OUTLOOK

AI-assisted proofreading of RNA splicing

Ángel Guerra-Moreno[1] and Juan Valcárcel[1,2,3]

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona 08003, Spain; [2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain; [3]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

RNA helicases orchestrate proofreading mechanisms that facilitate accurate intron removal from pre-mRNAs. How these activities are recruited to spliceosome/pre-mRNA complexes remains poorly understood. In this issue of *Genes & Development*, Zhang and colleagues (pp. 968–983) combine biochemical experiments with AI-based structure prediction methods to generate a model for the interaction between SF3B1, a core splicing factor essential for the recognition of the intron branchpoint, and SUGP1, a protein that bridges SF3B1 with the helicase DHX15. Interaction with SF3B1 exposes the G-patch domain of SUGP1, facilitating binding to and activation of DHX15. The model can explain the activation of cryptic 3′ splice sites induced by mutations in SF3B1 or SUGP1 frequently found in cancer.

One persistent conundrum in understanding the process of intron removal from messenger RNA precursors (pre-mRNA splicing) is how the sophisticated machinery of the spliceosome achieves high efficiency and single-nucleotide precision despite the limited conservation of the sequences that delineate intron/exon boundaries (splice sites). The problem is further complicated in multicellular organisms where flexibility in recognition of alternative splice sites is important for the regulation of cell differentiation and homeostasis. This delicate balance in the spliceosome's operations is also revealed by the activation of pathogenic cryptic splice sites in cancer cells harboring mutations in splicing factors.

For instance, SF3B1, a protein component of U2 small nuclear ribonucleoprotein (U2 snRNP) complexes that is important for recognition of an intronic branchpoint adenosine essential for the splicing reaction, is frequently mutated in a variety of tumors. These mutations induce the activation of alternative 3′ splice sites that are not normally used in noncancerous cells (Fig. 1A). In turn, these transcript alterations contribute to disease progression. Thus, frequent replacement of lysine by glutamic acid at SF3B1 amino acid 700 (K700E) in chronic lymphocytic leukemia (CLL) tumors correlates with unfavorable prognosis, linked with deregulation of B-cell receptor signaling and increased sensitivity to inhibitors of this pathway (Yin et al. 2019).

Mechanistically, cancer-associated mutations in SF3B1 disrupt interactions with other protein partners, including the DEAD-box helicase DDX46 (human ortholog of yeast Prp5) (Zhang et al. 2022a). DDX46 and at least seven other proteins of the same family of RNA-dependent ATPases empower splicing complexes with proofreading mechanisms that stall spliceosome assembly or subsequent conformational transitions leading to catalysis on mutant or suboptimal pre-mRNA substrates, thus preventing the processing of erroneous transcripts. These helicases can use the rate of ATP hydrolysis as a timing device to measure whether stable molecular interactions occur with proper kinetics or otherwise trigger alternative discard pathways, leading to disassembly of spliceosome/pre-mRNA complexes (Zhang et al. 2021).

Cancer-associated SF3B1 mutations also disrupt interaction with the protein SUGP1 (Zhang et al. 2019). Consistently, mutations in SUGP1 found in tumors also lead to similar changes in splice site selection (Liu et al. 2020; Alsafadi et al. 2021). While SUGP1 does not display helicase activity, its function in spliceosome assembly has been linked to the recruitment to U2 snRNP/branchpoint complexes of the DEAH-box RNA helicase DHX15 (human ortholog of yeast Prp43) (Fig. 1; Maul-Newby et al. 2022; Zhang et al. 2022b; Beusch et al. 2023; Feng et al. 2023). While DHX15 has known functions in ribosome assembly and in disassembly of post-splicing complexes, its role at early stages of spliceosome assembly remains poorly understood.

Despite detailed knowledge of multiple spliceosomal snapshots obtained by cryo-EM and X-ray crystallography in recent years, structural information about the interaction between SUGP1 and SF3B1 was lacking, suggesting its transient or unstable nature. In this issue of *Genes & Development*, Zhang et al. (2023) report a detailed structure/function analysis of the SF3B1–SUGP1 interaction that significantly improves our understanding of the mechanism of cancer-associated mutations and in addition provides an attractive model for how the interaction can help to recruit DHX15 to U2 snRNP/branchpoint

[*Keywords*: spliceosome mutations; AlphaFold-Multimer; DHX15; myelodysplastic syndromes; G-patch]
Corresponding author: juan.valcarcel@crg.es
Article published online ahead of print. Article and publication date are online at http://www.genesdev.org/cgi/doi/10.1101/gad.351373.123. Freely available online through the *Genes & Development* Open Access option.

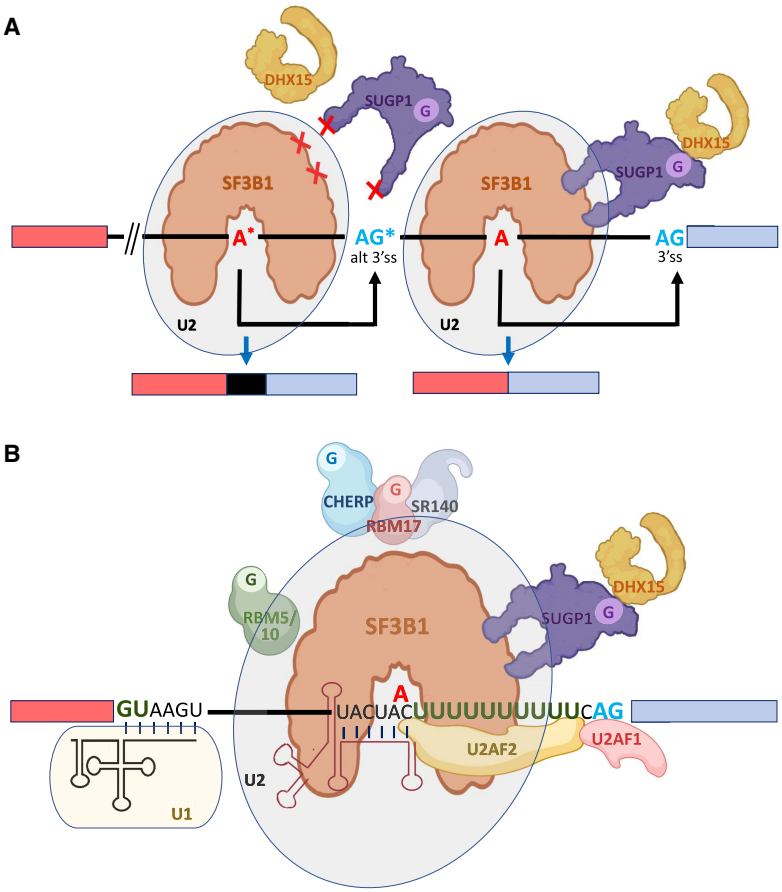

Figure 1. (*A*) Mechanistic model for the activation of cryptic (*) 3′ splice sites upon mutation of splicing factor SF3B1 or SUGP1. Recognition of the branchpoint adenosine by the U2 snRNP protein SF3B1 is important to define the 3′ splice site AG. Recruitment of the DEAH-box RNA helicase DHX15 via bridging interaction with SUGP1 provides a proofreading checkpoint that allows further assembly of spliceosomal components. Mutants in SF3B1 or SUGP1 that destabilize their interaction fail to recruit DHX15, allowing assembly of U2 snRNP complexes on cryptic branchpoints and the activation of alternative 3′ splice sites. (*B*) Schematic representation of molecular interactions leading to early recognition and proofreading of splice sites. Base-pairing interactions with the snRNA components of U1/U2 snRNPs define the 5′ splice site/branchpoint, while U2AF2 and U2AF1 recognize the polypyrimidine tract and 3′ splice site AG, respectively, and assist in U2 snRNP recruitment, in part through interaction with SF3B1. Helicases DDX46 and DHX15 provide proofreading checkpoints for U2 snRNP/branchpoint assembly. SUGP1 interacts with DHX15 through its glycine patch domain (G). Other G-patch-containing proteins such as RBM5, RBM10, RBM17, or CHERP (the latter two in complex with SR140) are associated with U2 snRNP and may help to recruit DHX15 and/or other helicases, which can influence 3′ splice site recognition and selection.

complexes. Zhang et al. (2023) generated a structural model by combining results from biochemical experiments with remarkably accurate, artificial intelligence-based protein complex structure prediction methods (AlphaFold-Multimer) (Evans et al. 2021). Two regions of interaction with SF3B1 were identified in SUGP1, flanking its glycine (G)-patch domain. G-patch domains activate RNA helicases by stabilizing conformations that facilitate their NTPase activity (Studer et al. 2020). Zhang et al. (2023) propose that simultaneous contacts of SF3B1 with regions of SUGP1 that flank the G-patch serve to loop out the domain and thus facilitate its interaction with DHX15 (Fig. 1B). Nicely correlating with the detailed structural predictions, a variety of SUGP1 mutations engineered to disrupt binding to SF3B1 not only reduce the interaction but also activate the use of cryptic 3′ splice sites, phenocopying the effects of cancer-associated mutations in SF3B1 or SUGP1 (Fig. 1A).

Interestingly, Damianov et al. (2023) have recently characterized a U2 snRNP complex purified from chromatin, tightly associated with pre-mRNA branchpoints, which contains not only SUGP1 but also several other G-patch domain-containing splicing regulatory factors such as RBM5, RBM10, RBM17, or CHERP. These observations raise the possibility that other G-patch-containing factors, interacting with SF3B1 or other components of U2 snRNP, serve functions similar to that of SUGP1 in the recruitment of DHX15 (or other helicases) to branchpoint regions (Fig. 1B). The combined/combinatorial effects of these factors might modulate proofreading of 3′ splice sites and differential splice site utilization.

An interesting variation on this theme is discussed by Zhang et al. (2023). Because SUGP1 also interacts with early factors SF1/BBP and U2AF2, which recognize the branchpoint and downstream polypyrimidine tract, respectively, before U2 snRNP assembly, they propose that SUGP1, through its interaction with SF3B1, plays a key role in recruiting U2 snRNP to the branchpoint, which can explain the dominant-negative effect of SUGP1 mutants that prevent the SUGP1–SF3B1 interaction. G-patch-containing RBM17 can also interact with SF1/BBP, U2AF2, and SF3B1 and could similarly favor U2 snRNP recruitment (Corsini et al. 2007).

Consistent with the dynamic interplay between SF3B1, SUGP1, and DHX15, Beusch et al. (2023) recently identified mutations in residues located in the G-patch of SUGP1 that prevent its interaction with DHX15 and confer resistance to pladienolide B, a SF3B1-targeting splicing inhibitor. These observations strengthen the concept that DHX15 kinetically proofreads 3′ splice site recognition and that this activity can help to modulate splice site usage under different conditions. This concept is also in line with recent results from Feng et al. (2023) showing extensive activation of suboptimal splice sites and cryptic introns upon knockdown of DHX15 or SUGP1 in human embryonic kidney (HEK293) cells.

Collectively, the results of Zhang et al. (2023) and other recent studies provide a structural and functional framework to understand proofreading mechanisms that ensure splicing accuracy and possibly also splice site selection and the activity of compounds that display antitumor activity. More generally, they illustrate the potential of AI methods to predict, with structural detail, key functional interactions between protein components of complex biomolecular assemblies.

Acknowledgments

We thank Andrey Damianov, Doug Black, and members of our laboratory for comments on the manuscript. Work in our laboratory is supported by laCaixa Health, the European Research Council, and the Spanish Ministerio de Ciencia, Innovación y Universidades.

References

Alsafadi S, Dayot S, Tarin M, Houy A, Bellanger D, Cornella M, Wassef M, Waterfall JJ, Lehnert E, Roman-Roman S, et al. 2021. Genetic alterations of SUGP1 mimic mutant-SF3B1 splice pattern in lung adenocarcinoma and other cancers. *Oncogene* **40:** 85–96. doi:10.1038/s41388-020-01507-5

Beusch I, Rao B, Studer MK, Luhovska T, Šukytė V, Lei S, Oses-Prieto J, SeGraves E, Burlingame A, Jonas S, et al. 2023. Targeted high-throughput mutagenesis of the human spliceosome reveals its in vivo operating principles. *Mol Cell* **83:** 2578–2594.e9. doi:10.1016/j.molcel.2023.06.003

Corsini L, Bonnal S, Basquin J, Hothorn M, Scheffzek K, Valcárcel J, Sattler M. 2007. U2AF-homology motif interactions are required for alternative splicing regulation by SPF45. *Nat Struct Mol Biol* **14:** 620–629. doi:10.1038/nsmb1260

Damianov A, Lin CH, Huang J, Zhou L, Jami-Alahmadi Y, Wohlschlegel J, Black DL. 2023. The apoptotic splicing regulators RBM5 and RBM10 are subunits of the U2 snRNP engaged with intron branch sites on chromatin. bioRxiv doi:10.1101/2023.09.21.558883

Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, Žídek A, Bates R, Blackwell A, Yim J, et al. 2021. Protein complex prediction with AlphaFold-Multimer. bioRxiv doi:10.1101/2021.10.04.463034

Feng Q, Krick K, Chu J, Burge CB. 2023. Splicing quality control mediated by DHX15 and its G-patch activator SUGP1. *Cell Rep* **42:** 113223. doi:10.1016/j.celrep.2023.113223

Liu Z, Zhang J, Sun Y, Perea-Chamblee TE, Manley JL, Rabadan R. 2020. Pan-cancer analysis identifies mutations in *SUGP1* that recapitulate mutant *SF3B1* splicing dysregulation. *Proc Natl Acad Sci* **117:** 10305–10312. doi:10.1073/pnas.1922622117

Maul-Newby HM, Amorello AN, Sharma T, Kim JH, Modena MS, Prichard BE, Jurica MS. 2022. A model for DHX15 mediated disasssembly of A-complex spliceosomes. *RNA* **28:** 583–595. doi:10.1261/rna.078977.121

Studer MK, Ivanović L, Weber ME, Marti S, Jonas S. 2020. Structural basis for DEAH-helicase activation by G-patch proteins. *Proc Natl Acad Sci* **117:** 7159–7170. doi:10.1073/pnas.1913880117

Yin S, Gambe RG, Sun J, Martinez AZ, Cartun ZJ, Regis FFD, Wan Y, Fan J, Brooks AN, Herman SEM, et al. 2019. A murine model of chronic lymphocytic leukemia based on B cell-restricted expression of Sf3b1 mutation and Atm deletion. *Cancer Cell* **35:** 283–296.e5. doi:10.1016/j.ccell.2018.12.013

Zhang J, Ali AM, Lieu YK, Liu Z, Gao J, Rabadan R, Raza A, Mukherjee S, Manley JL. 2019. Disease-causing mutations in SF3B1 alter splicing by disrupting interaction with SUGP1. *Mol Cell* **76:** 82–95.e7. doi:10.1016/j.molcel.2019.07.017

Zhang Z, Rigo N, Dybkov O, Fourmann JB, Will CL, Kumar V, Urlaub H, Stark H, Lührmann R. 2021. Structural insights into how Prp5 proofreads the pre-mRNA branch site. *Nature* **596:** 296–300. doi:10.1038/s41586-021-03789-5

Zhang X, Zhan X, Bian T, Yang F, Li P, Lu Y, Xing Z, Zhang QC, Shi Y. 2022a. Structural insights into branch site proofreading by human spliceosome. bioRxiv doi:10.1101/2022.11.07.515429

Zhang J, Huang J, Xu K, Xing P, Huang Y, Liu Z, Tong L, Manley JL. 2022b. DHX15 is involved in SUGP1-mediated RNA missplicing by mutant SF3B1 in cancer. *Proc Natl Acad Sci* **119:** e2216712119. doi:10.1073/pnas.2216712119

Zhang J, Xie J, Huang J, Liu X, Xu R, Tholen J, Galej WP, Tong L, Manley JL, Liu Z. 2023. Characterization of the SF3B1–SUGP1 interface reveals how numerous cancer mutations cause mRNA missplicing. *Genes Dev* (this issue). doi:10.1101/gad.351154.123

AI-assisted proofreading of RNA splicing

Ángel Guerra-Moreno and Juan Valcárcel

Genes Dev. 2023, **37:** originally published online December 13, 2023
Access the most recent version at doi:10.1101/gad.351373.123