SUPPLEMENTAL FIGURES

Identification of a non-canonical RNA binding domain in the U2 snRNP protein SF3A1

William Martelly[1, 3], Bernice Fellows[1], Kristen Senior[1], Tim Marlowe[2] and Shalini Sharma[1,*]

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

[2]Molecular Analysis Core, University of Arizona, College of Medicine-Phoenix, Phoenix, Arizona 85004, USA

[3]School of Life Sciences, Arizona State University, Tempe, Arizona 85287, USA

*To whom correspondence should be addressed: Tel: 1-602-827-2149; Fax: 1-602-827-2127; Email: shalinijs@email.arizona.edu.

Present Address: Shalini Sharma, Department of Basic Medical Sciences, University of Arizona, College of Medicine-Phoenix, Phoenix, Arizona 85004, USA

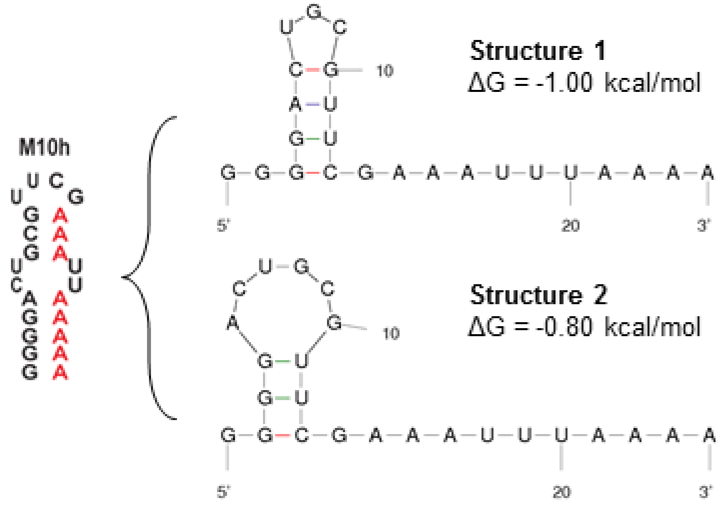

Figure S1. Secondary structure prediction of U1-SL4 M10h mutant. The two lowest free-energy (ΔG) structures for the M10h RNA as predicted by mfold are shown. The non-specific binding observed for this ligand in competitive EMSA experiments may be explained by the capacity to form short stem-loops *in vitro*.

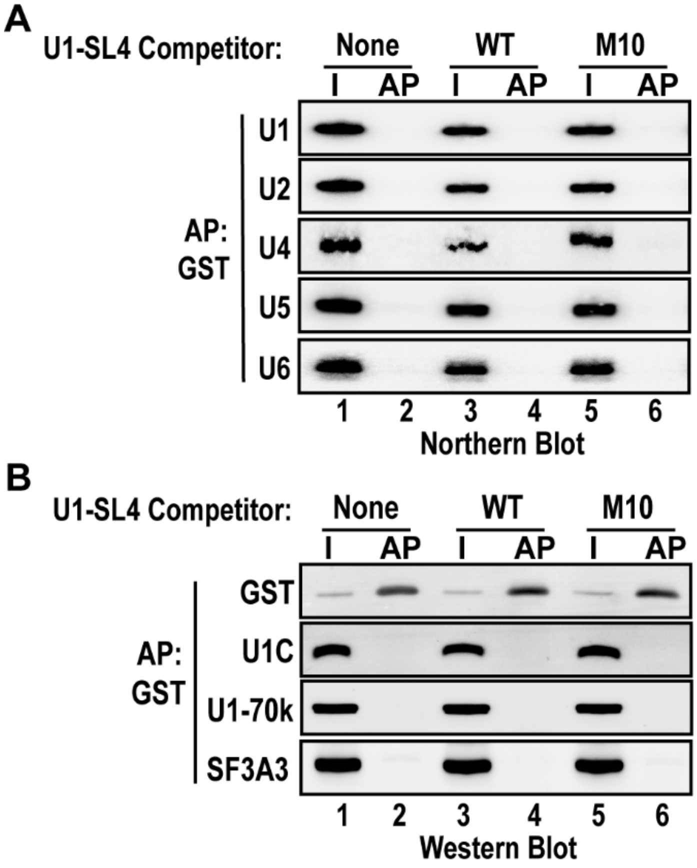

Figure S2. Analysis of U1 proteins and major spliceosomal RNAs present in control GST pull-down complexes. (A) Northern blotting of U1, U2, U4, U5, and U6 snRNA show no enrichment of any snRNAs in GST alone affinity-purification fractions. (B) Immunoblotting of proteins present in the GST affinity pull-down complexes demonstrating the lack of enrichment of U1C, U1-70k or SF3A3. I and AP indicate input and affinity pull-down complexes, respectively.

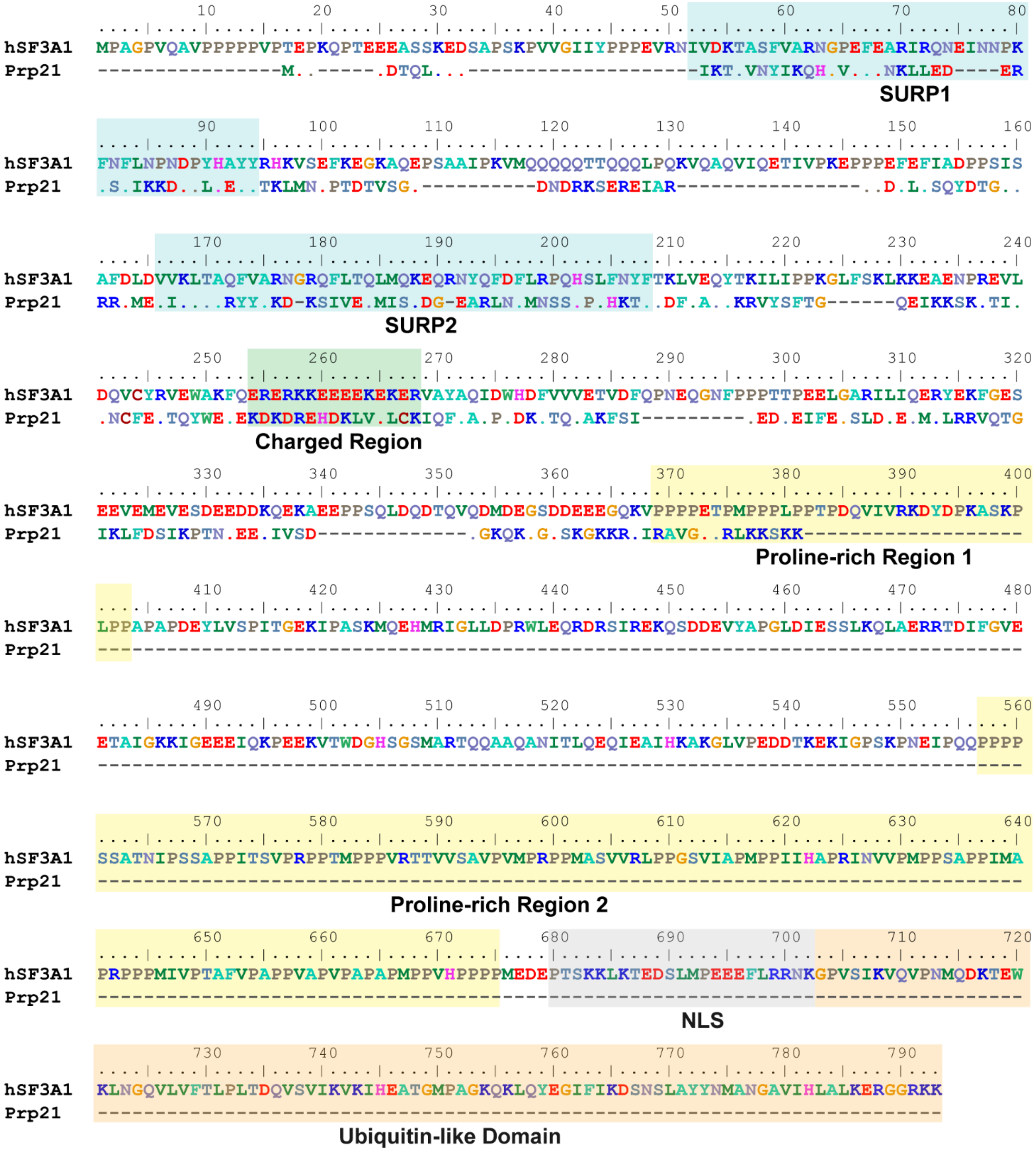

Figure S3. Alignment of full-length SF3A1 protein from human and yeast. SF3A1 from *Saccharomyces cerevisiae* (Prp21; UniProt ID: P32524) was aligned to human SF3A1 (hSF3A1; UniProt ID: Q15459) using ClustalW. The N-terminal features of SF3A1 are conserved between yeast and humans; including the two SURP domains and the small region of charged residues. Prp21 lacks the large proline rich regions, nuclear localization signal, and the UBL domain. In the aligned sequences, a dot indicates the presence of a residue identical to reference (hSF3A1) and a dash indicates a gap.

Table S1. Identity matrix comparing SF3A1 UBL domains in higher eukaryotes.

Identity Matrix	H. sapiens	M. musculus	D. rerio	D. melanogaster	C. elegans
H. sapiens	**1.0**	**0.978**	**0.934**	**0.472**	**0.373**
M. musculus		**1.0**	**0.912**	**0.461**	**0.373**
D. rerio			**1.0**	**0.450**	**0.351**
D. melanogaster				**1.0**	**0.406**
C. elegans					**1.0**

Table S2. Accession numbers of UBL domain containing proteins with SF3A1-like features.

Protein Name	UBL Family	UniProt Accession
SF3A1	UBL	Q15459
MAP2K5	PB1	Q13163
PARD6B	PB1	Q9BYG5
UBIQUITIN	UB	P0CG47
RPS27A	NEDD8	P62979
NEDD8	NEDD8	Q15843
ZFAND4	UBL	Q86XD8
URM1	UBL	Q9BTM9
UBD UBL-2	UBL	O15205
OASL UBL-2	UBL	Q15646
DDI1	UBL	Q8WTU0
DDI2	UBL	Q5TDH0
SQSTM1	PB1	Q13501
SNRNP25	UBL	Q9BV90
UHRF1	UBL	Q96T88
RAD23A	PIM	P54725
RAD23B	PIM	P54727
UBL5	UBL	Q9BZL1
PRKCI	PB1	P41743
UBD UBL-1	UBL	O15205
OASL UBL-1	UBL	Q15646
UBL7	UBL	Q96S82
UHRF2	PIM	Q96PU4
UBL4B	PIM	Q8N7F7
BAG6	PIM	A0A0G2JK23
MAP1LC3A	ATG8	Q9H492
MAP1LC3B	ATG8	Q9GZQ8
MAP1LC3B2	ATG8	A6NCE7
MAP1LC3C	ATG8	Q9BXW4
GABARAP	ATG8	O95166
GABARAPL1	ATG8	Q9H0R8
GABARAPL2	ATG8	P60520
PCGF1	RAWUL	Q9BSM1

TBK1	UBL	Q9UHD2
MIDN	UBL	Q504T8
PARK2	PIM	O60260
UBAC1	PIM	Q9BSL1
PARD6A	PB1	Q9NPB6
BAG1	PIM	Q99933
NUB1	PIM	Q9Y5A7
TDP-43	UBL	Q13148
SUMO-1	UBL	P63165

Table S3. Sequences of oligonucleotide probes used for Northern blotting.

Oligo Name	Sequence (5′→3′)
U1$_{27-46}$R	TGATCACGAAGGTGGTTTTC
U2$_{114-135}$R	AGATGGAATAGGAGCTTGCTCC
U4$_{86-106}$R	CCGTGACGACTTGCAATATAG
U5$_{50-68}$R	GATTTCCGTGGAGAGGAAC
U6$_{36-57}$R	ACGATACAGAGAAGATTAGCA