CANCER-RELEVANT SPlicing FACTor CAPERα ENGAGES THE ESSENTIAL SPlicing FACTor SF3B155 IN A SPECIFIC TERNARY COMPLEX*
Sarah Loerch†, Alexandre Maucuer1,2, Valérie Manceau†, Michael R. Green†, and Clara L. Kielkopf2
From the 1Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, the 2Howard Hughes Medical Institute and Programs in Gene Function and Expression and Molecular Medicine, University of Massachusetts Medical School, Worcester, MA 01605, USA.

†1Current address: Université Pierre et Marie Curie, 9 quai St Bernard, 75005 Paris, France. 2These authors contributed equally to this work.

3To whom correspondence may be addressed: Neuroscience Paris Seine, CNRS UMR8246, Inserm U1130, Université Pierre et Marie Curie Bat B, 6th, Box 12, 9 quai St. Bernard, 75005, Paris, FRANCE; Tel.: 331-44273447; Fax: 331-44272669; E-mail: alexandre.maucuer@inserm.fr

4To whom correspondence may be addressed: Dept. of Biochemistry & Biophysics, University of Rochester School of Medicine and Dentistry, 601 Elmwood Ave. Box 712, Rochester, NY, 14642. Tel.: 585-273-4799; Fax: 585-275-6007; E-mail: clara_kielkopf@urmc.rochester.edu

Running Title: CAPERα Recognizing the Splicing Factor SF3b155
*This work was supported by the National Institutes of Health: R01 GM070503 to C.L.K., R01 GM035490 to M.R.G., and S10 RR026501. The atomic coordinates and structure factors (codes 4OZ0 and 4OZ1 for apo- and ligand-bound CAPERα structures, respectively) have been deposited in the Protein Data Bank (http://www.wwpdb.org).

The abbreviations used are: CAPERα, Coactivator of Activating Protein-1 And Estrogen Receptors-α; ITC, isothermal titration calorimetry; SF1, Splicing Factor-1; SF3b155, Splicing Factor 3b 155 kDa; UHM, U2AF Homology Motif; ULM, U2AF Ligand Motif

Background: CAPERα contains a U2AF Homology Motif (UHM) that typically recognizes U2AF Ligand Motifs (ULM) of pre-mRNA splicing factors.

Results: Crystal structures reveal CAPERα UHM/SF3b155 ULM interactions. CAPERα preferentially associates with the SF3b155 ULM-containing domain.

Conclusion: CAPERα is recruited for pre-mRNA splicing via UHM/ULM interactions with SF3b155.

Significance: Knowledge of CAPERα/SF3b155 complexes will enhance understanding of angiogenic spliceoform regulation by CAPERα.

U2AF Homology Motifs (UHMs) mediate protein-protein interactions with U2AF Ligand Motifs (ULMs) of pre-mRNA splicing factors. The UHM-containing alternative splicing factor CAPERα regulates splicing of tumor-promoting VEGF isoforms, yet the molecular target of the CAPERα UHM is unknown. Here, we present structures of the CAPERα UHM bound to a representative SF3b155 ULM at 1.7 Å resolution, and for comparison, in the absence of ligand at 2.2 Å resolution. The prototypical UHM/ULM interactions authenticate CAPERα as a bona fide member of the UHM-family of proteins. We identify SF3b155 as the relevant ULM-containing partner of full-length CAPERα in human cell extracts. Isothermal titration calorimetry comparisons of the purified CAPERα UHM binding known ULM-containing proteins demonstrate that high affinity interactions depend on the presence of an intact, intrinsically-unstructured SF3b155 domain containing seven ULM-like motifs. The interplay among bound CAPERα molecules gives rise to the appearance of two high affinity sites in the SF3b155 ULM-containing domain. In conjunction with the previously-identified, UHM/ULM-mediated complexes of U2AF65 and SPF45 with SF3b155, this work demonstrates the capacity of SF3b155 to offer a platform for coordinated recruitment of UHM-containing splicing factors.
The Coactivator of Activating Protein-1 and Estrogen Receptors, CAPERα, (also known as HCC1.3, RBM39, F3AP59 and RNPC2) is a pre-mRNA splicing factor and transcriptional co-activator that has emerged as a candidate tumor suppressor in several human malignancies. First identified as an autoantigen in a hepatocellular carcinoma patient (1), overexpression of CAPERα reduces tumor vascularization and growth (2) and inhibits v-Rel mediated lymphocyte transformation (3). CAPERα is detected in human spliceosome complexes (4-7) and has been shown to function in the pre-mRNA splicing pathways of fission yeast, Drosophila, and humans (8-10). In fission yeast, the CAPERα homologue Rsd1 has been shown to connect the U1 small nuclear ribonucleoprotein particle (snRNP) at the 5'-splice site with Prp5 and the SF3b subunit of the U2 snRNP at the 3'-splice site (9). In humans, CAPERα promotes formation of the less angiogenic spliceforms of the vascular endothelial growth factor (VEGF), which could explain the reduced vascularization and growth of CAPERα-overexpressing tumors derived from Ewing sarcoma cells (2). Likewise, CAPERα expression levels are reduced in a CREBBP (11) mouse model of myelodysplastic syndrome (11). In addition to its importance for alternative splicing, CAPERα also functions as a transcriptional coactivator of hormone-sensitive genes (8,12). Despite an increasing knowledge of CAPERα’s physiological roles, the interaction partners of human CAPERα in the considerable splicing factor network (13) remain elusive.

Comparisons of protein sequences afford clues concerning the basis for CAPERα interactions with components of the pre-mRNA splicing machinery (Fig. 1). A similar domain organization, including an N-terminal RS-domain, central RNA Recognition Motifs (RRM1 and RRM2), and a C-terminal U2AF Homology Motif (UHM), mark CAPERα as a parologue of the essential splicing factor U2AF65 (14). However, the specific regulatory function of CAPERα in alternative splicing (2,8) differs from the general U2AF65 requirement for splicing of the major class of introns (14,15). CAPERα and U2AF65 are members of a broader U2AF65-like family of pre-mRNA splicing factors that also includes Puf60 (also called FIR) and CAPERβ (Fig. 1A). The presence of both Puf60 and U2AF65 together has been shown to cooperatively stimulate pre-mRNA splicing in vitro, and the relative ratios of Puf60 and U2AF65 can influence alternative splice site choice in HeLa cells (16). CAPERα is a close relative of CAPERβ at the level of primary sequence conservation (49% sequence identity between CAPERα and CAPERβ, compared with only 19% or 11% identity with U2AF65 or Puf60, respectively). Yet, CAPERα differs from CAPERβ in the addition of a distinctive CAPERα UHM. Functionally, the tissue-specific expression pattern of CAPERα is lower than CAPERβ in human kidneys, and the two proteins differ slightly in their abilities to influence alternative splicing and transactivation of regulated genes (8). These differences and the unknown pre-mRNA splicing partners highlight the importance of elucidating the structure and molecular interactions of the distinctive CAPERα UHM.

UHMs comprise a subfamily of RRM that have evolved specialized features to bind U2AF Ligand Motifs (ULM) (reviewed in (17)) as opposed to RNA. The distinctive UHM features include an acidic α-helix and an "RXF"-containing (RXF)-loop that embelish a core RRM-like fold. Candidate UHMs are found in a relatively small group of proteins, the majority of which function in pre-mRNA splicing (for example, Fig. 1A). The abilities of most UHM family members to bind ULMs have been confirmed, yet the partner(s) of the CAPERα UHM remain an open question. In contrast with the globular UHM domains, the consensus ULM comprises an intrinsically-unstructured, linear stretch of basic residues preceding a key tryptophan (Fig. 1B) that respectively engages the acidic UHM residues and stacks between the arginine and phenylalanine side chains of the RXF-loop. Matches with the relatively short and degenerate ULM consensus ([RK]-x(0,3)-W-[DN]-[QE]) are identified in more than 500 human proteins using ScanProSite (18). Nevertheless, functionally important UHM-targets have been established for only three ULM-containing pre-mRNA splicing factors, including single ULMs in each of the early stage splicing factors SF1 (19) and U2AF65 (20,21) and seven ULM-like motifs in the later stage SF3b155 subunit, five of which have been confirmed to bind UHMs (22-25). A
the GST-tag was removed by proteolytic cleavage and ion exchange chromatography prior to ITC. SF3b155 ULM5 (residues 333-342 of NCBI RefSeq NP_036565) or ULM5L (residues 333-355) are synthetic peptides of >98% purity (BioSynthesis Inc). Phosphorylated (P)SF1 was produced using KIS kinase as described (27,29).

Crystallization and structure determination—
For crystallization experiments, the CAPERα UHM was purified by size exclusion chromatography in 50 mM NaCl, 15 mM HEPES pH 7.4, 0.2 mM tris(2-carboxyethyl)phosphine (TCEP). All crystals were grown at 277 K by sitting drop vapor diffusion from equal-volume mixtures of protein and reservoir solutions (400 nL total volume). The apo-CAPERα UHM crystals were obtained from 24 mg/mL protein solution and reservoir composed of 0.1 M Na acetate pH 4.5, 0.1 M 2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol (Bis-Tris) pH 5.5, 25% (w/v) PEG 3350. The SF3b155-bound crystals were obtained from 18.8 mg/mL CAPERα UHM in the presence of four-fold excess SF3b155 ULM5 and layered with a reservoir solution of 0.8 M sodium phosphate, 0.8 M potassium phosphate, 0.1 M HEPES pH 7.5. In both states, the space groups were P2_1_2_1, and the unit cell dimensions were similar (Table 1). Crystals were coated with a 1:1 v/v mixture of Paratone-N and silicon oils and flash-cooled in liquid nitrogen. Crystallographic data were collected using an in-house microfocus sealed-tube X-ray generator equipped with CCD detector (λ=1.54 Å) and processed with the Proteum software package (Bruker AXS, Inc).

Initial attempts to use other UHM structures as search models for molecular replacement (MR) were unsuccessful. A MR solution was obtained using the unpublished structure of mouse apo-CAPERα UHM at 0.95 Å resolution, which meanwhile became available through the Joint Center for Structural Genomics/Partnership for T-Cell Biology (PDB ID: 3S6E). The resulting structures are similar (RMSD 0.4 Å between PDB ID: 3S6E compared with 0.25 Å maximum-likelihood coordinate error for the human apo-structure presented here). The CAPERα UHM/SF3b155 ULM5 structure was determined by difference Fourier keeping the same R_free set as the apo-counterpart. Phenix (31) was used for MR

EXPERIMENTAL PROCEDURES

Peptide preparation, protein purification and phosphorylation of recombinant proteins—
Human CAPERα UHM (isoform b, residues 411-524 of NCBI RefSeq NP_004893), the SF3b155 ULM-containing domain (residues 190-344 of NCBI RefSeq NP_036565), nearly full length SF1 (residues 1-255 of NCBI RefSeq NP_004621), SF1(14-132) comprising the ULM and adjoining coiled-coil (residues 14-132 of NCBI RefSeq NP_004621), and the U2AF65 ULM (residues 85-112 of NCBI RefSeq NP_0092010) were expressed as GST-fusions in E. coli. Following an initial glutathione affinity step, proteins were either dialyzed prior to GST-pull down assays or growing number of UHM/ULM structures have been determined, including the U2AF35 UHM/U2AF65 ULM heterodimer for 3'-splice site recognition (20) and the alternative splicing factor SPF45 UHM bound to the fifth ULM (ULM5) of SF3b155 (22), as well as an NMR characterization of an intramolecular UHM/ULM complex in the nuclear envelope protein Man1 (26). Most recently, the core U2AF65 UHM/SF1 ULM structure (19) showed a U2AF65 UHM interface with a coiled-coiled domain of SF1 (27,28), which in turn is phosphorylated by the UHM-containing KIS kinase (21,25,29). Despite emerging structural views of UHM/ULM complexes, the rules governing UHM/ULM selection and the functional interplay among the five contiguous SF3b155 ULMs remain to be defined.

Here, we demonstrate that CAPERα contains a bona fide UHM that engages in prototypical ULM interactions by determining 2.20 Å and 1.74 Å resolution structures of the apo-CAPERα UHM and its complex with a representative SF3b155 ULM (ULM5), respectively. We identify SF3b155 as the primary ULM-partner of full-length CAPERα in human cell extracts, and show that the CAPERα UHM specifically recognizes the SF3b155 ULM-containing domain with more than 50-fold higher affinity than other ULM-containing proteins. Together with the previously described association of SF3b155 ULMs with the UHMs of U2AF65 (23,30) and SPF45 (22), these findings suggest that the SF3b155 ULM-containing domain coordinates the recruitment of multiple UHM-containing proteins to the pre-mRNA splice site.

Cell Biology Center for Structural Genomics

For crystallization experiments, UHM/SF3b155 ULM crystals were obtained from 24 mg/mL protein solution and reservoir composed of 0.1 M Na acetate pH 4.5, 0.1 M 2,2-bis(hydroxymethyl)-2,2',2''-nitrilotriethanol (Bis-Tris) pH 5.5, 25% (w/v) PEG 3350. The SF3b155-bound crystals were obtained from 18.8 mg/mL CAPERα UHM in the presence of four-fold excess SF3b155 ULM5 and layered with a reservoir solution of 0.8 M sodium phosphate, 0.8 M potassium phosphate, 0.1 M HEPES pH 7.5. In both states, the space groups were P2_1_2_1, and the unit cell dimensions were similar (Table 1). Crystals were coated with a 1:1 v/v mixture of Paratone-N and silicon oils and flash-cooled in liquid nitrogen. Crystallographic data were collected using an in-house microfocus sealed-tube X-ray generator equipped with CCD detector (λ=1.54 Å) and processed with the Proteum software package (Bruker AXS, Inc).

Initial attempts to use other UHM structures as search models for molecular replacement (MR) were unsuccessful. A MR solution was obtained using the unpublished structure of mouse apo-CAPERα UHM at 0.95 Å resolution, which meanwhile became available through the Joint Center for Structural Genomics/Partnership for T-Cell Biology (PDB ID: 3S6E). The resulting structures are similar (RMSD 0.4 Å between PDB ID: 3S6E compared with 0.25 Å maximum-likelihood coordinate error for the human apo-structure presented here). The CAPERα UHM/SF3b155 ULM5 structure was determined by difference Fourier keeping the same R_free set as the apo-counterpart. Phenix (31) was used for MR
and refinement and Coot (32) for manual adjustments. The final structures were evaluated using MolProbity (33) and illustrated using Pymol (Schrödinger, LLC). Buried surface areas were calculated using NACCESS (34) as the sum of the solvent-accessible surface areas of the two molecules less that of the pair. The crystallographic data and refinement statistics are reported in Table 1.

**GST pull-down assays**—For pull-down from cell extracts (Fig. 4 and Fig. 7), HEK293 cells were grown in DMEM supplemented with 10% fetal calf serum. Each 80% confluent 15 cm plate was washed twice in PBS and cells were lysed in 3 mL of 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1% Nonidet P40, 1 mM DTT, and antiprotease mix (Roche Diagnostics). Cell extracts were then clarified by centrifugation at 3000 g for 5 min. To prepare [35S]-labelled CAPERα (used in Fig. 6), a CAPERα isoform b cDNA (NCBI RefSeq NM_004902) was amplified from reverse transcribed HEK293 total RNA using primers: 5'-ATAGGATCCGCCATGGATTAC
AAGGATGACGACGAT
ATTGATGGAC-3' (forward) and 5'-
ATACCTGAGTCATCGTCTACTTGAACCAG
-3' (reverse) and cloned in plasmid pGEM-T. Integrity of the sequence was verified and in vitro translation of CAPERα from the resulting construct was performed using the TnT® SP6 Coupled Reticulocyte Lysate System (Promega) and [35S]-methionine (NEG-709A, Perkin Elmer).

Cell extracts and in vitro translation products were treated with 10 µg/mL RNase I (R4875, Sigma-Aldrich) for 5 min at room temperature, in conditions that were optimized to remove RNA as monitored by agarose gel electrophoresis and ethidium bromide staining and stored at -80 °C before use. Immediately prior to pull-down assays, cell extracts and in vitro translation products were thawed and clarified by centrifugation at 20,000 g for 10 min. For each pull-down experiment, 40 pmol of purified GST fusion protein was used as bait and either 200 µL of HEK293 cell extract (Fig. 4) or 0.3 µL of diluted in vitro translation product in 250 µL of interaction buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1% NP40, 10% v/v glycerol and 1 µg/µL BSA as a nonspecific competitor) (Fig. 6). For pull-downs from increasing amounts of cell extract (Fig. 7), 40 pmol of GST-SF3b155 was incubated with the indicated volumes of cell extract. Comparisons of SF1 phosphorylation states (Fig. 4) used GST-SF1 protein incubated with or without purified KIS kinase in kinase assay conditions as described (27).

The interaction reactions were incubated for 90 min at 4 °C. Glutathione beads (10 µL) (GE Healthcare) were washed twice with interaction buffer, incubated with the interaction reaction for 30 min, washed rapidly 5 times with interaction buffer, and the retained proteins were separated by SDS-PAGE. For assays with in vitro translated proteins, gels were first stained with Coomassie blue and then radioactivity was quantified using a Phosphorimagery (Typhoon FLA9000, GE Healthcare). Quantification of the Coomassie blue staining was achieved using a 700 nm infrared laser scanner (Odyssey, Li-Cor). The fraction of [35S]-labelled proteins bound to the GST-fusion proteins in the interaction mixture was then calculated as the fraction of radioactivity recovered on the beads divided by the fraction of GST fusion proteins recovered on the beads. For experiments with cell extracts, bound proteins were analysed by SDS-PAGE, and immunoblot with anti-U2AF65 (mouse mAb, clone MC3, Sigma-Aldrich), anti-CAPERα (mouse mAb P14, Santa Cruz) and anti-GST (mouse mAb B14, Santa Cruz Biotech). A secondary 800 nm-IRDye-conjugated anti-mouse antibody (Rockland Immunocchemicals) was used for detection and the fluorescence signal was acquired with an infrared laser scanner (Odyssey, Li-Cor). Quantification was performed with ImageJ (35).

**Isothermal titration calorimetry**—Binding affinities of the CAPERα UHM for the SF3b155 ULM5, wild-type and mutant SF3b155, SF1[14-132], and U2AF65 ULM were measured using a VP-ITC (MicroCal, LLC). Both the CAPERα UHM and the respective ULM protein, except for the SF3b155 ULM5 and ULM5L peptides, which were diluted ≥250-fold into the dialysis buffer, were dialyzed against 50 mM NaCl, 25 mM HEPES pH 7.4, 0.2 mM TCEP prior to calorimetry and then extensively degassed. The CAPERα UHM was injected in 28 aliquots of 10 µL each at 2 sec µL-1 into 1.4 ml of the ULM binding partner at 30 °C with constant stirring at 307 rpm, 4 min relaxation time between injections and 15 µCal sec-1 reference power. A control
experiment titrating CAPERα UHM into buffer was used to correct the isotherms for the heats of CAPERα UHM dilution. The isotherms were fit using Origin v7.0 (MicroCal, LLC). The average values and standard deviations resulting from two independent titration experiments are given in Table 2 and representative isotherms are shown in supplemental Fig. 1.

Size exclusion chromatography—The SF3b155 ULM-containing domain was incubated with a seven-fold molar ratio of CAPERα UHM and then separated from excess UHM on a pre-packed Superdex-200 size exclusion column (GE Healthcare) in 50 mM NaCl, 25 mM HEPES pH 7.4, 0.2 mM TCEP. Elution profiles of the individual subunits were compared. Blue Dextran 2000 was used to determine the column void volume. $K_v = (V_e - V_0)/(V_c - V_0)$ was calculated from the void volume ($V_0$), total column volume ($V_c$), and elution volumes of molecular weight standards ($V_e$): chymotrypsinogen (25 kDa), ovalbumin (43 kDa), conalbumin (75 kDa), and aldolase (158 kDa). The linear fit of $K_v$ plotted against the logarithm of the molecular weights of the standard proteins was used to calculate the apparent molecular weights from the elution volumes of the experimental samples. The concentration of the eluted CAPERα/SF3b155 complex was approximately 5 μM assuming 2:1 stoichiometry.

Amino acid analysis—Following 24 hrs hydrolysis at 110 °C in the presence of 6N HCl and 1% Phenol, the amino acid contents of the purified SF3b155/CAPERα complex were quantified using a post-column derivatization technique with ninhydrin and an internal 2.0 nmol Norleucine standard, on a Hitachi L-8800 amino acid analyzer at the University of California-Davis Proteomics Core Facility.

Circular dichroism spectra—Separate CAPERα UHM, SF3b155 and the saturated complex were purified by size exclusion chromatography and dialyzed against 20 mM phosphate buffer pH 7.4 and 0.2 mM TCEP. Reference spectra of stoichiometric mixtures of the CAPERα UHM and SF3b155, isolated proteins and the saturated complex were collected on a Jasco J-815 spectropolarimeter in continuous mode with 1.0 nm bandwidth, 2 sec data integration time, accumulation of 3 scans and 50 nm min$^{-1}$ scan speed at room temperature using a 0.1 cm cell. Spectra were corrected for buffer absorbance.

RESULTS

Structure of the apo-CAPERα UHM—To confirm that the C-terminal domain of CAPERα (residues 411-526) adopts a canonical UHM fold, we determined the 2.20 Å resolution structure in the absence of ligand (Table 1, Fig. 2). Two independent molecules in the crystallographic asymmetric unit are similar, with RMSD’s for the human CAPERα molecules ranging from 0.5 Å between Cα atoms located in regular secondary structure to >3 Å in the RXXF-loops, which are engaged differently by crystal packing (overall RMSD 1.1 Å for 108 matching Cα atoms). The buried surface area at the interface between the two CAPERα molecules (800 Å$^2$) is less than would be expected for a weak homodimer (1620 ± 670 Å$^2$) and within the expected range for crystal packing (570 ± 520 Å$^2$) (36). Accordingly, the CAPERα UHM migrates as a monomer by size exclusion chromatography (Figure 5A).

The structure of the C-terminal domain matches of human CAPERα matches the core βαββββ topology of the parent RRM and specialized UHM-fold families (Fig. 2A). In a structural homology search using DaliLite (37), the tertiary structure of the CAPERα domain clusters with the UHMs of U2AF$^{65}$ (PDB ID: 4FXW, Z-score 11.0, 2.5 Å RMSD for 94 matching Cα atoms), SPF45 (PDB ID: 2PEH, Z-score 10.2, RMSD 2.4 Å for 89 Cu atoms), and Puf60 (PDB ID: 3DXB, Z-score 10.2, RMSD 2.5 Å for 92 Cu atoms), and stands apart from the U2AF$^{35}$ UHM (PDB ID: 1JMT, Z-score 7.8, RMSD 2.4 Å for 81 Cu atoms). An additional α-helix extends the RRM-like core of most known UHMs, including these four similar UHMs (CAPERα, U2AF$^{65}$, Puf60, SPF45). The amphipathic α-helix (residues 502-508 of CAPERα) packs against the antiparallel β-sheet (Fig. 2B), which in RRM5 offers consensus ribonucleoprotein motifs for RNA interactions (RNP1 and RNP2) (reviewed in (38)). Among UHMs, two key RNP residues are degenerate (Q421 at the second RNP2 and N467 at the third RNP1 positions of CAPERα). The fifth RNP1 position, which for RRM5 is an aromatic residue that stacks with bound nucleotide bases, is
conserved among the four CAPERα-like UHMs (Y469 of CAPERα) but is masked from solvent exposure and hence inadvertent RNA binding by hydrophobic residues in the appended UHM α-helix (L508, F509 and Y505 in CAPERα). Following this α-helix, the extended CAPERα C-terminus adopts a unique proline kink (P510) that alters the backbone trajectory and initiates a short 3_10 helix (residues 510-512). Finally, the CAPERα extension integrates between the exterior β-strand (β2) and adjacent α-helix (α1) via an L519-anchor of hydrophobic interactions. Altogether, the interface of this distinctive, 20-residue extension of the CAPERα RRM-like domain buries 1295 Å^2 of surface area, which approaches the average size of a protein-protein interface (1910 ± 760 Å^2) (36).

Structure of the CAPERα UHM bound to a prototypical ULM from SF3b155—We next investigated the nature of ULM interactions with the CAPERα UHM. Given that CAPERα is a parologue of U2AF^65, we initially attempted to co-crystallize the CAPERα UHM with the interaction domains of the U2AF^65 partner, SF1 (SF1[14-132], residues 14-132 (27)), in the phosphorylated and unphosphorylated forms. In parallel, we screened for co-crystallization of the CAPERα UHM with the U2AF^65 ULM (residues 85-112 (20)) based on the documented co-localization and co-immunoprecipitation of CAPERα and U2AF^65 (39). Lastly, we chose to co-crystallize the CAPERα UHM with the SF3b155 ULM5, the most conserved and highest-affinity U2AF^65 UHM-binding site out of the five confirmed SF3b155 ULMs (23). Although complexes of the CAPERα UHM with any of these ULM-containing regions could be isolated by size exclusion chromatography, only the SF3b155 ULM5 produced diffraction-quality co-crystals with the CAPERα UHM (Table 1).

The 1.74 Å resolution structure of the CAPERα UHM/SF3b155 ULM5 complex revealed strong difference density for the bound ULM in one of the two copies in the crystallographic asymmetric unit (Fig. 3A). A symmetry-related molecule occludes the binding site of the other CAPERα UHM. Apparent stoichiometries from size exclusion chromatography (Figure 5A) and isothermal titration calorimetry (ITC) (Table 2) provide respective support for a monomeric CAPERα UHM and one-to-one complex between the CAPERα UHM and SF3b155 ULM5 peptide. The bound copy of the CAPERα UHM is essentially preconfigured to fit the SF3b155 ULM5, with no detectable differences in key residue conformations at the binding site (RMSD 0.56 Å between 103 Cα of the corresponding bound and apo-polypeptide chains).

CAPERα UHM interactions with SF3b155 ULM5—The key interactions between the CAPERα UHM and the minimal SF3b155 ULM5 are structurally conserved compared with known UHM/ULM complexes (Fig. 3B-C). The SF3b155 ULM5 tryptophan (W338), which in general is essential for ULM/UHM interactions (19,20,23), is buried in a hydrophobic pocket between two α-helices of the CAPERα UHM. One surface of the indole ring stacks in a T-shape with F490 from the CAPERα RXF-loop, whereas the other is masked by a salt bridge between R488 of this loop and E447 from the opposing α-helix. Preceding the ULM tryptophan, a serine residue (S336) is located within a cluster of acidic CAPERα residues (D443, E446, and E447) at an analogous position as a serine of the SF1 ULM (S20), which interferes with U2AF^65 association following phosphorylation by protein kinase G-I (40).

Among bound ULMs, the pairwise Cα-Cα distances of the CAPERα-bound ULM5 most closely match the U2AF^65 ULM bound to the U2AF^35 UHM (20) (RMSD 0.13 Å) (Fig. 3C). This may be due to a key tryptophan at the “X” position of the RXF-loop that is a shared feature of the CAPERα and U2AF^35 UHMs and appears to influence the conformation of the C-terminal ULM residues. The U2AF^65 ULM sandwiches the U2AF^35 indole group between prolines near the C-terminus of the peptide. Although the C-terminus of the SF3b155 ULM5 peptide used for structural studies is truncated compared with the U2AF^65 counterpart, a C-terminal proline of the SF3b155 ULM5 neighbors the CAPERα-tryptophan (W489). In the context of the SPF45 UHM (22), SF3b155 ULM5 ligand shares a similar U-shaped conformation as the CAPERα-bound ULM5 (RMSD 0.35 Å) (Fig. 3C) and lies in similar proximity to an SPF45 tyrosine in an “RYF” motif. In contrast, the C-terminal trajectory of the SF1 ULM is more distant from and passes on the
CAPERα Recognizing the Splicing Factor SF3b155

opposite face of the U2AF65 RXF-loop (27), which rather than an aromatic residue offers a solvent-exposed lysine at the “X” position. A C-terminal proline in either the SF3b155 ULM5 or U2AF65 ULMs is preceded by a threonine (SF3b155 T341 and U2AF65 T103) that is phosphorylated in vivo (41). As such, phosphorylation of ULM “TP” motifs offers a potential switch for UHM/ULM regulation through the known influence of phospho-threonine on the cis/trans-isomerization of an adjacent proline (42).

CAPERα specifically recognizes SF3b155—To identify the preferred ULM-containing partner of CAPERα, we compared the abilities of endogenous, full-length CAPERα from human cell extracts to associate with GST-fusions of known ULM-containing splicing factors (Fig. 4A), including the SF3b155 ULM-containing domain, SF1 (residues 1-255 including the ULM, coiled-coil and RNA binding domains), or U2AF65. We tested SF1 either in the unphosphorylated state or following treatment with KIS kinase to introduce specifically phosphorylated serines that slightly enhance U2AF65 association (21). As a negative control, we included a “nonbinding” SF3b155 domain with all ULM tryptophans mutated to alanine (SF3b155-dW) (23). As a positive control, we re-probed the same pull-downs by immunoblots for endogenous U2AF65, which is known to function in a complex with phosphorylated SF1 (21,27). Whereas the pull-downs by U2AF65 or by SF1/phosphorylated SF1 respectively are weak or not detected, CAPERα exhibits significant association with SF3b155, consistent with the presence of CAPERα in B/Bnet spliceosome complexes that contain the SF3b subunits (6,7).

To directly confirm the apparent preference of CAPERα for binding SF3b155, we used ITC to quantify association of purified CAPERα UHM with various ULMs or ULM-containing splicing factors (Table 2, supplemental Fig. 1): the intact SF3b155 ULM-containing domain, the U2AF65 ULM, or the SF114-132 fragment comprising the ULM and adjoining coiled-coil domain known to recognize the U2AF65 UHM (27). The SF3b155 titrations with the CAPERα UHM were best fit by a model of two nonidentical types of sites (supplemental Fig. 1A-B, \( \chi^2 \) 9.6E4 for identical sites vs. 0.5E4 for nonidentical sites). The first of these two classes of sites comprised high affinity sites (Kd 58 ± 2 nM) with an apparent stoichiometry of approximately 2 CAPERα UHM:1 SF3b155 domain. The second class comprised lower affinity sites (Kd 330 ± 42 nM) with an apparent stoichiometry of three CAPERα UHM per SF3b155 domain. Thus, the total apparent stoichiometry in the ITC experiments was close to 5 CAPERα UHM:1 SF3b155, in agreement with five documented SF3b155 ULMs that detectably bind U2AF65 (23). The apparent equilibrium dissociation constants (Kd) of the CAPERα UHM for the U2AF65 ULM or SF114-132 were respectively 20- or 35-fold weaker than even the lower affinity class of SF3b155 sites. The substantially higher affinity of the CAPERα UHM for SF3b155 than for the U2AF65 ULM, which in turn was slightly greater than for SF1, provided quantitative evidence for direct interactions that corroborated the binding preferences in human cell extracts (Fig. 4).

SF3b155 simultaneously associates with two CAPERα UHMs—The presence of seven ULM-like candidates within SF3b155, five of which have been documented to bind UHMs (22-25), raised the question of whether the different SF3b155 sites would sterically compete or simultaneously bind multiple CAPERα molecules. The ITC fits suggested the presence of two high affinity and three lower affinity binding sites for the CAPERα UHM in the SF3b155 ULM-containing domain, however this apparent stoichiometry is subject to the fitting procedure and concentrations of active molecules. We first attempted to characterize the stoichiometry of a saturated CAPERα UHM/SF3b155 complex by size exclusion chromatography of a 7:1 mixture in comparison with the individual subunits and molecular weight standards (Fig. 5A). The CAPERα UHM eluted close to the expected monomeric molecular weight (MW) (18 kDa apparent MW compared with 13 kDa expected MW). However, the isolated SF3b155 domain eluted at a substantially larger apparent size than expected (44 kDa apparent MW compared with 17 kDa expected MW), consistent with the intrinsic disorder of apo-SF3b155 (23) that confers a large effective hydrodynamic radius (43). This discrepancy complicates interpretation of the CAPERα UHM/SF3b155 stoichiometry based on
the elution profiles. Assuming a globular shape and considering the expected MW’s of the subunits, the apparent MW of the complex (86 kDa) would be consistent with a 5:1 CAPERα UHM:1 SF3b155 ratio. In contrast, a stoichiometry of approximately 2:1 CAPERα UHM:1 SF3b155 is calculated using the apparent MW’s of the individual subunits and assuming that the SF3b155 subunit remains largely unfolded when bound to CAPERα.

We resolved these possible interpretations by use of three alternative methods that are independent of the hydrodynamic radius, and converged on the conclusion that the CAPERα UHM/SF3b155 complex most likely comprises a 2:1 stoichiometry following size exclusion chromatography. First, the Coomassie-Blue stained band intensities following SDS-PAGE of the purified complex in comparison with CAPERα UHM/SF3b155 mixtures of known composition agrees with a 2:1 complex (Fig. 5B). Second, amino acid analysis of the purified complex best matched the predicted composition of a 2:1 complex (Fig. 5C). Third, the circular dichroism (CD) spectrum of the purified complex was nearly indistinguishable from the CD spectrum of a known 2:1 mixture of the CAPERα UHM and SF3b155 subunits, but clearly differed from the CD spectra of a 5:1 mixture or the individual subunits (Fig. 5D). Altogether, the purified complex with the SF3b155 ULM-containing domain contains two CAPERα UHM molecules, in agreement with the apparent stoichiometry of the high affinity class of sites resulting from the ITC fits.

CAPERα weakly associates with SF3b155 variants containing individual ULMs—To aid structural interpretation and test this known ULM with highest affinity for U2AF65 (23) we first determined the affinity of the co-crystallized SF3b155 ULM5 for the CAPERα UHM using ITC (Table 2, supplemental Figure 1). The CAPERα UHM bound the minimal SF3b155 ULM5 with 7-fold reduced affinity than the weaker class of sites in the intact SF3b155 domain. To account for this disparity, we initially pursued two straightforward hypotheses, either that the bona fide ULM5 comprises additional bordering sequences or that CAPERα preferentially associates with other SF3b155 ULMs. To test these hypotheses, we leveraged our established SF3b155 variants in which all but one of the SF3b155 ULM tryptophans are replaced by alanines (23), leaving each variant with only a single ULM with integrity for UHM interactions. Although only five ULM-like tryptophans of SF3b155 have been documented to bind UHMs (22-24,30), it remains possible that the additional two tryptophans in poorly conserved motifs (W254 and W310) could bind the CAPERα UHM. As such, we included these tryptophans in our SF3b155 mutations. Both the wild-type SF3b155 domain and an SF3b155 variant with all tryptophans replaced by alanine (SF3b155-dW) share similar random-coil circular dichroism spectra (23), indicating that differences in UHM-binding are unlikely to result from disruption of a well-ordered SF3b155 structure.

To investigate the role of sequences surrounding SF3b155 ULM5, we used ITC to characterize the CAPERα UHM association with an SF3b155 variant containing only the single tryptophan (SF3b155-W338) corresponding to ULM5 (Table 2, supplemental Fig. 1). An apparent binding stoichiometry of 1:1 confirmed that the CAPERα UHM primarily associated with the remaining ULM5 without nonspecific binding to other mutated ULMs. However, the context of the SF3b155-W338 domain reduced the CAPERα UHM affinity by 5-fold (14.0 μM) relative to the isolated ULM5 (residues 333-342), indicating that local extension of ULM5 could not account for the high CAPERα affinity for wild-type SF3b155. Indeed, the SF3b155 domain context reduced the apparent ULM5 affinity for the CAPERα UHM, possibly due to steric hindrance and/or favorable interactions between the positively-charged N-terminus of the ULM5 peptide with the acidic UHM α-helix2, which are expected to be in close proximity. To confirm that sequences directly flanking the SF3b155 ULM5 had little impact on CAPERα UHM affinity, we compared titrations of the CAPERα UHM into a longer ULM5 peptide (ULM5L, residues 333-355) that contained C-terminal TP repeats, and found little improvement in affinity relative to the minimal ULM5 (respective K_d’s 2.3 and 2.4 μM, Table 2).

We next explored the potential contributions of other ULMs to SF3b155 association with CAPERα. To enable the relevant, full length CAPERα protein to be studied, and to circumvent
the large amount of material required for ITC (3 mg CAPERα UHM per experiment), we used pull-down methods to comprehensively compare levels of CAPERα association for all seven different tryptophan-containing sites within SF3b155 (Fig. 1C). Initial pull-downs using human cell extracts failed to show detectable co-precipitation of endogenous CAPERα with any SF3b155 variants containing single ULMs (data not shown), in contrast with the clear CAPERα association with the wild-type SF3b155 domain (Fig. 4). To improve the signal, we modified the assay as described (21) to input 35S-labeled CAPERα that was produced by in vitro translation for pull-down by GST-fused SF3b155 variants (Fig. 6). Using the S35-detection method, CAPERα detectably associates with an SF3b155 variants containing unmodified ULM1 (W200) and ULM5 (W338), and to a lesser extent with an SF3b155 variant containing only ULM4 (W293). However, the sum of the CAPERα pull-down by SF3b155 ULM-containing variants remained nearly an order of magnitude less than that by the intact SF3b155 domain.

We confirmed these results using ITC experiments to quantify the CAPERα UHM affinities for SF3b155 variants containing the tryptophans of either ULM1 (W200) or ULM4 (W293) (Table 2, supplemental Figure 1). We also had characterized the ULM5 (W338) variant by ITC as described above. The SF3b155-W293 variants bound the CAPERα UHM with extremely low affinity (68 μM) (Table 2), 200-fold weaker than the lower affinity class of wild-type SF3b155 sites. Although the SF3b155-W200 variant bound the CAPERα UHM with 6-fold higher affinity than SF3b155-W338, the affinity (2.3 μM) remained respectively 7- and 40-fold weaker than the lower- and higher-affinity sites of wild-type SF3b155.

Altogether, we concluded that none of SF3b155 ULM-like sites, either individually or in the context of a mutated SF3b155 region, could fully account for the strong association of the CAPERα UHM with the intact SF3b155 ULM-containing domain, thereby highlighting roles for different molecular effects in order to fully account for the preferential association of CAPERα with the SF3b155 partner in the spliceosome.

Multiple ULMs enhance CAPERα association with SF3b155—The observations that CAPERα detectably binds three different SF3b155 ULM-containing variants and that at least two CAPERα UHMs simultaneously associate with the SF3b155 ULM-containing domain raised the alternate explanation that interactions with multiple ULMs enhance CAPERα affinity for SF3b155. To investigate this possibility, we further assessed the SF3b155-binding characteristics of full length CAPERα using a series of carefully quantified pull-downs. Due to difficulties in expressing and manipulating in vitro translated CAPERα, the GST-fusion of the SF3b155 ULM-containing domain was incubated with endogenous CAPERα in increasing amounts of HEK 293 cell extract. The retained CAPERα was detected by immunoblot analysis (Fig. 7).

The resulting plot of retained CAPERα versus volume of cell extract exhibits a sigmoidal shape that is typical of cooperative binding. Considering our finding that two CAPERα UHM simultaneously associate with one SF3b155 domain (Fig. 5), the Hill coefficient of this plot (1.8 ± 0.3) suggests that the two apparent high affinity sites are likely to arise from strong positive cooperativity between SF3b155-bound CAPERα UHM molecules.

DISCUSSION

Here, our crystal structures reveal CAPERα as a bona fide member of the UHM family bound to a prototypical ULM ligand. The CAPERα UHM complex with the SF3b155 ULM5 is mediated by (i) a ULM tryptophan (W338 of SF3b155) inserted within a hydrophobic pocket between the UHM α-helices and (ii) preceding basic residues of the ULM interacting with an acidic α1-helix of the UHM. These central interactions are shared among known UHM/ULM structures (19,20,22,27,28), yet new themes and variations emerge from structural comparisons with the CAPERα UHM/SF3b155 ULM5 complex.

We now conclude that nearly all UHM structures append a C-terminal α-helix (α3 of the CAPERα UHM) to the core RRM-like topology (with the exception of the U2AF35 UHM, which is followed by structurally-uncharacterized zinc knuckle). This C-terminal α-helix packs against
the degenerate RNP2 and RNP1 motifs of the UHM, in particular concealing a singular conserved aromatic residue with the capacity to interact with RNA (CAPERα Y469). Notably, it was recently determined that a coiled-coil extension of the SF1 ULM interacts with the C-terminal α-helix of the U2AF^{65} UHM, stabilizing its position against the RNP motifs (27,28). The corresponding, C-terminal extension of the CAPERα UHM is remarkably long; the α-helix is followed by a 3_{10} helix and an extended polypeptide wrapped across the core UHM surface. By analogy, these observations raise the possibility that intact SF3b155 could selectively recognize the distinctive structure of the CAPERα C-terminal extension.

The conformation of the SF3b155 ULM5 bound to CAPERα adopts an inverted U-shape, which is shared by both the same ligand bound to SPF45 UHM (22) and the U2AF^{65} ULM bound to the U2AF^{35} UHM (20). In contrast, the trajectory of the SF1 ULM bound to U2AF^{65} is nearly linear and passes across the opposite face of the RXF-loop (19,27,28). The conformations of the CAPERα- or SPF45-bound SF3b155 ULM5's are secured by intramolecular electrostatic interactions between charged ULM residues flanking the central tryptophan (SF3b155 R337/E340), as well as by steric limits set by a bulky aromatic residue at the “X” position of the UHM RXF-loop (CAPERα W338 or SPF45 Y376). These SF3b155 ULM5 and CAPERα or SPF45 residues diverge from their counterparts in the SF1 ULM/U2AF^{65} UHM complex (respectively, SF1 R21/Q24 and U2AF^{65} K453).

Although the U2AF^{65} ULM lacks the intramolecular R337/E340 salt bridge of SF3b155 ULM5, its backbone conformation is nearly identical to the bound SF3b155 ULM5 ligands. As seen for the SF3b155 ULM5, the U2AF^{65} ULM is constrained sterically by U2AF^{35} W134 in the RXF-loop. This UHM tryptophan additionally engages a proline-rich extension of the U2AF^{65} ULM in a hydrophobic sandwich that is required for U2AF heterodimer formation (20). A C-terminal proline of the minimal, co-crystallized SF3b155 ULM5 abuts the RXF-loop, where in the context of the TP repeats found in intact SF3b155, it could interact with an exposed aromatic UHM residue in an analogous manner as the U2AF^{65} ULM/U2AF^{35} UHM complex. Nevertheless, comparison of a C-terminally extended SF3b155 ULM5 peptide shows little contribution of this region to CAPERα binding, at least in the absence of SF3b155 phosphorylation.

Importantly, we demonstrate that the second-step pre-mRNA splicing factor, SF3b155, is the ULM-containing partner of CAPERα, both for purified proteins and in human cell extracts. Our finding that CAPERα preferentially associates with SF3b155 agrees with mass spectrometry findings that CAPERα is associated with catalytically active spliceosomes (4-7). Furthermore, CAPERα is a known interaction partner of the SR-related alternative splicing factor, SRrp53, which is required for the second step of splicing (44).

We find that two major sites within the ULM-containing domain of SF3b155 can associate simultaneously with CAPERα molecules, and further demonstrate that the integrity of multiple ULMs are important for CAPERα selection of SF3b155 over off-target ULM-containing proteins. This finding expands the paradigm that most splicing factors function as single copy subunits during spliceosome assembly. For example, early hypotheses for multimeric assemblies of the polypyrimidine tract binding protein (PTB) were clarified by evidence for looping of the RNA bound to a PTB monomer (45,46). The UHM-containing U2AF^{65} parologue Puf60 forms weak homodimers (24,47) yet lacks detectable cooperativity for binding SF3b155 fragments containing either ULM1-ULM2 or ULM2-ULM3 (24). The formal prospect remains that the intact ULM-containing domain of SF3b155 is important for Puf60, as we observe here for CAPERα. Now, our establishment that at least two CAPERα molecules simultaneously bind one SF3b155 supports the possibility that multiple copies of CAPERα, and possibly other UHM-containing proteins, contribute to spliceosome assembly.

In a wider context, an open question is whether various UHM-containing splicing factors will compete for binding the multiple SF3b155 ULMs, or exhibit heterotypic positive cooperativity. The ULM-containing domain of SF3b155 has been shown to interact with several UHM-containing splicing factors, including U2AF^{65} (23,30), SPF45 (22), and Puf60 (24).
preferences of U2AF<sup>65</sup> (23) and here CAPERα for binding each of the separated tryptophan-containing SF3b155 sites have been characterized and now can be compared. Both U2AF<sup>65</sup> and CAPERα associate strongly with ULM5 (W338), yet CAPERα displays a distinct preference for SF3b155 ULM1 (W200) and to a lesser extent ULM4 (W293) (Fig. 6). Altogether, considering that U2AF<sup>65</sup> and CAPERα co-localize in cells (39) as well as bind distinct SF3b155 ULMs, and that the U2AF<sup>65</sup> and Puf60 proteins synergistically promote splicing (16), it is tempting to speculate that cross-cooperativity occurs among UHM-containing proteins. Although little is known concerning the RNA-sequence specificity of CAPERα, an assembly of multiple RNA binding proteins would assist recognition of longer cognate RNA sites. Certainly, a sigmoidal response resulting from cooperative UHM binding to SF3b155 would offer the means for subtle changes in splicing factor levels to invoke nearly all-or-none regulation of alternative splicing.

Altogether, we envision a tightly-controlled UHM/ULM network wherein several UHM-containing splicing factors, including U2AF<sup>65</sup> and CAPERα, recognize multiple, "weak" SF3b155 sites in a concerted fashion, and selectively enhance or disrupt recognition of specific splice sites (Fig. 8). SF3b155 is the target of several lead compounds such as spliceostatin A that confer cytotoxicity and anti-tumor effects (reviewed in (48)). Furthermore, mutations in the SF3b1 gene encoding SF3b155 occur in approximately 75% of myelodysplastic syndromes with ring sideroblasts (49-51), as well as chronic lymphocytic leukemia (reviewed in (52)) and cancers (53-55). Although the affected residues lie outside the SF3b155 ULM-containing region <i>per se</i>, future studies will determine whether these pre-leukemia mutations or natural product binding sites in SF3b155 could have an indirect, allosteric influence on the interplay among multiple, bound UHM-containing proteins. Such downstream effects potentially could include inhibition of CAPERα functions as a tumor suppressor controlling the expression of angiogenic VEGF spliceoforms (2).

**Acknowledgments**— We thank Dr. Ben Miller for calorimeter access and Dr. James McGarrah for access to the CD spectrometer at SUNY-Geneseo. We acknowledge Dr. Jermaine Jenkins for ITC and crystallography guidance. We acknowledge Dr. John Schulze for amino acid analysis at UC-Davis.
REFERENCES

1. Imai, H., Chan, E. K., Kiyosawa, K., Fu, X. D., and Tan, E. M. (1993) Novel nuclear autoantigen with splicing factor motifs identified with antibody from hepatocellular carcinoma. *J. Clin. Invest.* **92**, 2419-2426
2. Huang, G., Zhou, Z., Wang, H., and Kleinerman, E. S. (2012) CAPERα alternative splicing regulates the expression of vascular endothelial growth factor 165 in Ewing sarcoma cells. *Cancer* **118**, 2106-2116
3. Dutta, J., Fan, G., and Gelin, C. (2008) CAPERα is a novel Rel-TAD-interacting factor that inhibits lymphocyte transformation by the potent Rel/NF-kappaB oncoprotein v-Rel. *J. Virol.* **82**, 10792-10802
4. Behzadnia, N., Golas, M. M., Hartmuth, K., Sander, B., Deckert, J., Dube, P., Will, C. L., Urlaub, H., Stark, H., and Luhrmann, R. (2007) Composition and three-dimensional EM structure of double affinity-purified, human prespliceosomal A complexes. *EMBO J.* **26**, 1737-1748
5. Hartmuth, K., Urlaub, H., Vornlocher, H. P., Will, C. L., Gentzel, M., Wilm, M., and Luhrmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. USA* **99**, 16719-16724
6. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H., and Luhrmann, R. (2008) Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846-850
7. Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M. M., Sander, B., Will, C. L., Urlaub, H., Stark, H., and Luhrmann, R. (2010) Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**, 2384-2403
8. Dowhan, D. H., Hong, E. P., Auboeuf, D., Dennis, A. P., Wilson, M. M., Berget, S. M., and O'Malley, B. W. (2005) Steroid hormone receptor coactivation and alternative RNA splicing by U2AF65-related proteins CAPERα and CAPERβ. *Mol Cell* **17**, 429-439
9. Shao, W., Kim, H. S., Cao, Y., Xu, Y. Z., and Query, C. C. (2012) A U1-U2 snRNP interaction network during intron definition. *Mol. Cell. Biol.* **32**, 470-478
10. Park, J. W., Parisky, K., Celotto, A. M., Reenan, R. A., and Graveley, B. R. (2004) Identification of alternative splicing regulators by RNA interference in Drosophila. *Proc. Natl. Acad. Sci. USA* **101**, 15974-15979
11. Lemieux, M. E., Cheng, Z., Zhou, Q., White, R., Cornell, J., Kung, A. L., and Rebel, V. I. (2011) Inactivation of a single copy of CREBBP selectively alters pre-mRNA processing in mouse hematopoietic stem cells. *PLoS One* **6**, e24153
12. Jung, D. J., Na, S. Y., Na, D. S., and Lee, J. W. (2002) Molecular cloning and characterization of CAPER, a novel coactivator of activating protein-1 and estrogen receptors. *J Biol Chem* **277**, 1229-1234
13. Smith, D. J., Query, C. C., and Konarska, M. M. (2008) "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol. Cell* **30**, 657-666
14. 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
15. Zamore, P. D., and Green, M. R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. USA* **86**, 9243-9247
16. Hastings, M. L., Allemand, E., Duelli, D. M., Myers, M. P., and Krainer, A. R. (2007) Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF65. *PLoS One* **2**, e538
17. Kielkopf, C. L., Lucke, S., and Green, M. R. (2004) U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* **18**, 1513-1526
18. de Castro, E., Sigrist, C. J., Gattiker, A., Buhl, V. L., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and
ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362-365

19. Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A., and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. Mol Cell 11, 965-976

20. 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

21. Manceau, V., Swenson, M., Le Caer, J. P., Sobel, A., Kielkopf, C. L., and Maucuer, A. (2006) Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF65. FEBS J. 273, 577-587

22. Corsini, L., Bonnal, S., Basquin, J., Hothorn, M., Scheffzek, K., Valcarcel, J., and Sattler, M. (2007) U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. Nature Struct. Mol. Biol. 14, 620-629

23. Thickman, K. R., Swenson, M. C., Kabogo, J. M., Gryczynski, Z., and Kielkopf, C. L. (2006) Multiple U2AF65 binding sites within SF3b155: Thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. J Mol Biol 356, 664-683

24. Corsini, L., Hothorn, M., Stier, G., Rybin, V., Scheffzek, K., Gibson, T. J., and Sattler, M. (2009) Dimerization and protein binding specificity of the U2AF homology motif of the splicing factor Puf60. J Biol Chem 284, 630-639

25. Manceau, V., Kielkopf, C. L., Sobel, A., and Maucuer, A. (2008) Different requirements of the kinase and UHM domains of KIS for its nuclear localization and binding to splicing factors. J Mol Biol 381, 748-762

26. Konde, E., Bourgeois, B., Tellier-Lebegue, C., Wu, W., Perez, J., Caputo, S., Attanda, W., Gasparini, S., Charbonnier, J. B., Gilquin, B., Worman, H. J., and Zinn-Justin, S. (2010) Structural analysis of the Smad2-MAN1 interaction that regulates transforming growth factor-β signaling at the inner nuclear membrane. Biochemistry 49, 8020-8032

27. Wang, W., Maucuer, A., Gupta, A., Manceau, V., Thickman, K. R., Bauer, W. J., Kennedy, S. D., Wedekind, J. E., Green, M. R., and Kielkopf, C. L. (2013) Structure of phosphorylated SF1 bound to U2AF65 in an essential splicing factor complex. Structure 21, 197-208

28. Zhang, Y., Madl, T., Bagdiul, I., Kern, T., Kang, H. S., Zou, P., Mausbacher, N., Sieber, S. A., Kramer, A., and Sattler, M. (2013) Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3′splice site recognition. Nucleic Acids Res. 41, 1343-1354

29. Maucuer, A., Le Caer, J. P., Manceau, V., and Sobel, A. (2000) Specific Ser-Pro phosphorylation by the RNA-recognition motif containing kinase KIS. Eur. J. Biochem. 267, 4456-4464

30. Cass, D. M., and Berglund, J. A. (2006) The SF3b155 N-terminal domain is a scaffold important for splicing. Biochemistry 45, 10092-10101

31. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 58, 1948-1954

32. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126-2132

33. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12-21

34. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS computer program., Department of Biochemistry and Molecular Biology, University College London, London
35. Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004) Image Processing with ImageJ. Biophotonics Int. 11, 36-42
36. Janin, J., Bahadur, R. P., and Chakrabarti, P. (2008) Protein-protein interaction and quaternary structure. Q Rev Biophys 41, 133-180
37. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res. 38, W545-549
38. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J. 272, 2118-2131
39. Ellis, J. D., Lleres, D., Denegri, M., Lamond, A. I., and Caceres, J. F. (2008) Spatial mapping of splicing factor complexes involved in exon and intron definition. J. Cell Biol. 181, 921-934
40. Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P. J., Kramer, A., and Robinson, P. J. (1999) Phosphorylation of splicing factor SF1 on Ser20 by cGMP-dependent protein kinase regulates spliceosome assembly. EMBO J 18, 4549-4559
41. Hornbeck, P. V., Kornhauser, J. M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. 40, D261-270
42. Wulf, G., Finn, G., Suizu, F., and Lu, K. P. (2005) Phosphorylation-specific prolyl isomerization: is there an underlying theme? Nature cell biology, 7, 435-441
43. Uversky, V. N. (2012) Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. Methods Mol. Biol. 896, 179-194
44. Cazalla, D., Newton, K., and Caceres, J. F. (2005) A novel SR-related protein is required for the second step of Pre-mRNA splicing. Mol. Cell. Biol. 25, 2969-2980
45. Lamichhane, R., Daubner, G. M., Thomas-Crusells, J., Auweter, S. D., Manatschal, C., Austin, K. S., Valniuk, O., Allain, F. H., and Rueda, D. (2010) RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression. Proc. Natl. Acad. Sci. USA 107, 4105-4110
46. Auweter, S. D., and Allain, F. H. (2008) Structure-function relationships of the polypyrimidine tract binding protein. Cell. Mol. Life Sci. 65, 516-527
47. Page-McCaw, P. S., Amonlirdviman, K., and Sharp, P. A. (1999) PUF60: a novel U2AF65-related splicing activity. RNA 5, 1548-1560
48. Webb, T. R., Joyner, A. S., and Potter, P. M. (2013) The development and application of small molecule modulators of SF3b as therapeutic agents for cancer. Drug Discov. Today 18, 43-49
49. Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., Chalkidis, G., Suzuki, Y., Shiosaka, M., Kawahata, R., Yamaguchi, T., Otsu, M., Obara, N., Sakata-Yanagimoto, M., Ishiyama, K., Mori, H., Nolte, F., Hofmann, W. K., Miyawaki, S., Sugano, S., Haferlach, C., Koeffler, H. P., Shih, L. Y., Haferlach, T., Chiba, S., Nakauchi, H., Miyano, S., and Ogawa, S. (2011) Frequent pathway mutations of SF3b related protein is required for the second step of Pre-mRNA splicing activity. RNA 5, 1548-1560
50. Graubert, T. A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C. L., Shao, J., Kryski, K., Harris, C. C., Koboldt, D. C., Larson, D. E., McLellan, M. D., Dooling, D. J., Abbott, R. M., Fulton, R. S., Schmidt, H., Kalicki-Veizer, J., O’Laughlin, M., Grillot, M., Baty, J., Heath, S., Frater, J. L., Nasim, T., Link, D. C., Tommasson, M. H., Westervelt, P., DiPersio, J. F., Mardis, E. R., Ley, T. J., Wilson, R. K., and Walter, M. J. (2012) Frequent pathway mutations of the U2AF1 splicing factor in myelodysplasia. Nature 478, 64-69
51. Papaemmanuil, E., Cazzola, M., Bouloukas, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainio, J., Sasse, E., Hestholm, L., Lindberg, E., Gammack, C., Godfrey, K. L., Vlachos, I., Cvejic, A., Rance, R., McGee, C., Ellis, P., Mudie, L. J., Stephens, P. J., McLaren, S., Massie, C. E., Tarpey, P. S., Varella, I., Nik-Zainal, S., Davies, H. R., Shlien, A., Jones, D., Raine, K., Hinton, J., Butler, A. P., Teague, J. W., Baxter, E. J., Scose, J., Galli, A., Della Porta, M. G., Travaglini, E., Groves, M., Tauro, S., Munshi, N. C., Anderson, K. C., El-Naggar, A., Fischer,
A., Mustonen, V., Warren, A. J., Cross, N. C., Green, A. R., Futreal, P. A., Stratton, M. R., Campbell, P. J., and Chronic Myeloid Disorders Working Group of the International Cancer Genome, C. (2011) Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. New Engl. J. Med. 365, 1384-1395

52. Mori, J., Takahashi, Y., and Tanimoto, T. (2012) SF3B1 in chronic lymphocytic leukemia. The New England journal of medicine 366, 1057; author reply 1057-1058

53. Harbour, J. W., Roberson, E. D., Anbunathan, H., Onken, M. D., Worley, L. A., and Bowcock, A. M. (2013) Recurrent mutations at codon 625 of the splicing factor SF3B1 in uveal melanoma. Nat. Genet. 45, 133-135

54. Cancer Genome Atlas, N. (2012) Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70

55. Biankin, A. V., Waddell, N., Kassahn, K. S., Gingras, M. C., Muthuswamy, L. B., Johns, A. L., Miller, D. K., Wilson, P. J., Patch, A. M., Wu, J., Chang, D. K., Cowley, M. J., Gardiner, B. B., Song, S., Harliwong, I., Idrisoglu, S., Nourse, C., Nourbakhsh, E., Manning, S., Wani, S., Gongora, M., Pajic, M., Scarlett, C. J., Gill, A. J., Pinho, A. V., Rooman, I., Anderson, M., Holmes, O., Leonard, C., Taylor, D., Wood, S., Xu, Q., Nones, K., Fink, J. L., Christ, A., Bruxner, T., Cloonan, N., Kolle, G., Newell, F., Pines, M., Mead, R. S., Humphris, J. L., Kaplan, W., Jones, M. D., Colvin, E. K., Nagrial, A. M., Humphrey, E. S., Chou, A., Chin, V. T., Chantrill, L. A., Mawson, A., Samra, J. S., Kench, J. G., Lovell, J. A., Daly, R. J., Merrett, N. D., Toon, C., Epari, K., Nguyen, N. Q., Barbour, A., Zeps, N., Australian Pancreatic Cancer Genome, I., Kakkar, N., Zhao, F., Wu, Y. Q., Wang, M., Muzny, D. M., Fisher, W. E., Brunicardi, F. C., Hodges, S. E., Reid, J. G., Drummond, J., Chang, K., Han, Y., Lewis, L. R., Din, H., Buhay, C. J., Beck, T., Timms, L., Sam, M., Begley, K., Brown, A., Pai, D., Panchal, A., Buchner, N., De Borja, R., Denroche, R. E., Yung, C. K., Serra, S., Oneto, N., Mukhopadhyay, D., Tsao, M. S., Shaw, P. A., Petersen, G. M., Gallinger, S., Hrub, R. H., Maitra, A., Iacobuzio-Donahue, C. A., Schullick, R. D., Wolfgang, C. L., Morgan, R. A., Lawlor, R. T., Capelli, P., Corbo, V., Scardoni, M., Tortora, G., Tempero, M. A., Mann, K. M., Jenkins, N. A., Perez-Mancera, P. A., Adams, D. J., Largaeupada, D. A., Wessels, L. F., Rust, A. G., Stein, L. D., Tuveson, D. A., Copeland, N. G., Musgrove, E. A., Scarpa, A., Eshleman, J. R., Hudson, T. J., Sutherland, R. L., Wheeler, D. A., Pearson, J. V., McPherson, J. D., Gibbs, R. A., and Grimmond, S. M. (2012) Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. Nature 491, 399-405
**FIGURE LEGENDS**

**FIGURE 1.** Domains of UHM and ULM-containing proteins relevant to this study. A, Human CAPERα (NCBI RefSeq NP_004893) compared with human paralogues CAPERβ (NP_060577), U2AF65 (NP_009210), Puf60 (NP_001258027), and SF45 (NP_001139019). B, Human ULM-containing splicing factors SF1 (NP_004621) and SF3b155 (NP_036565). Circled “P”, phosphorylated SF1 SPSP motif. HEAT, helical repeats; KH-QUA2, K-Homology Quaking-Homology-2; RS, arginine-serine-rich; RRM, RNA recognition motif (blue); UHM, U2AF homology motif (cyan); ULM, U2AF ligand motif (magenta); Zn, zinc knuckle. Sequence boundaries of domains relevant to this study and the residue numbers of C-termini are indicated above. C, ULM “consensus” compared with known sequences of human splicing factor ULMs. ULM tryptophans are highlighted yellow.

**FIGURE 2.** Conserved structure of the CAPERα UHM. A, Secondary structure elements of the human CAPERα UHM (residues 412-524 from NCBI RefSeq NP_004893) aligned with its primary sequence and those of representative homologues from *Mus musculus* (NP_573505), *Xenopus laevis* (NP_001086950), *Drosophila melanogaster* (NP_609095), Arabidopsis thaliana (AAO30095), *Caenorhabditis elegans* (NP_74130), Dictyostelium discoideum (XP_638966), and *Schizosaccharomyces pombe* (NP_594422) using Clustal Omega (www.clustal.org). Below, a structure-based alignment among known UHM structures is given: U2AF65 (PDB ID: 4FXW), Puf60 (PDB ID: 3DXB), SF45 (PDB ID: 2PEH), and U2AF35 (PDB ID: 1JMT). Sequence identity is colored: <40% white; 40%-54% yellow, 55%-69% chartreuse, 70%-84% lime, green 85%-99% and 100% forest. UHM/ULM contacts are marked by symbols (red): S, side chain hydrogen bond; M, main chain hydrogen bond; and π, interaction with an aromatic residue. RNP motifs are highlighted by black lines and consensus sequence indicated: o, aliphatic, ϕ, aromatic; x, any residue, L, leucine; G, glycine; +, positively-charged residue. B, Apo-CAPERα UHM structure. Top, view into the RNP face. Below, surface and ball-and-stick representations of the C-terminal extension viewed following a 90° rotation about the y-axis.

**FIGURE 3.** Structure of the CAPERα UHM/SF3b155 ULM5 complex. A, Structure of the CAPERα UHM (dark teal) bound to the SF3b155 ULM5 peptide. The σF-weighted |Fσ|-|Fc| OMIT electron-density map for the SF3b155 ULM5 is shown in magenta at 2.5 σ contour level. B, Expanded view of the CAPERα UHM (cyan)/SF3b155 ULM5 (magenta) complex. A schematic representation of the interactions is expanded to the right. Residues involved in hydrogen bonds or salt bridges are enclosed in ovals and connected by solid lines. Residue interactions in hydrophobic and aromatic interactions are enclosed in squares and connected by solid lines. C, Superposition of the CAPERα UHM/SF3b155 ULM5 structure with other known UHM/ULM structures, including U2AF65 UHM/SF1 ULM (green, PDB ID: 4FXW), SPF45 UHM/SF3b155 ULM5 (yellow, PDB ID: 2PEH), and U2AF35 UHM/U2AF65 ULM (red, PDB ID: 1JMT). Labels mark the “TP” motifs of the ULMs and “RXF” loop of the UHM (where CAPERα W489 is the representative “X” shown).

**FIGURE 4.** Specific association of CAPERα with SF3b155. A, GST-fusions of the ULM-containing proteins SF1 residues 1-255, KIS-phosphorylated (P)SF1, SF3b155 residues 190-344, or SF3b155-dW with all tryptophans mutated to alanines were used for pull-down assays from HEK293 cells of endogenous CAPERα, or as a well-studied control for comparison, U2AF65. The retained CAPERα and U2AF65 proteins were detected by immunoblot. B, The mean values of three experiments with standard deviations show clear, preferential CAPERα co-precipitation with SF3b155 over the other GST fusion proteins, as well as corroborate U2AF65 co-precipitation with SF1 and SF3b155 and slight preference for (P)SF1.

**FIGURE 5.** Stoichiometry of the CAPERα UHM/SF3b155 complex. A, Size exclusion chromatograms of the individual CAPERα UHM (cyan) and SF3b155 ULM-containing domain (residues 190-344) (magenta) overlaid with that of a 7:1 molar mixture of the CAPERα UHM/SF3b155 complex (green, CAPERα UHM/SF3b155SAT). Right, the linear fit of K_in, the linear fit of K_in, the logarithm for the molecular weight standards compared with experimental samples. B, Coomassie blue-stained SDS-PAGE loaded from left-to-right with: equimolar amounts of individual CAPERα (13 kDa) and SF3b155 (17 kDa) subunits, standard mixtures of 1:1, 3:1, and 5:1 molar ratios, and CAPERα:SF3b155SAT purified by size exclusion chromatography (peak *e* in A), either loaded directly or following two-fold dilution. STD, molecular weight markers. The linear fit of known stoichiometry vs. intensity ratios of integrated, background-
corrected CAPERα:SF3b155 bands for the standard mixtures compared with the purified complex (both direct and diluted data points are shown, green). C, The experimental amino acid composition of the purified complex compared with compositions calculated for 1:1, 2:1 and 5:1 CAPERα UHM:SF3b155 stoichiometries. The amino acids are shown that remain intact following hydrolysis and have >0.5% expected change among different stoichiometries. D, Circular dichroism spectra of the CAPERα UHM, the SF3b155 ULM-containing domain and the purified CAPERα UHM/SF3b155SAT complex, compared with standard CAPERα UHM/SF3b155 mixtures of known respective 2:1 (black) or 5:1 (black dashed line) molar ratios. The calculated sum of 2 CAPERα UHM spectra and 1 SF3b155 spectrum (grey dashed line) also is shown for comparison and is similar to the CAPERα UHM/SF3b155SAT complex.

FIGURE 6. Multiple SF3b155 ULM tryptophans are involved in CAPERα association. The indicated GST or GST-fusions of SF3b155 variants containing all (wild-type), no (dW), or single ULM tryptophans were used in pull-down experiments with (35S)-labeled CAPERα produced by in vitro translation. The top and middle panels show a representative autoradiograph of retained protein at 1x (top) and 20x (middle) exposure levels following SDS-PAGE. The Coomassie-blue stain of the same gel is shown in the lower panel. The average percent retention of the input CAPERα and standard deviations of three independent experiments are: Wild-type 24.6±5.5; W200 (in ULM1) 1.0±0.1; W293 (in ULM4) 0.5±0.1; W338 (in ULM5) 0.9±0.1. No CAPERα was detected bound to the dW, W218, W232, W254, or W310 variants of the SF3b155 ULM-containing domain.

FIGURE 7. CAPERα appears to associate cooperatively with SF3b155. A, The GST-fusion of SF3b155 residues 190-344 was incubated with increasing amounts of HEK 293 cell extract and the retained CAPERα was detected by immunoblot. B, “Direct” plot of retained CAPERα (normalized relative to the amount of CAPERα detected in 1 µL of cell extract) vs. volume of cell extract in the binding reaction (Vext) is exponential indicative of cooperativity rather than hyperbolic. C, The Hill plot of (B) results in a Hill coefficient of 1.8 ± 0.3. The mean values and standard deviations of three experiments are plotted.

FIGURE 8. CAPERα binds the ULM-containing domain of SF3b155. A double-headed arrow represents homotypic cooperativity between CAPERα molecules. Whether heterotypic cooperativity or competition occurs among CAPERα and other SF3b155-bound UHM-containing proteins, and the pre-mRNA sequence recognized by CAPERα, remains for future work ("?"). The SF3b155 ULMs (black) and two tryptophan-containing sites without known UHM partners (gray) are shown as “W”s; the SF3b155 phosphoryled state is represented by “P”.
### TABLE 1
Crystallographic Data and Refinement Statistics
RMSD, root mean square deviation.

|                      | Apo-CAPERα UHM | CAPERα UHM/ SF3b155 ULM5 |
|----------------------|----------------|--------------------------|
| Wavelength (Å)       | 1.54           | 1.54                     |
| No. reflections      | 10858/11319    | 20649/21207               |
| Space group          | P2₁2₁2₁        |                          |
| Unit cell (Å)        | \(a = 44.7, b = 52.4, c = 85.1\) |
| Resolution (Å)       | 42.65 - 2.20   | 33.99 - 1.74             |
| Redundancy\(^a\)     | 5.6 (1.6)      | 6.9 (2.9)                |
| Completeness\(^a\)   | 95.9 (80.8)    | 97.5 (88.7)              |
| \(R_{\text{sym}}\)\(^a,b\) (%) | 10.1 (20.7)   | 7.0 (30.5)               |
| \(<I/\sigma(I)>\)\(^a\) | 12.4 (3.4)    | 17.0 (2.7)               |
| No. atoms:           |                |                          |
| Protein              | 1712           | 1806                     |
| Ions                 | 1              | 2                        |
| Water                | 189            | 307                      |
| \(R_{\text{cryst}}/R_{\text{free}}\)\(^c,d\) | 19.6/22.3      | 14.2/18.1                |
| RMSD from ideality:  |                |                          |
| Bond lengths (Å)     | 0.004          | 0.010                    |
| Bond angles (°)      | 0.895          | 1.310                    |
| \(<B>-factors\) (Å²): |                |                          |
| Protein              | 14.0           | 15.6                     |
| Solvent              | 16.7           | 26.8                     |
| Ramachandran plot:   |                |                          |
| Outliers (%)         | 0.0            | 0.0                      |
| Allowed (%)          | 0.9            | 1.8                      |
| Favored (%)          | 99.1           | 98.2                     |

\(^a\) Highest-resolution shell: apo 2.29-2.20 Å and ULM5 complex 1.80-1.74 Å. \(^b\) \(R_{\text{sym}} = \Sigma_{hkl} \Sigma_{i} I_{i} / \Sigma_{hkl} \Sigma_{i} I_{i} \) where \(I_{i}\) is an intensity \(I\) of the \(i\)th measurement of a reflection with indices \(hkl\) and \(<I>\) is the weighted mean of all measurements of \(I\). \(^c\) \(^d\) \(R_{\text{cryst}} = \Sigma |F_{\text{obs}} - F_{\text{calc}}| / \Sigma F_{\text{obs}}\) where \(F_{\text{obs}}\) and \(F_{\text{calc}}\) are the observed and calculated structure factor amplitudes, respectively. \(R_{\text{cryst}}\) and \(R_{\text{free}}\) respectively were calculated using the working set and a 10% test set of randomly-selected reflections that were excluded from the refinement.
TABLE 2
Isothermal titration calorimetry of CAPERα UHM binding ULMs or ULM-containing proteins
Average values and standard deviations of two independent experiments. ∆G° was calculated using ∆G° = -RT ln (K_D) and -T∆S° was calculated using ∆G° = ∆H° - T∆S°, T = 303 K. Values for each class of multiple sites in the wild-type SF3b155 domain describe binding of one CAPERα-UHM to one of these SF3b155 sites. Representative isotherms are given with c-values in supplemental Fig. 1. Boundaries of SF3b155, -W293, -W338 are residues 190-344; ULM5 is residues 333-342 (KRKSRWDETP); ULM5L is residues 333-355 (KRKSRWDETPASQMGGSTPVLTP).

| ULM ligand     | K_D (nM) | Apparent Stoichiometry | ∆G° (kcal mol⁻¹) | ∆H° (kcal mol⁻¹) | -T∆S° (kcal mol⁻¹) |
|----------------|---------|------------------------|------------------|------------------|------------------|
| SF1 (14-132)   | 11000 ± 700 | 1.0 ± 0.1               | -6.9 ± 0.1       | -11.5 ± 0.2      | 4.6 ± 0.2        |
| U2AF₆⁵ ULM     | 6500 ± 1100 | 1.1 ± 0.0               | -7.2 ± 0.1       | -14.5 ± 0.4      | 7.3 ± 0.5        |
| SF3b155        |          |                        |                  |                  |                  |
| high-affinity site | 58 ± 2  | 1.7 ± 0.1               | -10.1 ± 0.1      | -16.3 ± 0.8      | 6.3 ± 0.8        |
| low-affinity site | 330 ± 4 | 3.2 ± 0.1               | -9.0 ± 0.1       | 0.5 ± 0.2        | -9.5 ± 0.2       |
| SF3b155-W200   | 2300 ± 100  | 1.0 ± 0.1               | -7.8 ± 0.1       | -17.4 ± 0.2      | 9.6 ± 0.3        |
| SF3b155-W293   | 68000 ± 1900 | 1.1 ± 0.2               | -5.8 ± 0.1       | -16.0 ± 0.7      | 10.2 ± 0.8       |
| SF3b155-W338   | 14000 ± 1800 | 1.0 ± 0.1               | -6.7 ± 0.9       | -15.3 ± 0.1      | 8.6 ± 0.1        |
| SF3b155 ULM5   | 2400 ± 1    | 0.9 ± 0.1               | -7.8 ± 0.1       | -21.9 ± 1.0      | 12.1 ± 1.0       |
| SF3b155 ULM5L  | 2300 ± 200  | 1.0 ± 0.1               | -7.8 ± 0.1       | -18.0 ± 0.4      | 10.1 ± 0.3       |
FIGURE 1.

A  Representative UHM-containing proteins:

- **CAPER α**
- **CAPER β**
- **U2AF**
- **PUF60**
- **SPF45**

B  ULM-containing splicing factors (in addition to U2AF):

- **SF1**
- **SF3b155**

C  ULM Consensus:

| Protein | Consensus | Length |
|---------|-----------|--------|
| U2AF$^{65}$ | W92       | 84     |
| SF1     | W22       | 34     |
| SF3b155-W200 (ULM1) | 192 PPSKRKKRWDNAQTD 204 |
| SF3b155-W218 (ULM2) | 210 ATPKKSVERQAE 222 |
| SF3b155-W232 (ULM3) | 224 PGHTPSSLRDETP 236 |
| SF3b155-W254  | 246 GATPGSKIWDDETP 258 |
| SF3b155-W293 (ULM4) | 285 TSSARKRKRWDETP 297 |
| SF3b155-W310  | 302 DTPGHGSAETP 314 |
| SF3b155-W338 (ULM5) | 330 GASKRRKRWDETP 342 |
FIGURE 2.

**A**

- **H. sapiens** 412 VQPLA\(\alpha\)QTO SQSQRGQ 433
- **H. musculus** 418 VQPLA\(\alpha\)QTO SQSQRGQ 439
- **X. laevis** 428 QPLA\(\alpha\)QTO SQSQRGQ 449
- **D. melanogaster** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 343 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 273 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 630 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 500 ...\(\alpha\)QTO SQSQRGQ ...

**CAPERα**

- **H. sapiens** 463 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**CAPERα**

- **H. sapiens** 464 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**U2Aβ\(^{31}\)**

- **H. sapiens** 464 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**Puβ60**

- **H. sapiens** 464 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**SPP45**

- **H. sapiens** 464 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**U2Aβ35**

- **H. sapiens** 464 ...\(\alpha\)QTO SQSQRGQ ...
- **H. musculus** 469 ...\(\alpha\)QTO SQSQRGQ ...
- **X. laevis** 479 ...\(\alpha\)QTO SQSQRGQ ...
- **D. melanogaster** 528 ...\(\alpha\)QTO SQSQRGQ ...
- **A. thaliana** 480 ...\(\alpha\)QTO SQSQRGQ ...
- **C. elegans** 322 ...\(\alpha\)QTO SQSQRGQ ...
- **D. discoideum** 680 ...\(\alpha\)QTO SQSQRGQ ...
- **S. pombe** 551 ...\(\alpha\)QTO SQSQRGQ ...

**B**

- **RNP2**
- **RNP1**
- **N**
- **3'**
- **5'**

**References**

[1] TAN VQPLA\(\alpha\)QTO SQSQRGQ 433
[2] H. musculus 418 VQPLA\(\alpha\)QTO SQSQRGQ 439
[3] X. laevis 428 QPLA\(\alpha\)QTO SQSQRGQ 449
[4] D. melanogaster 480 ...\(\alpha\)QTO SQSQRGQ ...
[5] A. thaliana 343 ...\(\alpha\)QTO SQSQRGQ ...
[6] C. elegans 273 ...\(\alpha\)QTO SQSQRGQ ...
[7] D. discoideum 630 ...\(\alpha\)QTO SQSQRGQ ...
[8] S. pombe 500 ...\(\alpha\)QTO SQSQRGQ ...

**CAPERα Recognizing the Splicing Factor SF3b155**
FIGURE 3.

A

B

C

Superposition of:
CAPERα / SF3b155 ULM5
U2AF35 / SF1
SF45 / SF3b155 ULM5
U2AF35 / U2AF35 ULM

Interactions:
D443 - A491 - W489 - R488 - G492
K335 - S336 - R337 - W338 - D339 - E340 - T341 - P342
M425 - E447 - L485 - F490

CAPERα UHM / SF3b155 ULM5
FIGURE 4.
FIGURE 5.

A. Size Exclusion Chromatography

B. Comparison of Coomassie-Staining

C. Amino Acid Analysis

D. Circular Dichroism Spectra
FIGURE 6.

pull-down

SF3b155 variant

Input 50% GST wild-type dW W200 W218 W232 W254 W293 W310 W338

\( ^{(35)S}\text{CAPER}\alpha \)

x1

x20

Coomassie blue
FIGURE 7.

CAPERα Recognizing the Splicing Factor SF3b155

- Pull-down assay with cell extract (2 μL) and various volumes (50, 100, 150, 200 μL) of cell extract.
- Western blots showing anti-CAPERα and anti-GST antibody reactivity.

Graphs showing the relationship between the volume of cell extract and the recovery of CAPERα relative to 1 μL of extract:

- Linear relationship expressed as $\ln(f) = 1.8(\ln(V_{ext}) - 8.5)$.
FIGURE 8.
Cancer-Relevant Splicing Factor CAPERα Engages the Essential Splicing Factor SF3b155 in a Specific Ternary Complex
Sarah Loerch, Alexandre Maucuer, Valérie Manceau, Michael R. Green and Clara L. Kielkopf

J. Biol. Chem. published online May 2, 2014

Access the most updated version of this article at doi: 10.1074/jbc.M114.558825

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2014/05/02/M114.558825.DC1