the wild-type protein is higher, it is possible that the enhancement of SF1's pre-mRNA affinity by the presence of U2AF⁶⁵ may be even more dramatic than previously published (20-fold enhancement for the SF1(Arg19Gly) variant).

The different free energies driving formation of the SF1/U2AF⁶⁵-UHM and SF3b155r/U2AF⁶⁵-UHM complexes were dissected to examine the enthalpic and entropic contributions. Both binding events were enthalpically driven to overcome entropically unfavorable terms. The estimated entropic penalty for association of U2AF⁶⁵-UHM with SF1 was $\sim 5 \text{ kcal mol}^{-1}$ greater than for SF3b155r, suggesting that SF1 undergoes a larger transition than SF3b155r from an unfolded to a folded state upon association with U2AF⁶⁵. This observation is superficially remarkable given that the SF3b155r domain is the unstructured as demonstrated above, compared with the structurally well-characterized fold of a maxi-KH RNA binding domain within SF1³⁷. However, U2AF⁶⁵ binding may affect the conformation of a ~ 100 -residue linker following the U2AF⁶⁵ binding site, which is predicted to be intrinsically unstructured despite a highly phylogenetically conserved sequence¹². In contrast, the relatively small entropic cost for U2AF⁶⁵ association with SF3b155 may indicate that SF3b155 folding remains locally constrained at the individually bound sites. The enthalpic contributions that drive association of U2AF⁶⁵-UHM with SF1 or SF3b155r also differ by a substantial $8.1 \text{ kcal mol}^{-1}$, suggesting that SF1 satisfies more of the potential hydrogen bonds and salt bridges buried by complex formation with the U2AF⁶⁵-UHM. Accordingly, the local SF1p site offers six basic residues to interact with the seven acidic residues available on the U2AF⁶⁵-UHM α -helix, compared with an average of two basic residues per tryptophan-containing SF3b155r site.

No heats of binding to the U2AF⁶⁵-UHM were observed when all of the SF3b155r tryptophans were replaced with alanines in SF3b155-dW. Thus, in the absence of the SF3b155 tryptophans the apparent equilibrium dissociation constant, K_d is weaker than experimentally accessible range of the calorimeter ($\sim 1 \text{ mM}$), a >350 -fold reduction in U2AF⁶⁵-binding affinity for SF3b155-dW compared with the wild-type SF3b155r. The lack of U2AF⁶⁵-UHM binding with SF3b155-dW establishes the critical nature of the SF3b155 tryptophans for U2AF⁶⁵ recognition, consistent with the dependence of the SF1p/U2AF⁶⁵-UHM complex on the presence of the SF1 tryptophan¹³. The absence of detectable U2AF⁶⁵-UHM interactions with SF3b155-dW indicates that the SF3b155-W1 through W7 variants, in which alanine replaces all but a single tryptophan, are likely to exhibit single site binding at working concentrations.

Structural and functional integrity of U2AF⁶⁵-dFW

In preparation for use of the U2AF⁶⁵-dFW variant for fluorescence assays, the secondary structural content and ability of U2AF⁶⁵-dFW to bind SF1 were compared with the wild-type U2AF⁶⁵-UHM. As judged by CD spectroscopy, both the U2AF⁶⁵-UHM and U2AF⁶⁵-dFW proteins were folded (Figure 4(c)). Based upon the SF1p/U2AF⁶⁵-UHM structure¹³, a 14% decrease in α -helical content was predicted for unfolding of the entire C-terminal α -helix of the U2AF⁶⁵-UHM, which is normally anchored to the UHM β -sheet surface by aromatic stacking with the deleted C-terminal Trp475 and Phe474 residues. The α -helical content of U2AF⁶⁵-dFW, estimated from the mean residue ellipticity of the CD spectrum at 222 nm³⁹, decreases by 8% compared with wild-type U2AF⁶⁵-UHM. Given that interpretation of CD spectra is somewhat qualitative, it is likely that some or all of the U2AF⁶⁵-UHM C-terminal α -helix unfolds in the absence of the Trp475 and Phe474 residues. However, it cannot be ruled out that removing the aromatic C-terminal residues indirectly destabilizes regions of the first or second UHM α -helices.

To further compare the ability of U2AF⁶⁵-dFW to bind tryptophan-containing sites with the wild-type U2AF⁶⁵-UHM, the thermodynamics of U2AF⁶⁵-dFW or U2AF⁶⁵-UHM association with SF1 or a representative SF3b155 variant containing an individual tryptophan site (SF3b155-W7) were compared using ITC (Table 1, representative binding isotherms are shown in Figure 6). As discussed in detail in subsequent sections, both the U2AF⁶⁵-UHM and U2AF⁶⁵-dFW significantly interact with SF3b155-W7. Interestingly, a slightly greater energetic penalty was observed for U2AF⁶⁵-dFW association with SF1 (4.7-fold reduction) than with the SF3b155 variant (1.3-fold reduction). The greater overall energetic penalty for U2AF⁶⁵-dFW binding SF1 compared with SF3b155-W7 arises from unfavorable effects on the enthalpic (1.3 kcal mol⁻¹), and to a lesser extent, the entropic terms (0.5 kcal mol⁻¹). The distinct thermodynamic effects of deleting the C-terminal U2AF⁶⁵ residues on binding SF1 compared with SF3b155 indicate that the interactions of the two complexes differ in detail, although both depend on presence of a key tryptophan. Given that the C-terminus of U2AF⁶⁵ is spatially separated from the SF1p binding site, longer-range interactions with the SF1 protein may contribute to the greater thermodynamic impact of altering the U2AF⁶⁵ C-terminus on binding SF1 compared with SF3b155. Regardless, U2AF⁶⁵-dFW provides a useful tool for monitoring U2AF⁶⁵ interactions with SF3b155 by intrinsic tryptophan fluorescence.

Heat capacity change of the SF3b155-W7/U2AF⁶⁵-dFW complex

The heat capacity of binding (ΔC_p°), or enthalpy change upon variation of the temperature, is the thermodynamic parameter with the most straightforward relationship to the structure of the complex. We chose to calculate the heat capacity change of the SF3b155-W7/U2AF⁶⁵-dFW complex for comparison with fluorescence experiments, which were conducted at fluctuating ambient temperatures rather than the controlled, 30°C temperature of the ITC experiments. The thermodynamic parameters for U2AF⁶⁵-dFW association with SF3b155-W7 were obtained at three different temperatures (Table 2), and the heat capacity was calculated from the slope of ΔH° plotted against temperature. The SF3b155-W7/U2AF⁶⁵-dFW partners have a relatively large heat capacity of -0.24 kcal mol⁻¹ K⁻¹. The heat capacity change is directly related to the burial of polar and nonpolar solvent-accessible surface area associated with the binding event⁴⁰. The polar and nonpolar components of the solvent accessible surface area buried by formation of the SF3b155-W7/U2AF⁶⁵-dFW complex were calculated from our homology model of SF3b155-W7p/U2AF⁶⁵-dFW, which was based upon the SF1p/U2AF⁶⁵-UHM structure. In the SF3b155-W7p/U2AF⁶⁵-dFW model, 59% (770 Å²) of the total buried surface area was hydrophobic and the remaining 41% (533 Å²) was polar or charged. The amount of predicted buried hydrophobic surface area is comparable to the lower limit expected to stabilize subunit association, in general, by the hydrophobic effect alone (700 Å²), but is less than the average fraction of nonpolar surface observed at subunit interfaces ($65 \pm 4\%$)⁴¹. A

predicted heat capacity change of $-0.21 \text{ kcal mol}^{-1} \text{ K}^{-1}$ is close to the observed value of $-0.24 \text{ kcal mol}^{-1} \text{ K}^{-1}$. The close match between the calculated and experimental heat capacity changes strongly supports the idea that individual tryptophan-containing SF3b155 sites, particularly SF3b155-W7, interact with the U2AF⁶⁵-UHM in a similar manner to that of the known SF1p/U2AF⁶⁵-UHM structure.

Intrinsic tryptophan fluorescence changes reveal five SF3b155 sites for U2AF⁶⁵-binding

The fluorescence emission spectra of mutant SF3b155 proteins containing single tryptophan fluorophores were examined for changes in intensity and wavelength of maximum emission upon addition of U2AF⁶⁵-dFW (Figure 7). The fluorescence emission spectra of the indole group of tryptophan is highly sensitive to its local environment, most notably blue spectral shifts and enhancement of emission intensity result from burial of the side chain within a hydrophobic binding pocket³⁵. Dramatic changes in the intrinsic tryptophan fluorescence spectra for SF3b155/U2AF⁶⁵ association were expected based upon similarity with SF1, in which a solvent-exposed tryptophan is buried within a hydrophobic crevice of the U2AF⁶⁵-UHM.

Five of the seven individual tryptophan-containing variants of SF3b155 (SF3b155-W1, W2, W3, W5, and W7) showed significant spectral changes upon addition of U2AF⁶⁵-dFW, comparable to a control titration with SF1. The total emission intensity increased by ~two-fold upon saturation of the SF3b155 variants or SF1 with U2AF⁶⁵-dFW, and the wavelength of maximum emission intensity (λ_{max}) shifted to considerably lower wavelengths (Figure 7(a) and Table 3). This significant spectral change (15–19 nm blue shift) is consistent with transfer of the tryptophan side chain from a polar aqueous environment to a hydrophobic pocket, such as that observed in the SF1p/U2AF⁶⁵-UHM structure. In contrast, the fourth and sixth tryptophan-containing SF3b155 variants (SF3b155-W4 and W6) lacked obvious changes in the wavelength of maximum emission or emission intensity, even following addition of 26-fold molar excess U2AF⁶⁵-dFW (Figure 7(b)). The absence of spectral changes for SF3b155-W4 (Trp254) and SF3b155-W6 (Trp310) indicates that these potential binding sites either do not interact with U2AF⁶⁵-dFW beyond the lower detectable limit of the experiment (~1mM), or weakly interact in a distinct manner that does not alter the local environment of the key tryptophan residues. Thus, a tryptophan-containing sequence alone is insufficient for substantial U2AF⁶⁵ binding in the absence of preceding positively charged residues.

Following titration of SF1 or the five ‘binding’ SF3b155 variants with U2AF⁶⁵-dFW, the changes in emission intensity were fit to obtain apparent dissociation constants (K_d) for each unique tryptophan-containing site, assuming a 1:1 binding stoichiometry (Figure 7(c) and Table 2). A relatively large experimental variability is consistent with ambient temperature fluctuations coupled with large heat capacities of the interaction extrapolated from the SF3b155-W7/U2AF⁶⁵-dFW complex. For example, a 30% affinity difference for a 5°C temperature change is expected based on the experimentally determined $-0.24 \text{ kcal mol}^{-1} \text{ K}^{-1}$ heat capacity change of the SF3b155-W7/U2AF⁶⁵-dFW complex. The affinity of the SF3b155-W7 protein for U2AF⁶⁵-dFW was among the highest of the other SF3b155 variants, whereas SF3b155-W3 protein was several-fold weaker (6-fold weaker compared with SF3b155-W7). Both SF3b155-W3, which weakly interacts with U2AF⁶⁵-dFW, and SF3b155-W4, which entirely lacks detectable interactions with U2AF⁶⁵-dFW, contain a single basic arginine or lysine preceding the key tryptophan. However, the basic residues are located at distinct positions in the homology models of the two complexes; the SF3b155-W3 arginine stacks with a phenylalanine of the tryptophan-binding pocket as well as satisfying electrostatic interactions with the U2AF⁶⁵-UHM, whereas the SF3b155-W4 lysine is separated from the tryptophan by a bulky isoleucine and faces towards solvent. The SF3b155-W3 arginine corresponds to Arg21 of SF1, which is required for the SF1/U2AF⁶⁵ interaction¹³. In contrast,

the SF3b155-W4 lysine corresponds to a neutral Ser20 of SF1. Rather than a general requirement for positively charged patches on the UHM-ligands, the spatial locations of the basic side chains perform distinct roles in complex assembly.

Two acidic residues that follow the SF3b155-W3 tryptophan provide a second sequence feature that may contribute to the higher U2AF⁶⁵-affinity compared with SF3b155-W4. These aspartate and glutamate residues are also observed in SF3b155-W5 and SF3b155-W7, sites that display the highest approximate U2AF⁶⁵-affinity. However, the acidic residues are replaced with their neutral amide counterparts in SF1 (Asn23 and Gln24), which are energetically unimportant in site-directed mutagenesis experiments¹³. Given the dramatically higher U2AF⁶⁵-affinity of SF1 compared with SF3b155, the relative importance of the conserved SF3b155 aspartates and glutamates for U2AF⁶⁵ binding remains to be further tested by future site-directed mutagenesis experiments.

Although U2AF⁶⁵ recognizes five of the seven tryptophan-containing SF3b155r sites during separate fluorescence experiments with individual tryptophan-containing SF3b155 variants, the SF3b155r/U2AF⁶⁵-UHM binding isotherm was best fit with a stoichiometry of 3:1 rather than concurrent association with all five binding sites (Figure 5(a)). In addition to deviations from the model of identical, independent sites made for curve-fitting, steric or other unfavorable interactions may prevent all five SF3b155r sites from being occupied simultaneously. For example, the N-terminal three U2AF⁶⁵-binding sites (SF3b155-W1, W2, and W3) were uncomfortably close even in a molecular model with the SF3b155r polypeptide in a fully extended, β -strand conformation (13 Å closest approach of UHMs, Figure 2). Since the U2AF⁶⁵-UHM domain is highly charged (predicted net charge of -9.5 at pH 7.4), such close proximities would generate large repulsive effects. In the assembling spliceosome, as opposed to our *in vitro* experiments, specific U2AF⁶⁵ interaction sites may be occluded or exposed by conformational rearrangements of SF3b155 and associated proteins or RNA.

Thermodynamic comparison of U2AF⁶⁵ -UHM complexes with isolated SF3b155 and SF1 sites

To further explore the thermodynamic parameters underlying formation of the U2AF⁶⁵ complexes, ITC was used to compare U2AF⁶⁵-UHM complexes with SF1 to a representative SF3b155 variant containing an individual tryptophan site (SF3b155-W7). Since the thermodynamic values for SF3b155r represent average values for three U2AF⁶⁵-binding sites that are assumed to be identical and independent, a slightly higher U2AF⁶⁵-UHM affinity for the SF3b155-W7 site compared with SF3b155r was expected by ITC, given that SF3b155-W7 displays one of the highest U2AF⁶⁵-UHM affinities by intrinsic tryptophan fluorescence (Table 3). However, the free energy for U2AF⁶⁵-UHM association with SF3b155-W7 was comparable to SF3b155r (1.1-fold higher affinity). At least two explanations for the similar SF3b155r and SF3b155-W7 affinities are conceivable, although a more thorough thermodynamic analysis of each site would be required to distinguish the relative importance of these possible contributions. Since the best fit for the SF3b155r/U2AF⁶⁵-UHM binding isotherm was obtained using a stoichiometry of three U2AF⁶⁵-UHM per one SF3b155r, only a subset of the higher affinity sites may have been observed during the ITC experiment. Alternatively, if the SF3b155 sites are nonidentical and slightly cooperative rather than independent, the increased apparent affinity of an average site within the SF3b155r domain would mask the lower affinity sites.

Large but balanced changes in the entropic and enthalpic terms contribute to the nearly equivalent free energies of binding; the enthalpy of the SF3b155-W7/U2AF⁶⁵-UHM interaction is more favorable by 5.5 kcal mol⁻¹ K⁻¹, whereas the entropy is 5.4 kcal mol⁻¹ less favorable when compared with the average SF3b155r site. The enthalpic benefit, yet entropic cost of substituting alanine for all but a single tryptophan suggests that the SF3b155-W7 variant

becomes more ordered when bound to U2AF⁶⁵ than the average site of wild-type SF3b155r. For example, it is possible that the alanine mutations disrupt a collapsed SF3b155r conformation, which is normally stabilized by burial of the hydrophobic tryptophans. Such a difference would not be apparent upon comparison of the SF3b155r and SF3b155-dW far-ultraviolet CD spectra, which only reflect ordered secondary structures³³.

Next, to evaluate the contribution of long-range interactions to complex formation, sixteen-residue peptides with sequences corresponding to the individual tryptophan-containing sites of SF1 (SF1p) and SF3b155-W7 (SF3b155-W7p) were used in ITC experiments with the U2AF⁶⁵-UHM. The borders of the peptides were chosen to match that of the structurally characterized SF1 site for U2AF⁶⁵-UHM¹³ plus three additional C-terminal residues to include 'TP' phosphorylation sites of SF3b155²⁸. Relative to the SF3b155-W7 domain, the U2AF⁶⁵ affinity of the truncated SF3b155-W7p site increases 5.7-fold, due to a ~ 1 kcal mol⁻¹ reduction in the entropic cost of binding. However, both the enthalpy and entropy of the SF3b155-W7p interaction changes relative to an average site of wild-type SF3b155r, suggesting that binding is coupled with folding in a similar manner to the SF3b155-W7 variant. The higher affinity and the similar thermodynamic components of U2AF⁶⁵-interactions with SF3b155-W7p compared with the SF3b155-W7 indicates that the energetically favorable interactions with SF3b155 are local, with no requirement for functional groups that are distant in the linear sequence.

In contrast to the increased U2AF⁶⁵ affinity upon shortening the SF3b155 domain, the U2AF⁶⁵ affinity of the minimal SF1p binding site decreased by two-fold compared with the SF1 protein. The enthalpy of the SF1p interaction with U2AF⁶⁵ was less favorable by 6.4 kcal mol⁻¹, and slightly outweighed a 6.0 kcal mol⁻¹ reduction in the entropic penalty for binding (Table 1). These thermodynamic differences suggest the SF1 protein undergoes a larger folding transition upon U2AF⁶⁵ binding than does the SF1p peptide, in contrast to the increased folding upon binding of the isolated SF3b155-W7 or W7p sites compared with the average site of the SF3b155r domain. Unlike SF3b155, the higher U2AF⁶⁵ affinity of the larger SF1 protein compared with the isolated SF1p indicates that the SF1/U2AF⁶⁵ interface is more extensive than the established linear SF1p peptide. This is consistent with the nearly five-fold reduction in binding affinity for SF1 compared with SF3b155 produced by the U2AF⁶⁵-dFW truncation, which is located far from the SF1p interaction site. As noted above, SF1 contains a conserved ~ 100 -residue region adjacent to the minimal SF1p sequence, which is predicted to be intrinsically unstructured and may contribute to the observed thermodynamic effects by folding upon association with U2AF⁶⁵. The idea that additional SF1 regions contributing to U2AF⁶⁵-binding is lent further support by our unpublished observation that distant post-translational modifications of SF1 enhance association with U2AF⁶⁵ (A. Maucuer, INSERM & C. Kielkopf, unpublished results).

Discussion

The ATP-dependent exchange of SF1/U2AF⁶⁵ for SF3b155/U2AF⁶⁵ protein complexes during stable association of the U2 snRNP with the pre-mRNA is a critical, regulated process that commits a particular pair of splice sites to use⁸. Our results indicate that the U2AF⁶⁵ exchange of SF1 for SF3b155 is energetically unfavorable. The energy released by ATP-dependent RNP unwindases would be needed to break apart the SF1/U2AF⁶⁵ complex and allow association of U2AF⁶⁵ with SF3b155 ($\Delta\Delta G^\circ$ 3.3 kcal mol⁻¹ for SF1/U2AF⁶⁵ $\rightarrow$ SF3b155/U2AF⁶⁵, compared with -7.3 kcal mol⁻¹ released per ATP hydrolysis under standard conditions). The association of the U2 snRNP with the pre-mRNA would be stabilized by further interactions, including contacts between SF3b155 and the pre-mRNA¹⁹, as well as annealing of the BPS with the U2 snRNA, the N-terminal arginine-serine rich domain of U2AF⁶⁵⁴², and the U2 snRNP protein, p14²⁰. Thus, beyond the established role of SF1 to couple U2AF⁶⁵/Py-tract

association with a directly adjacent BPS^{9; 10}, SF1 provides a thermodynamic barrier that blocks premature recruitment of the U2 snRNP (Figure 1(a)); only in the presence of ATP and the appropriate RNP unwindase (such as UAP56¹⁸) would the U2 snRNP, U2AF⁶⁵, and pre-mRNA be primed to form an active splicing complex.

The heat capacity change associated with formation of the SF3b155-W7/U2AF⁶⁵-dFW complex was consistent with a model based on the previously described SF1p/U2AF⁶⁵-UHM structure¹³. Moreover, few thermodynamic differences were observed when the SF3b155-W7 was reduced to the sixteen-residue SF3b155-W7p peptide. In contrast, replacing SF1 with the minimal SF1p peptide dramatically altered both the entropy and enthalpy of binding U2AF⁶⁵-UHM. Further, the effect of removing the C-terminal residues of the U2AF⁶⁵-UHM on the free energy of binding SF1 was more dramatic than for binding SF3b155-W7. While a short, linear interaction site adequately described the U2AF⁶⁵-UHM binding site of SF3b155-W7, our thermodynamic results indicate that UHM binding by SF1 is more complicated. One possibility is that the highly conserved linker of SF1, C-terminal to the minimal U2AF⁶⁵-binding site, is involved in specifying the U2AF⁶⁵-UHM.

Using intrinsic tryptophan fluorescence spectroscopy, we showed that U2AF⁶⁵ binds five of the seven tryptophan-containing SF3b155 sites. Molecular models and sequence similarities suggested that formation of the SF3b155/U2AF⁶⁵ complex would depend on the SF3b155 tryptophans and preceding basic residues in a similar manner to the SF1p/U2AF⁶⁵-UHM complex. Consistent with dependence on key tryptophan residues, no interactions were detected between U2AF⁶⁵ and an SF3b155-dW variant lacking tryptophan residues, and titrations with SF3b155 variants containing single tryptophan residues were adequately fit using one-to-one binding models. Significant 'blue shifts' in the $\lambda_{\max}$ of the SF3b155/U2AF⁶⁵ complexes indicated that the tryptophan was transferred to a highly nonpolar environment upon binding³⁵, such as that observed for the SF1p/U2AF⁶⁵-UHM structure¹³. However, the tryptophan alone was unable to elicit high affinity recognition by the U2AF⁶⁵-UHM, since U2AF⁶⁵-UHM association with two of the tryptophan-containing SF3b155 sites was undetectable.

Up to four basic residues (arginine or lysine) were present in the SF3b155 sites with significant affinity for the U2AF⁶⁵-UHM, but conspicuously absent from non-binding sites. Furthermore, SF3b155-W3, which contains only a single arginine directly preceding the tryptophan, displayed a considerably weaker affinity for U2AF⁶⁵ than the other SF3b155 sites. The SF3b155-W3 arginine is strictly conserved in all the 'binding' SF3b155 and SF1 sites. In contrast, a single lysine of the 'non-binding' SF3b155-W4 mutant corresponds to a serine of SF1, with similar residues in only one other SF3b155 site (SF3b155-W1, Arg198). Consistent with our structure-based models and previous site-directed mutagenesis analysis of SF1/U2AF⁶⁵ interactions¹³, these results demonstrate the importance of basic residues, in particular an arginine directly preceding the tryptophan, for formation of the SF3b155/U2AF⁶⁵ complex.

One potential difference between the key interactions mediating the SF3b155/U2AF⁶⁵ and SF1/U2AF⁶⁵ complexes is suggested by the presence of phylogenetically conserved acidic residues (Asp201, Asp219, Asp233/Glu234, Asp294/Glu295, or Asp339/Glu340) following the SF3b155 tryptophans in all of the sites of U2AF⁶⁵-binding (Figure 3). In our structural models, the conserved SF3b155 aspartates were positioned to interact favorably with Lys453, between the second α -helix and last β -strand of the U2AF⁶⁵-UHM. Strikingly, U2AF⁶⁵-Lys453 is replaced by a tryptophan in the U2AF³⁵-UHM (Trp134). This U2AF³⁵-UHM tryptophan is specifically recognized by unique, C-terminal prolines of the tryptophan-containing U2AF⁶⁵ ligand in the U2AF heterodimer. The SF1p residues corresponding to the SF3b155 aspartates and glutamates (SF1-Asn23 and Gln24) are uncharged, and make little contribution to SF1/U2AF⁶⁵ affinity¹³. Nevertheless, SF1 thermodynamically involves binding surfaces beyond

the minimal SF1p peptide, and upon further investigation may support a general mechanism for specifying sequence variations within the α/β loop of the many different UHM-containing proteins by using C-terminal extensions of a core positively-charged/tryptophan-containing ligand.

The free energy for U2AF⁶⁵-UHM binding the local, SF1p site, which is 4 kcal mol⁻¹ more favorable than binding the SF3b155-W7p peptide, raises the question of whether this difference can be rationalized in terms of primary sequences. Although ion pairs often destabilize protein folding due to desolvation effects, the removal of one member of a buried ion pair is generally expected to be destabilizing⁴³. Remarkably, arginines are one of the most abundant residues observed at subunit interfaces⁴¹. The guanidinium side chain potentially contributes four times the number of hydrogen bonds of lysine, and participates in ionic networks such as those thought to stabilize hyperthermophilic proteins⁴⁴. Five of the seven SF1p arginine or lysine residues are shared by the SF3b155-W7p site, although several basic residues are partially solvent exposed and lack direct salt bridges with the acidic UHM α -helix in the SF1p/U2AF⁶⁵-UHM structure. The loss of two basic charges is more than sufficient to account for the energetic differences given an estimated 3 kcal mol⁻¹ penalty for disrupting one partner of a buried salt bridge⁴⁵, especially considering that the conserved aspartate and glutamate of SF3b155-W7 are likely to contribute favorably to the U2AF⁶⁵ interaction. Thus, specific residues framing the ligand tryptophans contribute to affinity as well as specific, unidirectional recognition of the UHM, although hydrophobic interactions with the tryptophan are required for stability of the complexes^{40; 46}.

The observation of five U2AF⁶⁵-binding sites within SF3b155 leads us to speculate concerning their possible functions during pre-mRNA splicing. Given that SF1 uses only a single U2AF⁶⁵-binding site to specify a BPS near a Py-tract, it is unlikely that several U2AF⁶⁵ molecules are concurrently associated with the 3' splice site and hence available for binding SF3b155. Accordingly, dynamic light scattering, ITC, and gel filtration chromatography experiments showed a 1:1 stoichiometry for complexes of the U2AF⁶⁵-RNA binding domain with a poly-uridine dodecamer at concentrations in the range of the apparent K_d (C. Kielkopf, unpublished results). It is also unlikely that alternative complexes between U2AF⁶⁵ and different SF3b155 sites would be required for the 3' splice site complex to adapt to variable spacings among BPS and Py-tract consensus sequences, since a BPS must be located within a few nucleotides of a Py-tract to be used efficiently during splicing^{7; 19}. Recent electron cryomicroscopy structures indicate that SF3b155 undergoes a dramatic structural rearrangement upon integration of the U2 snRNP into higher order spliceosomal complexes^{21; 22}. In contrast, SF1, with a single U2AF⁶⁵-binding site, is not thought to undergo conformational rearrangements such as those required for SF3b155 function. Hence, it is a fascinating possibility that the conformational changes of spliceosome assembly may relocate U2AF⁶⁵ among different binding sites of SF3b155. Ultimately, transfer to one of the low affinity SF3b155 sites might contribute to dissociation of U2AF⁶⁵ prior to the first catalytic step of splicing^{17; 47}.

The multiple U2AF⁶⁵-binding sites of SF3b155 are likely to provide a mechanism for regulating pre-mRNA splicing. For example, the efficiency of recruiting the U2 snRNP to a particular 3' splice site may be fine-tuned by exposing or phosphorylating different subsets of the U2AF⁶⁵ binding sites of SF3b155. Several regulatory proteins contain UHM domains with unknown protein ligands, including PUF60^{48; 49}, SPF45⁵⁰, and KIS kinase⁵¹ may mask specific subsets of SF3b155 tryptophan-containing sites, thereby forcing U2AF⁶⁵ to associate with different SF3b155 sites during alternative splicing. Accordingly, U2AF-like alternative splicing factors¹⁴ and phosphorylation-dependent splicing factors such as NIPP1⁵² are known to regulate splice site choice in response to cell division or differentiation. The increasing number of tryptophan-containing sites as SF3b155 homologues ascend the evolutionary ladder

may reflect an increased demand for regulating alternative splicing in light of multicellular complexity (Figure 3(a): one SF3b155r tryptophan in Baker's yeast; two in fission yeast; five in worms; six in plants; seven in flies, amphibians, mice, and humans). The only observed SF3b155 homologue with a single as opposed to multiple U2AF⁶⁵-binding sites is from *Saccharomyces cerevisiae*, which is one of the few eukaryotic organisms in which the U2AF⁶⁵ homologue (Mud2p) is not essential for viability⁵³, and also lacks the need to extensively regulate pre-mRNA splicing due to its low frequency of intron-containing genes⁵⁴.

Like established IUPs, the SF3b155r domain was strongly predicted to lack a well-defined three-dimensional fold in the absence of binding partners, and lacked detectable secondary structure when examined by CD spectroscopy. IUPs are estimated to be prevalent among higher eukaryotes⁵⁵, and play established roles in a growing list of essential processes, including transcriptional regulation, translation, and signal transduction⁵⁶. Coupled IUP folding and binding enables different conformations of a single ligand to interact with multiple targets, and often exposes the side chains of the unbound IUP for post-translation modification²⁹. Accordingly, our calorimetry results reveal that SF3b155 likely couples folding and binding. Additionally, both SF1 and SF3b155 are regulated by protein kinases; phosphorylation of SF1 on Ser20 inhibits binding of U2AF⁶⁵²⁷, and SF3b155 is a substrate for phosphorylation by CDKs including cyclin-E/CDK2²⁸. Thus, the unstructured conformation SF3b155r domain may facilitate recognition by protein kinases or other regulatory factors to influence alternative splicing.

Establishing the presence of five U2AF⁶⁵-binding sites within the SF3b155 N-terminus represents an important step towards understanding critical events of spliceosome assembly, which are often mis-regulated during human genetic disease⁵⁷. Many issues raised by this result will require further investigation; chiefly, whether the five U2AF⁶⁵ binding sites of SF3b155 are functionally important targets for U2AF⁶⁵ *in vivo*, and whether specific, tryptophan-containing sites of SF3b155 are regulated by other UHM-containing proteins. By comparing the sequence characteristics that enable the U2AF⁶⁵-UHM to discriminate among the seven, tryptophan SF3b155r containing sites, we have confirmed and extended a consensus sequence for predicting optimal U2AF⁶⁵-UHM ligands: (R/K)_nXRW(DE), where X is any residue and parentheses surround residues that strengthen but are not required for binding. This consensus represents a small step towards a major goal of the post-genomic era, the ability to predict protein interaction networks from genomic sequences.

Materials and Methods

Protein and peptide preparation

Variants of the human SF3b155 domain (residues 190–344), a fragment of human SF1 containing U2AF⁶⁵-interacting and the BPS-interacting domains (residues 1–255), human U2AF⁶⁵-UHM (residues 375–475), and a shortened variant lacking the terminal aromatic residues (U2AF⁶⁵-dFW, residues 375–473) were expressed as glutathione-S-transferase (GST) proteins from pGEX-6p vectors (Amersham Biosciences) in *Escherichia coli* BL21(Rosetta) (Novagen) using standard procedures. The QuikChange site-directed mutagenesis procedure (Stratagene) was used to replace the SF3b155 tryptophans with alanines, to reintroduce individual tryptophans, and to correct a glycine mutation resulting from an error in the SF1 template⁹ to the wild-type arginine (Arg19). Proteins were purified by glutathione-affinity chromatography using the manufacturer's protocol, then further purified by ion-exchange chromatography following removal of the GST tag by treatment with Precision Protease. Prior to fluorescence experiments, U2AF⁶⁵-dFW was further purified by size-exclusion chromatography (Amersham Biosciences chromatography resins). Synthetic peptides of >95% purity, with sequences corresponding to human SF1 (residues 13–28) and SF3b155 (residues

329–344) were purchased from Biosynthesis. With the exception of SF3b155-dW, protein and peptide concentrations were calculated based upon the absorbance at 280 nm and estimated extinction coefficients using the ProtParam tool on the ExPASy website (<http://us.expasy.org/tools/protparam.html>). Concentrations of SF3b155-dW were estimated based upon the Bradford colorimetric assay⁵⁸. The isoelectric point of U2AF⁶⁵-UHM was also calculated using the ProtParam tool on the ExPASy website.

Circular dichroism spectroscopy

CD spectra were recorded at concentrations of 1.4–2.0 absorbance units at 222 nm in 10 mM Na₂HPO₄ (pH 7.2), 0.2 mM tris-carboxyethyl-phosphine (TCEP), using a 0.1 cm cell on a Jasco J-810 spectropolarimeter in continuous mode with 1 nm bandwidth, 1 sec response time, and 100 nm min⁻¹ scan speed at room temperature. All spectra represent the average of three separate scans, and are plotted in units of mean molar residue ellipticity following subtraction of buffer scans.

Fluorescence measurements

Fluorescence measurements were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer at ambient temperatures (~20°C), with a Corning O-54 cut-off filter to reduce light scattering. All samples were prepared as for ITC (see below). Samples were excited at 305 nm, a wavelength where excitation of tyrosine and phenylalanine is minimal, and emission intensities recorded from 315 to 440 nm with a 1 nm step size. Excitation and emission slit bandpass settings were 10 nm with 900 V excitation beam voltage, which was reduced to 800 V for titrations with SF3b155r.

To calculate the association constant (K_a), we considered the equilibrium between free SF3b155 (or SF1) and the SF3b155/U2AF⁶⁵-dFW (or SF1/U2AF⁶⁵-dFW) complexes as a function of U2AF⁶⁵-dFW concentration. In our experiment, the addition of U2AF⁶⁵-dFW during titration significantly diluted the sample. Therefore, for each titration point, the observed relative concentrations of free SF3b155 (or SF1), the complexes of SF3b155/U2AF⁶⁵-dFW (or SF1/U2AF⁶⁵-dFW), and the fluorescence intensity were corrected for the dilution effect. Let us consider an initial volume of SF3b155 (or SF1) to be V_0 and the total volume of added U2AF⁶⁵-dFW in x titration steps V_x . Where the initial concentration of SF3b155 (or SF1) is called $[S_T(0)]$, the initial concentration of U2AF⁶⁵-dFW is called $[U_T(0)]$, and the observed intensity is called $I_{obs}(x)$, the apparent concentrations of SF3b155 (or SF1) (called $[S_T(x)]$), U2AF⁶⁵-dFW (called $[U_T(x)]$), and volume-corrected intensity (called I_{corr}) will be:

$$[S_T(x)] = [S_T(0)] \left(\frac{V_0}{V_0 + V_x} \right) \quad [U_T(x)] = [U_T(0)] \left(\frac{V_x}{V_0 + V_x} \right) \quad I_{corr} = I_{obs}(x) \left(\frac{V_0 + V_x}{V_0} \right) \quad (1)$$

Based on equation (1), the concentration of SF3b155 (or SF1)/U2AF⁶⁵-dFW complex, $[SU(x)]$ would be calculated with the following equation, where n is the estimated number of U2AF⁶⁵-dFW-binding sites within SF3b155, assuming identical and independent sites:

$$[SU(x)] = \frac{(nK_a[S_T(x)] + K_a[U_T(x)] + 1) - \sqrt{(nK_a[S_T(x)] + K_a[U_T(x)] + 1)^2 - (4nK_a^2[S_T(x)][U_T(x)])}}{2K_a} \quad (2)$$

The derivation of equation (2) is given in the Supplementary Data. For each titration point, emission intensity was integrated from 320.0 – 410.0 nm to obtain the observed total emission intensity, $I_{obs}(x)$. Before minimization, $I_{obs}(x)$ was corrected for dilution to obtain I_{corr} as shown in equation (1). Only free and bound (in complex) SF3b155 (or SF1) contribute to the

fluorescence signal. If the fluorescence intensity of free form is I_F and bound form is I_B , the fluorescence intensity may be calculated using the following equation:

$$I_{corr} = \left(\frac{n[S_T(x)] - [SU(x)]}{n[S_T(x)]} \right) I_F + \left(\frac{[SU(x)]}{n[S_T(x)]} \right) I_B \quad (3)$$

Both the K_a and the total emission intensity of the bound complex (I_B) were determined by least squares minimization of equation (3) using MathCad 2001i Professional. The dissociation constants ($K_d = K_a^{-1}$) shown in Table 3 are average values of three experiments.

Isothermal titration calorimetry

The heats generated by addition of U2AF⁶⁵-UHM to SF1 or SF3b155 variants in a sample cell were measured at 30°C (with the exception of U2AF⁶⁵-dFW to SF3b155-W7, which was also measured at 20°C and 25°C) using a VP-ITC calorimeter (Microcal). Samples were extensively dialyzed into 50 mM NaCl, 25 mM HEPES [4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid], 0.2 mM TCEP, pH 7.2 before ITC and fluorescence experiments. The heat of U2AF⁶⁵ dilution was negligible, so linear regressions of buffer-buffer titrations were subtracted to correct for heats of binding. Data were processed and thermodynamic parameters obtained using the least squares fitting routines available in the Origin v7.0 software (Microcal). Values in Tables 1 and 2 are averages of at least two experiments.

Sequence analysis and Molecular modeling

SF3b155 sequences shown in Figure 3 were aligned using ClustalW⁵⁹. Models of SF3b155/U2AF⁶⁵ complexes were energy minimized using the program Modeller⁶⁰, and illustrated using Pymol v0.97 (<http://www.pymol.org>). The polar and nonpolar components of the buried solvent accessible surface area (ASA) were calculated for the SF3b155-W7p/U2AF⁶⁵-dFW model using the program NACCESS⁶¹, using a probe radius of 1.4 Å and slice width of 0.05 Å. The SF3b155-W7p/U2AF⁶⁵-dFW model buries 770 Å² of nonpolar surface, and 533 Å² of polar surface. Empirically, the heat capacity change (ΔC_p) associated with complex formation obeys the following equation⁴⁰:

$$\Delta C_p = \Delta ASA_{nonpolar} \Delta C_{p,nonpolar} + \Delta ASA_{polar} \Delta C_{p,polar}$$

$\Delta C_{p,nonpolar}$ and $\Delta C_{p,polar}$ are the elementary ΔC_p contributions, and have been estimated from solid model peptide dissolution studies as 0.45 ± 0.02 cal (K mol Å²)⁻¹ or -0.26 ± 0.03 cal (K mol Å²)⁻¹, respectively.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank S. Lindley and X. Yuan for assistance with protein production; A. Armstrong for assistance with NACCESS; M. Green (U2AF⁶⁵) and J. Berglund (SF1) for kind gifts of PCR templates; and B. Garcia-Moreno, K. Fleming, V. Frasca, and T. Record for insightful discussions. We are especially grateful to R. McMacken, C. Pickart, J. Lorsch, and S. Muzamil for critically reading the manuscript. The laboratory of C.L.K. is supported by the National Institutes of Health (NIH, GM070503-01). The Center for Fluorescence Spectroscopy (Z.G.) is supported by Center grant RR-08119. K.R.T. was supported in part by a Training Grant (T32 GM08403) from the NIH.

References

1. Mazroui R, Puoti A, Kramer A. Splicing factor SF1 from Drosophila and Caenorhabditis: presence of an N-terminal RS domain and requirement for viability. RNA 1999;5:1615–31. [PubMed: 10606272]

2. Kanaar R, Roche SE, Beall EL, Green MR, Rio DC. The conserved pre-mRNA splicing factor U2AF from *Drosophila*: requirement for viability. *Science* 1993;262:569–573. [PubMed: 7692602]
3. Banerjee H, Rahn A, Gawande B, Guth S, Valcarcel J, Singh R. The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF⁶⁵) is essential *in vivo* but dispensable for activity *in vitro*. *RNA* 2004;10:240–253. [PubMed: 14730023]
4. Maniatis T, Tasic B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 2002;418:236–243. [PubMed: 12110900]
5. Jurica MS, Moore MJ. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 2003;12:5–14. [PubMed: 12887888]
6. Brow DA. Allosteric cascade of spliceosome activation. *Annu Rev Genet* 2002;36:333–60. [PubMed: 12429696]
7. Reed R. The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* 1989;3:2113–2123. [PubMed: 2628164]
8. Lim SR, Hertel KJ. Commitment to splice site pairing coincides with A complex formation. *Mol Cell* 2004;15:477–83. [PubMed: 15304226]
9. Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF⁶⁵ and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* 1998;12:858–67. [PubMed: 9512519]
10. Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 1997;89:781–7. [PubMed: 9182766]
11. Zamore PD, Green MR. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* 1989;86:9243–9247. [PubMed: 2531895]
12. Rain JC, Rafi Z, Rhani Z, Legrain P, Kramer A. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* 1998;4:551–65. [PubMed: 9582097]
13. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. Structural basis for the molecular recognition between human splicing factors U2AF⁶⁵ and SF1/mBBP. *Mol Cell* 2003;11:965–76. [PubMed: 12718882]
14. Kielkopf CL, Lucke S, Green MR. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 2004;18:1513–26. [PubMed: 15231733]
15. Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* 1997;16:277–282. [PubMed: 9207794]
16. Rutz B, Seraphin B. Transient interaction of BBP/ScSF1 and Mud2 with the splicing machinery affects the kinetics of spliceosome assembly. *RNA* 1999;5:819–31. [PubMed: 10376880]
17. Bennett M, Michaud S, Kingston J, Reed R. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev* 1992;6:1986–2000. [PubMed: 1398075]
18. Fleckner J, Zhang M, Valcarcel J, Green MR. U2AF⁶⁵ recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes Dev* 1997;11:1864–72. [PubMed: 9242493]
19. Gozani O, Potashkin J, Reed R. A potential role for U2AF-SAP155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* 1998;18:4752–4760. [PubMed: 9671485]
20. Query CC, Strobel SA, Sharp PA. Three recognition events at the branch-site adenine. *EMBO J* 1996;15:1392–402. [PubMed: 8635472]
21. Golas MM, Sander B, Will CL, Luhrmann R, Stark H. Molecular architecture of the multiprotein splicing factor SF3b. *Science* 2003;300:980–4. [PubMed: 12738865]
22. Golas MM, Sander B, Will CL, Luhrmann R, Stark H. Major conformational change in the complex SF3b upon integration into the spliceosomal U11/U12 di-snRNP as revealed by electron cryomicroscopy. *Mol Cell* 2005;17:869–83. [PubMed: 15780942]
23. Will CL, Schneider C, Reed R, Luhrmann R. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 1999;284:2003–2005. [PubMed: 10373121]
24. Will CL, Schneider C, MacMillan AM, Katopodis NF, Neubauer G, Wilm M, Luhrmann R, Query CC. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* 2001;20:4536–46. [PubMed: 11500380]

25. Maris C, Dominguez C, Allain FH. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 2005;272:2118–31. [PubMed: 15853797]
26. Kielkopf CL, Rodionova NA, Green MR, Burley SK. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF³⁵/U2AF⁶⁵ heterodimer. *Cell* 2001;106:595–605. [PubMed: 11551507]
27. Wang X, Bruderer S, Rafi Z, Xue J, Milburn PJ, Kramer A, Robinson PJ. Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. *EMBO J* 1999;18:4549–59. [PubMed: 10449420]
28. Seghezzi W, Chua K, Shanahan F, Gozani O, Reed R, Lees E. Cyclin E associates with components of the pre-mRNA splicing machinery in mammalian cells. *Mol Cell Biol* 1998;18:4526–36. [PubMed: 9671462]
29. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;6:197–208. [PubMed: 15738986]
30. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure Fold Des* 2003;11:1453–9. [PubMed: 14604535]
31. Dosztanyi Z, Csizmok V, Tompa P, Simon I. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J Mol Biol* 2005;347:827–39. [PubMed: 15769473]
32. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop Genome Inform* 1999;10:30–40.
33. Woody, RW. Theory of circular dichroism in proteins. In: Fasman, GD., editor. *Circular Dichroism and the Conformational Analysis of Biomolecules*. Plenum Press; New York: 1996. p. 35–67.
34. Bochicchio B, Tamburro AM. Polyproline II structure in proteins: identification by chiroptical spectroscopies, stability, and functions. *Chirality* 2002;14:782–92. [PubMed: 12395395]
35. Lakowicz, JR. *Principles of Protein Fluorescence*. Plenum Publishers; New York: 1999.
36. Abovich N, Rosbash M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 1997;89:403–12. [PubMed: 9150140]
37. Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 2001;294:1098–102. [PubMed: 11691992]
38. O'Brien, R.; Ladbury, JE.; Chowdhry, BZ. Isothermal titration calorimetry of biomolecules. In: Harding, SE.; Chowdhry, BZ., editors. *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*. Oxford University Press; New York: 2001. p. 263–286.
39. Morrisett JD, David JS, Pownall HJ, Gotto AM Jr. Interaction of an apolipoprotein (apoLP-alanine) with phosphatidylcholine. *Biochemistry* 1973;12:1290–9. [PubMed: 4348832]
40. Freire E. Structural thermodynamics: prediction of protein stability and protein binding affinities. *Arch Biochem Biophys* 1993;303:181–4. [PubMed: 8512307]
41. Janin J, Miller S, Chothia C. Surface, subunit interfaces and interior of oligomeric proteins. *J Mol Biol* 1988;204:155–64. [PubMed: 3216390]
42. Valcarcel J, Gaur RK, Singh R, Green MR. Interaction of U2AF⁶⁵ RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* 1996;273:1706–9. [PubMed: 8781232]
43. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 2000;10:153–159. [PubMed: 10753808]
44. Xiao L, Honig B. Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol* 1999;289:1435–1444. [PubMed: 10373377]
45. Serrano L, Kellis JTJ, Cann P, Matouschek A, Fersht AR. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* 1992;224:783–804. [PubMed: 1569557]
46. Sheinerman FB, Norel R, Honig B. Electrostatic aspects of protein-protein interactions. *Curr Opin Struct Biol* 2000;10:153–9. [PubMed: 10753808]
47. Chiara MD, Palandjian L, Feld Kramer R, Reed R. Evidence that U5 snRNP recognizes the 3' splice site for catalytic step II in mammals. *EMBO J* 1997;16:4746–59. [PubMed: 9303319]

48. Page-McCaw PS, Amonlirdviman K, Sharp PA. PUF60: a novel U2AF⁶⁵-related splicing activity. *RNA* 1999;5:1548–1560. [PubMed: 10606266]
49. Van Buskirk C, Schupbach T. Half pint regulates alternative splice site selection in *Drosophila*. *Dev Cell* 2002;2:343–53. [PubMed: 11879639]
50. Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J. Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 2002;109:285–96. [PubMed: 12015979]
51. Maucuer A, Ozon S, Manceau V, Gavet O, Lawler S, Curmi P, Sobel A. KIS is a protein kinase with an RNA recognition motif. *J Biol Chem* 1997;272:23151–6. [PubMed: 9287318]
52. Boudrez A, Beullens M, Waelkens E, Stalmans W, Bollen M. Phosphorylation-dependent interaction between the splicing factors SAP155 and NIPPI. *J Biol Chem* 2002;277:31834–41. [PubMed: 12105215]
53. Abovich N, Liao XC, Rosbash M. The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition. *Genes Dev* 1994;8:843–54. [PubMed: 7926772]
54. Spingola M, Grate L, Haussler D, Ares M Jr. Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 1999;5:221–34. [PubMed: 10024174]
55. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 2004;337:635–45. [PubMed: 15019783]
56. Vucetic S, Obradovic Z, Vacic V, Radivojac P, Peng K, Iakoucheva LM, Cortese MS, Lawson JD, Brown CJ, Sikes JG, Newton CD, Dunker AK. DisProt: a database of protein disorder. *Bioinformatics* 2005;21:137–40. [PubMed: 15310560]
57. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev* 2003;17:419–37. [PubMed: 12600935]
58. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;72:248–54. [PubMed: 942051]
59. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–80. [PubMed: 7984417]
60. Sali A, Blundell TL. Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815. [PubMed: 8254673]
61. Hubbard, SJ.; Thornton, JM. Department of Biochemistry and Molecular Biology. University College London; London: 1993.

Figure 1.

Schematic diagrams and nomenclature of splicing factors and variants used in this study. **(a)** Schematic diagram of the exchange of SF1 for SF3b155 complexes with U2AF⁶⁵. The U2 snRNA, branch point sequence (BPS), poly-pyrimidine tract (Py-tract), and 3' splice site (3' SS), and intron-exon boundary are labeled. Tryptophan residues (W) of splicing factors SF1 (yellow), U2 snRNP subunits (green), U2AF⁶⁵ (blue) are indicated. The U2AF³⁵ subunit is shown in red. **(b)** SF3b155 domain organization. The domain labeled 'p14-interaction' is an α -helical domain thought to surround p14 in the electron cryomicroscopy structure (pink). SF3b155 variants used in this study are shown below: SF3b155r, tryptophan-containing 'repeat' domain (W, grey); SF3b155-W1, W2, W3, W4, W5, W6, W7 contain alanine

substitutions of all SF3b155r tryptophans other than the one indicated; SF3b155-dW, all SF3b155r tryptophans are substituted with alanine. **(c)** SF1 domain organization. Domains responsible for protein-protein interactions with U2AF⁶⁵ (W, grey), BPS-binding (pink), and the C-terminal proline-rich domain (grey) are indicated. **(d)** U2AF⁶⁵ domain organization. Domains responsible for heterodimerization with U2AF³⁵ (W, grey), Py-tract binding (pink), and protein-protein interactions with SF1 and SF3b155 (UHM, blue) are indicated. The terminal phenylalanine and tryptophan residues were deleted in the U2AF⁶⁵-dFW variant.

Figure 2.

Structural models of U2AF⁶⁵-UHM complexes with the SF3b155r domain. SF3b155 is modeled in an extended β -strand conformation, with U2AF⁶⁵-UHMs bound to all seven tryptophans based upon the SF1p/U2AF⁶⁵-UHM peptide structure¹³, using the program Modeller⁶⁰. The space-filling U2AF⁶⁵-UHM models are colored according to interactions with the SF3b155 site, blue for detectable binding and red for no detectable binding. The tryptophan side chains are colored green. An expanded view of the model of U2AF⁶⁵-UHM with SF3b155-W7p (including residues 337–340) is inset. The C-terminal U2AF⁶⁵ Phe474 and Trp475 side chains are shown in yellow in the inset. Acidic U2AF⁶⁵-UHM residues that interact with SF1 are colored red, basic residues (Arg452 and Lys453) near the acidic SF3b155 residues (Asp339 and Glu340 in SF3b155-W7p) are blue. The SF3b155-W7p tryptophans are colored green. Basic residues immediately preceding the tryptophans are colored purple, and acidic residues following the tryptophans are shown in red.

Figure 3.

Sequence alignments of the SF3b155 'repeat' domain. (a) Sequence alignment of SF3b155r from representative organisms, including: *Homo sapiens* (AAC97189), *Mus musculus*, (NP_112456), *Rattus norvegicus* (AAG01404), *Xenopus laevis* (AAH97718), *Drosophila melanogaster* (AAV55123), *Arabidopsis thaliana* (BAB09858), *Caenorhabditis elegans* (CAA90775), *Schizosaccharomyces pombe* (CAA9328), *Saccharomyces cerevisiae* (CAA89786). Phylogenetically identical residues (*) or residues with conserved expected charge (#) among six or more of the nine homologues are indicated above the alignment. Conserved tryptophans are shaded green. Basic residues immediately preceding the tryptophans are colored purple, and acidic residues following the tryptophans are shown in red to match the colors of the expanded view of Figure 2. (b) Local sequence alignment of the human SF3b155 tryptophan-containing sites and the U2AF⁶⁵-binding site of human SF1. Each

SF3b155 site is identified by the order of the tryptophan in the sequence from N to C, according to nomenclature used in the text (SF3b155-W1, W2, W3, W4, W5, W7), with the residue number of the tryptophan in the following parenthesis. These labels are colored according to interactions with U2AF⁶⁵, blue for detectable binding and red italics for no detectable binding. Residues are colored as shown in (a).

Figure 4.

Unstructured nature of the SF3b155r domain. (a) Predictions of disorder by three different algorithms, DisEMBL³⁰ (—), IUPred³¹ (---), and PONDR³² (· · ·). A probability of 0.5 or greater indicates significant disorder. (b) Far-UV circular dichroism (CD) spectra of SF3b155r (■) and SF3b155-dW (○). (c) Far-UV CD spectra of U2AF⁶⁵-UHM (■) and U2AF⁶⁵-UHM-dFW (○).

Figure 5.

Representative isotherms from ITC experiments with protein domains. With the exception of SF3b155r, the best fit of the binding stoichiometry (n) was one within experimental error. All titrations were measured at 30°C. (a) Titration of U2AF⁶⁵-UHM into SF3b155r. The best fit with stoichiometry of nearly three U2AF⁶⁵-UHM per SF3b155r molecule ($n=2.9$) is indicated with the solid line (—). Fitted curves with fixed n values of one (●●●) or five (---) are shown for comparison. Relative χ^2 values are also given, normalized to one for $n=2.9$. (b) Titration of U2AF⁶⁵-UHM into SF1. (c) Titration of U2AF⁶⁵-UHM into SF1(R19G). (d) Titration of U2AF⁶⁵-UHM into SF3b155-dW.

Figure 6.

Representative isotherms from ITC experiments with individual sites. The best fit of the binding stoichiometry (n) was one within experimental error. All titrations were measured at 30°C with the exception of (b). (a) Titration of U2AF⁶⁵-UHM into SF3b155-W7. (b) Titration of U2AF⁶⁵-dFW into SF3b155-W7 at 20°C. (c) Titration of U2AF⁶⁵-UHM into SF3b155-W7p. (d) Titration of U2AF⁶⁵-UHM into SF1p.

Figure 7.

Tryptophan fluorescence spectroscopy of SF3b155 variants. **(a)** Changes in the emission spectra of a representative SF3b155 variant (SF3b155-W7) upon titration with U2AF⁶⁵-dFW. The emission intensity increases and λ_{max} shifts toward shorter wavelengths as more U2AF⁶⁵-dFW is added. Emission intensity is plotted in arbitrary units (a.u.). **(b)** Emission spectra of a representative SF3b155 variant (SF3b155-W4) that lacks changes upon titration with U2AF⁶⁵-dFW. All spectra are corrected for changes in volume using equation (1) as described in *Methods*. **(c)** Agreement between the observed total emission intensities of SF3b155-W7 upon addition of U2AF⁶⁵-dFW with fitted data points (■) calculated using equations (2) and (3) as described in *Methods*.

Table 1

Thermodynamic parameters determined using ITC

Interaction	K_d (nM)	ΔH° (kcal mol ⁻¹)	ΔG° ^a (kcal mol ⁻¹)	$-T\Delta S^\circ$ ^b (kcal mol ⁻¹)
SF3b155r + U2AF ⁶⁵ -UHM ^c	2833 ± 23	-9.4 ± 0.7	-7.7 ± 0.01	1.7 ± 0.7
SF3b155-W7 + U2AF ⁶⁵ -UHM	2502 ± 115	-14.9 ± 1.5	-7.8 ± 0.03	7.1 ± 1.5
SF3b155-W7 + U2AF ⁶⁵ -dFW	3215 ± 14	-10.8 ± 0.1	-7.6 ± 0.01	3.2 ± 0.1
SF3b155-W7p + U2AF ⁶⁵ -UHM	449 ± 75	-15.0 ± 1.4	-8.8 ± 0.1	6.2 ± 1.4
SF3b155-dW + U2AF ⁶⁵ -UHM	>10 ⁷		<i>no detectable binding</i>	
SF1 + U2AF ⁶⁵ -UHM ^d	11.8 ± 5.1	-17.5 ± 0.7	-11.0 ± 0.3	6.5 ± 0.4
SF1 (R19G) + U2AF ⁶⁵ -UHM	40.5 ± 5.5	-19.2 ± 2.2	-10.3 ± 0.1	9.0 ± 2.1
SF1 + U2AF ⁶⁵ -dFW	55.6 ± 1.3	-12.1 ± 1.5	-10.1 ± 0.01	2.0 ± 1.6
SF1p + U2AF ⁶⁵ -UHM	23.8 ± 3.8	-11.1 ± 0.4	-10.6 ± 0.1	0.5 ± 0.5

With the exception of SF3b155r, the stoichiometry of binding (n) was determined to be one within experimental error. For SF3b155r titrated with U2AF⁶⁵-UHM, n=2.9 ± 0.2. The mean parameters and standard deviations of 2–4 experiments are reported. The standard state (°) for these experiments is defined as 1 M concentrations of each protein at 30°C.

^a Calculated using the equation $\Delta G^\circ = -RT \ln(K_d^{-1})$.

^b Calculated using the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

^c These values describe binding of one U2AF⁶⁵-UHM with one of the SF3b155 sites, assuming three independent and identical sites.

^d These values are estimates due to the difficulty of accurately fitting ITC binding data for very tight interactions, as observed for this binding event.

Table 2
Heat capacity change for SF3b155-W7/U2AF⁶⁵-dFW

Interaction	K _d (nM)	ΔH° (kcal mol ⁻¹)	ΔG° ^a (kcal mol ⁻¹)	-TΔS° ^b (kcal mol ⁻¹)
SF3b155-W7 + U2AF ⁶⁵ -dFW (30° C)	3215 ± 14	-10.8 ± 0.1	-7.6 ± 0.01	3.2 ± 0.1
SF3b155-W7 + U2AF ⁶⁵ -dFW (25° C)	1807 ± 93	-11.6 ± 0.5	-7.8 ± 0.03	3.8 ± 0.5
SF3b155-W7 + U2AF ⁶⁵ -dFW (20° C)	1782 ± 65	-8.4 ± 0.2	-7.7 ± 0.02	0.7 ± 0.2
Heat capacity change (ΔC _p) ^c		-0.24 kcal mol ⁻¹ K ⁻¹		

The stoichiometry of binding (n) was determined to be one within experimental error. For SF3b155r titrated with U2AF⁶⁵-UHM. For each temperature, the mean parameters and standard deviations of 3 experiments are reported. The standard state (°) for these experiments is defined as 1 M concentrations of each protein at the indicated temperature.

^a Calculated using the equation $\Delta G^\circ = -RT \ln(K_d^{-1})$.

^b Calculated using the equation $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$.

^c Calculated from the slope of the linear best fit of ΔH° plotted versus temperature.

Table 3Apparent U2AF⁶⁵-binding affinities of SF3b155 mutants using intrinsic tryptophan fluorescence

Cell Sample ^a	Blue shift in $\lambda_{\max}$ (nm)	K_d (μ M) ^b
SF3b155-W1	18.4 $\pm$ 3.7	15.8 $\pm$ 8.7
SF3b155-W2	19.0 $\pm$ 2.9	21.0 $\pm$ 13.5
SF3b155-W3	14.7 $\pm$ 2.1	36.5 $\pm$ 22.6
SF3b155-W4	0.1 $\pm$ 1.4	No detectable binding
SF3b155-W5	17.5 $\pm$ 2.7	9.5 $\pm$ 9.2
SF3b155-W6	4.0 $\pm$ 2.8	No detectable binding
SF3b155-W7	18.8 $\pm$ 3.0	6.0 $\pm$ 4.7
SF3b155r	18.4 $\pm$ 4.6	Not determined ^c
SF1	15.8 $\pm$ 1.7	Not determined ^d

^aIn all cases the titrant is U2AF⁶⁵-dFW at ambient temperature (~20°C). The mean parameters and standard deviations of three experiments are reported.

^b K_d is the apparent equilibrium dissociation constant estimated for an individual site. Based upon the stoichiometries of binding obtained from the ITC fitting procedure, the stoichiometry (n) is assumed to be one during the fitting procedure.

^cThe sensitivity of spectral changes to indirect binding events prohibited reliable estimation of this K_d by fluorescence techniques.

^dThe high affinity of SF1 for U2AF⁶⁵-dFW prohibited reliable estimation of this K_d by fluorescence techniques.