# Rewards of divergence in sequences, 3-D structures and dynamics of yeast and human spliceosome SF3b complexes

Arangasamy Yazhini, Sankaran Sandhya, Narayanaswamy Srinivasan [*]

*Molecular Biophysics Unit, Indian Institute of Science, Bangalore, Karnataka, 560012, India*

## ARTICLE INFO

## ABSTRACT

The evolution of homologous and functionally equivalent multiprotein assemblies is intriguing considering sequence divergence of constituent proteins. Here, we studied the implications of protein sequence divergence on the structure, dynamics and function of homologous yeast and human SF3b spliceosomal subcomplexes. Human and yeast SF3b comprise of 7 and 6 proteins respectively, with all yeast proteins homologous to their human counterparts at moderate sequence identity. SF3b6, an additional component in the human SF3b, interacts with the N-terminal extension of SF3b1 while the yeast homologue Hsh155 lacks the equivalent region. Through detailed homology studies, we show that SF3b6 is absent not only in yeast but in multiple lineages of eukaryotes implying that it is critical in specific organisms. We probed for the potential role of SF3b6 in the spliceosome assembled form through structural and flexibility analyses. By analysing normal modes derived from anisotropic network models of SF3b1, we demonstrate that when SF3b1 is bound to SF3b6, similarities in the magnitude of residue motions (0.86) and inter-residue correlated motions (0.94) with Hsh155 are significantly higher than when SF3b1 is considered in isolation (0.21 and 0.89 respectively). We observed that SF3b6 promotes functionally relevant 'open-to-close' transition in SF3b1 by enhancing concerted residue motions. Such motions are found to occur in the Hsh155 without SF3b6. The presence of SF3b6 influences motions of 16 residues that interact with U2 snRNA/branchpoint duplex and supports the participation of its interface residues in long-range communication in the SF3b1. These results advocate that SF3b6 potentially acts as an allosteric regulator of SF3b1 for BPS selection and might play a role in alternative splicing. Furthermore, we observe variability in the relative orientation of SF3b4 and in the local structure of three β-propeller domains of SF3b3 with reference to their yeast counterparts. Such differences influence the inter-protein interactions of SF3b between these two organisms. Together, our findings highlight features of SF3b evolution and suggests that the human SF3b may have evolved sophisticated mechanisms to fine tune its molecular function.

## 1. Introduction

Homologous proteins arise through divergent evolution and perform similar functions in different organisms with some differences in functional features (Povolotskaya and Kondrashov, 2010; Tatusov et al., 1997). It is generally accepted that sequence variations may confer functional specializations in terms of substrate specificity (Kim et al., 2000), efficiency (Dermitzakis and Clark, 2002) and allosteric regulation (Chan et al., 2017). Here, we define homologous multiprotein complexes as functionally equivalent molecular assemblies in two different

organisms with most of the components in the two assemblies being evolutionarily related. In homologous multiprotein complexes, in which assembled proteins work together, it is unclear how divergent evolution of the constituent proteins influences the function of such assemblies in different organisms. One such example of a homologous multiprotein complex that differs in the composition and yet conserves function is the spliceosome subcomplex SF3b (Sun, 2020; Will et al., 1999).

SF3b in yeast comprises of 6 proteins namely Hsh155, Cus1, Rse1, Hsh49, Rds3 and Ysf3 (Fig. 1A). Corresponding homologues in the human SF3b are SF3b1, SF3b2, SF3b3, SF3b4, SF3b14b and SF3b5 (Sun,

**A**



**Fig. 1. SF3b complex in yeast and human.** A) Shown is the surface representation of yeast (left) and human (right) SF3b complexes. Each protein component is colored differently, with labels pointing to their location in both complexes. A cross symbol indicates the absence of SF3b6. B) Cartoons depicting Pfam domains present in the yeast (top) and human (bottom) SF3b proteins. The length scale of the cartoon reflects protein sequence length. Domain names and boundaries are marked for SF3b proteins that are labelled as per human nomenclature. Pfam domains referred as SF3b1 (Pfam id: PF08920) and SAP (Pfam id: PF02037) are assigned uniquely to human SF3b1 and SF3b2 proteins, respectively.

**B**



2020). In addition, the human SF3b contains another integral protein SF3b6 (or p14) for which there is no recognizable yeast homologue (van der Feltz and Hoskins, 2019) (Fig. 1A). SF3b complex is a major constituent of U2 and U11/U12 snRNPs (Golas et al., 2003; Will et al., 1999; Zhang et al., 2020). In the major class spliceosome, it participates in early stages of splicing starting from complex A to activated B spliceosome (B$^{act}$) assembly to i) recognize branch-point sequence (BPS), ii) stabilize U2 snRNA/BPS duplex and iii) prevent pre-mature transesterification step (Gozani et al., 1996; Lardelli et al., 2010; Will and Lührmann, 2011). Mutations in the SF3b complex are observed to disrupt the spliceosome assembly (Wang and Rymond, 2003) and affect the selection of BPS (Cretu et al., 2016; Darman et al., 2015; Tang et al., 2016). SF3b1 is one of the largest proteins in the human SF3b with a long N-terminal extension (490aa) and 20 HEAT repeats that adopt a superhelical structure (Cretu et al., 2016). Frequently observed mutations in several cancer conditions such as myelodysplastic syndrome, chronic lymphocytic leukemia and uveal melanoma are located in the HEAT repeats of SF3b1 (Alsafadi et al., 2016; Yoshida and Ogawa, 2014). The N-terminal extension of SF3b1 acts as a scaffold to interact with other spliceosomal proteins U2AF65, CAPERα, PUF60, SF3b6 etc (Loerch and Kielkopf, 2016), that are not uniformly conserved across eukaryotes (Stelzer et al., 2016). Interestingly, in yeast, the equivalent N-terminal extension of human SF3b1 and an interacting partner SF3b6 are both absent

(Dziembowski et al., 2004; van der Feltz and Hoskins, 2019; Zhang et al., 2018), suggesting significant divergence in the evolution of SF3b complex.

SF3b6 a component of SF3b seen in humans but not in yeast harbours a RNA recognition motif (RRM). Some studies have reported that SF3b6 directly interacts with branch-point adenosine to recognize BPS (Mac-Millan et al., 1994; Query et al., 1996; Will et al., 2001). On the other hand, some others have demonstrated that SF3b6 lacks the specificity for the BPS and neither discriminates adenosine monophosphate over adenine nor single stranded RNA over double stranded RNA/DNA (Perea et al., 2016; Spadaccini et al., 2006). Also, structural studies show that the RNA binding region in the RRM of SF3b6 is impeded by its own helix α3 and C-terminal tail as well as by an interacting region from SF3b1 (Schellenberg et al., 2006, 2011). We observe from the available major spliceosome structures that Tyr22 of SF3b6 which is suggested to directly interact with bulged branch-point adenosine is spatially distal (~54 Å) from the branch-point adenosine (Fig. S1). Although the pre-mRNA binding site of SF3b6 is occluded and the distance is large, it is unclear whether inherent dynamics and conformational transitions of U2 snRNP (van der Feltz and Hoskins, 2019) could contribute to pre-mRNA binding. Therefore, it is interesting to examine the role of SF3b6 when the SF3b complex is integrated into the spliceosome in humans, especially since it is absent in yeast and seems to have a unique role.

In this study, we compare yeast and human SF3b complexes based on sequence, 3-D structure and dynamics of their individual proteins. Firstly, using sequence analysis, we probed for homologues of individual components of the complex in all organisms and compared their sequences. We find that the yeast and human SF3b proteins are moderately conserved. SF3b6 is universally absent in the Saccharomyces genus, which includes yeast, as well as in several other lineages across eukaryotes. We note that the N-terminal region of SF3b1 homologues in such species have evolved to compensate for the loss of SF3b6. Secondly, we probed for the plausible role of SF3b6 in the spliceosome assembled form, particularly in the B^act state, using *in silico* anisotropic normal modes, perturbation-response scanning and structural network analyses. Our studies suggest a possible role for SF3b6 in allosteric regulation of SF3b1, with potential relevance in alternative splicing. Finally, we analyzed B^act spliceosome assembly structures to study the implications of sequence divergence of constituent proteins in the yeast and human SF3b complexes. Here, our results highlight inherently variable features within the homologous SF3b structures. We find that such variations are primarily contributed by differences in the relative orientation of human SF3b4 *vis-a-vis* its yeast homologue Hsh49 as well as in the interaction patterns of SF3b3-SF3b5 and SF3b3-SF3b1 interfaces. Together, these observations provide insights into the implication of species-specific sequence diversity of the various components in the homologous SF3b complexes and their impact on structure and function.

## 2. Materials and methods

### 2.1. In silico structure preparation for SF3b1 and SF3b1-SF3b6 complex

In all the 9 cryo-electron microscopy (cryo-EM) atomic models of SF3b complex representing early stages of human spliceosome assemblies (Bertram et al., 2017; Charenton et al., 2019; Cretu et al., 2016; Finci et al., 2018; Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2018) (Table S1), the N-terminal extension of SF3b1 is only partly resolved due to its intrinsically disordered nature. Interface identification (Tina et al., 2007) of these complexes and previous experimental studies on SF3b1 and SF3b6 interaction (Perea et al., 2016; Schellenberg et al., 2006; Spadaccini et al., 2006; Will et al., 2001) have revealed that the N-terminal residues from 373 to 415 of SF3b1 is the binding site for SF3b6. However, atomic positions for certain residues in this region are not available. Among the 9 assembly structures of spliceosome, conformational differences in the SF3b1 as well as SF3b6 structures are small (average Cα root mean square deviation (RMSD): 1.1 Å and 1.3 Å respectively, refer Table S2). Therefore, we selected activated spliceosome assembly structure or B^act which corresponds to the Protein Data Bank (https://www.rcsb.org/, PDB) entry 5Z58 for the analysis (Berman et al., 2002; Zhang et al., 2018). This structure contains structural information for the entire SF3b complex at a better resolution (4.9 Å) with coordinates for side chains. We generated an *in silico* model for SF3b1 bound to SF3b6 using comparative modeling (Šali and Blundell, 1993) and built missing regions in the SF3b6 binding site. Keeping SF3b1-SF3b6 complex structure from human B^act entry as the reference (PDB code: 5Z58), we used X-ray crystal structure of SF3b6 bound to SF3b1 fragment as a template to build missing segments in the N-terminal extension of SF3b1 and SF3b6 (PDB code: 2F9D) (Schellenberg et al., 2006). Of the 100 generated models, the best model based on DOPE score was selected for analysing SF3b1-SF3b6 complex. To study SF3b1 in the uncomplexed state, SF3b6 was removed *in silico* from the generated model of the complex structure. Likewise, missing regions were built for SF3b1 structure obtained from the isolated SF3b complex (PDB code: 5IFE) (Cretu et al., 2016), using B^act structure as a template (PDB code: 5Z58).

### 2.2. Normal mode analysis and perturbation-response scanning analysis

Intrinsic dynamics of a protein can be studied by representing the

protein molecule as an elastic network model and by calculating their normal modes using normal mode analysis (NMA). 3-D models of SF3b1 and SF3b1-SF3b6 complex as well as cryo-EM structure of the yeast homologue Hsh155, belonging to the same functional B^act assembly state (PDB code: 5GM6) (Yan et al., 2016) were used as inputs for NMA. Anisotropic network models were built at Cα level for all three structures with a distance cut-off of 15 Å to connect Cα atoms by virtual springs following which normal modes were calculated. Network construction, normal mode calculation and related analysis were performed using the ProDy package (Atilgan et al., 2001). To compare features of dynamics among SF3b1 in isolation, SF3b1-SF3b6 complex and Hsh155, mean square fluctuations of residues and cross-correlation matrices were calculated. This would capture the magnitude of residue motions and coupling between fluctuations of each residue pair. In addition, collectivity which corresponds to the proportion of residues moving in a correlated fashion was calculated. The top 20 lowest frequency modes were used to derive cross-correlations between residue fluctuations of all possible pairs within a structure. Similarity in the strength of correlations between two cross-correlation matrices was measured using $R_V$ coefficient. This value computes the degree of closeness between two sets of multivariate data by generalizing Pearson correlation coefficient (Smilde et al., 2009).

To study the effect of SF3b6 binding on SF3b1, we performed perturbation-response scanning analysis (Atilgan and Atilgan, 2009) on SF3b1 using ProDy package. Here, systematic directed forces are applied on a selected residue over 1000 times in randomly chosen directions. For each perturbation, the magnitude and direction of displacement of every other residue present in the anisotropic network model were measured, and the average over 1000 perturbations was calculated to determine the effectiveness of the residue.

### 2.3. Network analysis of SF3b1

Long-range residue-residue communication can facilitate allostery (del Sol et al., 2007). Here, to examine whether SF3b6 binding causes an allosteric effect, we performed network analysis using webPSN (Felline et al., 2020) for SF3b1 considered in isolation and in complex with SF3b6. The method integrates graph-based protein structure network and elastic network model-based NMA to combine inherent structural communications and their dynamics properties. Based on cross-correlation between two residue motions (≥0.6) and their frequency of association (≥5%) in all possible structural communications between two extreme ends in the structure, the method identifies a metapath (Seeber et al., 2015). Metapath provides information about the most recurrent long-range communication present within a protein.

### 2.4. Structural variability in the constituent proteins of SF3b complex

We screened a number of entries corresponding to spliceosome structures available in the PDB (Berman et al., 2002) to identify yeast and human spliceosomes having an intact SF3b complex. We found that the SF3b complex structures are also available in a spliceosome unintegrated form and in the context of 17S U2 snRNP but only for human (Cretu et al., 2016; Finci et al., 2018; Zhang et al., 2020), and therefore such structures were not considered for comparative studies with yeast. In addition, conformational differences among multiple structures of SF3b proteins in the distinct functional states of spliceosome assembly were examined (Konagurthu et al., 2006; Ye and Godzik, 2004). These structures showed average Cα RMSDs of 1.2 Å for humans and 1.0 Å for yeast SF3b proteins (Table S2). Since conformational differences are small among distinct functional states, two cryo-EM structures (PDB codes: 5GM6 from yeast (Yan et al., 2016) and 5Z58 from human (Zhang et al., 2018)) corresponding to B^act assembly were selected as reference structures for comparing structural features of homologous SF3b complexes. Structural comparison of SF3b homologues was performed by superposing individual proteins as well as the entire SF3b complex (Schrödinger, 2015).

This was performed to determine structural variability contributed by differences in the position and orientation of SF3b proteins within the complex. Here, Cα distance of residue pairs that are aligned in the sequence and superposed in the structure was calculated. Residues showing Cα distance above 2.0 Å with respect to their equivalent residues in the homologue are defined as structurally variable regions. This threshold was decided based on the structural difference observed due to experimental conditions (described in the next section). Further, we used 5 Å distance cut-off to select topologically equivalent residues for the calculation of global Cα RMSDs when superposed in isolation. The same set of residue pairs were considered in the global Cα RMSD calculations when the entire SF3b complex was superposed. Intrinsically disordered regions were predicted using three algorithms (Ishida and Kinoshita, 2007; Jones and Cozzetto, 2015; Mészáros et al., 2018) and a region is considered disordered only when at least two algorithms predict it to be disordered. Sequence similarity and identity between yeast and human SF3b homologues were calculated using a global sequence alignment algorithm (Needleman and Wunsch, 1970).

### 2.4.1. Inherent structural difference in cryo-EM structures

To account for structural differences resulting from different experimental conditions, we retrieved cryo-EM structures of protein macromolecular complexes deposited in the PDB (accessed on November 13, 2019). We selected the available cryo-EM structures which fell within the resolution range of 3.0–6.5 Å. The dataset of 2668 cryo-EM structures was pruned further to obtain pairs of the same protein structures (identical sequence) with the same oligomeric state and same protein composition but obtained from different cryo-EM experiments. Uncertain positions were excluded from the analysis based on the B-factor value (above 200). As a result, 1814 pairs of structures corresponding to 162 unique proteins were compared. We estimated that an average Cα RMSD of proteins among multiple cryo-EM structures of the same assembly is 0.8 Å with a standard deviation of 0.4 Å. Also, the Cα distance between every superposed residue pair was computed for 412908 topologically equivalent positions. We found that the average distance of Cα atoms is 0.7 Å with a standard deviation of 0.6 Å. Hence, we employed 2.0 Å as a threshold to identify genuine structural variability between yeast and human SF3b homologues. Distributions of global RMSDs and Cα distances in this control dataset are provided in Fig. S2.

### 2.5. Recognition of inter-protein, protein-RNA interactions, molecular cavity and calculations of protein interface area

Interactions among SF3b proteins were determined using Protein Interactions Calculator (PIC) (Tina et al., 2007). Interactions between proteins and U2 snRNA/pre-mRNA were obtained from NUCPLOT and hbplus programs (Luscombe et al., 1997; McDonald and Thornton, 1994). V-cleft cavity present in the SF3b3/Rse1 between BPA and BPC domains was analyzed using CASTp (Tian et al., 2018). Inter-protein interface area was calculated using the total solvent accessible area (SA) of each protein partner and protein complex (interface area = SA$_{proteinA}$+SA$_{proteinB}$-SA$_{ABcomplex}$). Solvent accessible surface area calculations were performed using NACCESS program (Hubbard and Thornton, 1993).

## 3. Results and discussion

### 3.1. Overview of sequence divergence in the yeast and human SF3b complexes

SF3b is a conserved multiprotein spliceosomal subcomplex with a varied number of components in humans and yeast. Fig. 1A shows integral components of human and yeast SF3b complexes in which six proteins viz. SF3b1/Hsh155, SF3b2/Cus1, SF3b3/Rse1, SF3b4/Hsh49, SF3b14b/Rds3 and SF3b5/Ysf3 are homologous while the SF3b6 is unique to humans. In the heart-shaped SF3b complex, SF3b1/Hsh155,

SF3b14b/Rds3 and SF3b5/Ysf3 form the core and the remaining proteins occur on the surface. Pairwise-sequence alignments (Needleman and Wunsch, 1970) between six yeast and human SF3b homologues show that the percentage sequence identity lies in the range of 16%–53% with the mean value of 29% (Table S3). Of the six homologues, SF3b14b and Rds3, which are located at the core, share the highest sequence identity (53%), followed by SF3b1 and Hsh155 that interact with U2 snRNA/BPS duplex (37%). SF3b2 and Cus1, which are located at the surface, share the lowest sequence identity and show a high length difference of 459aa. This can be viewed in the light of mean, minimum and maximum sequence identities of 58%, 26% and 100%, respectively observed between 80 conserved yeast and human proteins that make up the ribosome assembly (protein list was retrieved from RPG database (Nakao et al., 2004)). A similar comparison for 29 proteins that form RNA polymerase assemblies (I, II and III) in yeast and human (Cramer et al., 2008) showed sequence identities ranging from 16% to 71% with the mean value of 38%. While SF3b, RNA polymerase and ribosome are essential, we find from the distribution of sequence identities that SF3b proteins have relatively lower sequence conservation than components of the other two complexes. Sequence domain assignment based on Pfam database (Eddy, 2009; El-Gebali et al., 2019) indicates that domain content at the N-terminus of SF3b1 and SF3b2 varies from their corresponding yeast homologues (Fig. 1B). Moreover, three human SF3b proteins viz. SF3b1, SF3b2, SF3b4 are longer and have a higher proportion of intrinsically disordered regions compared to their yeast counterparts (Table S3). Such features hint that these SF3b proteins are relatively versatile and may elicit add-on functional roles (Chen and Moore, 2014). Indeed, SF3b1 and SF3b4 are shown to play other biological functions as well (Sun, 2020).

SF3b6, an additional protein in humans, may have been present in the eukaryotic ancestor (Collins and Penny, 2005). Hence, the loss of SF3b6 in the yeast SF3b complex implies that the protein may have been deselected during evolution. To understand whether the loss of SF3b6 is specific to yeast, we probed for SF3b6 homologues in eukaryotes. Our homology searches (refer Supplementary Text S1 and Table S4) reveal that SF3b6 was not identified in 215 species that cover members of Saccharomyces genus as well as others spanning several lineages of eukaryotes, including a few metazoans and infectious Trypanosoma parasites (Fig. S3A and Table S4B). It is important to note that homologues of the other SF3b proteins are observed in these species. Further, we identified the interaction interface region between SF3b1 and SF3b6 using human SF3b structures (Table S1) and studied the extent of sequence conservation (refer Supplementary Text S2). The interface region resides largely in the N-terminal extension of SF3b1. Our result shows that the N-terminus of SF3b1 has diverged extensively in species that lack SF3b6 such as yeast (Fig. S3B). This suggests that SF3b1 homologues have evolved to counter the loss of SF3b6. Taken together, SF3b6 is absent in 215 eukaryotic species in a lineage-specific manner and the SF3b complex in these organisms may function like their yeast counterpart (Table S4B). Our dataset of their distribution in various eukaryotes will serve as a useful reference set to probe and predict the organization of SF3b complexes across eukaryotes using the currently available 3-D structures of yeast and humans.

### 3.2. SF3b6 influences the global motions of SF3b1

We investigated the role of SF3b6 in the human SF3b complex based on dynamical features. Since SF3b6 primarily interacts with the SF3b1, we analyzed the influence of SF3b6 binding on the intrinsic dynamics of SF3b1. We first compared the available 9 SF3b1 structures and 7 SF3b6 structures representing early stages of spliceosome assemblies (Bertram et al., 2017; Charenton et al., 2019; Cretu et al., 2016; Finci et al., 2018; Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2018) to determine the conformational differences between different functional stages. We observed that Cα RMSDs for SF3b1 and SF3b6 are 1.1 Å (804 residues) and 1.3 Å (97 residues), respectively (Table S2). This indicates that there

is no significant conformational change in SF3b1 and SF3b6 structures among pre-B, pre-catalytic B and B$^{act}$ spliceosome assemblies. We used B$^{act}$ assembly structure as a representative for the analysis since it has structural information for the entire SF3b complex. Missing regions at the SF3b6 binding site of SF3b1 (positions 373–415) were built using comparative modeling. The final *in silico* prepared structure includes SF3b6 binding and HEAT repeat regions (370–1304) of the SF3b1 (refer Materials and Methods).

We performed anisotropic elastic network model-based NMA on SF3b1 in isolation (henceforth referred as SF3b1$^{iso}$) and on SF3b1 bound to SF3b6 (henceforth referred as SF3b1$^{SF3b6-bound}$). Global motions derived from NMA and characterized by low-frequency normal modes represent high collective large scale motions and explain the mechanistic basis of protein functions (Atilgan et al., 2001). Analysis of the two lowest frequency normal modes shows that they represent 'open-to-close' and 'twist' movements in the SF3b1$^{SF3b6-bound}$. We measured square inner product (SIP) for the two lowest frequency modes between SF3b1$^{SF3b6-bound}$ and SF3b1$^{iso}$ to find similarity in the residue mean square fluctuation profiles, where SIP value of 1 implies identical profiles. SIP of 0.2 indicates that the magnitude of residue fluctuations of SF3b1 is remarkably different between the first two global motions in the SF3b6 bound and unbound forms (Fig. 2A). We find that collectivity, a measure to capture the proportion of residues concertedly involved in

each large scale motion, is far lower for SF3b1$^{iso}$ in the first two global motions (0.14 and 0.05) than that for SF3b1$^{SF3b6-bound}$ (0.58 and 0.64) (Fig. 2B). We find that such a poor collectivity in SF3b1$^{iso}$ is on account of large scale motions concentrated only in the N-terminal extension without the involvement of HEAT repeats (Movie S1). This suggests that when SF3b1 is not bound to SF3b6, its N-terminal extension is flexible and large scale motions of HEAT repeats are not observed in the first two global motions. However, SF3b1$^{iso}$ does experience large scale motions of HEAT repeats exhibiting 'open-to-close' and 'twist' movements in the 5th and 6th modes, respectively (Fig. 2B). This observation suggests that the mode preference of these global motions in SF3b1 has changed in the absence of SF3b6. The 'twist' movement in SF3b1$^{iso}$ has collectivity value of 0.62, which is similar to that of SF3b1$^{SF3b6-bound}$ (0.64) indicating that the proportion of residues involved in this motion is comparable between SF3b6 bound and unbound forms (Movie S2).

Supplementary video related to this article can be found at https://doi.org/10.1016/j.crstbi.2021.05.003.

In the case of 'open-to-close' movement, SF3b1$^{iso}$ has a poor collectivity value of 0.24 compared to that of SF3b1$^{SF3b6-bound}$ (0.58). Earlier studies have shown that when SF3b complex is not associated with the spliceosome assembly, such as in the isolated SF3b complex or in the 17S U2 snRNP structures, SF3b1 adopts an 'open' conformation relative to the conformation observed in the spliceosome assemblies (Maji et al.,
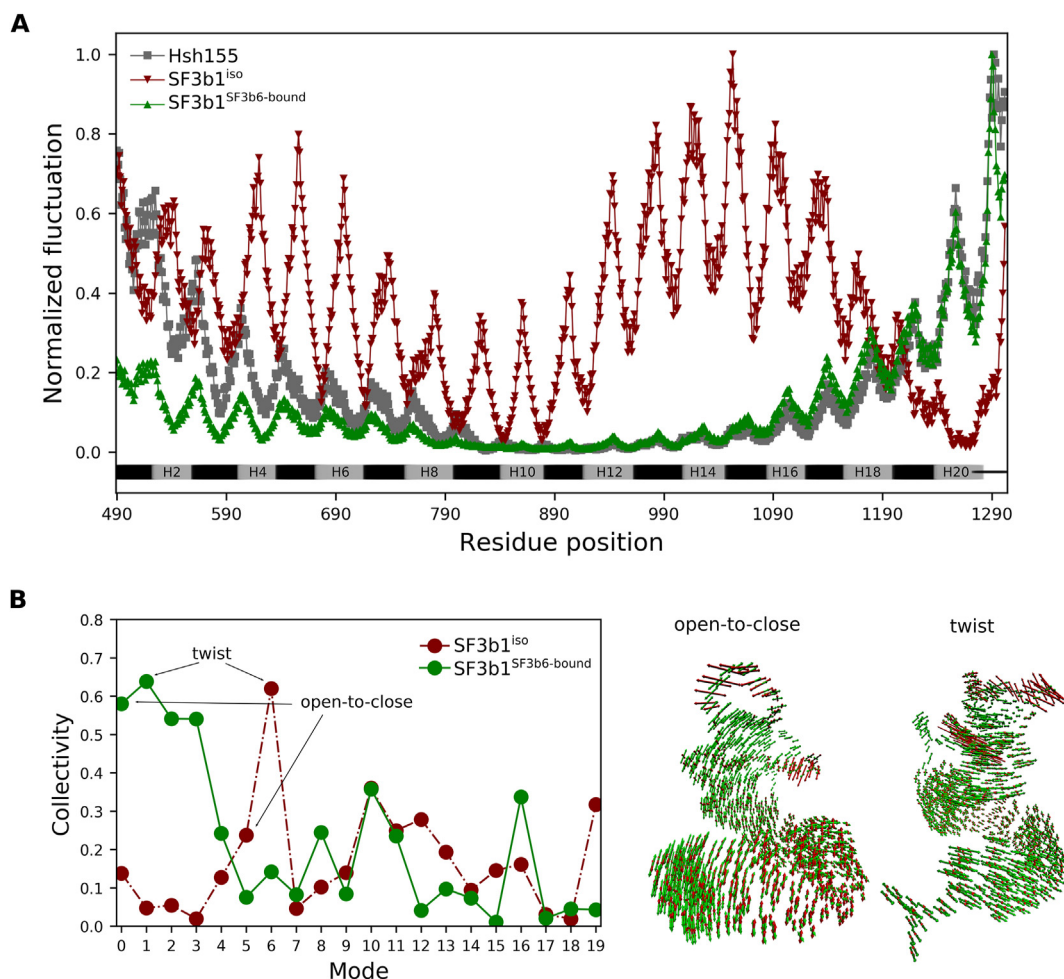


**Fig. 2. Comparison of residue fluctuation profiles among Hsh155 (grey), SF3b1 in isolation (SF3b1$^{iso}$, maroon) and SF3b1 bound to SF3b6 (SF3b1$^{SF3b6-bound}$, green).** Results are shown only for the common region that includes 20 HEAT repeats and C-terminus. A) Residue mean square fluctuations are normalized by the maximum residue fluctuation observed in the protein anisotropic elastic network model and residues corresponding to sequence insertions/deletions are excluded from the comparison. B) Collectivity of top 20 normal modes (left panel) and vector representation (right panel) of the magnitude and direction of residue motions of SF3b1$^{iso}$ (maroon) and SF3b1$^{SF3b6-bound}$ (green) in the first (left) and second (right) global motions. In the collectivity profile, 'open-to-close' and 'twist' movements are labelled.

2019; Rauhut et al., 2016; Zhang et al., 2020). This implies that the 'open-to-close' movement is relevant to the association of SF3b1 with the spliceosome assembly. Upon SF3b6 binding, our study shows that this movement, involving more residue participation is observed to be the lowest frequency global motion for SF3b1.

### 3.3. SF3b6 binding modulates the coordinated inter-residue motions of SF3b1

Inter-residue communications within a tertiary structure are essential for coordinating large scale motions in a protein. We compared inter-residue communications, in 'open-to-close' and 'twist' movements between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$, in terms of coupling in the residue fluctuations represented as cross-correlation matrices. Heatmap in Fig. 3A shows the difference in correlated motions of the Cα atoms for all residue pairs between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$. Positive values ($>0$ to 1) indicate that the correlation of residue motions is stronger in SF3b1$^{SF3b6-bound}$ than in SF3b1$^{iso}$ while negative values ($<0$ to $-1$) indicate the opposite. The value 0 indicates that the strength of correlation for a given residue pair is the same between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$. As shown in the top triangle of Fig. 3A, SF3b6 binding significantly affects the correlation of residue motions in the 'open-to-close' and 'twist' movements of SF3b1. When we extend the analysis to the 20 generated modes, the influence of SF3b6 binding is found to be substantial, not only at the N-terminal extension where it binds but also at many other regions of the 20 HEAT repeats (lower triangle in Fig. 3A). To understand the effect of SF3b6 binding on the U2 snRNA and pre-mRNA interacting regions, we focused on changes in the correlated motions between SF3b6 binding site and RNA binding residues when SF3b6 interacts with SF3b1. Results from the normal modes of SF3b1$^{SF3b6-bound}$ show that residues at the SF3b6 binding site of SF3b1 *viz.* Arg395-Asp405, Lys413-Pro417, Tyr421, Arg425 and Ala428 have higher correlated motions with pre-mRNA binding residues than in SF3b1$^{iso}$ (Fig. 3B). These pre-mRNA binding residues are present in the HEAT repeats of H1, H3, H4 and H6 (Lys496, Leu500, Ala514, Ala515, Gln518, Lys522, Tyr588, Arg590, Glu622, Arg630, Gln699 and Lys700). In contrast, the correlation of the same SF3b6 binding residues with other pre-mRNA binding residues located in the H6, H9 and H15–H19

(Glu714, Arg831, Lys1070-1071, Arg1074-1075, Gln1104-Gln1107, Agr1109, Asn1142, Lys1149, Val1183, Gln1186 and His1225) is lower in SF3b1$^{SF3b6-bound}$ compared to that of SF3b1$^{iso}$. Hence, SF3b6 binding modulates the coordinated inter-residue motions between its binding site and pre-mRNA interacting regions.

Likewise, in the presence of SF3b6, a few U2 snRNA interacting residues (Pro509, Lys513, Leu516, Arg517 and Arg558) located in the H1 and H2 regions experience moderately enhanced correlated motions while residues His1069, Ser1223, Pro1224, Pro1257, Ala1258 and Arg1259 located in the H15, H19 and H20 regions exhibit decreased correlated motions with the SF3b6 binding site (Fig. 3C). Among them, His1069 shows strong correlations with SF3b6 binding residues in SF3b1$^{iso}$ and the strength of the correlations is weakened upon SF3b6 binding. It interacts with a guanine base at the position 31 of the U2 snRNA that base pairs with cytosine of pre-mRNA at position 147 within the U2 snRNA/BPS duplex. We infer from these observations that SF3b6 binding influences the motions of His1069, which could potentially affect its interaction with the U2 snRNA/BPS duplex. Taken together, SF3b6 modulates the magnitude and coherency of residue motions, including pre-mRNA and U2 snRNA binding residues in the SF3b1.

Furthermore, as SF3b1 functions in the spliceosome assembly, we studied the influence of other protein interactions in the context of spliceosome. Here, we first recognized the other spliceosomal proteins within the complex that interact with SF3b1. Based on the maximum distance criteria (7 Å) used in the PIC (Tina et al., 2007) and results from PISA server (Krissinel and Henrick, 2007), we find that SF3b1 interacts with 16 other proteins in addition to SF3b6 in the spliceosome assembly in one or other functional states (Table S5). To account for the effect of these proteins, we systematically studied changes in features of the dynamics of SF3b1-SF3b6 complex in their presence. For this, we compared inter-residue correlated motions of SF3b1 in the binary (SF3b1-SF3b6 complex) and ternary complex (SF3b1-SF3b6-SF3b1 interacting protein complex), which includes SF3b1 interacting proteins. It should be noted that since structures of some of the SF3b1 interacting proteins contain long missing segments (Table S5), we considered domains/regions proximal to SF3b1 alone, to generate reliable elastic network models. Differences in the cross-correlation matrices of SF3b1 between these complexes indicate that Smad1, Dhx16, SF3b3, Bud13, Prp8, Rnf113a,
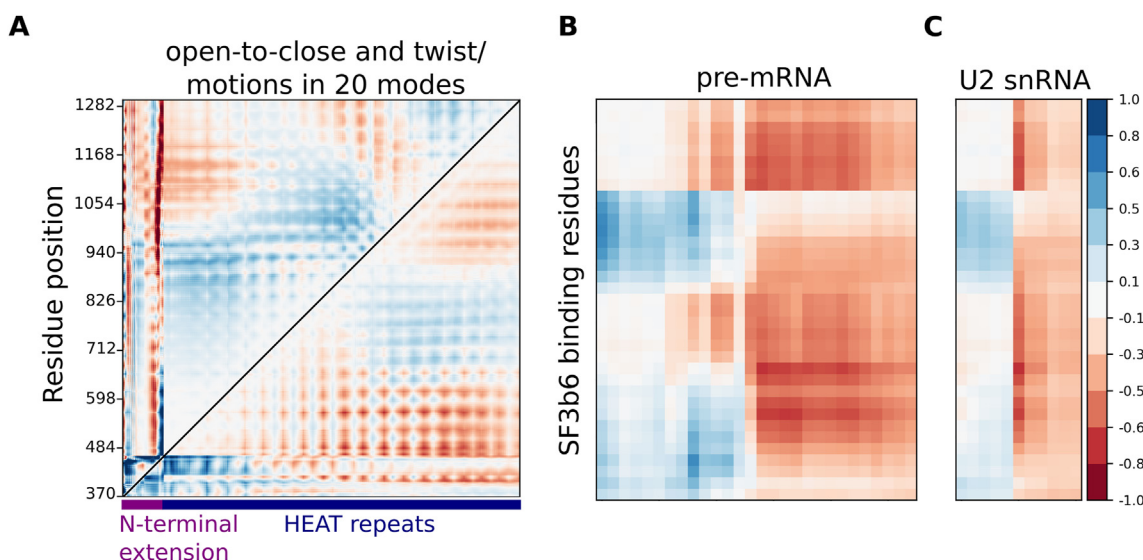


**Fig. 3. SF3b binding modulates inter-residue correlated motions in the SF3b1.** A) Heatmaps show difference in cross-correlation matrices between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$ that represents the strength of correlated motions of Cα atoms for all residue pairs. Left panel shows difference in the cross-correlations matrices of 'open-to-close' and 'twist' movements (upper triangle) as well as difference in the cross-correlation matrices of 20 modes between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$ (lower triangle). B) and C) show difference in the cross-correlation values of SF3b6 binding site with pre-mRNA binding sites and U2 snRNA binding sites between SF3b1$^{iso}$ and SF3b1$^{SF3b6-bound}$. Positive value ($>0$) indicates stronger correlation for a given residue pair in the SF3b1$^{SF3b6-bound}$ compared to SF3b1$^{iso}$ while negative value ($<0$) indicates stronger correlation in the SF3b1$^{iso}$ compared to SF3b1$^{SF3b6-bound}$.

SF3b14b and Snw1 have a substantial effect on the inter-residue correlation motions of SF3b1 (Fig. S4, left panel in the paired heatmaps). Given these effects, we next examined the effect of SF3b6 on the intrinsic dynamics of SF3b1 in the presence of other SF3b1 interacting proteins. This was done by taking the binary complex of SF3b1 bound to SF3b1 interacting protein (SF3b1-SF3b1 interacting protein) and comparing its cross-correlated residue motions with the ternary complex having SF3b6. For example, cross-correlation matrices of SF3b1 were compared in the context of SF3b1-Smad1 and SF3b1-Smad1-SF3b6 complexes. We observed that in most cases, SF3b6 binding significantly alters the correlated residue motions even when another SF3b1 interacting protein is bound (Fig. S4, right panel in the paired heatmaps). Therefore, SF3b6 exerts a strong influence on the intrinsic dynamics of SF3b1 even in the presence of other spliceosomal proteins.

### 3.4. Intrinsic dynamics of SF3b1 is similar to that of yeast Hsh155 when SF3b6 binds at the N-terminus

We have shown that SF3b6 alters residue motions of SF3b1 and facilitates its coherent 'open-to-close' movement. The N-terminal extension of SF3b1 that forms the SF3b6 binding site is absent in the yeast homologue Hsh155 (Fig. S3B). The structural difference between SF3b1 and Hsh155 is observed to be small with Cα RMSD of 1.5 Å when considering 20 HEAT repeats and the C-terminal tail (Table S3). Given the similarity in overall topology, the intrinsic dynamics of these two homologous proteins is expected to be similar when compared in isolation. However, we observed a contrasting result. The magnitude of residue fluctuations of Hsh155 in the first two lowest frequency normal modes, corresponding to 'open-to-close' and 'twist' movements, are closer to that of SF3b1$^{SF3b6-bound}$ (SIP 0.86) as compared to SF3b1$^{iso}$ (SIP 0.21) (Movie S3). This is further evident from the residue mean square fluctuations profile of Hsh155 HEAT repeats H6–H15, which aligns well with that of SF3b1$^{SF3b6-bound}$ (grey line in Fig. 2A). To compare the extent of correlation in the inter-residue motions of Hsh155 to that of SF3b1$^{SF3b6-bound}$ and SF3b1$^{iso}$, we used R$_V$ coefficient, which captures the extent of similarity between two cross-correlation matrices (Smilde et al., 2009). Here again, the R$_V$ coefficient of Hsh155 and SF3b1$^{SF3b6-bound}$ cross-correlation matrices of 20 generated normal modes is 0.94, which is

higher than that of Hsh155 and SF3b1$^{iso}$ (0.89). These results indicate that SF3b6 binding influences SF3b1 such that the extent of similarity in the concerted residue motions of homologous SF3b1 and Hsh155 is increased.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.crstbi.2021.05.003.

Nevertheless, detailed comparisons of the dynamical features of Hsh155 dynamics were also performed and we observed many differences from SF3b1$^{SF3b6-bound}$. Inter-residue correlated matrices reveal considerable differences in the strength of residue cross-correlations between Hsh155 and SF3b1$^{SF3b6-bound}$. For example, inter-residue correlated motions between H1–H5 and H7–H11 are higher in SF3b1$^{SF3b6-bound}$ as compared to Hsh155 (Fig. 4A, upper triangle). In contrast, region H7–H13 exhibits a stronger correlation with H16–H20 and the C-terminal tail of Hsh155 as compared to SF3b1$^{SF3b6-bound}$. Similar differences are also observed when Hsh155 is compared to SF3b1$^{iso}$ (Fig. 4A, lower triangle). Since the N-terminal extension is unique to SF3b1, we extended our studies to probe for its role in the observed differences in the intrinsic dynamics. For this, we used the human SF3b1 structure as a control in which the N-terminal extension and the binding partner SF3b6 are missing (PDB code: 5IFE). Again, we compared the features of dynamics of HEAT repeats between Hsh155 and this structure. The R$_V$ coefficient of cross-correlation matrices between Hsh155 and SF3b1 (without the N-terminal) is 0.99, indicating that inter-residue concerted motions are similar between them. As can be seen in Fig. 4B, the difference between cross-correlation matrices of Hsh155 and SF3b1 (without the N-terminal) is smaller than the differences observed between Hsh155 and SF3b1$^{SF3b6-bound}$/SF3b1$^{iso}$. This finding is further supported by the normalized mean square fluctuation profiles of Hsh155 and SF3b1 (without the N-terminal) that have a SIP value of 0.95, indicating high similarity (Fig. S5). These results demonstrate that the extent of similarity between Hsh155 and SF3b1 (without the N-terminal) is higher than that between Hsh155 and SF3b1$^{SF3b6-bound}$ or SF3b1$^{iso}$. We therefore reckon that the presence of highly flexible N-terminal extension modulates the dynamical nature of superhelical HEAT repeats in SF3b1. When SF3b6 is bound, the flexibility of the N-terminal extension is reduced because of which the SF3b1 motions are similar to those of Hsh155 from yeast which lacks the long N-terminal extension and SF3b6.
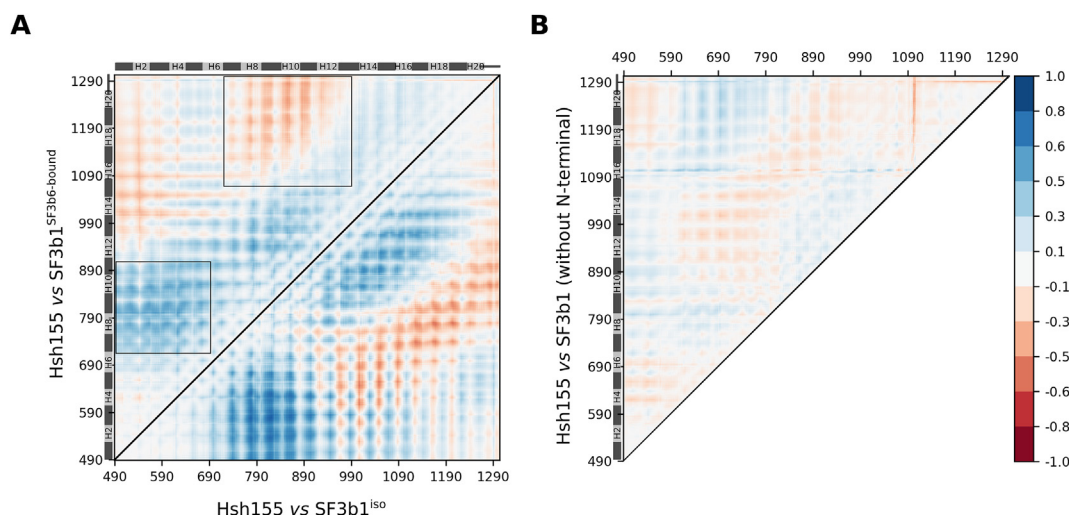


**Fig. 4. Comparison of inter-residue correlation matrices between Hsh155 and SF3b1$^{iso}$/SF3b1$^{SF3b6-bound}$/SF3b1 (without N-terminal).** A) Lower triangle indicates difference in the cross-correlation matrices between Hsh155 and SF3b1$^{iso}$. Upper triangle indicates difference in the cross-correlation matrices between Hsh155 and SF3b1$^{SF3b6-bound}$. Black boxes highlight notable differences in the strength of residue correlations. B) Shown in the upper triangle is difference in the cross-correlation matrices between Hsh155 and SF3b1 (without N-terminal). In both figures, positive value indicates residue correlation is stronger in SF3b1$^{iso}$/SF3b1$^{SF3b6-bound}$/SF3b1 without N-terminal, whereas negative value indicates the residue correlation is stronger in Hsh155. Regions corresponding to 20 HEAT repeats are labelled by numerical numbers.

### 3.5. SF3b6 influences U2 snRNA/BPS duplex binding site in SF3b1: potential relevance in alternative splicing

SF3b1 is directly involved in BPS recognition and hence mutations or small molecules binding in the SF3b1 are known to affect BPS selection, leading to alternative splicing (Alsafadi et al., 2016; Corrionero et al., 2011; Darman et al., 2015; Maguire et al., 2015; Seiler et al., 2018; Zhang and Meng, 2020). Further, SF3b1 contributes to the regulatory mechanism of alternative splicing by serving as a scaffold for the binding of alternative splicing modulators at its N-terminal extension viz. CAPERα, PUF60 and SPF45 (Corsini et al., 2007, 2009, 2009; Loerch et al., 2014). Therefore, any perturbations to SF3b1 may potentially influence BPS selection. Since our results from NMA show that SF3b6 binding impacts the dynamical behaviour of SF3b1, including distal regions from the binding site (Fig. 3), they point to the possibility of SF3b6 playing an allosteric role. To probe for this, we performed perturbation-response scanning analysis on SF3b1. In the analysis, directed forces were applied on SF3b6 binding residues 1000 times and average changes in the atomic positions of all other residues in SF3b1 were calculated to determine effectiveness of the perturbed residues. Effectiveness is a measure to quantify the magnitude and direction of atomic displacement (refer Materials and Methods). Based on the effectiveness profile, residues that are sensitive to SF3b6 binding can be identified. Our results show that among the 21 identified SF3b6 interacting residues, perturbations of 10 residues viz. Asp405-Pro409, Tyr412, Ile472, Asp476, Lys477 and Lys496 elicit maximum responses in the SF3b1 structure (Fig. 5A). Effectiveness of these residues upon perturbations is higher than that of residues adjacent to SF3b6 binding region (Arg390 and Ser400), that were considered as background effectiveness profiles (cyan lines in Fig. 5A). Indeed, these changes are largely reflected in the N- and C-terminus of the superhelical structure formed by the 20 HEAT repeats. The high impact SF3b6 binding residues are clustered near HEAT repeats H1 and H2 (Fig. 5B).

To understand the significance of this effect, we considered perturbation effects at Lys700 in SF3b1. Substitution of lysine to glutamate at 700th position is the most frequently occurring mutation in several cancers and has been shown to drive the selection of alternative BPS (Cretu et al., 2016; Darman et al., 2015). Lys700 interacts non-covalently with the uracil base 158 of pre-mRNA (Fig. S6). Upon glutamate substitution, this interaction is lost and creates unfavourable negatively charged environment for uracil phosphate anion. Such a change might contribute to the altered BPS selection, leading to cancers. Our analysis reveals that perturbation at Lys700 impacts all 20 HEAT repeats in the SF3b1 protein (green line in Fig. 5A), pointing to a critical role for the residue in the network of inter-residue communications. Therefore,

effectiveness profile of Lys700 was taken as a positive control and compared with the effectiveness profiles of SF3b6 binding residues. We observe that perturbations of SF3b6 binding residues exert significant effect on the pre-mRNA binding residues (Lys496, Leu500, Ala514, Ala515, Gln518, Lys522, Tyr588 and Arg590) and U2 snRNA binding residues (Pro509, Lys513, Leu516, Arg517, Arg558 and Pro1257-Arg1259), similar to the effect seen for the Lys700 perturbation (vertical dashed lines in Fig. 5A).

Furthermore, we performed a structural network analysis of SF3b1$^{SF3b6-bound}$, to study the effect of SF3b6 binding in the long-range communication occurring in the SF3b1 (refer Materials and Methods). Analysis of metapath representing the most recurrent long-range communication in the SF3b1 indicates that one of the pre-mRNA interacting residues (Leu500) is involved in long-range communication. When SF3b6 is bound, its interface residues in the SF3b1 viz. Asn396, Phe408, Ile472, Tyr474 and Pro537 participate in the long-range communication (Fig. S7A). Perturbation-response scanning analysis shows that Phe408 and Ile472 have a significant impact on the pre-mRNA and U2 snRNA binding residues. However, when we compare these results with that of SF3b1$^{iso}$, we find that in the absence of SF3b6, these residues do not participate in the long-range communication (Fig. S7B). Hence, SF3b6 binding appears to influence the interface residues into participating in the long-range communication passing through a pre-mRNA interacting residue within the SF3b1.

It has been shown by earlier mutational studies that BPS recognition is largely conserved between yeast and humans and proofreading of BPS selection is likely regulated in multiple stages in the spliceosome (Carrocci et al., 2017; Kaur et al., 2020; Tang et al., 2016). Here, we have investigated the potential role of SF3b6, that is unique to humans and binds at the N-terminal extension of SF3b1, during pre-B to B$^{act}$ stages of spliceosome through various analyses such as NMA, perturbation-response scanning and structural networks. Our studies demonstrate that SF3b6 has a substantial influence on the intrinsic dynamics of SF3b1 and notably on the U2 snRNA/BPS duplex binding site which is distal to SF3b6 binding site. Such effects in other protein complexes have been associated with allosteric regulations (Dobbins et al., 2008; Swapna et al., 2012; Tandon et al., 2021). Therefore, we predict that SF3b6 acts as an allosteric regulator of human SF3b1 that modulate BPS selection during alternative splicing. Further biochemical experiments and transcriptome studies under SF3b6 knockout conditions could shed light on these findings.

### 3.6. Structural variability in homologous SF3b complexes

We next extended our study to understand the inherent variability
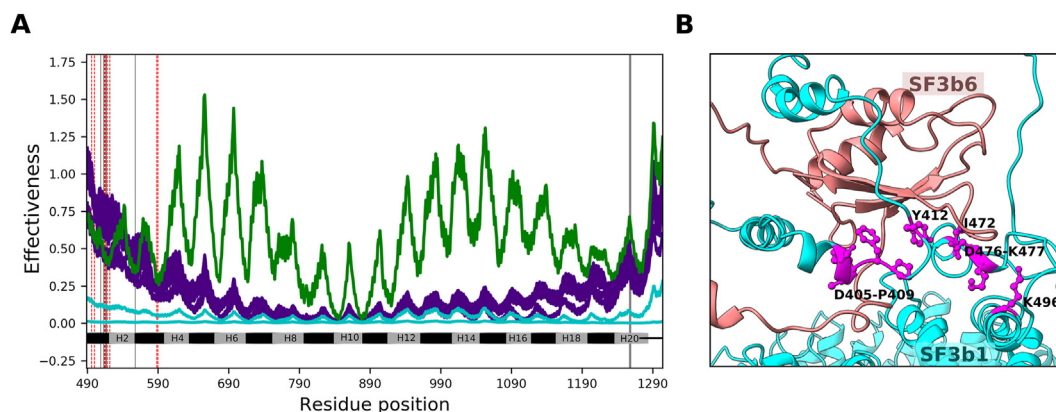


**Fig. 5. Effect of perturbation of SF3b6 binding site on SF3b1.** A) shown are effectiveness profiles of SF3b6 binding residues (purple lines), effectiveness profiles of residues (Arg390 and Ser400) adjacent to SF3b6 binding sites, considered as background signal (cyan lines) and effectiveness profile of Lys700 (green line). Regions of 20 heat repeats are labelled. B) Spatial position of selected SF3b6 binding residues (shown as ball and stick representation in magenta) in the SF3b1 structure whose perturbations exhibit higher effect on the SF3b1 than the background signal as shown in the left panel (A).

present in the SF3b complex by comparing structural features of yeast and human SF3b proteins. To recognize regions of significant structural variability, we carried out a control study in which deviation in the global structure and Cα position of equivalent residues was calculated in multiple structures of the same protein determined through independent cryo-EM studies (refer Materials and Methods). This was done to quantify conformational differences among the cryo-EM structures of a given protein. We found that the distribution of Cα deviation of identical residues among multiple cryo-EM structures of same proteins is $0.7 \pm 0.6$ Å (mean $\pm$ standard deviation). Hence, we used 2 Å as a threshold distance to identify local regions in the cryo-EM structures of SF3b proteins whose structure differ between yeast and human homologues.

The structural comparison shows that although the overall architecture of SF3b complex is conserved (Fig. 6A), only HEAT repeats of SF3b1/Hsh155 and SF3b14b/Rds3 share the highest structural similarity between yeast and human with Cα RMSDs of 1.3 Å and 1.2 Å, respectively (Table S3). This observation emphasizes that in terms of structural features, SF3b1/Hsh155 and SF3b14b/Rds3 are the most conserved proteins among the six common SF3b proteins. Notably, sequences of these two proteins are relatively better conserved between yeast and human than other SF3b proteins. SF3b1/Hsh155 is a large protein that acts as a structural scaffold for the SF3b complex. SF3b14b/Rds3 is located at the core of SF3b complex and its knotted zinc-finger motif interacts with the pre-mRNA segment that succeeds U2 snRNA/BPS duplex (Fig. 6A). Further, we find that the extent of structural similarity is higher between homologues when they are superposed in isolation as compared to when they are superposed in the context of the whole SF3b complex (Fig. 6B). For example, the structural difference between SF3b14b and Rds3 is increased (Cα RMSD 1.6 Å) when superposed them in the context of the whole SF3b complex. Likewise, the Cα RMSD of SF3b3/Rse1 is increased from 3.0 Å to 4.5 Å upon superposing the whole complex. In both the contexts, local variability in the structure was recognized by identifying regions with Cα distances of ≥2 Å between their topologically equivalent residues (refer Materials and Methods). Fig. 6B highlights local regions with structural variation observed by superposing respective homologous structures in isolation (thickened region) as well as by superposing them with the entire complex (colored red). We observed that the HEAT

repeats of SF3b1/Hsh155 are well conserved between yeast and humans with structural deviations in a few loops, N-terminal helix and C-terminal tail. Also, despite being a well-conserved protein, SF3b14b/Rds3 shows a structurally variable region encompassing Cys61 that coordinates with a Zn atom in the SF3b complex (Fig. 6A). Such zinc coordination in the unusual knotted fold of SF3b14b/Rds3 functions as a platform for SF3b assembly (Van Roon et al., 2008). SF3b5/Ysf3, another core protein, shows that 22% of their structure is variable. Together, it indicates that core proteins harbour considerable structural variations between homologous SF3b complexes.

In the case of the surface proteins, the availability of structural coordinates for SF3b2/Cus1 is low and we observe no significant structural variations in these regions. SF3b4/Hsh49 has two RRMs, of which RRM1 shows Cα RMSD of 1.5 Å when compared in isolation. Interestingly, when we superpose the structures in the context of the whole SF3b complex, it shows a large structural deviation having Cα RMSD of 14.6 Å for the same set topologically equivalent residues (Fig. S8A). This implies that the relative orientation of RRM1 differs in the yeast and human SF3b complexes. Further, structural coordinates of RRM2 and a 202aa long disordered C-terminal tail of human SF3b4 are unavailable in the structures of early stages spliceosome assemblies (Table S1). However, the RRM2 structure is available in the isolated 17S U2 snRNP structure (Zhang et al., 2020), suggesting it may adopt flexible conformation when present within the spliceosome assemblies. In contrast, both RRMs of yeast Hsh49 assume a stable and globular structure in the available spliceosome structures including B$^{act}$ assembly (Fig. S8B) suggesting that it adopts a more defined conformation than SF3b4 within the spliceosome assembly. Next, we find that SF3b3/Rse1 shows the largest structural variation covering 68% of the topologically equivalent regions. Given such large structural differences, we performed a detailed assessment of the extent of structural variations in SF3b3/Rse1 in the context of protein-protein interactions and protein association in the assembled form.

### 3.7. SF3b3 and Rse1 interact differently with their SF3b protein partners

SF3b3 comprises of three intertwined seven-bladed β-propeller
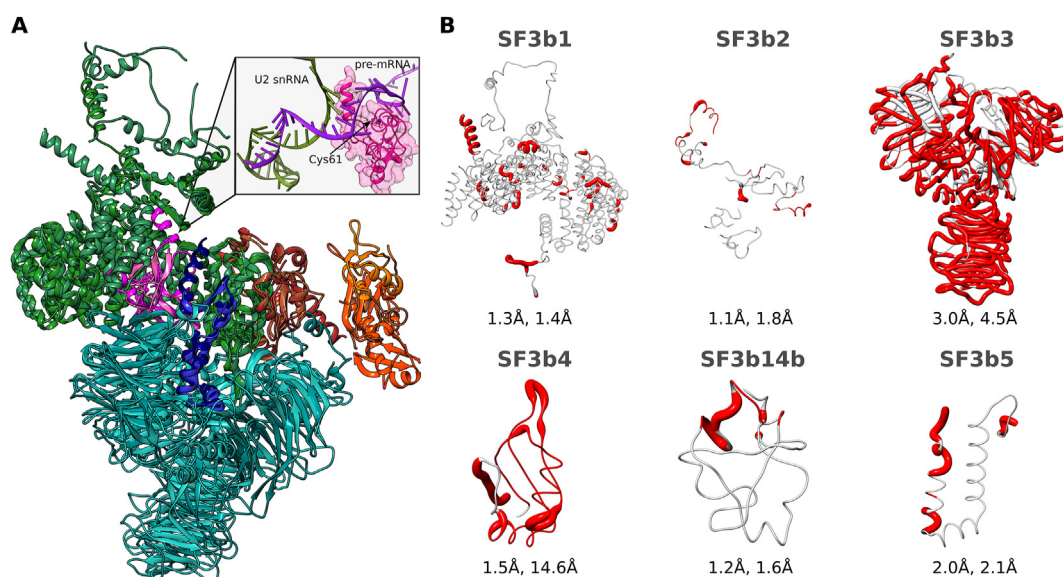


**Fig. 6. Structure comparison of yeast and human SF3b complexes.** A) Superposition of yeast and human SF3b complexes. Color represents different constituent proteins of SF3b and homologous proteins are shown in gradients of the same color. Shown within the box is the interaction of SF3b14b/Rds3 with pre-mRNA that succeeds U2 snRNA/BPS duplex. Cys61 that coordinated with the Zn atom is colored blue (referred yeast structure, PDB code: 5GM6). B) Shown are structural differences between yeast and human SF3b proteins. Regions with Cα distance above 2.0 Å between equivalent residues are highlighted by thickness as well as color of cartoon representation. Thickened region indicates regions with structural differences when superposed in isolation. Color red indicates structural differences observed when superposed along with the whole SF3b complex. Given below the cartoon representations are Cα RMSD values of homologous protein structures when superposed in isolation (left) and with the entire complex (right) separated by ','.

domains referred to as BPA, BPB and BPC that require 6°, 5°, 22° twist, respectively to superimpose well with their equivalents in Rse1 (Movie S4). Since B$^{act}$ structure was chosen as the reference, we analyzed whether the observed structural differences in SF3b3/Rse1 are specific to B$^{act}$ assembly or indicate inherent structural variability. To address this, we examined SF3b3/Rse1 structures from all available early stages spliceosome assembly structures of yeast and humans, as well as human SF3b structure solved independently and not in the context of spliceosome assembly (Bai et al., 2018; Bertram et al., 2017; Charenton et al., 2019; Cretu et al., 2016; Finci et al., 2018; Haselbach et al., 2018; Plaschka et al., 2017, 2018, 2018; Rauhut et al., 2016; Yan et al., 2016; Zhan et al., 2018; Zhang et al., 2018) (Table S1). Here, we find that the centroids of BPA, BPB and BPC domains show 0.8 Å, 9.6 Å and 6.7 Å deviation, respectively, in the overlaid structures of SF3b3 and Rse1 (Fig. 7A). The dihedral angle between BPA and BPC centroids differs by 14° while that between BPB and BPC differs by 6°, suggesting that the positions and orientation of these three domains differ between SF3b3 and Rse1 within the SF3b complex, irrespective of the functional states of spliceosome assembly.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.crstbi.2021.05.003.

Further, analysis of local structures reveals that eight regions vary in the secondary structural conformation between SF3b3 and Rse1 (Fig. 7B). Of these, five regions are associated with sequence insertions/deletions (Fig. S9). Notably, Rse1 has a 15-residue long insertion of helical structure at position 1234 that dislodges a successive helix by ~23 Å with 45° tilt, as compared to its equivalent helix in SF3b3 (Movie S4). These variations result in differences in the size of V-cleft cavity observed at the junction of BPA and BPC domains (Fig. S10) that is occupied by

SF3b5/Ysf3 and the C-terminal tail of SF3b1/Hsh155. Interface areas of SF3b3 with SF3b5 and SF3b1 in the human SF3b complex are larger by 491 Å$^2$ and 531 Å$^2$ than the respective interfaces in the yeast homologue.

The C-terminal tail of SF3b1 has additional interactions with SF3b3 and the interface involves insertions of 13 residues in SF3b3 and a single residue in SF3b1 with reference to their respective yeast homologues (Fig. 7C). Likewise, we observe that the C-terminal region of SF3b5 is disordered while its equivalent region adopts helical conformation in Ysf3 (Fig. 7D, right bottom panel). The interface between SF3b5 and SF3b3 contains a 10-residue insertion in SF3b3 compared to Rse1. To examine the influence of these factors on SF3b3/Rse1 and SF3b5/Ysf3 interaction, we generated an *in silico* chimeric complex where SF3b5 is replaced by Ysf3 using structural superposition within the V-cleft cavity of SF3b3 (Fig. 7D). However, several short contacts involving 9 residues in Ysf3 were observed in the resultant chimeric model identified using WHATIF (Vriend, 1990) (Fig. 7D, right bottom panel). A steric clash was also observed between the Cα atoms of Gly79 in Ysf3 and Ala201 in SF3b3 (Fig. 7D, right top panel). This suggests that Ysf3 as it occurs in yeast, cannot be accommodated within the V-cleft cavity of human SF3b3. Clearly, sequence variations of SF3b3/Rse1 influence the local conformations of its interacting partners.

To quantify the effect of such sequence variation at the interface regions, we calculated residue propensity at the interfaces (refer Supplementary Text S3). We find that cysteine, methionine, phenylalanine, threonine, tryptophan, proline and valine differ in their interface propensities in yeast and human SF3b complexes (Fig. S12). Subsequently, we compared the interaction energies of interfaces to understand the influence of altered residue propensity and unique interaction patterns (refer Supplementary Text S4). As anticipated, we observe that the
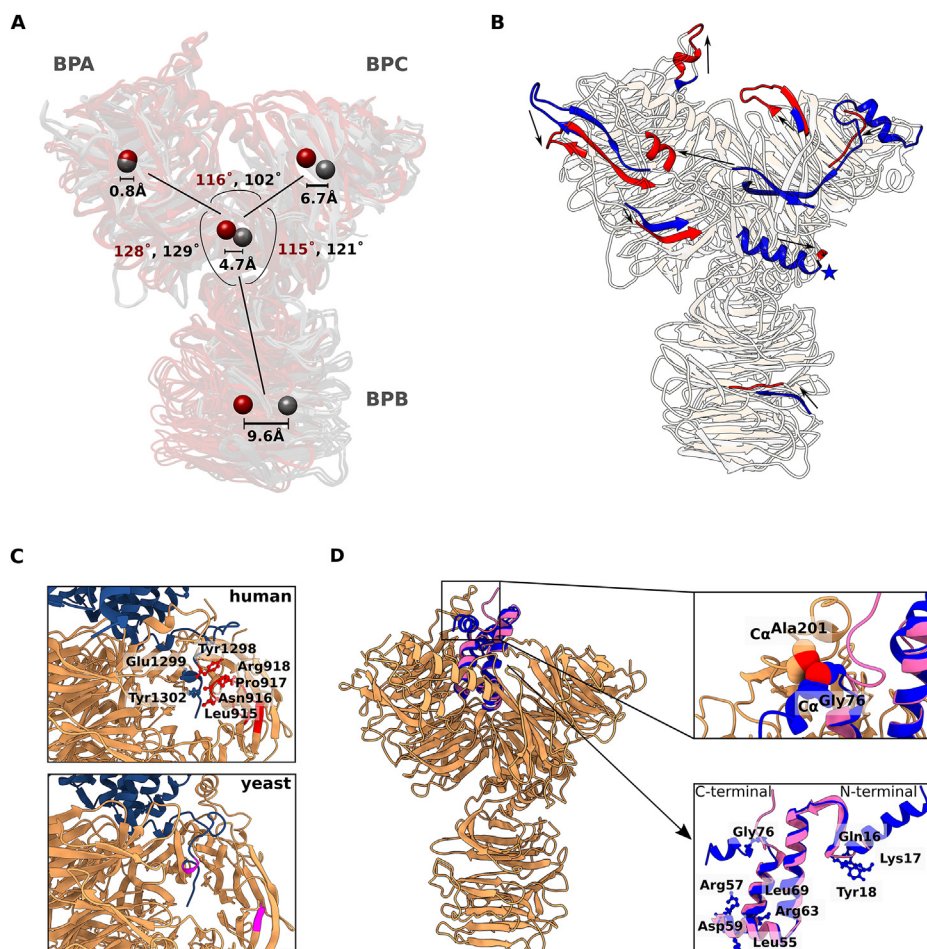


**Fig. 7. Local structural variability in SF3b3/Rse1.** A) Centroids corresponding to BPA, BPB and BPC domains of human SF3b3 (maroon)/yeast Rse1 (grey) are shown with details on centroid distance between respective domains of two homologues and inter-domain angles among three domains. Centroids were defined from multiple structure superposition. For structure superposition, 9 available structures were used for SF3b3 and 5 available structures were used for Rse1. B) Regions with difference in the conformation of secondary structures between SF3b3 (red) and Rse1 (blue) are highlighted and symbol '*' indicates 15aa long helix insertion in Rse1. C) Shown are interactions between C-terminal tail of SF3b1 (navy) and SF3b3 (orange, top panel). Sequence insertion is highlighted in red and interacting residues are shown in ball and stick representation. In bottom panel, the topologically equivalent regions in yeast are highlighted (magenta). D) Cartoon representation of SF3b3 (orange) associated with SF3b5 (pink) and Ysf3 (blue). In the right panel, bottom figure shows the superposed structure of Ysf3 on to SF3b5 and residues that have short contacts with SF3b3 are highlighted in ball and stick representation. Top figure depicts a short contact observed between Cα atoms of Ala201 in SF3b3 and Gly76 in Ysf3.

interaction energies of the interfaces also show differences between SF3b complexes of yeast and humans (Table S6). The observed energy difference is largely contributed by sidechain hydrogen bonds, van der Waals interactions and hydrophobic effect. Collectively, these variations show that inter-protein interactions between these homologous SF3b complexes are tailor-made for each system. Our reports on the characterization of interfaces and species-specific interaction patterns are useful for modeling and determining accurate 3-D structures of homologous SF3b complexes of other species based on human or yeast SF3b structures.

Furthermore, we examined the relevance of such structural variations in the context of spliceosome assembly. In the human B$^{act}$ spliceosome assembly, SF3a3 interacts with BPC domain of SF3b3 sharing interface area of 655 Å$^2$ (Fig. S11A). However, we observed that Prp9 (homologue of SF3a3) is missing in the corresponding yeast B$^{act}$ assembly though the protein is observed in other early stages namely A, pre-B and B of yeast spliceosome assemblies. Likewise, Brr2 closely interacts with BPB domain of Rse1, whereas the homologue (SNRNP200) shows no detectable interaction with SF3b3 in the human B$^{act}$ assembly (Fig. S11B). These observations imply that protein-protein interactions involving SF3b3/Rse1 are also different between yeast and human B$^{act}$ assembly that could influence the local structure of SF3b3/Rse1. Possibly such variations could explain differences in the mechanism of their spliceosome activation, as suggested elsewhere (Kastner et al., 2019). Also, it is worth noting that the nature of interaction protein partners is markedly different between yeast and human SF3b proteins (refer Supplementary Text S5 and Table S7). Given that they perform versatile roles by interacting with a diverse set of protein partners (Sun, 2020), the difference in their protein-protein interaction networks may be a significant factor in conferring unique sequence and structural signatures at the interfaces of SF3b complex, as observed in this study.

Taken together, this investigation highlights local structural variability between homologous yeast and human SF3b complexes. Since protein-protein interfaces are known to be well conserved in evolution (Choi et al., 2009; Mintseris and Weng, 2005; Swapna et al., 2012), our observation of structural variability, altered residue propensity and variations in the interaction strength indicates that interfaces in the SF3b complex have evolved distinctly and considerably. As SF3b is a potential anticancer drug target (Kotake et al., 2007), the interface variations reported here could be exploited for modulating SF3b assembly through small molecules targeting conserved or variable regions for therapeutic purposes (Bonnal et al., 2012; Chen et al., 2017; Rosell and Fernández-Recio, 2018).

## 4. Conclusions

SF3b complex in the spliceosome functions to screen and recognize BPS in the nascent pre-mRNA, one of the key primary steps in splicing that governs the expression of protein isoforms. The complex is ancient in eukaryotes and hence is anticipated to capture adaptive evolutionary changes. This is evident in yeast and human SF3b proteins that differ considerably in their length and sequence domain content. Also, the presence of an additional SF3b6 makes the human SF3b complex unique from its yeast counterpart. In this study, we examined the overall differences between these two homologous complexes with a view to understanding the extent and role of sequence and structural variation in the assembly of SF3b. As a result, we were able to recognize the points of inherent variability between these complexes that could have a direct bearing on the assembly, molecular function and interactions of various proteins within the complex.

Firstly, we show that SF3b proteins are moderately conserved. SF3b6 is absent in all members of the Saccharomyces genus considered here including yeast and in several lineages of eukaryotes. Our normal mode and network analyses demonstrate that SF3b6 promotes 'open-to-close' and 'twist' movements of SF3b1 by enhancing collective residue motions. SF3b6 modulates the motions of its binding site and distal regions,

encompassing residues that interact with U2 snRNA/BPS duplex. Also, upon SF3b6 binding, the interface residues participate in long-range residue communication in the SF3b1. These results point to an allosteric role for SF3b6, with potential implications in alternative splicing, and such a role may be critical for only a specific set of eukaryotes in which it occurs. Further, our comprehensive structural comparison shows that the relative orientation of human SF3b4/yeast Hsh49 is different within the SF3b complex. Sequence indels confer species-specific interface interaction patterns for human SF3b3 and yeast Rse1. Also, interface regions of these two homologous SF3b complexes differ substantially based on the propensities of certain residues and the interaction strength between protein partners. Such variations may help tailor inter-protein interactions of homologous yeast and human SF3b complexes.

Our results on these structural variations will help to select appropriate templates for homology modeling or model fitting in cryo-EM maps, to build more accurate 3-D structures of SF3b complexes in other organisms. Similar analyses on ribosomal proteins revealed that lineage-specific structural variabilities reflect the organism lifestyle and are associated with the repurposing of structural elements for multiple functions (Melnikov et al., 2018). This may well be the case here since SF3b proteins *viz.* SF3b1, SF3b3, SF3b4 and SF3b5 are associated with multiple functions (Sun, 2020) and interact with diverse protein partners. For example, the disordered C-terminal tail is known to mediate SF3b4 interactions with other proteins namely p180 and bone marrow PRI-A, which seem to be absent in yeast (Camacho et al., 2009; Ueno et al., 2019; Watanabe et al., 2007). We find that the presence of C-terminal tail influences the conformation of RRM2 in SF3b4 while such effect is non-existent in Hsh49. Our reports on differences in the 3-D structure and interface regions within the SF3b complex highlight local regions in the proteins that may be associated with their multi-functional roles. Indeed, such findings will guide future structural studies of SF3b through X-ray crystallography or cryo-EM techniques in other organisms. In summary, with the limited available data from yeast and humans, the integrative analysis performed here serves as a starting point to gain insights into evolutionary excursions of the conserved SF3b complex. Further, more biochemical and structural studies from diverse eukaryotes will be required to study the functional implications of SF3b divergence. We believe that similar analysis on other protein systems could deepen our understanding of the functional specialization of homologous complexes/assemblies, which is now feasible with the burgeoning availability of cryo-EM assembly structures.

## CRediT authorship contribution statement

**Arangasamy Yazhini:** Investigation, Formal analysis, Writing – original draft, designed and performed computational investigations and analyzed the results, wrote the first version of the manuscript. **Sankaran Sandhya:** Investigation, Writing – review & editing, analyzed the results, reviewed and edited the manuscript. **Narayanaswamy Srinivasan:** Formal analysis, Project administration, Writing – review & editing, conceived the project, analyzed the results, reviewed and edited the manuscript.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2021.05.003.

## References

Alsafadi, S., Houy, A., Battistella, A., Popova, T., Wassef, M., Henry, E., Tirode, F., Constantinou, A., Piperno-Neumann, S., Roman-Roman, S., et al., 2016. Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. Nat. Commun. 7, 1–12.

Atilgan, C., Atilgan, A.R., 2009. Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein. PLoS Comput. Biol. 5, e1000544.

Atilgan, A.R., Durell, S.R., Jernigan, R.L., Demirel, M.C., Keskin, O., Bahar, I., 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys. J. 80, 505–515.

Bai, R., Wan, R., Yan, C., Lei, J., Shi, Y., 2018. Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation. Science 360, 1423–1429.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., et al., 2002. The protein data bank. Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 899–907.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., Stark, H., 2017. Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell 170, 701–713 e11.

Bonnal, S., Vigevani, L., Valcárcel, J., 2012. The spliceosome as a target of novel antitumour drugs. Nat. Rev. Drug Discov. 11, 847–859.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinf. 10, 421.

Carrocci, T.J., Zoerner, D.M., Paulson, J.C., Hoskins, A.A., 2017. SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic Acids Res. 45, 4837–4852.

Chan, Y.H., Venev, S.V., Zeldovich, K.B., Matthews, C.R., 2017. Correlation of fitness landscapes from three orthologous TIM barrels originates from sequence and structure constraints. Nat. Commun. 8, 1–12.

Charenton, C., Wilkinson, M.E., Nagai, K., 2019. Mechanism of 5' splice site transfer for human spliceosome activation. Science 364, 362–367.

Chen, W., Moore, M.J., 2014. The spliceosome: disorder and dynamics defined. Curr. Opin. Struct. Biol. 24, 141–149.

Chen, K., Xiao, H., Zeng, J., Yu, G., Zhou, H., Huang, C., Yao, W., Xiao, W., Hu, J., Guan, W., et al., 2017. Alternative splicing of EZH2 pre-mRNA by SF3B3 contributes to the tumorigenic potential of renal cancer. Clin. Canc. Res. 23, 3428–3441.

Choi, Y.S., Yang, J.S., Choi, Y., Ryu, S.H., Kim, S., 2009. Evolutionary conservation in multiple faces of protein interaction. Proteins Struct. Funct. Bioinforma. 77, 14–25.

Collins, L., Penny, D., 2005. Complex spliceosomal organization ancestral to extant eukaryotes. Mol. Biol. Evol. 22, 1053–1066.

Corrionero, A., Miñana, B., Valcárcel, J., 2011. Reduced fidelity of branch point recognition and alternative splicing induced by the anti-tumor drug spliceostatin A. Genes Dev. 25, 445–459.

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.

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

Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., et al., 2008. Structure of eukaryotic RNA polymerases. Annu. Rev. Biophys. 37, 337–352.

Cretu, C., Schmitzová, J., Ponce-Salvatierra, A., Dybkov, O., De Laurentiis, E.I., Sharma, K., Will, C.L., Urlaub, H., Lührmann, R., Pena, V., 2016. Molecular architecture of SF3b and structural consequences of its cancer-related mutations. Mol. Cell 64, 307–319.

Darman, R.B., Seiler, M., Agrawal, A.A., Lim, K.H., Peng, S., Aird, D., Bailey, S.L., Bhavsar, E., Chan, B., Colla, S., et al., 2015. Cancer-associated SF3B1 hotspot mutations induce cryptic 3' splice site selection through use of a different branch point. Cell Rep. 13, 1033–1045.

Dermitzakis, E.T., Clark, A.G., 2002. Evolution of transcription factor binding sites in mammalian gene regulatory regions: conservation and turnover. Mol. Biol. Evol. 19, 1114–1121.

Dobbins, S.E., Lesk, V.I., Sternberg, M.J.E., 2008. Insights into protein flexibility: the relationship between normal modes and conformational change upon protein–protein docking. Proc. Natl. Acad. Sci. U.S.A. 105, 10390–10395.

Dziembowski, A., Ventura, A.-P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprévote, O., Séraphin, B., 2004. Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. EMBO J. 23, 4847–4856.

Eddy, S.R., 2009. A new generation of homology search tools based on probabilistic inference. Genome Inform. 23, 205–211.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., et al., 2019. The Pfam protein families database in 2019. Nucleic Acids Res. 47, D427–D432.

Felline, A., Seeber, M., Fanelli, F., 2020. webPSN v2.0: a webserver to infer fingerprints of structural communication in biomacromolecules. Nucleic Acids Res. 48, W94–W103.

van der Feltz, C., Hoskins, A.A., 2019. Structural and functional modularity of the U2 snRNP in pre-mRNA splicing. Crit. Rev. Biochem. Mol. Biol. 54, 443–465.

Finci, L.I., Zhang, X., Huang, X., Zhou, Q., Tsai, J., Teng, T., Agrawal, A., Chan, B., Irwin, S., Karr, C., et al., 2018. The cryo-EM structure of the SF3b spliceosome complex bound to a splicing modulator reveals a pre-mRNA substrate competitive mechanism of action. Genes Dev. 32, 309–320.

Golas, M.M., Sander, B., Will, C.L., Lührmann, R., Stark, H., 2003. Molecular architecture of the multiprotein splicing factor SF3b. Science 300, 980–984.

Gozani, O., Feld, R., Reed, R., 1996. Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev. 10, 233–243.

Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., Stark, H., 2018. Structure and conformational dynamics of the human spliceosomal Bact complex. Cell 172, 454–464 e11.

Hubbard, S.J., Thornton, J.M., 1993. NACCESS. Comput. Progr. 2.

Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res. 35, W460–W464.

Jones, D.T., Cozzetto, D., 2015. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857–863.

Kastner, B., Will, C.L., Stark, H., Lührmann, R., 2019. Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. Cold Spring Harb. Perspect. Biol. 11, a032417.

Kaur, H., Groubert, B., Paulson, J.C., McMillan, S., Hoskins, A.A., 2020. Impact of cancer-associated mutations in Hsh155/Sf3b1 HEAT repeats 9-12 on pre-mRNA splicing in Saccharomyces cerevisiae. PloS One 15, e0229315.

Kim, D.-Y., Stauffacher, C.V., Rodwell, V.W., 2000. Dual coenzyme specificity of Archaeoglobus fulgidus HMG-CoA reductase. Protein Sci. 9, 1226–1234.

Konagurthu, A.S., Whisstock, J.C., Stuckey, P.J., Lesk, A.M., 2006. MUSTANG: a multiple structural alignment algorithm. Proteins Struct. Funct. Genet. 64, 559–574.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., Mizui, Y., 2007. Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nat. Chem. Biol. 3, 570–575.

Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. J. Mol. Biol. 372, 774–797.

Lardelli, R.M., Thompson, J.X., Yates, J.R., Stevens, S.W., 2010. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. RNA 16, 516–528.

Loerch, S., Kielkopf, C.L., 2016. Unmasking the U2AF homology motif family: a bona fide protein-protein interaction motif in disguise. RNA 22, 1795–1807.

Loerch, S., Maucuer, A., Manceau, V., Green, M.R., Kielkopf, C.L., 2014. Cancer-relevant splicing factor CAPERα engages the essential splicing factor SF3b155 in a specific ternary complex. J. Biol. Chem. 289, 17325–17337.

Luscombe, N.M., Laskowski, R.A., Thornton, J.M., 1997. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. Nucleic Acids Res. 25, 4940–4945.

MacMillan, A.M., Query, C.C., Allerson, C.R., Chen, S., Verdine, G.L., Sharp, P.A., 1994. Dynamic association of proteins with the pre-mRNA branch region. Genes Dev. 8, 3008–3020.

Maguire, S.L., Leonidou, A., Wai, P., Marchiò, C., Ng, C.K.Y., Sapino, A., Salomon, A.V., Reis-Filho, J.S., Weigelt, B., Natrajan, R.C., 2015. SF3B1 mutations constitute a novel therapeutic target in breast cancer. J. Pathol. 235, 571–580.

Maji, D., Grossfield, A., Kielkopf, C.L., 2019. Structures of SF3b1 reveal a dynamic Achilles heel of spliceosome assembly: implications for cancer-associated abnormalities and drug discovery. Biochim. Biophys. Acta - Gene Regul. Mech. 1862, 194440.

McDonald, I.K., Thornton, J.M., 1994. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 238, 777–793.

Melnikov, S., Manakongtreecheep, K., Söll, D., 2018. Revising the structural diversity of ribosomal proteins across the three domains of life. Mol. Biol. Evol. 35, 1588–1598.

Mészáros, B., Erdös, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46, W329–W337.

Mintseris, J., Weng, Z., 2005. Structure, function, and evolution of transient and obligate protein-protein interactions. Proc. Natl. Acad. Sci. U.S.A. 102, 10930–10935.

Nakao, A., Yoshihama, M., Kenmochi, N., 2004. RPG: the ribosomal protein gene database. Nucleic Acids Res. 32, D168–D170.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

Perea, W., Schroeder, K.T., Bryant, A.N., Greenbaum, N.L., 2016. Interaction between the spliceosomal pre-mRNA branch site and U2 snRNP protein p14. Biochemistry 55, 629–632.

Plaschka, C., Lin, P.C., Nagai, K., 2017. Structure of a pre-catalytic spliceosome. Nature 546, 617–621.

Plaschka, C., Lin, P.C., Charenton, C., Nagai, K., 2018. Prespliceosome structure provides insights into spliceosome assembly and regulation. Nature 559, 419–422.

Povolotskaya, I.S., Kondrashov, F.A., 2010. Sequence space and the ongoing expansion of the protein universe. Nature 465, 922–926.

Query, C.C., Strobel, S.A., Sharp, P.A., 1996. Three recognition events at the branch-site adenine. EMBO J. 15, 1392–1402.

Rauhut, R., Fabrizio, P., Dybkov, O., Hartmuth, K., Pena, V., Chari, A., Kumar, V., Lee, C.T., Urlaub, H., Kastner, B., et al., 2016. Molecular architecture of the Saccharomyces cerevisiae activated spliceosome. Science 353, 1399–1405.

Van Roon, A.M.M., Loening, N.M., Obayashi, E., Yang, J.C., Newman, A.J., Hernández, H., Nagai, K., Neuhaus, D., 2008. Solution structure of the U2 snRNP protein Rds3p reveals a knotted zinc-finger motif. Proc. Natl. Acad. Sci. U.S.A. 105, 9621–9626.

Rosell, M., Fernández-Recio, J., 2018. Hot-spot analysis for drug discovery targeting protein-protein interactions. Expet Opin. Drug Discov. 13, 327–338.

Šali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Lührmann, R., Glover, J.N.M., MacMillan, A.M., 2006. Crystal structure of a core spliceosomal protein interface. Proc. Natl. Acad. Sci. U.S.A. 103, 1266–1271.

Schellenberg, M.J., Dul, E.L., MacMillan, A.M., 2011. Structural model of the p14/SF3b155·branch duplex complex. RNA 17, 155–165.

Schrödinger, 2015. The PyMol (TM) Moelcular Graphics System. 1 Schrödinger, LLC, Version 1.7.2.

Seeber, M., Felline, A., Raimondi, F., Mariani, S., Fanelli, F., 2015. WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. Bioinformatics 31, 779–781.

Seiler, M., Peng, S., Agrawal, A.A., Palacino, J., Teng, T., Zhu, P., Smith, P.G., Caesar-Johnson, S.J., Demchok, J.A., Felau, I., et al., 2018. Somatic mutational landscape of splicing factor genes and their functional consequences across 33 cancer types. Cell Rep. 23, 282–296 e4.

Smilde, A.K., Kiers, H.A.L., Bijlsma, S., Rubingh, C.M., Van Erk, M.J., 2009. Matrix correlations for high-dimensional data: the modified RV-coefficient. Bioinformatics 25, 401–405.

del Sol, A., Araúzo-Bravo, M.J., Amoros, D., Nussinov, R., 2007. Modular architecture of protein structures and allosteric communications: potential implications for signaling proteins and regulatory linkages. Genome Biol. 8, R92.

Spadaccini, R., Reidt, U., Dybkov, O., Will, C., Frank, R., Stier, G., Corsini, L., Wahl, M.C., Lührmann, R., Sattler, M., 2006. Biochemical and NMR analyses of an SF3b155-p14-U2AF-RNA interaction network involved in branch point definition during pre-mRNA splicing. RNA 12, 410–425.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., et al., 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr. Protoc. Bioinforma. 54, 1.30.1–1.30.33.

Sun, C., 2020. The SF3b complex: splicing and beyond. Cell. Mol. Life Sci. 1–13.

Swapna, L.S., Mahajan, S., De Brevern, A.G., Srinivasan, N., 2012. Comparison of tertiary structures of proteins in protein-protein complexes with unbound forms suggests prevalence of allostery in signalling proteins. BMC Struct. Biol. 12, 6.

Tandon, H., de Brevern, A.G., Srinivasan, N., 2021. Transient association between proteins elicits alteration of dynamics at sites far away from interfaces. Structure 29, 371–384.

Tang, Q., Rodriguez-Santiago, S., Wang, J., Pu, J., Yuste, A., Gupta, V., Moldón, A., Xu, Y.Z., Query, C.C., 2016. SF3B1/Hsh155 HEAT motif mutations affect interaction with the spliceosomal ATPase Prp5, resulting in altered branch site selectivity in pre-mRNA splicing. Genes Dev. 30, 2710–2723.

Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. Science 278, 631–637.

Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., 2018. CASTp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Res. 46, W363–W367.

Tina, K.G., Bhadra, R., Srinivasan, N., 2007. PIC: protein interactions calculator. Nucleic Acids Res. 35, W473–W476.

Ueno, T., Taga, Y., Yoshimoto, R., Mayeda, A., Hattori, S., Ogawa-Goto, K., 2019. Component of splicing factor SF3b plays a key role in translational control of polyribosomes on the endoplasmic reticulum. Proc. Natl. Acad. Sci. U.S.A. 116, 9340–9349.

Vriend, G., 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. 8, 52–56.

Wang, Q., Rymond, B.C., 2003. Rds3p is required for stable U2 snRNP recruitment to the splicing apparatus. Mol. Cell Biol. 23, 7339–7349.

Watanabe, H., Shionyu, M., Kimura, T., Kimata, K., Watanabe, H., 2007. Splicing factor 3b subunit 4 binds BMPR-IA and inhibits osteochondral cell differentiation. J. Biol. Chem. 282, 20728–20738.

Will, C.L., Lührmann, R., 2011. Spliceosome structure and function. Cold Spring Harb. Perspect. Biol. 3, 1–2.

Will, C.L., Schneider, C., Reed, R., Lührmann, R., 1999. Identification of both shared and distinct proteins in the major and minor spliceosomes. Science 284, 2003–2005.

Will, C.L., Schneider, C., Macmillan, A.M., Katopodis, N.F., Neubauer, G., Wilm, M., Lührmann, R., Query, C.C., 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. EMBO J. 20, 4536–4546.

Yan, C., Wan, R., Bai, R., Huang, G., Shi, Y., 2016. Structure of a yeast activated spliceosome at 3.5 Å resolution. Science 353, 904–912.

Ye, Y., Godzik, A., 2004. FATCAT: a web server for flexible structure comparison and structure similarity searching. Nucleic Acids Res. 32, W582–W585.

Yoshida, K., Ogawa, S., 2014. Splicing factor mutations and cancer. Wiley Interdiscip. Rev. RNA 5, 445–459.

Zhan, X., Yan, C., Zhang, X., Lei, J., Shi, Y., 2018. Structures of the human pre-catalytic spliceosome and its precursor spliceosome. Cell Res. 28, 1129–1140.

Zhang, D., Meng, F., 2020. A comprehensive overview of structure-activity relationships of small-molecule splicing modulators targeting SF3B1 as anticancer agents. ChemMedChem 15, 2098–2120.

Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., Shi, Y., 2018. Structure of the human activated spliceosome in three conformational states. Cell Res. 28, 307–322.

Zhang, Z., Will, C.L., Bertram, K., Dybkov, O., Hartmuth, K., Agafonov, D.E., Hofele, R., Urlaub, H., Kastner, B., Lührmann, R., et al., 2020. Molecular architecture of the human 17S U2 snRNP. Nature 583, 310–313.