PNAS

Supporting Information for

DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer

Jian Zhang[a], Ji Huang[a], Ke Xu[a], Peiqi Xing[b], Yue Huang[b,c], Zhaoqi Liu[b,c], Liang Tong[a], and James L. Manley[a,*]

[a]Department of Biological Sciences, Columbia University, New York, NY 10027
[b]Chinese Academy of Sciences Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China
[c]University of Chinese Academy of Sciences, Beijing 100049, China

*Corresponding author: James L. Manley
Email: jlm2@columbia.edu

This PDF file includes:

Supporting Information Text
Figures S1 to S3
Table S1
SI References

Other supporting materials for this article include the following:

Datasets S1 and S2

Supporting Information Text

Supplemental Materials and Methods

Affinity Purification Using Two Tags. The small-scale protocol for two rounds of affinity purification using FLAG and His6 tags was previously described (1). The small-scale two-round affinity purification using FLAG and GST tags was carried out as follows. HEK293T cells were seeded in 10-cm plates with 1.5 million cells per plate. On the next day, 4 μg of expression plasmid DNA (or a mixture of two plasmids with 2 μg each for co-expression) were transfected to each plate of cells using Lipofectamine 2000 (Thermo Fisher Scientific). At 24 h post transfection, cells were trypsinized and re-seeded evenly in four 10-cm plates. After 48 h of growth, cells were harvested by scraping and then lysed with 1.3 mL of a lysis buffer containing 30 mM Tris-Cl (pH 7.4), 300 mM NaCl (except that in the experiment involving co-expression of His6-FLAG tagged SF3B1 and GST-tagged SUGP1, we used 150 mM NaCl instead), 1 mM EDTA, 0.5% Triton X-100, 10 mM Sodium Orthovanadate, 10 mM Sodium Fluoride, protease inhibitor cocktail (Roche), and PhosSTOP (Roche), in the presence of 200 μg RNase A. After incubation at 4°C for 20 min and centrifugation at 21,130 × g for 10 min, cell extracts (1.2 mL of the supernatant) were incubated with 10 μL (0.5 μg/μL) anti-DYKDDDDK antibody (GenScript, A00187) at 4 °C for 30 min, followed by incubation at 4 °C for 3 h with 50 μL Pierce ProteinA/G Magnetic Beads (Thermo Fisher Scientific) and precipitation by a magnetic stand. After four washes with a wash buffer (30 mM Tris-Cl, pH 7.4, 150 mM NaCl, and 0.5% Triton X-100), proteins were eluted from the magnetic beads using four iterations of elution, each by incubation at 4 °C for 30 min with 300 mL Elution Buffer (30 mM Tris-Cl, pH 7.4, 150 mM NaCl, 0.5% Triton X-100, and 150 ng/μL 3X FLAG peptide). The four eluates were then combined (1.2 mL in total) and subjected to a second-round of affinity purification by incubation at 4 °C overnight with 40 μL Glutathione Sepharose 4B beads (GE Healthcare). After four washes with the wash buffer, proteins were eluted from the beads using 1X SDS loading buffer. The large-scale affinity purification using FLAG and GST tags was carried out essentially the same as described in the small-scale protocol above, except that the starting materials were doubled and that we used 150 mM NaCl (instead of 300 mM) in the lysis buffer. Instead of transfecting one 10-cm plate of HEK293T cells per plasmid, we transfected two 10-cm plates of HEK293T cells with the plasmid construct expressing FLAG-GST-tagged SUGP1 (aa 543–645). After two rounds of affinity purification as described above (except with 150 mM NaCl in the lysis buffer), the eluted proteins were resolved by SDS-PAGE, followed by staining with QC Colloidal Coomassie Stain (Bio-Rad). The relevant gel region (~85–100 KD) was excised, and the proteins in this gel section were identified by mass spectrometry (Taplin Mass Spectrometry Facility at Harvard University).

Recombinant Protein Purification and In Vitro Protein-Protein Interaction. N-terminally His6-HA-tagged SUGP1 was cloned in pET26b (Novagen) using NdeI and BamHI sites. Its G574A-G582A mutant construct was generated by replacing the WT fragment between AgeI and BamHI sites with the G574A-G582A mutant fragment double-digested from the HA-tagged SUGP1 G574A-G582A mutant construct (1). N-terminally His6-FLAG-MBP-tagged DHX15 (aa 113–795) construct was cloned as follows. Overlap extension PCR was used to generate a long PCR product that combines His6, FLAG, and MBP tags with DHX15 (aa

113–795, codon optimized), as well as a short fragment (containing an XbaI site) from the vector pET26b (Novagen). This PCR product was then double-digested with XbaI and BamHI and cloned in pET26b (Novagen). Protein expression was induced in *Escherichia coli* Rosetta cells at 14°C for 24 h by 0.2 mM isopropyl b-D-1-thiogalactopyranoside. Bacterial cells were lysed using a lysis buffer containing 50 mM Na_2HPO_4 (pH 8.0), 500 mM NaCl, 0.1% Triton X-100, and protease inhibitor cocktail (Roche). Recombinant proteins were then purified using TALON Superflow Metal Affinity Resin (Clontech). Protein concentration was determined by measuring the optical density absorbance at 280 nm. In vitro protein-protein interaction assay by co-immunoprecipitation was carried out as described (1), except that 20 pmol each purified protein was used. Briefly, 20 pmol His6-FLAG-MBP-tagged DHX15 (aa 113–795) was mixed with 20 pmol His6-HA-tagged SUGP1 or with its G574A-G582A mutant in a binding buffer (50 mM Tris-Cl, pH 7.4, 150 mM NaCl, and 0.5% Triton X-100), followed by incubation at 4°C for 30 min. Each reaction was then incubated with 10 μL (0.5 μg/μL) anti-DYKDDDDK antibody (GenScript, A00187) at 4°C for another 30 min, followed by incubation with 50 μL Pierce Protein A/G Magnetic Beads (Thermo Fisher Scientific) at 4°C overnight. After four washes with the binding buffer, proteins were eluted from the magnetic beads with 30 μL (1 μg/μL) 3X FLAG peptide.

Computational Identification of Cryptic 3′ Splice Sites. Cryptic 3′ss were computationally identified using a previously described method (1). This method enables highly sensitive detection of novel splice junctions and is therefore specifically designed for identification of cryptic 3′ss and 5′ss. Briefly, we first downloaded the FASTQ files of the RNA-seq data (GEO accession GSE156390) of HeLa cells with or without DHX15 knockdown (2), and then aligned the sequencing reads to the human genome (hg19) using STAR 2.7.4 (3), with a splice junction database (4). Counts of junction reads were obtained from the STAR output file (SJ.out.tab) and low-abundance junctions with fewer than 10 reads (summed up across all samples) were filtered out. For the remaining junctions, we compared each of them to the known splice junctions in the database provided, and defined it as a novel cryptic junction if only one of its two ends (either 3′ or 5′) is annotated in the database. The associated canonical junction is then identified if it shares the same annotated 3′ or 5′ end with the cryptic junction and if the other end is also annotated in the database. We then determined the relative position of each cryptic 3′ss or 5′ss to the associated canonical 3′ss or 5′ss based on their locations on the same strand of the transcript. Next, the PSI (percent spliced in) values were calculated from the raw read counts. We then used t-tests with PSI values (instead of raw read counts) to identify differential usage of cryptic 3′ss between DHX15 knockdown and control, and the resulting *P* values were further adjusted by Benjamini–Hochberg multiple test correction to obtain *Q* values. Differences in mean PSI values were also calculated as important metrics for differential splicing. We used two threshold parameters (*P* value < 0.05 and located closer than 50 nt upstream of the canonical 3′ss) to identify differentially used cryptic 3′ss upon DHX15 knockdown.

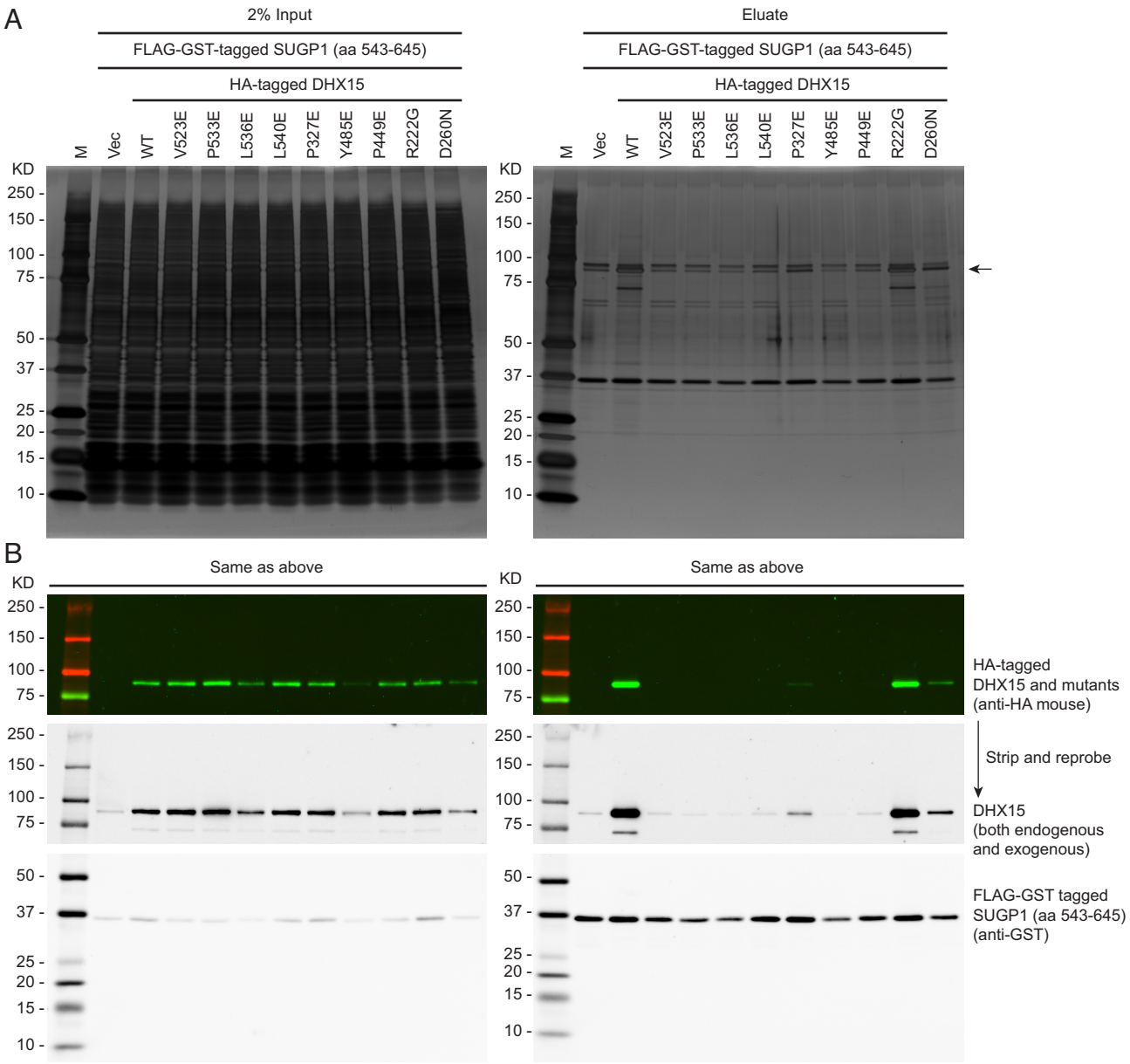

Fig. S1. *DHX15* mutations affect the interaction of DHX15 with the SUGP1 G-patch. (*A* and *B*) Expression plasmid for FLAG-GST-tagged SUGP1 (aa 543–645) was co-transfected with empty vector plasmid (Vec) or expression plasmid for HA-tagged DHX15 (WT or one of the mutants as indicated), followed by small-scale affinity purification using FLAG and GST tags. Two equal aliquots of the purified proteins were resolved in two SDS-polyacrylamide gels, one for silver staining (*A*) and the other for western blotting (*B*). The arrow in (*A*) points to the protein bands with the expected size of DHX15. M, Precision Plus Protein marker (Bio-Rad).

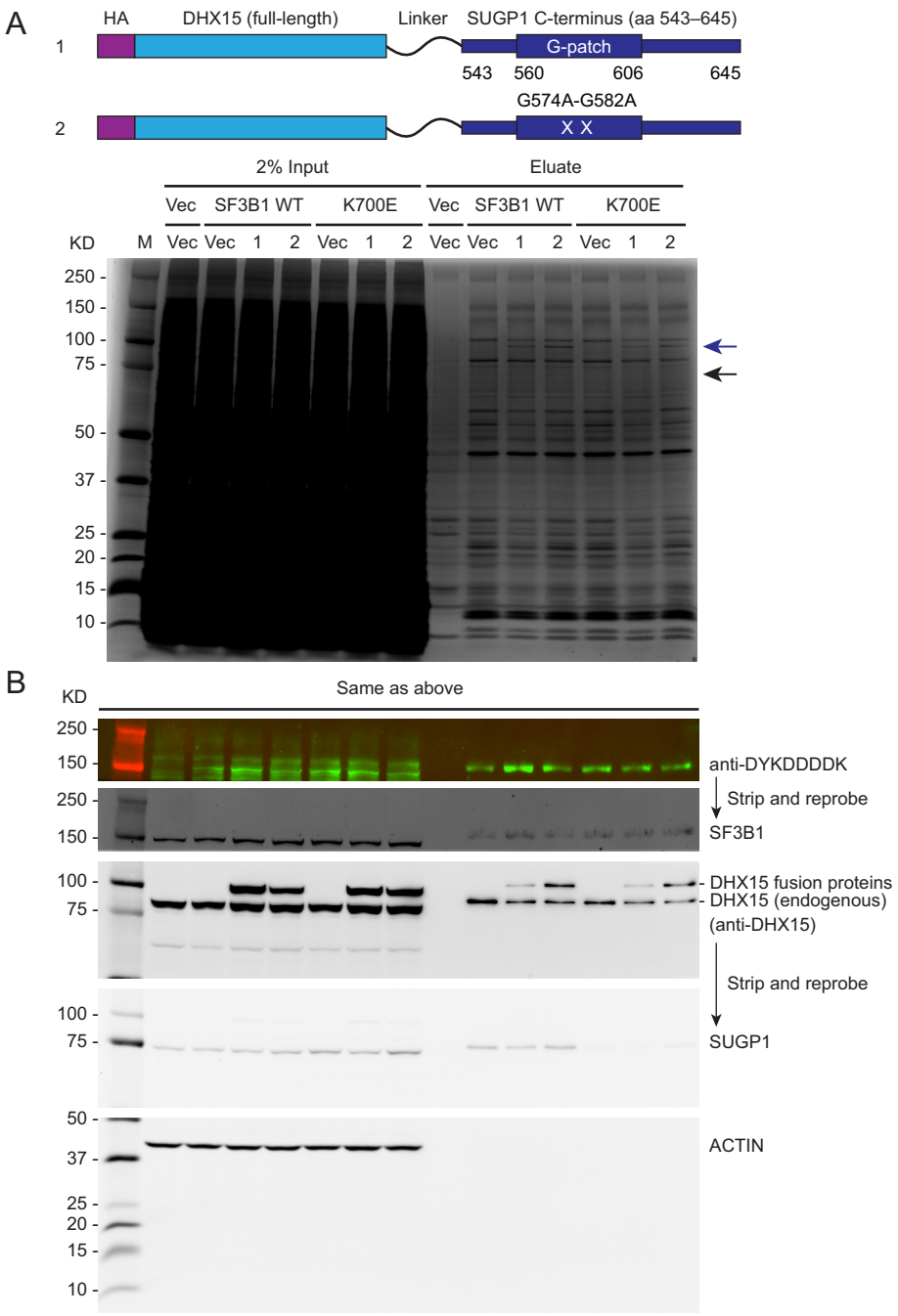

Fig. S2. DHX15-SUGP1 G-patch fusion protein incorporates into the spliceosome. (*A, Upper*) Schematic representation of plasmid constructs (1 and 2) expressing HA-tagged DHX15-SUGP1 (aa 543–645) fusion protein and its G574A-G582A mutant, respectively. (*Lower*) Vec or expression plasmid for His6-FLAG-tagged SF3B1 (WT or K700E) was co-transfected with Vec or one of the two constructs (as indicated) in HEK293T cells, followed by small-scale affinity purification using FLAG and His6 tags. Two equal aliquots of the purified proteins were subjected to SDS-PAGE, followed by silver staining (*A, Lower*) or western blotting (*B*). The blue and black arrows in (*A, Lower*) point to the protein bands with the expected sizes of the fusion protein and SUGP1, respectively. M, Precision Plus Protein marker (Bio-Rad).

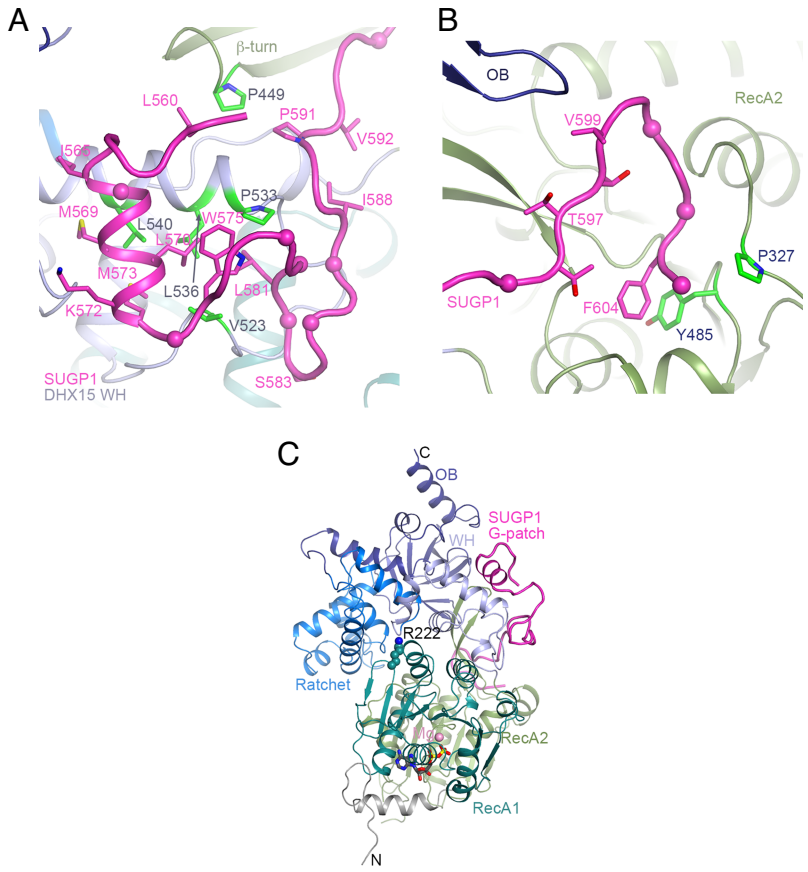

Fig. S3. Structural insights into the interaction between DHX15 and the SUGP1 G-patch. (*A*) Interaction of the N-terminal part of the SUGP1 G-patch with the DHX15 WH domain. Side chains in the SUGP1 G-patch (in magenta) making large contributions to the interface are shown as sticks. DHX15 residues that were selected for mutation (*SI Appendix*, Fig. S1) are shown as sticks (green) and labeled. The Cα atoms of glycine residues are indicated by spheres. (B) Interaction of the C-terminal part of the SUGP1 G-patch with the DHX15 RecA2 domain. (C) Residue R222 is located in a β-reverse turn in the RecA1 domain of DHX15, making no contact with the SUGP1 G-patch.

Table S1. Summary of crystallographic information

Structure	Human DHX15-SUGP1 G-patch complex	Human SUGP1 (433–577)	Human SUGP1 (433–586)
Data Collection			
Space group	$C222_1$	$P6_1$	$P6_1$
Cell dimensions			
a, *b*, *c* (Å)	82.3, 91.0, 213.2	59.6, 59.6, 178.4	60.2, 60.2, 177.8
α, β, γ (º)	90, 90, 90	90, 90, 120	90, 90, 120
Resolution (Å)*	58.7-1.8 (1.91-1.8)	51.7-2.4 (2.48-2.4)	44.9-2.8 (2.91-2.8)
R_{merge} (%)	6.6 (132.4)	6.2 (44.3)	4.7 (64.1)
$CC_{1/2}$	0.999 (0.696)	0.997 (0.835)	0.999 (0.87)
I/σI	13.2 (1.1)	16.9 (3.1)	27.5 (2.9)
Completeness (%)	99.9 (99.6)	99.9 (99.9)	99.8 (98.9)
No. of reflections	74658 (11895)	13999 (1405)	8896 (887)
Redundancy	6.7 (6.4)	5.2 (4.7)	7.8 (7.8)
Refinement			
Resolution (Å)	58.7-1.8 (1.86-1.8)	51.7-2.4 (2.48-2.4)	44.9-2.8 (2.98-2.8)
R_{work} (%)	18.9 (40.8)	20.2 (27.2)	21.2 (29.5)
R_{free} (%)	22.4 (43.0)	25.2 (36.9)	28.0 (34.4)
Number of atoms	6133	2115	2058
Protein	5808	2026	2058
Ligand/Ion	52	-	-
Water	273	55	-
B-factors (Å^2)			
Protein	45.5	52.6	86.2
Ligand	52.8	-	-
Water	45.8	53.6	-
r.m.s.d.			
Bond lengths (Å)	0.013	0.008	0.010
Bond angles (°)	1.2	1.0	1.0
Ramachandran plot			
Favored (%)	97.6	97.8	98.2
Allowed (%)	2.4	2.2	1.8
Outliers (%)	0.0	0.0	0.0

*The numbers in parentheses are for the highest resolution shell.

SI References

1. J. Zhang *et al.*; Disease-causing mutations in *SF3B1* alter splicing by disrupting interaction with SUGP1. *Mol. Cell* **76**, 82–95.e1–e7 (2019).
2. A. Duchemin *et al.*; DHX15-independent roles for TFIP11 in U6 snRNA modification, U4/U6.U5 tri-snRNP assembly and pre-mRNA splicing fidelity. *Nat. Commun.* **12**, 6648 (2021).
3. A. Dobin *et al.*; STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
4. C. DeBoever *et al.*; Transcriptome sequencing reveals potential mechanism of cryptic 3′ splice site selection in SF3B1-mutated cancers. *PLoS Comput. Biol.* **11**, e1004105 (2015).

Other supporting materials for this article include the following:

Dataset S1 (separate file). Summary of the mass spectrometry data.
Dataset S2 (separate file). Cryptic 3′ splice sites misregulated upon DHX15 knockdown.