PNAS
www.pnas.org

Supplementary Information for

Sequence-specific RNA recognition by an RGG motif connects U1 and U2 snRNP for spliceosome assembly

Tebbe de Vries[1,5], William Martelly[2,5], Sébastien Campagne[1,5], Kevin Sabath[3], Chris P. Sarnowski[4], Jason Wong[2], Alexander Leitner[4], Stefanie Jonas[3,*] Shalini Sharma[2,*] and Frédéric H.-T. Allain[1,*]

[1]Institute of Biochemistry, Department of Biology, ETH Zürich, Hönggerbergring 64, CH-8093 Zürich, Switzerland.

[2]Department of Basic Medical Sciences, University of Arizona, College of Medicine-Phoenix, Phoenix, Arizona 85004, USA.

[3]Institute of Molecular Biology and Biophysics, Department of Biology, ETH Zürich, Otto-Stern-Weg 5, CH-8093, Zürich, Switzerland.

[4]Institute of Molecular Systems Biology, Department of Biology, ETH Zürich, Otto-Stern-Weg 3, 8093 Zürich, Switzerland.

[5]These authors contributed equally

*Correspondence should be addressed to:
stefanie.jonas@mol.biol.ethz.ch or shalinijs@email.arizona.edu or allain@bc.biol.ethz.ch

This PDF file includes:

SI Materials and Methods

Protein preparation. All proteins were expressed for 4 hours at 37°C after induction at an OD600 of 0.6-0.8 with 1 mM isopropyl-β-d-thiogalactopyransoide (IPTG). For NMR studies, we expressed proteins in M9-minimal medium containing (1 g/L) $^{15}NH_4Cl$ and (4 g/L) α-D-glucose (for ^{15}N-labeled protein) or (1 g/L) $^{15}NH_4Cl$ and (2 g/L) ^{13}C-α-D-glucose (for ^{15}N- and ^{13}C-labeled protein). For ^{13}C-labelling of ILV methyl groups of SF3A1-UBL, the protein was expressed in 99% D_2O (Sigma-Aldrich), $^{15}NH_4Cl$ and α-D-glucose supplemented with ($^{13}C5$, 98%; 3-D1, 98%) α-ketoisovaleric acid (100 mg/L) and (methyl-^{13}C, 99%; 3,3-D2, 98%) α ketobutyric acid (60 mg/L; both Cambridge Isotope Laboratories, CIL). For cross-linking and crystallography samples, cells were grown in LB-medium (DIFCO™ LB-Broth, Fisher Scientific). The media were complemented with kanamycin and chloramphenicol. Cells were harvested by centrifugation at 4°C, 10 min at 5,000 x *g*, and the cell pellet was resuspended in buffer A (20 mM Tris pH 8, 1 M NaCl, 10 mM Imidazole) supplemented with Complete-EDTA-Free Protease inhibitor (Roche). Cells were lysed using a microfluidizer (Microfluidics) and the lysate was clarified by centrifugation at 17,000 x *g* (30 minutes at 4°C). The protein was purified by Ni-affinity chromatography using an imidazole gradient of buffer B (20 mM Tris pH 8, 0.25 M NaCl, 500 mM Imidazole) on an ÄKTA Prime purification system (Amersham Biosciences) with 5 mL HisTrap column (GE Healthcare) or step-wise by gravity flow using Ni-NTA beads (QIAGEN). The fusion protein was dialysed against buffer C (20 mM Tris pH 8, 0.25 M NaCl, 2.5 mM β-mercaptoethanol) and cleaved with 6x-His-tag-TEV (purified in house) at 4°C overnight leaving three additional N-terminal residues (GlyGlySer). The cleaved protein was passed over Ni-NTA beads to remove the GB1-6xHis and the His-TEV protease. The cleaved protein was subsequently incubated with RNaseOUT (Invitrogen) for 15 minutes and further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 pg (GE) in the NMR buffer (10 mM sodium phosphate pH 6, 50 mM NaCl). The protein was concentrated in the NMR buffer with a 5 kDa molecular mass cut-off centrifugal filter device (Vivascience) at 15°C.

NMR spectroscopy. NMR titrations of protein-RNA complexes were performed by adding unlabeled RNA (5-10 mM) to ^{15}N or ^{13}C-^{15}N labeled protein (0.2-1.5 mM) to follow protein resonances by 2D ^{1}H-^{15}N-HSQC, or by adding unlabeled protein to unlabeled or ^{13}C-^{15}N labelled RNA (0.2 mM) to follow RNA resonances by 1D ^{1}H NMR, 2D ^{1}H-^{1}H-TOCSY (tm 50 ms) and 2D ^{1}H-^{15}N-HSQC at 303 K and 283 K. ^{13}C-labelled methyl groups of SF3A1-UBL were monitored upon addition of U1-SL4, U1 snRNA and *in vitro* reconstituted U1 snRNP (1). NMR fingerprints of ^{13}C-labelled ILV methyl groups of SF3A1-UBL were collected using 2D ^{13}C-^{1}H HMQC in 10 mM sodium phosphate pH 6.8, NaCl 100 mM, DTT 5 mM. Chemical shift changes were calculated and normalized using Excel.

Protein backbone resonance assignments of SF3A1-UBL, free and in complex, were performed using a combination of triple resonance experiments (3D HNCACB, 3D CBCA(CO)NH and 3D HNCO). Side chain assignments were obtained using TOCSY experiments (3D H(CCO)NH, 3D (H)C(CO)NH and

3D HCCH-TOCSY) and NOESY experiments (3D ^{15}N- ^{1}H HSQC NOESY, 3D $^{13}C_{ali}$-^{1}H HSQC NOESY and 3D $^{13}C_{aro}$-^{1}H HSQC NOESY) recorded with 80 ms mixing time.

Nonexchangeable RNA proton resonances, free and in complex, were assigned using 2D ^{1}H-^{1}H TOCSY, 2D ^{1}H-^{1}H NOESY, 2D ^{13}C-^{1}H HSQC D_2O. Exchangeable proton resonances in the RNA were assigned using a 2D ^{1}H-^{1}H NOESY spectrum (200 ms mixing time) at 283 K.

The {^{1}H}-^{15}N heteronuclear NOE experiments were recorded in an interleaved fashion, recording alternatively one increment for the reference and one for the NOE spectrum. A relaxation delay of 2 s and a ^{1}H presaturation delay of 3 s were used in the NOE experiment while a 5 s relaxation delay was used in the reference experiment. Intermolecular NOEs were obtained by comparing a 2D ^{1}H-^{1}H F2-filtered NOESY and a 2D ^{1}H-^{1}H F1,F2-filtered NOESY recorded in D_2O using a mixing of 120 and 100 ms, respectively.

Crystallization and structure determination. For phasing via native single-wavelength anomalous dispersion (native SAD) (2), a highly redundant dataset (8 x 360° at 5° χ increments) was recorded on the same crystal at a wavelength of 2.07505 Å on a PILATUS 6M detector and DA+ software at the PXIII beamline of the Swiss Light Source at 100 K. Data were processed using XDS (version Jan 31, 2020) and all datasets commonly scaled with XSCALE (version Jan 31, 2020) (3). Diffraction of the crystal extended to 1.564 Å for the native dataset and 2.005 Å for the anomalous dataset. Phasing was achieved in two steps: First molecular replacement was carried out using models of the SF3A1-UBL (residues 704-789, PDB ID: 1ZKH) and an UUCG tetraloop (4) in Phaser MR (version 2.8.3) (5). Second, the molecular replacement solution was used to find anomalous sites in the native SAD dataset and calculate an experimental density map with Phaser EP (version 2.8.3) (5) followed by solvent flattening with Phenix RESOLVE (6). The resulting electron density map was of excellent quality allowing building of the remainder of SF3A1-UBL and U1-SL4. Iterative cycles of model building in COOT (version 0.9) (7) and refinement performed in Phenix (version 1.15.2) (8) against the high resolution native dataset were then carried out to finalize the structure. The final structure contains residues 704-791 of SF3A1-UBL (plus N-terminal GlySer sequence from the TEV cleavage site) and nucleotides 139-162 of U1-SL4, with 98.86% of residues in the favored regions of the Ramachandran plot and no residues in the disallowed regions. The bottom of the U1-SL4 stem makes a crystal contact by forming a continuous stacked double helix with a neighboring U1-SL4, however due to flexibility in this part of the RNA, and resulting averaging of presumably slightly different conformations, the density of the last two base pairs (G140:C161, G139:C162) is poorly defined.

CLIR-MS/MS. For CLIR-MS/MS experiments, SF3A1-UBL was added at 1:1.1 molar ratio to U1-SL4 at a concentration of 50 μM (25 nmol). Samples of SF3A1-UBL bound to U1 snRNP for CLIR-MS/MS were reconstituted in two steps. First, U1 snRNP lacking U1-C was reconstituted and purified by anion

exchange chromatography as described previously (1). Then, SF3A1-UBL and U1-C were added in 1:1.5 molar ratio in 10 mM HEPES pH 7.5, 50 mM KCl, 1 mM TCEP. Both samples were cross-linked at 0.8 J/cm^2 at a concentration of 50 µM (U1-SL4 sample) and 0.75 µM (0.12 mg/mL, 5 nmol, U1 snRNP sample), respectively. Protease and RNase digestion for U1-SL4 samples was performed as described previously (9). To ensure protease and RNase digestion of the SF3A1-UBL/U1 snRNP complex, the sample was denatured with 6 M guanidine hydrochloride (GHCl), diluted to 1 M GHCl and incubated with 7.5 µg LysC (1:100 w/w) for 3 h at 37°C, followed by addition of 30 µg trypsin (1:24 w/w) and incubation over night at 37°C prior to RNAse digestion with 6 µg RNase T1, 6 U RNase A and 1500 U benzonase for 3 hours at 37°C, 30 min at 52°C. Following digestion, SF3A1-UBL/U1 snRNP complex samples were split into 4 equal aliquots for the enrichment protocol. Samples were purified using solid phase extraction (SPE, SepPak 100 mg tC18 cartridges, Waters) and dried in a vacuum centrifuge. Metal oxide affinity chromatography (MOAC) was used to enrich protein-RNA cross-links from the digested samples, as described previously (9). In brief, dried samples were resuspended in 100 µL MOAC loading buffer. For SF3A1-UBL, loading buffer composition was 50% acetonitrile (ACN), 0.1% trifluoroacetic acid (TFA), 10 mg/ml lactic acid; for SF3A1-UBL/U1 snRNP samples, the composition was identical except 300 mg/ml lactic acid was used instead. Samples were then added to 5 mg of pre-equilibrated TiO_2 beads (10 µm Titansphere PhosTiO, GL Sciences) and incubated on a Thermomixer (Eppendorf) at 1,200 rpm for 30 min at 25°C. Beads were centrifuged (10,000 x *g*, 1 min) and the supernatant removed, discarded, and replaced with 100 µL fresh loading buffer. Samples were again incubated on a Thermomixer (1,200 rpm, 15 min, 25°C) and beads centrifuged, and the supernatant was removed, discarded, and replaced with 100 µL MOAC washing buffer (50% ACN, 0.1% TFA). Samples were again incubated on a Thermomixer (1,200 rpm, 15 min, 25°C), centrifuged, and the supernatant was removed, discarded, and replaced with 75 µL MOAC elution buffer (50 mM $(NH_4)_2HPO_4$, pH 10.5). Finally, samples were incubated on a Thermomixer (1,400 rpm, 15 min, 25 °C) and beads settled by centrifugation. Supernatant was collected and stored on ice, and the elution repeated a second time. The eluate from the second step was combined with that from the first, and immediately acidified to pH 2-3 using TFA.

Enriched samples were cleaned up using Stage tip C_{18} solid phase extraction (10). Briefly, C18 filter (Empore, 3M) was packed into a 200 µL tip (Axygen), washed with 100% ACN, 80% ACN with 0.1% formic acid (FA), then twice equilibrated with 5% ACN with 0.1% FA. The sample was applied to the tips, flow through collected, and loaded again. Tips were washed 3 times using 5% ACN with 0.1% FA. Peptide-RNA adducts were then eluted from tips using 50% ACN with 0.1% FA into LoBind tubes (Eppendorf). At this stage, previously divided SF3A1-UBL/U1 snRNP samples were recombined. Samples were dried in a vacuum centrifuge, and resuspended in 20 µL 5% ACN with 0.1% FA. 5 µL of each sample was injected for LC-MS/MS analysis, and each sample analyzed in duplicate. LC-MS/MS was carried out using an Easy-nLC 1200 HPLC device (Thermo Fisher Scientific), connected to an Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was

equipped with a Nanoflex nanoflow electrospray ionization source (Thermo Fisher Scientific). Peptide-RNA adducts were separated using a PepMap RSLC column (250 mm x 75 µm, 2 µm particles, Thermo Fisher Scientific); the gradient for HPLC separation was 6-40% mobile phase B over 60 minutes (A =ACN:water:FA, 2:98:0.15 (v/v/v); B =ACN:water:FA, 80:20:0.15 (v/v/v)). Flow rate during separation was 300 nL/min. The mass spectrometer was operated in data-dependent acquisition mode, and the Orbitrap mass analyser was used both for acquisition of precursor ion spectra (resolution=120,000) and fragment ion spectra (resolution=30,000). Fragmentation was performed using higher-energy collisional dissociation (HCD) with stepped collision energies of 21.85%, 23% and 24.15%, and ions with charge states from +2 to +7 were selected for MS2 using the quadrupole mass analyzer (isolation window=1.2 m/z, cycle time=3 s). A dynamic exclusion period of 30 s was used.

Data files produced by the mass spectrometer (Thermo Fisher RAW format) were converted with msconvert.exe (ProteoWizard version 3.0.9393). Outputs were centroided mzXML files, which were searched against a FASTA sequence database containing only the SF3A1-UBL sequence for the SF3A1-UBL-SL4 sample, or a database containing the SF3A1-UBL and U1 protein sequences used in the sample for the SF3A1-UBL/U1 snRNP samples. The search was carried out using xQuest (11) 2.1.5 (available at https://gitlab.ethz.ch/leitner_lab/xquest_xprophet), with all amino acids specified as possible cross-linking sites. All possible nucleotide adducts (1-4 nt in length) based on theoretical complete nucleolysis of the U1-SL4 sequence, and a set of neutral loss products from these, were specified as possible monolink modifications, with expected differences between their respective $^{12}C/^{14}N$ ('light') and $^{13}C/^{15}N$ ('heavy') masses specified as expected isotopic spectral pairs. For spectral pairing, a retention time tolerance window of 60 sec and mass tolerance of +/- 15 ppm was used. Further parameters used for xQuest search (described further in (11)): enzyme = trypsin, maximum number of missed cleavages = 2, MS1 and MS2 mass tolerance = 10 ppm. Cross-link spectrum matches were further refined by ld.Score (>20) and mass error (with 95% of the mean value of all high scoring identifications). For data plotting, cross-link spectrum matches for U1-SL4 and U1 snRNP samples were filtered for counts >2 and >1, respectively. Identifications from xQuest were plotted using custom Python 3.7 scripts. RAW mass spectrometry data files and search engine outputs from xQuest are accessible through PRIDE with the dataset identifier PXD027189.

Surface Plasmon Resonance. SPR experiments were performed using a Pioneer FE Surface Plasmon Resonance System and streptavidin coupled biosensor chips (Sartorius: SADH Biosensors) as described previously (12). The 5′-biotinylated U1-SL4 wildtype (WT) RNA was folded by first heating to 65°C for 5 minutes and then cooling to room-temperature. The folded RNA was immobilized on the surface of flow cells one and three of the biosensor chip (Sartorius: SADH Biosensors) by injecting 10 nM RNA in SPR buffer (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5% glycerol, 62.5 µg/mL bovine serum albumin, 125 µg/mL tRNA, 1 mM dithiothreitol, and 0.05% tween-20) at a flow rate of 10µl/min until the density of 78-150 RU was achieved. Flow cell 2 was left blank to serve as a reference. In two

replicates (Experiments 2 and 3), protein stocks were diluted to a 7.5 μM (wild type), 10 μM (G780I), 15 μM (K717A), 20 μM (F763A), 25 μM (K765A, KK2AA, R788A, R788K) , 40 μM (K754A) or 50 μM (K786A, RGG2AAA, G789I) concentration in SPR running buffer and 9 twofold serial dilutions and were injected over all flow cells. Another replicate (Experiment 1) was performed using protein stocks diluted to 1 μM (wild type) and 20 μM (mutants) concentration and 4-5 twofold serial dilutions. Proteins were injected for a total of 100 μl at a flow rate of 75 or 150 μl/min and subsequently allowed to dissociate for 60 or 180s at a temperature of 25°C. A blank injection was performed every four protein injections. After dissociation, >99% of the biosensor surface was found to be regenerated, therefore, a separate regeneration step was not included. Kinetic parameters were collected at a 40 Hz sampling rate using the OneStep injection setting, and data were fit to a simple 1:1 interaction model using the global data analysis option available within the Qdat analysis software. Dissociation constants (K_d) shown in Table 2 are averages from three experiments (Table S2) and fold change was calculated for mutant proteins relative to WT. The observed K_d values for the binding of SF3A1-UBL to U1-SL4 by SPR here are higher than the previously reported values that were determined by SPR and EMSAs (12). This discrepancy is most likely due to the presence of the GST-tag in the constructs that were used in our previous study and in the EMSAs shown here in Fig. 3. GST has been shown to dimerize (13) and therefore may influence the binding properties of the fusion protein.

Electrophoretic Mobility Shift Assays. EMSA experiments were performed as described previously (12). Briefly, binding reactions containing 10 nM 5′-Cy5-labeled U1-SL4 RNA WT or mutant (Integrated DNA Technologies; see SI Appendix, Table S5 for sequences), 2.2 mM $MgCl_2$, 60% buffer DG, 0.5 mM ATP, and varying concentrations of purified GST-SF3A1-UBL. After incubation for 30 minutes at RT, binding reactions were loaded onto horizontal 6% native-PAGE gels and run at 100V for 45 minutes at 4°C. Native gels were visualized using the Typhoon FLA 9500 Imager.

***In vivo* splicing assays.** The three-exon/two-intron reporter pDUP51p and the U1 snRNA expression plasmid pNS6U1 have been described previously (14). The construct expressing U1-5a snRNA carrying U1-SL3 mutation M1d was generated by PCR and was verified by DNA sequencing. The sequences of the oligonucleotides used for U1-5a mutagenesis are provided in SI Appendix, Table S4. To express N-terminal FLAG-tagged SF3A1 protein in transfected HeLa cells, the full-length SF3A1 CDS containing upstream Kozak and FLAG-tag (DYKDDDDK) sequences was cloned into the *Hind*III and *Not*I restriction sites of the mammalian expression vector pcDNA3.1 by In-Fusion cloning (Takara Bio). Primers used to introduce silent point mutations in SF3A1 generating the RNAi resistant SF3A1 clone, in addition to those used to introduce point mutations into the C-terminal UBL domain, are reported in SI Appendix, Table S4.

HeLa cells, originally purchased from ATCC, were a gift from Kurt Gustin (College of Medicine-Phoenix, University of Arizona) and were authenticated by the University of Arizona Genomics Core.

They were cultured in DMEM containing 10% FBS and antibiotics (100 U/mL penicillin, 100 µg/mL streptomycin) and tested negative for mycoplasma by PCR of DNA harvested from culture supernatants (15). All transfections were performed using Lipofectamine 2000 reagent. For the siRNA knockdown and FLAG-RNAiR-SF3A1 rescue experiments using the Dup51p reporter minigene, HeLa cells were first plated the day before transfection into 12-well plates and allowed to grow to 50% confluency (0.2 $x10^5$ cells/well). HeLa cells were then co-transfected with 50 nM siSF3A1 (16) and 125 ng of pcDNA3.1 plasmid harboring WT and mutant FLAG-RNAiR-SF3A1 constructs using the Lipofectamine 2000 reagent (Thermo Fisher Scientific). After 24 hours, cells were co-transfected with the Dup51p reporter and U1 snRNA expression plasmids (pNS6U1) and total RNA was extracted using TRIzol reagent after an additional 24-hour incubation. In the rescue experiments, a pNS6U1-5a/SL3-M1d/SL4-WT or pNS6U1-5a/SL3-M1d/SL4-M10r to Dup51p ratio of 7.5:1 (1.5 µg pNS6U1-5a/SL3-M1d and 0.2 µg pDup51p) was used along with the addition of pNS6U1-WT (0.3 µg) to maintain the total levels of U1 expression plasmid consistent to that used in previous studies. 1-2 µg of total RNA harvested from transfections were used in primer extension reactions using ^{32}P-labeled Dup3r primer (SI Appendix, Table S5) and exon 2 inclusion of Dup51p was monitored by separation of reaction products on 10% urea-PAGE gels. Resolved gels were dried and were exposed to phosphor screens which we scanned using the Typhoon FLA 9500 imager. All statistical comparisons were performed using the two-tailed Student t-test in Microsoft Excel. Data obtained from three independent experiments was used for analysis and $p < 0.05$ was considered statistically significant. All siRNAs, including the non-targeting control (siNT; siGENOME Non-Targeting Pool #1) were purchased from Horizon Discovery and the sequence of siSF3A1 is reported in SI Appendix, Table S5.

To validate expression of RNAi resistant SF3A1 protein in HeLa cells under knockdown conditions, 50% confluent 12-well plates were co-transfected with 50 nM siNT or siSF3A1 and 250 ng pcDNA3.1 plasmid harboring WT or mutant FLAG-RNAiR-SF3A1 constructs. After 48 hours incubation under standard growth conditions, HeLa cells were scraped from wells, pelleted, and washed with 1X PBS. Total lysate was extracted using RIPA lysis buffer and protein yield was quantified by the BCA assay (Thermo Scientific). Approximately 30 µg of protein was boiled for 5 min at 95°C in 1x SDS-PAGE sample buffer (150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 8.0, 0.1% SDS, 0.5 mM sodium deoxycholate, and 1% NP-40) and loaded onto 8% SDS-PAGE gels. Proteins were transferred onto Immobilon-FL PVDF membranes (Millipore Sigma) and blocked with 5% blocking buffer (5% fat-free milk powder in 1X PBS supplemented with 0.1% Tween-20) for 1 hour at RT or overnight at 4°C. Polyclonal antibodies against the N- and C-terminus of SF3A1 were raised in rabbits and have been described previously (14). Anti-α-Tubulin mouse monoclonal antibody (Calbiochem; CP06-DM1A), and anti-FLAG monoclonal M2 antibody (Sigma-Aldrich; F1804) were obtained commercially. Secondary anti-mouse and anti-rabbit antibodies conjugated to Cy3 and Cy5 fluorophores were purchased from GE Healthcare and probed blots were scanned using the Typhoon FLA 9500 laser scanner and protein bands were quantified by densitometric scanning using the ImageQuant software.

Western blots were first probed with anti-FLAG and anti-C-terminal SF3A1 antibody to detect expression of RNAiR-SF3A1. Several of the point mutations introduced into UBL domain of SF3A1 interfere with SF3A1 detection by the C-terminal SF3A1 antibody, and since the anti-FLAG antibody interferes with the binding of the anti-N-terminal SF3A1 antibody, blots were stripped and re-probed with the anti-N-terminal SF3A1 antibody to confirm detection of all WT and mutant RNAi resistant SF3A1 proteins with SF3A1-specific antibodies. PVDF membranes were stripped by incubating blots with rotation at 50°C for 10 min in harsh stripping buffer (2% SDS, 62.5 mM Tris-HCl, pH = 6.8, 0.8% β-ME) followed by a 10 min wash in deionized water at 50°C with rotation. Blots were rinsed several times with deionized water at RT then blocked again in blocking buffer for 1 hour at RT before re-probing. Expression of the FLAG-tagged SF3A1 mutants was calculated relative to WT and normalized to α-Tubulin (± s.d., n = 3) (SI Appendix, Fig. S6E).

To evaluate assembly of the FLAG-tagged SF3A1 proteins into U2 snRNPs, we performed immunoprecipitation (IP) with an antibody against SF3B1, another U2 specific protein, and probed the IP-complexes with anti-FLAG antibody. For this, HeLa cells co-transfected with siSF3A1 and pcDNA3.1 control or plasmid harboring WT or F763A mutant FLAG-RNAiR-SF3A1 constructs were harvested as described above. Cells from three wells were pooled and resuspended in 300 µL of HKM buffer (20 mM HEPES, pH 7.5, 5 mM KCl, and 0.5 mM $MgCl_2$) containing 1x Halt Protease Inhibitor Cocktail without EDTA (Thermo Fisher Scientific) and then disrupted by sonication (two 15s pulses with 1-min interval) using Sonic Dismembrator (Thermo Fisher Scientific). The lysates were cleared by centrifugation at 14000 rpm for 10 min at 4°C and the supernatant was added to 10 µl packed-volume of GammaBind Sepharose beads (Cytiva) that were pre-bound with 5 µg of anti-SF3B1 antibody (14). The IP reactions were incubated for 2 hours at 4°C with rotation. After incubation, the beads were washed four times with 150 µL of HKM buffer. The bound complexes were eluted with 1x SDS-PAGE sample buffer and probed with anti-SF3B1 and anti-FLAG antibodies. The FLAG:SF3B1 ratios were calculated after normalization to the input samples.

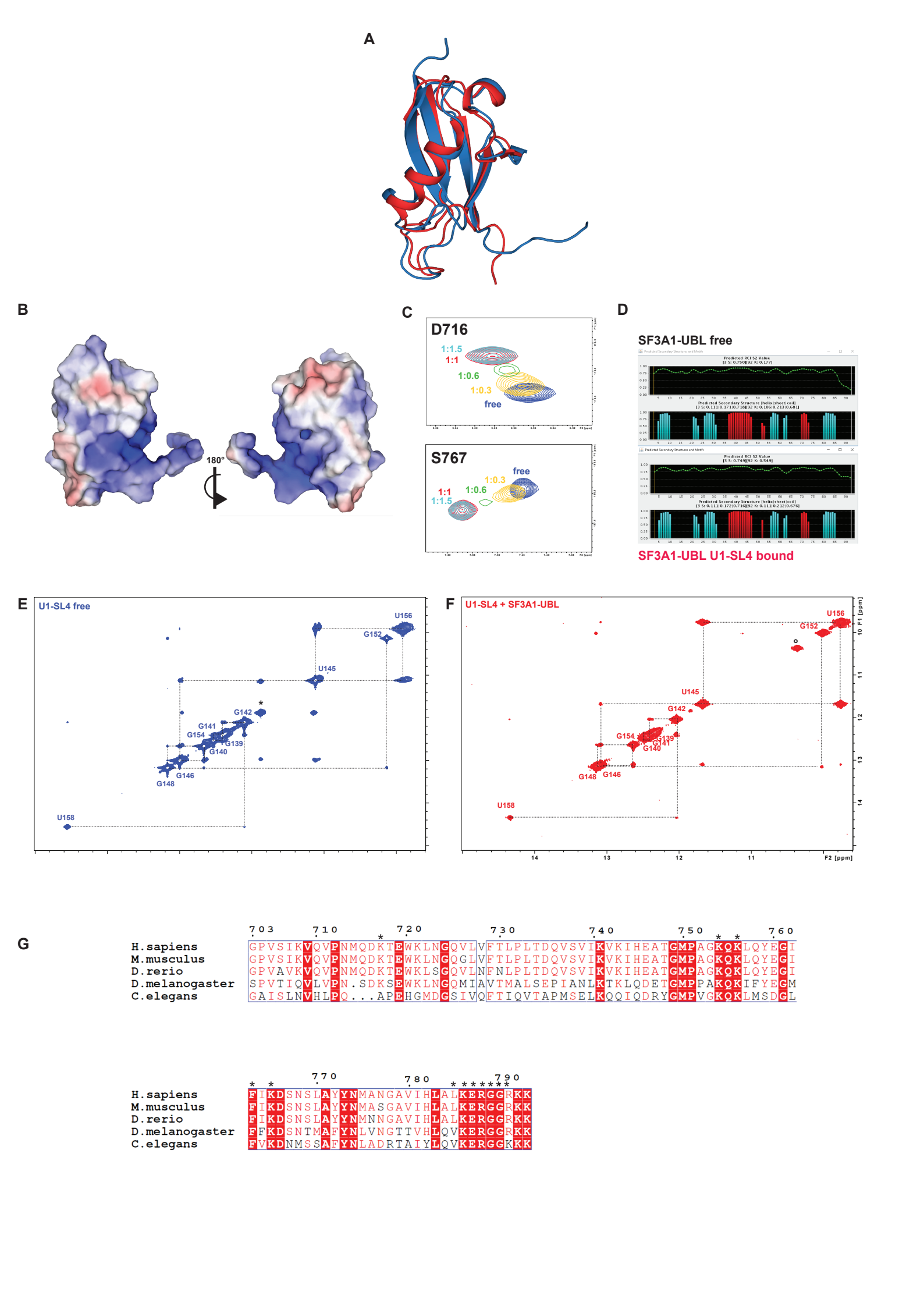

A
B
180°
C
D716
1:1.5
1:1
1:0.6
1:0.3
free
S767
free
1:0.3
1:1
1:0.6
1:1.5
D
SF3A1-UBL free
SF3A1-UBL U1-SL4 bound
E
U1-SL4 free
U156
G152
U145
G142
G141
G154
G139
G140
G146
G148
U158
F
U1-SL4 + SF3A1-UBL
U156
G152
U145
G142
G154
G139
G141
G140
G148
G146
U158
F2 [ppm]
G
H.sapiens
M.musculus
D.rerio
D.melanogaster
C.elegans
H.sapiens GPVSIKVQVPNMQDKTEWKLNGQVLVFTLPLTDQVSVIKVKIHEATGMPAGKQKLQYEGI
M.musculus GPVSIKVQVPNMQDKTEWKLNGQGLVFTLPLTDQVSVIKVKIHEATGMPAGKQKLQYEGI
D.rerio GPVAVKVQVPNMQDKTEWKLSGQVLNFNLPLTDQVSVIKVKIHEATGMPAGKQKLQYEGI
D.melanogaster SPVTIQVLVPN.SDKSEWKLNGQMIAVTMALSEPIANLKTKLQDETGMPAKQKIFYEGM
C.elegans GAISLNVHLPQ...APEHGMDGSIVQFTIQVTAPMSELKQQIQDRYGMPVGKQKLMSDGL
H.sapiens FIKDSNSLAYYNMANGAVIHLALKERGGRKK
M.musculus FIKDSNSLAYYNMASGAVIHLALKERGGRKK
D.rerio FIKDSNSLAYYNMNNGAVIHLALKERGGRKK
D.melanogaster FFKDSNTMAFYNLVNGTTVHLQVKERGGRKK
C.elegans FVKDNMSSAFYNLADRTAIYLQVKERGGKKK

Fig. S1. Interaction of SF3A1-UBL and U1-SL4. (A) Comparison of the solution structure of the free SF3A1-UBL determined previously (PDB ID: 1ZKH, red) and the structure determined in this study (blue). (B) Electrostatic surface potential of free SF3A1-UBL (lowest energy structure). Left panel oriented as in Fig. 1, right panel turned 180°. On its surface, SF3A1-UBL displays a very large electropositive region formed by seven lysines (Lys741, Lys754, Lys756, Lys765, Lys786, Lys792, Lys793), one histidine (His745) and two arginines (Arg788 and Arg791). (C) Representative fast to intermediate exchange residues of SF3A1-UBL upon addition of U1-SL4. Free protein (blue), protein:RNA ratio of 1:0.3 (green), 1:06 (yellow), 1:1 (red) and 1:1.5 (cyan). (D) Secondary structure predictions of SF3A1-UBL in free (upper panel) and U1-SL4 bound form (lower panel) using talos+. Blue bars indicate the β-strands, while red bars indicate α-helices. (E) ^{1}H-^{1}H-imino-NOESY of free and SF3A1-UBL bound U1-SL4 (F). (G) Amino acid sequence alignment of SF3A1-UBL domains from human, mouse, zebrafish, fly and worm generated with multalign (17) and ESPript (18). Asterisks indicate residues that are involved in RNA contacts based on the crystal structure.

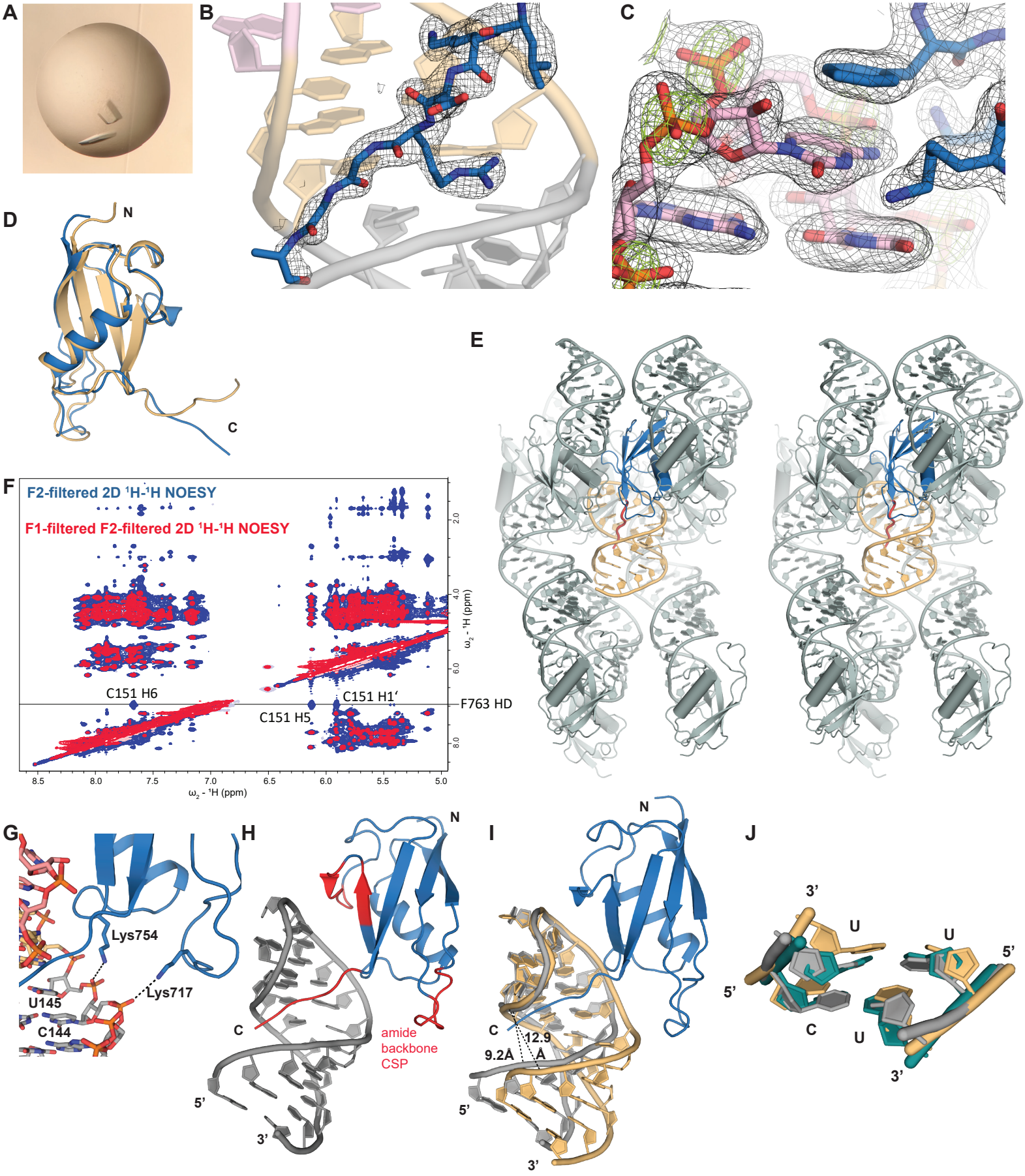
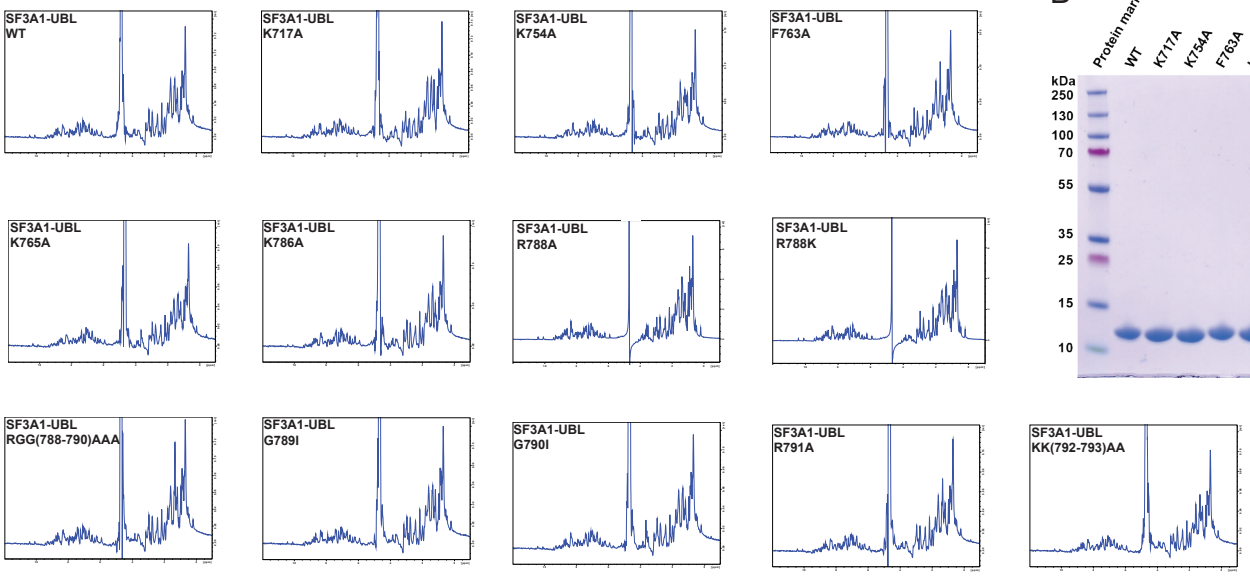
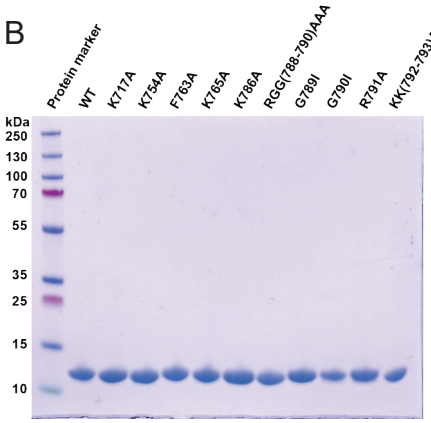

A
B
C
D
N
C
E
F
F2-filtered 2D ¹H-¹H NOESY
F1-filtered F2-filtered 2D ¹H-¹H NOESY
C151 H6
C151 H1′
C151 H5
F763 HD
ω2 - ¹H (ppm)
ω1 - ¹H (ppm)
G
Lys754
Lys717
U145
C144
H
N
C
amide
backbone
CSP
5’
3’
I
N
C
12.9
Å
9.2Å
5’
3’
J
3’
U
5’
C
U
U
5’
3’

Fig. S2. SF3A1-UBL in complex with U1-SL4. (A) Crystals of SF3A1-UBL/U1-SL4 complex. (B) Electron density ($2F_o$-F_c, black) of the SF3A1-UBL C-terminus (blue) contoured at 1.0σ. The U1-SL4 RNA is shown in cartoon representation and colored as in Fig. 2A. (C) Electron density ($2F_o$-F_c, black) of the U1-SL4 tetraloop (red) and the SF3A1-UBL interacting with this region (blue). The anomalous difference density of the native SAD experiment (light green) contoured at 4.0σ shows the location of backbone phosphorus atoms. (D) Overlay of free (yellow) and U1-SL4 bound SF3A1-UBL (blue). (E) Stereo-image of the crystal packing of SF3A1-UBL (blue) in complex with U1-SL4 (light yellow). The SF3A1-UBL C-terminus (highlighted in red) is not involved in contacts with neighbouring complex molecules (grey). (F) Detection of intermolecular NOEs. Overlay of F2-filtered (blue) with F1-filtered F2-filtered 2D ^{1}H-^{1}H NOESYs (red). (G) K717 and K754 side chains form salt bridges to the phosphate backbone of U1-SL4. (H) Amide CSP detected by NMR spectroscopy shown on SF3A1-UBL/U1-SL4 structure. (I) Comparison of free and SF3A1-bound U1-SL4. Superposition of the terminal nucleotides (146-155, RMSD 0.610) of the free (yellow) (19) and SF3A1-UBL bound U1-SL4 (grey). (J) Superposition of the tandem non-canonical 5'-CU-3'/5'-UU-3' internal loop of free U1-SL4 (yellow, PDB ID: 6QX9), SF3A1-UBL bound U1-SL4 (grey) and of a stem-loop from Poliovirus 3'-UTR (cyan, PDB ID: 1N66). The base pairing of U-U in the SF3A1-UBL bound U1-SL4 is identical to the polioviral internal loop involving U145N3-U156O2/U145O4-U156O3. The base pairing of C-U is different with two hydrogen bonds C144N4H-U157O4/C144N3-U157N3. This base pairing is more similar to a CU base pair observed elsewhere (20). Similarly as reported for the polioviral stem-loop containing a 5'-CU-3'/5'-UU-3' internal loop (21), the internal loops of free and SF3A1-UBL bound U1-SL4 show reduced C1'-C1' cross-strand distances (about 8.5-9.0 Å for both base pairs) compared to regular A-form helix conformation (about 11 Å).

A


SF3A1-UBL WT
SF3A1-UBL K717A
SF3A1-UBL K754A
SF3A1-UBL F763A
SF3A1-UBL K765A
SF3A1-UBL K786A
SF3A1-UBL R788A
SF3A1-UBL R788K
SF3A1-UBL RGG(788-790)AAA
SF3A1-UBL G789I
SF3A1-UBL G790I
SF3A1-UBL R791A
SF3A1-UBL KK(792-793)AA

B


Protein marker
WT
K717A
K754A
F763A
K765A
K786A
RGG(788-790)AAA
G789I
G790I
R791A
KK(792-793)AA
kDa
250
130
100
70
55
35
25
15
10

C

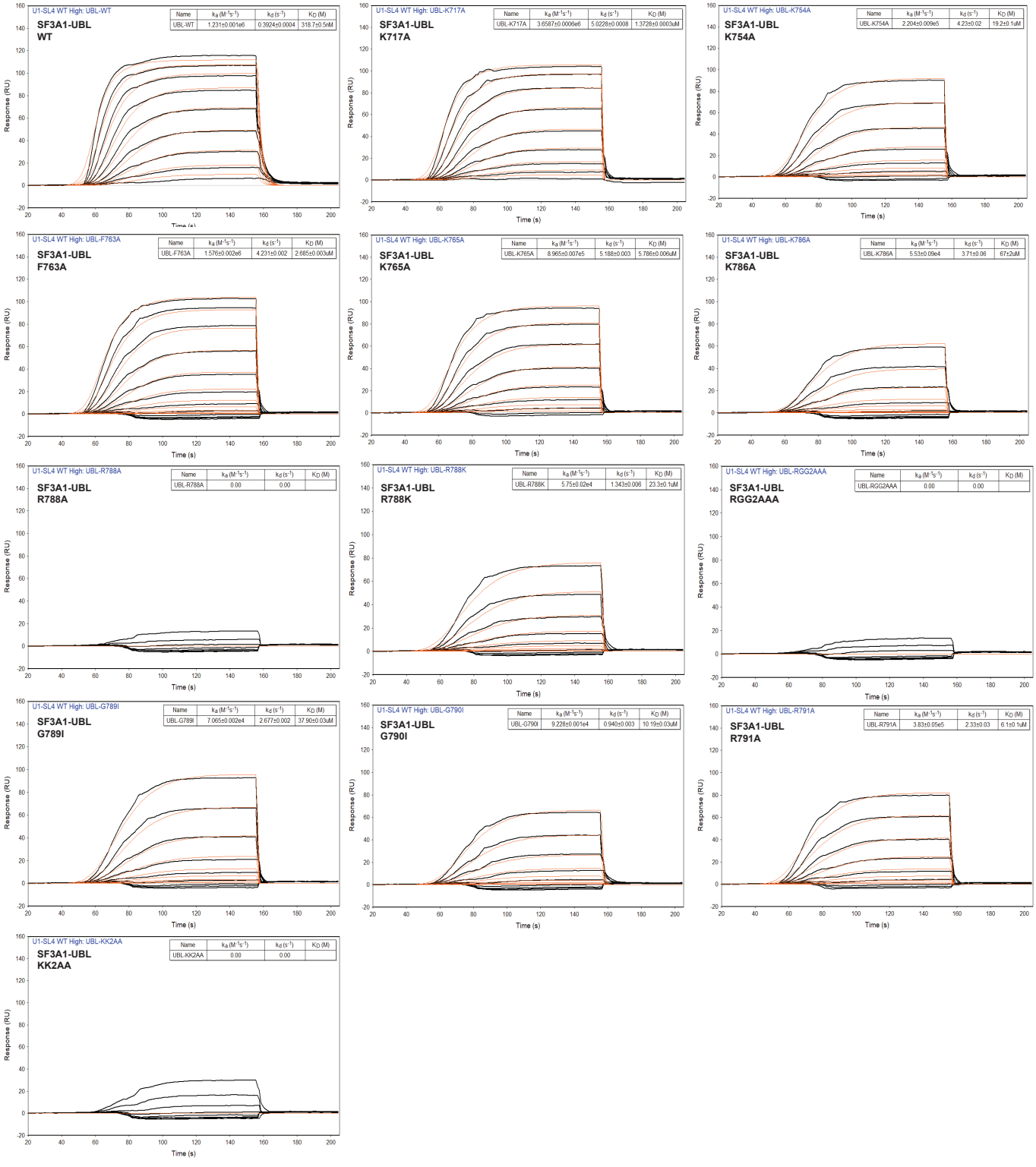

U1-SL4 WT High: UBL-WT
SF3A1-UBL WT
Name | ka (M-1s-1) | kd (s-1) | KD (M)
UBL-WT | 1.231±0.001e6 | 0.3924±0.0004 | 318.7±0.5nM
U1-SL4 WT High: UBL-K717A
SF3A1-UBL K717A
UBL-K717A | 3.6587±0.0006e6 | 5.0228±0.0008 | 1.3728±0.0003uM
U1-SL4 WT High: UBL-K754A
SF3A1-UBL K754A
UBL-K754A | 2.204±0.009e5 | 4.23±0.02 | 19.2±0.1uM
U1-SL4 WT High: UBL-F763A
SF3A1-UBL F763A
UBL-F763A | 1.576±0.002e6 | 4.231±0.002 | 2.685±0.003uM
U1-SL4 WT High: UBL-K765A
SF3A1-UBL K765A
UBL-K765A | 8.965±0.007e5 | 5.188±0.003 | 5.786±0.006uM
U1-SL4 WT High: UBL-K786A
SF3A1-UBL K786A
UBL-K786A | 5.53±0.09e4 | 3.71±0.06 | 67±2uM
U1-SL4 WT High: UBL-R788A
SF3A1-UBL R788A
UBL-R788A | 0.00 | 0.00
U1-SL4 WT High: UBL-R788K
SF3A1-UBL R788K
UBL-R788K | 5.75±0.02e4 | 1.343±0.006 | 23.3±0.1uM
U1-SL4 WT High: UBL-RGG2AAA
SF3A1-UBL RGG2AAA
UBL-RGG2AAA | 0.00 | 0.00
U1-SL4 WT High: UBL-G789I
SF3A1-UBL G789I
UBL-G789I | 7.065±0.002e4 | 2.677±0.002 | 37.90±0.03uM
U1-SL4 WT High: UBL-G790I
SF3A1-UBL G790I
UBL-G790I | 9.228±0.001e4 | 0.940±0.003 | 10.19±0.03uM
U1-SL4 WT High: UBL-R791A
SF3A1-UBL R791A
UBL-R791A | 3.83±0.05e5 | 2.33±0.03 | 6.1±0.1uM
U1-SL4 WT High: UBL-KK2AA
SF3A1-UBL KK2AA
UBL-KK2AA | 0.00 | 0.00
Response (RU)
Time (s)

Fig. S3. Mutational analyses of SF3A1-UBL mutants on U1-SL4 binding (A) Confirmation of correct folding of protein mutants by ^{1}H 1D NMR spectra by comparison to wild-type (WT). (B) Confirmation of purity of protein mutants by SDS PAGE. (C) Representative sensograms from SPR experiment #3 for measuring binding of SF3A1-UBL WT and mutants to U1-SL4. The highest protein concentration is indicated in the sensograms.

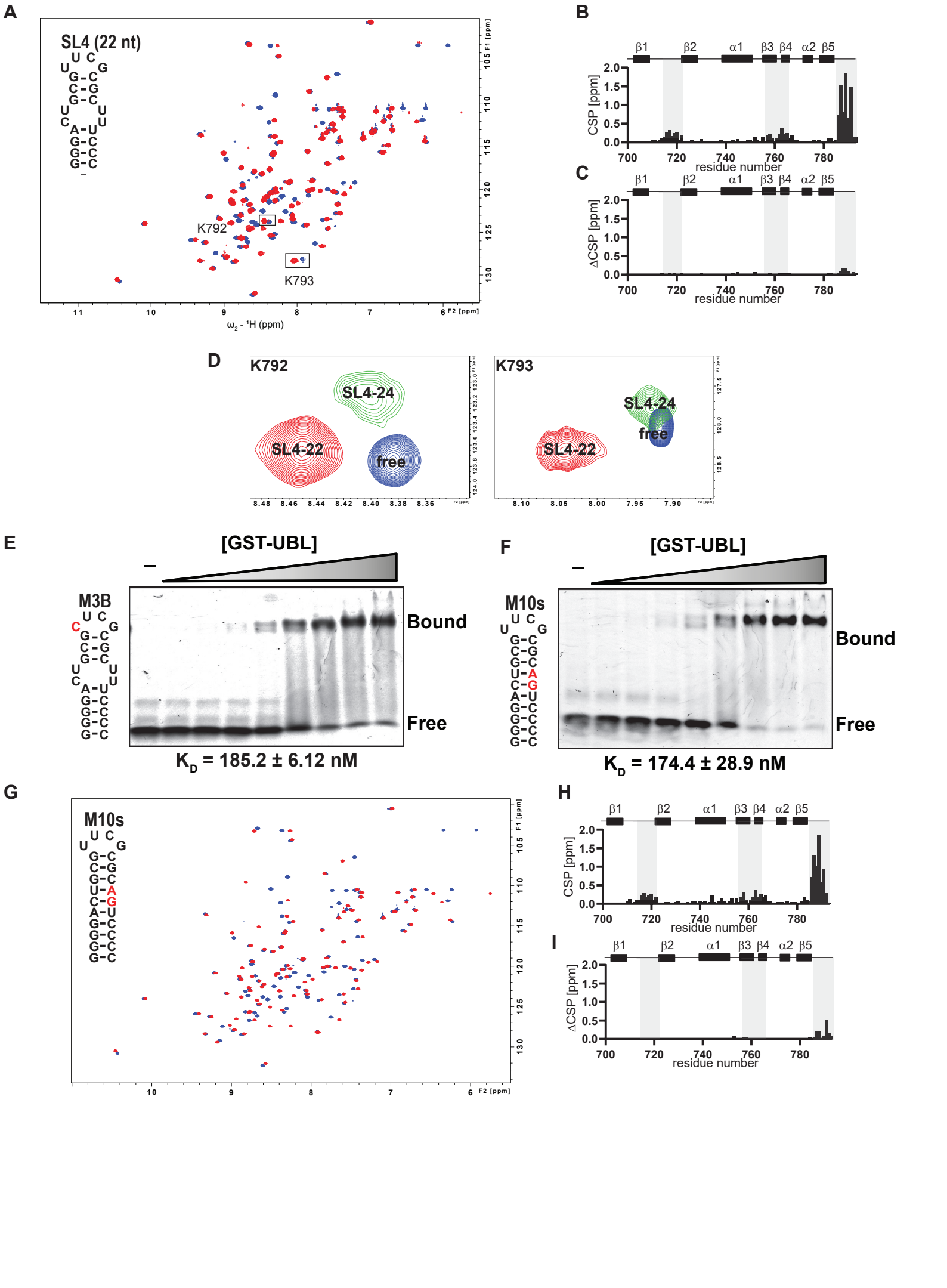

A
SL4 (22 nt)
K792
K793
ω2 - 1H (ppm)
F2 [ppm]
F1 [ppm]
B
β1 β2 α1 β3 β4 α2 β5
CSP [ppm]
residue number
C
ΔCSP [ppm]
residue number
D
K792
SL4-24
SL4-22
free
K793
SL4-24
free
SL4-22
E
[GST-UBL]
M3B
Bound
Free
K_D = 185.2 ± 6.12 nM
F
[GST-UBL]
M10s
Bound
Free
K_D = 174.4 ± 28.9 nM
G
M10s
F2 [ppm]
F1 [ppm]
H
β1 β2 α1 β3 β4 α2 β5
CSP [ppm]
residue number
I
β1 β2 α1 β3 β4 α2 β5
ΔCSP [ppm]
residue number

Fig. S4. Effect of RNA mutations on SF3A1-UBL binding. (A) Overlay of 2D ^{1}H-^{15}N HSQC spectra of free ^{15}N-labeled SF3A1-UBL (blue) and bound to a U1-SL4 construct lacking the bottom G-C base pair (22 nt U1-SL4) in a 1:1 complex (red). (B) Combined chemical shift differences of SF3A1-UBL with U1-SL4 construct lacking the bottom G-C base pair (22 nt U1-SL4) compared to its free form. (C) Differences in combined amide CSP (ΔCSP) of SF3A1-UBL in complex with 22 nt U1-SL4 as compared to WT U1-SL4 (24 nt). (D) Lys792 and Lys793 show altered amide CSP upon addition of 22 nt U1-SL4 compared to WT U1-SL4. (E and F) EMSA with U1-SL4 mutants (M3B and M10s). (G) Overlay of 2D ^{1}H-^{15}N HSQC spectra of free ^{15}N-labeled SF3A1-UBL (blue) and bound to U1-SL4 mutant M10s in a 1:1 complex (red). (H) Combined chemical shift differences of SF3A1-UBL with U1-SL4 mutant M10s compared to its free form. (I) Differences in combined amide CSP (ΔCSP) of SF3A1-UBL in complex with U1-SL4 mutant M10s as compared to WT U1-SL4 (24 nt).

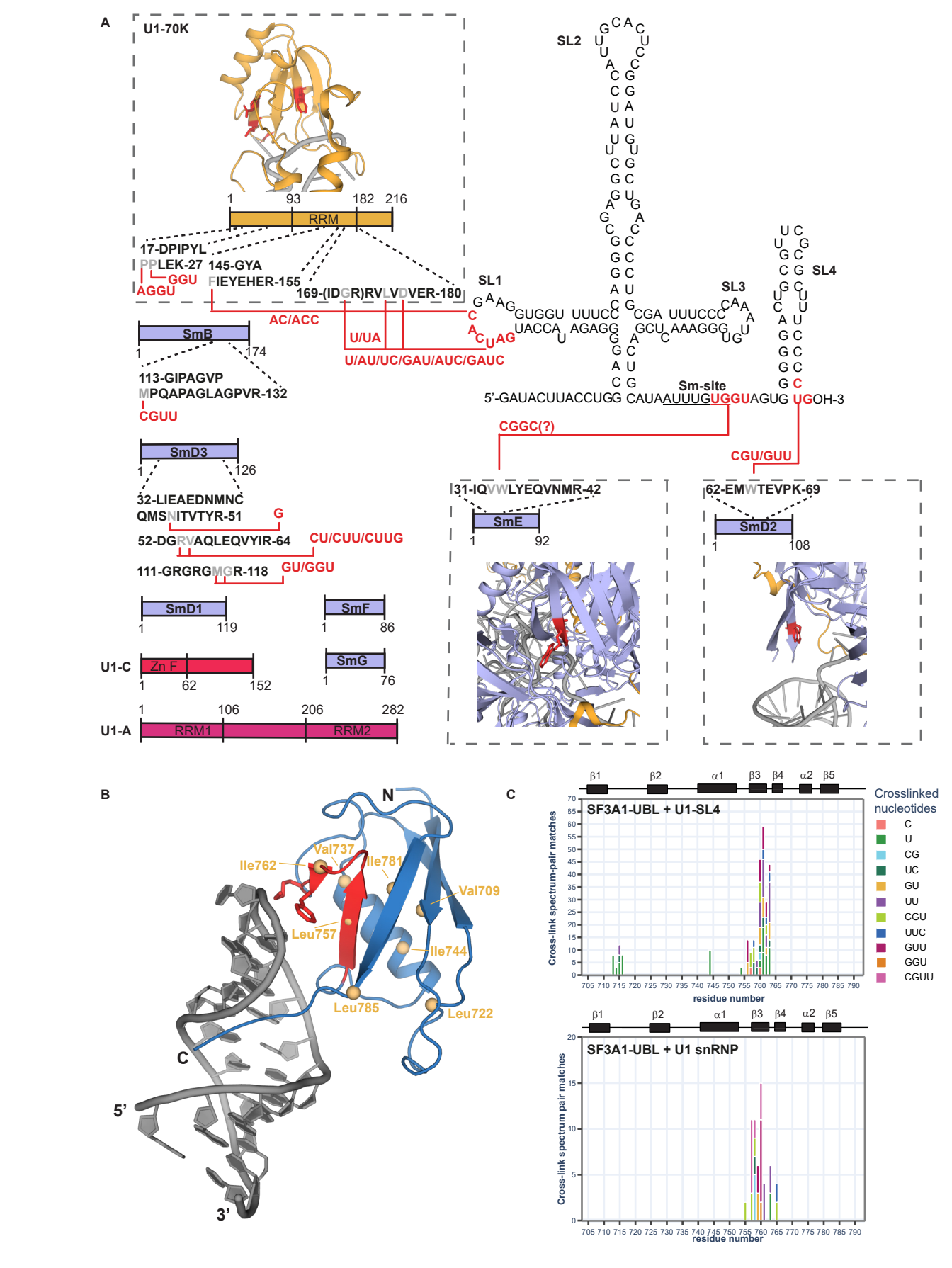
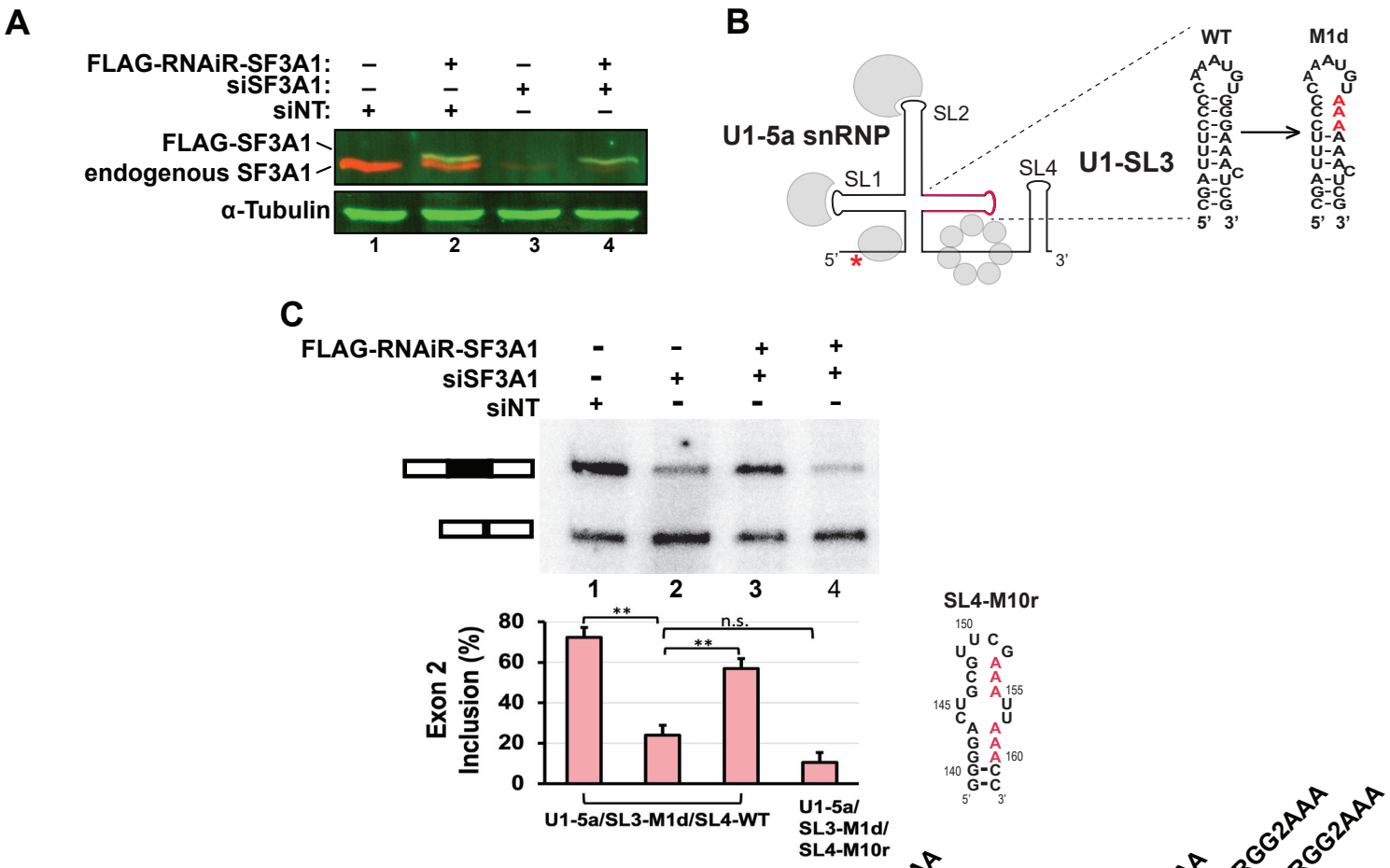
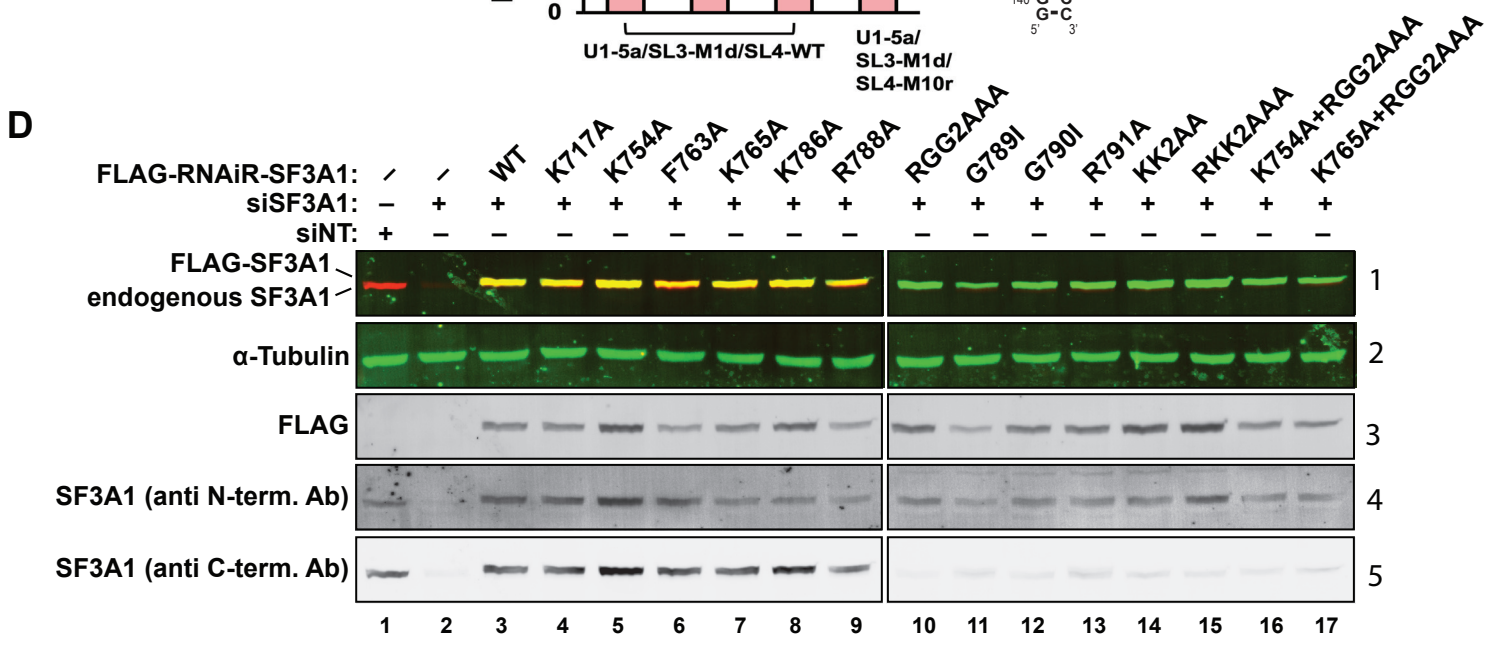
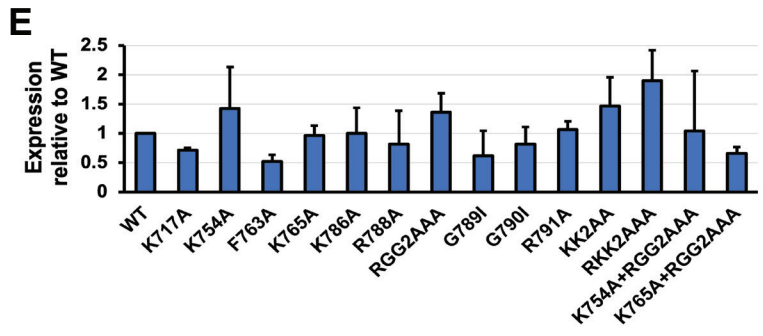
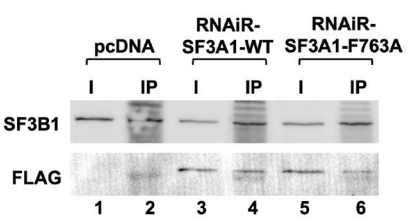

A
U1-70K
1
93
182
216
RRM
17-DPIPYL
PPLEK-27
145-GYA
GGU
AGGU
FIEYEHER-155
169-(IDGR)RVLVDVER-180
AC/ACC
U/UA
U/AU/UC/GAU/AUC/GAUC
SmB
1
174
113-GIPAGVP
MPQAPAGLAGPVR-132
CGUU
SmD3
1
126
32-LIEAEDNMNC
QMSNITVTYR-51
G
52-DGRVAQLEQVYIR-64
CU/CUU/CUUG
111-GRGRGMGR-118
GU/GGU
SmD1
1
119
SmF
1
86
U1-C
Zn F
1
62
152
SmG
1
76
U1-A
1
106
206
282
RRM1
RRM2
SL1
SL2
SL3
SL4
5'-GAUACUUACCUGG CAUAAUUUGUGGUAGUG UGOH-3
Sm-site
CGGC(?)
CGU/GUU
31-IQVWLYEQVNMR-42
SmE
1
92
62-EMWTEVPK-69
SmD2
1
108
B
N
C
5'
3'
Val737
Ile762
Ile781
Val709
Leu757
Ile744
Leu785
Leu722
C
β1
β2
α1
β3
β4
α2
β5
SF3A1-UBL + U1-SL4
Cross-link spectrum-pair matches
residue number
SF3A1-UBL + U1 snRNP
Crosslinked nucleotides
C
U
CG
UC
GU
UU
CGU
UUC
GUU
GGU
CGUU

Fig. S5. SF3A1-UBL in complex with U1 snRNP. (A) CLIR-MS/MS analysis identified protein-RNA cross-links for five of ten U1 snRNP proteins of which U1-70k, SmE and SmD2 could be mapped on U1 snRNA based on the structure of U1 snRNP. Protein-RNA XL are shown on the U1 snRNP structure (19). (B) Protein-RNA XL identified for SF3A1-UBL bound to U1 snRNP (red) and the strongest CSP shown on the crystal structure of the complex (yellow spheres). (C) All protein-RNA cross-links (ld.score > 20) for SF3A1-UBL in complex with U1-SL4 (upper panel) and U1 snRNP (lower panel) plotted on the sequence of SF3A1-UBL.


A
FLAG-RNAiR-SF3A1:
siSF3A1:
siNT:
FLAG-SF3A1
endogenous SF3A1
α-Tubulin
1 2 3 4
B
U1-5a snRNP
SL1
SL2
SL4
U1-SL3
WT
M1d
5' *
3'
C
FLAG-RNAiR-SF3A1
siSF3A1
siNT
1 2 3 4
Exon 2 Inclusion (%)
**
n.s.
**
U1-5a/SL3-M1d/SL4-WT
U1-5a/ SL3-M1d/ SL4-M10r
SL4-M10r


D
FLAG-RNAiR-SF3A1:
siSF3A1:
siNT:
WT
K717A
K754A
F763A
K765A
K786A
R788A
RGG2AAA
G789I
G790I
R791A
KK2AA
RKK2AAA
K754A+RGG2AAA
K765A+RGG2AAA
FLAG-SF3A1
endogenous SF3A1
α-Tubulin
FLAG
SF3A1 (anti N-term. Ab)
SF3A1 (anti C-term. Ab)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17


E
Expression relative to WT
WT
K717A
K754A
F763A
K765A
K786A
R788A
RGG2AAA
G789I
G790I
R791A
KK2AA
RKK2AAA
K754A+RGG2AAA
K765A+RGG2AAA


F
pcDNA
RNAiR-SF3A1-WT
RNAiR-SF3A1-F763A
I IP I IP I IP
SF3B1
FLAG
1 2 3 4 5 6

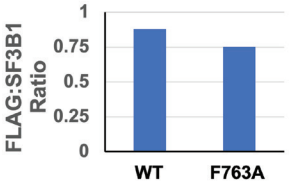

G
FLAG:SF3B1 Ratio
WT
F763A

Fig. S6. Expression of siRNA resistant SF3A1 mutants. (A) An siRNA resistant, FLAG-tagged SF3A1 clone (FLAG-RNAiR-SF3A1) can be expressed under endogenous SF3A1 knockdown conditions in HeLa cells. Western blotting of transfected HeLa cells demonstrates efficient expression of FLAG-RNAiR-SF3A1 under non-targeting control siRNA (siNT) treatment (lane 2) or when cells are treated with endogenous SF3A1 targeting siRNA (siSF3A1, lane 4). (B) Schematic of the U1 snRNA showing the mutations in SL3 used in reporter assays. (C) Primer extension analysis to monitor the inclusion of exon 2 in RNA isoforms of the Dup51p minigene reporter. The mRNA products are shown schematically to the left of the gel image. HeLa cells were transfected with siNT or siSF3A1 and a plasmid harboring WT or mutant FLAG-RNAiR-SF3A1. Knockdown of SF3A1 by siSF3A1 treatment reduces exon 2 inclusion, whereas siNT treatment does not have any effect (lanes 1 and 2). Expression of the RNAi resistant SF3A1 WT clone rescues exon 2 inclusion when the Dup51p reporter is co-transfected with a U1-5a harboring SL4-WT but not with an SL4 mutant M10r (a scheme is shown on the right). Percent exon 2 inclusion (n =3; * = $p< 0.05$, ** = $p< 0.01$, *** = $p< 0.001$, n.s. = not significant) is plotted below the gel. (D) WT and mutant FLAG-RNAiR-SF3A1 protein expression in transfected HeLa cells. Several point mutations were introduced into the FLAG-RNAiR-SF3A1 clone and Western blotting confirmed the expression of all proteins after 48-hours co-transfection with siSF3A1. A merged image of immunoblots probed with anti-FLAG antibody and anti-SF3A1 C-terminus targeting antibody (anti C-Term. Ab) is shown in the top panel. Endogenous SF3A1 (red) is efficiently silenced and many of the mutants are detectable using both antibodies (yellow). Point mutations introduced closer to the C-terminus of SF3A1 interfere with SF3A1 recognition by anti C-term. Ab, but are detectable with anti-FLAG antibody (green). Western blots were stripped and re-probed with N-terminus targeting SF3A1 antibody (anti N-term. Ab) and confirm that all mutant FLAG-RNAiR-SF3A1 proteins are detected using an SF3A1-specific antibody. The individual scans from each antibody used in the merged image is shown in the lower panels and α-Tubulin was used as a loading control. (E) Expression of FLAG-tagged SF3A1 mutants was compared to WT in siSF3A1 treated cells. Mutant protein expression (± s.d., n = 3) relative to WT after normalization to α-Tubulin is shown. (F) Immunoprecipitation experiments with anti-SF3B1 antibody followed by immunoblotting with anti-FLAG antibody demonstrates that the exogenously expressed N-terminal FLAG-tagged SF3A1 are assembled into U2 snRNPs. (G) Quantification of the incorporation of FLAG-tagged SF3A1 by comparison to SF3B1. Ratio of FLAG to SF3B1 in the immunoprecipitated complexes (IP) was calculated after normalization to input (I).

Table S1. NMR and refinement statistics for free SF3A1-UBL protein structure.

	SF3A1-UBL
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	3,890
Intra-residue	685
Inter-residue	
Sequential ($\|i-j\|=1$)	914
Medium-range ($\|i-j\|<4$)	819
Long-range ($\|i-j\|>5$)	1443
Hydrogen bonds	29
Total dihedral angle restraints	
ϕ	75
ψ	75
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (> 0.4 Å) (Å)	2.7 +/- 1.9
Dihedral angle constraints (> 5 °) (º)	3.4 +/- 10.7
Max. dihedral angle violation (º)	8.37 +/- 4.83
Max. distance constraint violation (Å)	0.57 +/- 0.30
Deviations from idealized geometry	
Bond lengths (Å)	0.0033
Bond angles (º)	1.347
Average pairwise r.m.s. deviation[a] (Å)	
Heavy	0.29 +/- 0.07 [a]
Backbone	0.09 +/- 0.02 [a]
Ramachandran statistics[b]	
Most favored regions	91.8 %
Additionally favored regions	8.1 %
Generously allowed regions	0.1 %
Disallowed regions	0.0 %
Clashscore[c]	8

[a] Pairwise RMSD calculated among the 20 NMR structures using residues 704-785.
[b] Ramachandran statistics determined using PROCHECK (22).
[c] Clashscore calculated by the PDB validation service.

Table S2. Dissociation constants obtained from SPR experiments.

SF3A1-UBL construct	K_d (μM)		
	Experiment 1	**Experiment 2**	**Experiment 3**
WT	0.398±0.0005	0.279±0.0008	0.3187±0.0006
K717A	2.39±0.03	0.994±0.0007	1.3728±0.0003
K754A	18.0±0.4	13.0±0.1	19.2±0.1
F763A	3.34±0.06	2.267±0.007	2.685±0.003
K765A	13.1±0.1	4.78±0.01	5.786±0.006
K786A	> 20	> 50	> 50
R788A	n.d.	> 25	> 25
R788K	n.d.	> 25	23.3±0.1
RGG2AAA	> 20	> 50	> 50
G789I	37.9±0.8	45.9±0.8	37.90±0.03
G790I	17.5±0.2	14.6±0.3	10.19±0.03
R791A	6.41±0.06	5.39±0.04	6.1±0.1
KK2AA	> 20	> 50	> 50

n.d. – not determined

Table S3. Primers used for site-directed mutagenesis of SF3A1-UBL.

SF3A1-UBL mutant	Forward Primer (5′→3′)	Reverse Primer (5′→3′)
K717A	CAG GAT GCG ACG GAA TGG AAA CTG AAT GGG	CCG TCG CAT CCT GCA TGT TGG GCA CCT G
K754A	GCA GGG GCG CAG AAG CTA CAG TAT GAG GGT ATC	CTT CTG CGC CCC TGC AGG CAT GCC TGT GGC TTC
F763A	GGT ATC GCG ATC AAA GAT TCC AAC TCA CTG GCT	CTT TGA TCG CGA TAC CCT CAT ACT GTA GCT TCT G
K765A	CAT CGC GGA TTC CAA CTC ACT GGC TTA CTA C	GAA TCC GCG ATG AAG ATA CCC TCA TAC TGT AGC
K786A	CCC TCG CGG AGA GAG GCG GGA GGA AGA AG	TCT CTC CGC GAG GGC CAG GTG GAT GAC TGC
R788A	CAA GGA GGC GGG CGG GAG GAA GAA GTA GCT CGA G	CGC CCG CCT CCT TGA GGG CCA GGT GGA TG
R788K	C AAG GAG AAA GGC GGG AGG AAG AAG TAG CTC	C GCC TTT CTC CTT GAG GGC CAG GTG GAT G
RGG2AAA	CTC AAG GAG GCG GCT GCA AGG AAG AAG TAG CTC GAG CAC CAC CAC CAC CAC CAC CAC TGA G	CTT CCT TGC AGC CGC CTC CTT GAG GGC CAG GTG GAT GAC TGC GCC ATT GGC CAT GTT G
G789I	GAG AGA ATT GGG AGG AAG AAG TAG CTC GAG CAC	CTC CCA ATT CTC TCC TTG AGG GCC AGG TGG
G790I	AGA GGC ATT AGG AAG AAG TAG CTC GAG CAC CAC	CTT CCT AAT GCC TCT CTC CTT GAG GGC CAG G
R791A	GGC GGG GCG AAG AAG TAG CTC GAG CAC CAC	CTT CTT CGC CCC GCC TCT CTC CTT GAG GGC
KK2AA	GGG AGG GCG GCG TAG CTC GAG CAC CAC CAC CAC CAC CAC	GCT ACG CCG CCC TCC CGC CTC TCT CCT TGA GGG CC

Table S4. Sequence of primers used for SF3A1 and U1-5a snRNA mutagenesis.

Final Clone Name	Template Plasmid	Forward Primer	Reverse Primer
pcDNA3.1:FLAG-RNAiR-SF3A1-WT	pcDNA3.1:FLAG-SF3A1-WT	GCcCCgTCcAAGCCAGTTGTGGGGATTATTTAC	cGAgTCtTCCTTTGAAGATGCTTCTTCTTCTGTGG
pcDNA3.1:FLAG-RNAiR-SF3A1 K717A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCGACGGAATGGAAACTGAATGGGC	ATCCTGCATGTTGGGCACCTGGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 K754A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCACAGAAGCTACAGTATGAGGGTATCTTCATC	CCCTGCAGGCATGCCTGTGGCTTC
pcDNA3.1:FLAG-RNAiR-SF3A1 F763A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	CATCAAAGATTCCAACTCACTGGCTTACTAC	GCGATACCCTCATACTGTAGCTTCTGTTTCC
pcDNA3.1:FLAG-RNAiR-SF3A1 K765A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCAGATTCCAACTCACTGGCTTACTACAAC	GATGAAGATACCCTCATACTGTAGCTTCTG
pcDNA3.1:FLAG-RNAiR-SF3A1 K786A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCGGAGAGAGGCGGGAGGAAGAAGTAGG	GAGGGCCAGGTGGATGACTGCGCCATTG
pcDNA3.1:FLAG-RNAiR-SF3A1 R788A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GGCGGGAGGAAGAAGTAGGCGGCCG	TGCCTCCTTGAGGGCCAGGTGGATGACTGC
pcDNA3.1:FLAG-RNAiR-SF3A1 RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTAGGAAGAAGTAGGCGGCCGCTC	TGCCTCCTTGAGGGCCAGGTGGATGACTGC
pcDNA3.1:FLAG-RNAiR-SF3A1 G789I	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	ATCGGGAGGAAGAAGTAGGCGGCCGCTC	TCTCTCCTTGAGGGCCAGGTGGATGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 G790I	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GGCATCAGGAAGAAGTAGGCGGCCGCTC	TCTCTCCTTGAGGGCCAGGTGGATGAC
pcDNA3.1:FLAG-RNAiR-SF3A1 R791A	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCCAAGAAGTAGGCGGCCGCTCGAGTCTAGAG	CCCGCCTCTCTCCTTGAGGGCCAG
pcDNA3.1:FLAG-RNAiR-SF3A1 KK2AA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTTAGGCGGCCGCTCGAGTCTAGAGG	CCTCCCGCCTCTCTCCTTGAGGG
pcDNA3.1:FLAG-RNAiR-SF3A1 RKK2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTGCTGCTTAGGCGGCCGCTCGAGTCTAGAG	CCCGCCTCTCTCCTTGAGGGCCAG
pcDNA3.1:FLAG-RNAiR-K754A+RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-RGG2AAA	GCACAGAAGCTACAGTATGAGGGTATCTTCATC	CCCTGCAGGCATGCCTGTGGCTTC
pcDNA3.1:FLAG-RNAiR-K765A+RGG2AAA	pcDNA3.1:FLAG-RNAiR-SF3A1-RGG2AAA	GCAGATTCCAACTCACTGGCTTACTACAAC	GATGAAGATACCCTCATACTGTAGCTTCTG
pcDNA3.1:FLAG-RNAiR-SF3A1 Y772C	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCTACAACATGGCCAATGG	AAGCCAGTGAGTTGGAATC
pcDNA3.1:FLAG-RNAiR-SF3A1 Y773C	pcDNA3.1:FLAG-RNAiR-SF3A1-WT	GCAACATGGCCAATGGCGC	AGTAAGCCAGTGAGTTGGAATCTTTG

pNS6:U1-5a/SL3-M1d	pNS6:U1-5a/WT	GTAAAAAACTCGACTGCATAATTTGTGGTAGTG	ATTTGGGGAAATCGCAGGGGTCAGCACATCCGGAG

Table S5. Miscellaneous sequences used for SPR, EMSA, and primer extension.

Oligo name	Sequence (5′→3′)	Technique
Dup3r	AACAGCATCAGGAGTGGACAGATCCC	Primer extension
siSF3A1	GGAGGAUUCUGCACCUUCUUU	RNAi
Biotin-U1-SL4-WT	Bi-GGGGACUGCGUUCGCGCUUUCCCC	SPR
Biotin-U1-SL4-M10b	Bi-GGGGACUUAUUUCGAUAUUUCCCC	SPR
Cy5-U1-SL4-WT	Cy5-GGGGACUGCGUUCGCGCUUUCCCC	EMSA
Cy5-U1-SL4-M3A	Cy5-GGGGACUGCGGAAACGCUUUCCCC	EMSA
Cy5-U1-SL4-M3B	Cy5-GGGGACUGCGCUCGCGCUUUCCCC	EMSA
Cy5-U1-SL4-M10d	Cy5-GGGGACUCGCUUCGGCGUUUCCCC	EMSA
Cy5-U1-SL4-M10b	Cy5-GGGGACUUAUUUCGAUAUUUCCCC	EMSA
Cy5-U1-SL4-M10b-2	Cy5-GGGGACUAUAUUCGUAUUUUCCCC	EMSA
Cy5-U1-SL4-M10s	Cy5-GGGGACUGCGUUCGCGCAGUCCCC	EMSA

SI References

1. S. Campagne, *et al.*, An in vitro reconstituted U1 snRNP allows the study of the disordered regions of the particle and the interactions with proteins and ligands. *Nucleic Acids Res.*, 1–13 (2021).
2. V. Olieric, *et al.*, Data-collection strategy for challenging native SAD phasing. *Acta Crystallogr. Sect. D* **72**, 421–429 (2016).
3. W. Kabsch, XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–132 (2010).
4. F. H. T. Allain, G. Varani, Structure of the P1 helix from group I self-splicing introns. *J. Mol. Biol.* **250**, 333–353 (1995).
5. A. J. McCoy, *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
6. T. C. Terwilliger, Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr. D. Biol. Crystallogr.* **57**, 1755–1762 (2001).
7. P. Emsley, K. Cowtan, Coot: Model-building tools for molecular graphics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
8. P. D. Adams, *et al.*, PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
9. G. Dorn, *et al.*, Structural modeling of protein-RNA complexes using crosslinking of segmentally isotope-labeled RNA and MS/MS. *Nat. Methods* **14**, 487–490 (2017).
10. J. Rappsilber, Y. Ishihama, M. Mann, Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
11. T. Walzthoeni, *et al.*, False discovery rate estimation for cross-linked peptides identified by mass spectrometry. *Nat. Methods* **9**, 901–903 (2012).
12. W. Martelly, B. Fellows, K. Senior, T. Marlowe, S. Sharma, Identification of a noncanonical RNA binding domain in the U2 snRNP protein SF3A1. *Rna* **25**, 1509–1521 (2019).
13. R. Fabrini, *et al.*, Monomer-dimer equilibrium in glutathione transferases: A critical re-examination. *Biochemistry* **48**, 10473–10482 (2009).
14. S. Sharma, S. P. Wongpalee, A. Vashisht, J. A. Wohlschlegel, D. L. Black, Stem–loop 4 of U1 snRNA is essential for splicing and interacts with the U2 snRNP-specific SF3A1 protein during spliceosome assembly. *Genes Dev.* **28**, 2518–2531 (2014).
15. R. V Pisal, *et al.*, Detection of Mycoplasma Contamination Directly from Culture Supernatant Using Polymerase Chain Reaction. *Folia Biol. (Praha).* **62**, 203–206 (2016).
16. G. Tanackovic, A. Krämer, Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo. *Mol. Biol. Cell* **16**, 1366–1377 (2005).
17. F. Corpet, Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890 (1988).
18. X. Robert, P. Gouet, Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.* **42**, 320–324 (2014).
19. C. Charenton, M. E. Wilkinson, K. Nagai, Mechanism of 5' splice site transfer for human spliceosome activation. *Science (80-.).* **364**, 362–367 (2019).
20. O. Ohlenschläger, *et al.*, The Structure of the Stemloop D Subdomain of Coxsackievirus B3 Cloverleaf RNA and Its Interaction with the Proteinase 3C. *Structure* **12**, 237–248 (2004).
21. E. M. H. P. Lescrinier, *et al.*, Structure of the pyrimidine-rich internal loop in the poliovirus 3′-UTR: The importance of maintaining pseudo-2-fold symmetry in RNA helices containing two adjacent non-canonical base-pairs. *J. Mol. Biol.* **331**, 759–769 (2003).
22. R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, J. M. Thornton, AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477–486 (1996).