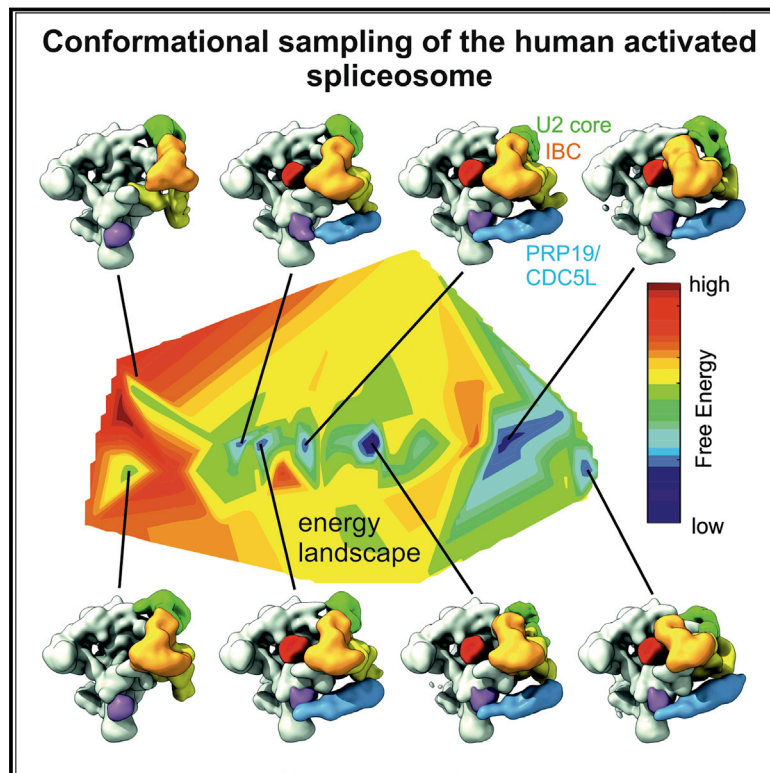


# Structure and Conformational Dynamics of the Human Spliceosomal B<sup>act</sup> Complex

## Graphical Abstract



## Authors

David Haselbach, Ilya Komarov, Dmitry E. Agafonov, ..., Berthold Kastner, Reinhard Lührmann, Holger Stark

## Correspondence

reinhard.luehrmann@mpi-bpc.mpg.de (R.L.),  
hstark1@gwdg.de (H.S.)

## In Brief

A new approach to analyzing cryo-EM data reports on conformational dynamics in the human spliceosome.

## Highlights

- High-resolution cryo-EM structure of the human B<sup>act</sup> spliceosome.
- Energy landscape reveals low energy barriers between conformations
- Thermal energy can drive large conformational changes of the spliceosome
- Large conformational changes are required during spliceosome assembly



# Structure and Conformational Dynamics of the Human Spliceosomal B<sup>act</sup> Complex

David Haselbach,<sup>1,5</sup> Ilya Komarov,<sup>2</sup> Dmitry E. Agafonov,<sup>2</sup> Klaus Hartmuth,<sup>2</sup> Benjamin Graf,<sup>1</sup> Olexandr Dybkov,<sup>2</sup> Henning Urlaub,<sup>3,4</sup> Berthold Kastner,<sup>2</sup> Reinhard Lührmann,<sup>2,\*</sup> and Holger Stark<sup>1,6,\*</sup>

<sup>1</sup>Department for Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>2</sup>Department for Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>3</sup>Bioanalytic Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>4</sup>Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, Robert-Koch-Str. 40, 37073 Göttingen, Germany

<sup>5</sup>Present address: Research Institute of Molecular Pathology, Campus-Vienna-Biocenter 1, 1030 Vienna, Austria

<sup>6</sup>Lead Contact

\*Correspondence: reinhard.luehrmann@mpi-bpc.mpg.de (R.L.), hstark1@gwdg.de (H.S.)

<https://doi.org/10.1016/j.cell.2018.01.010>

## SUMMARY

The spliceosome is a highly dynamic macromolecular complex that precisely excises introns from pre-mRNA. Here we report the cryo-EM 3D structure of the human B<sup>act</sup> spliceosome at 3.4 Å resolution. In the B<sup>act</sup> state, the spliceosome is activated but not catalytically primed, so that it is functionally blocked prior to the first catalytic step of splicing. The spliceosomal core is similar to the yeast B<sup>act</sup> spliceosome; important differences include the presence of the RNA helicase aquarius and peptidyl prolyl isomerases. To examine the overall dynamic behavior of the purified spliceosome, we developed a principal component analysis-based approach. Calculating the energy landscape revealed eight major conformational states, which we refined to higher resolution. Conformational differences of the highly flexible structural components between these eight states reveal how spliceosomal components contribute to the assembly of the spliceosome, allowing it to generate a dynamic interaction network required for its subsequent catalytic activation.

## INTRODUCTION

Pre-mRNA splicing is catalyzed by the spliceosome, a highly dynamic and complex molecular machine that contains pre-mRNA, the U1, U2, U4/U6, and U5 snRNPs, and numerous non-snRNP proteins. Interactions of U1 and U2 snRNPs with the 5' splice site (ss) and branch site (BS) of the intron, respectively, pave the way for the stable integration of the U4/U6.U5 tri-snRNP, yielding the spliceosomal B complex (Wahl et al., 2009). However, although the tri-snRNP brings in the necessary catalytic components, the spliceosomal B complex still needs to be converted into a catalytically active complex, a process that requires extensive structural rearrangements. Activation is initiated by unwinding the U4/U6 duplex (by the RNA helicase Brr2), which destabilizes U4 snRNA; U6 then is free to form short duplexes with U2 (U2/U6 helix Ia and helix Ib) as well as an internal stem loop (U6 ISL),

which together coordinate the metal ions important for splicing catalysis (Fica et al., 2013; Galej et al., 2016; Hang et al., 2015; Yan et al., 2017). The catalytic RNA-RNA network that is ultimately established during spliceosome activation is very similar to the catalytic core of group II self-splicing introns (Fica et al., 2014; Hang et al., 2015). The resulting activated but pre-catalytic B (B<sup>act</sup>) complex is subsequently converted into a catalytic spliceosome (designated B\*) by the RNA helicase PRP2 (Kim and Lin, 1996). In humans, the B-to-B\* transition also requires the ATPase activity of the AQR RNA helicase (De et al., 2015), which is not found in spliceosomes from the yeast *S. cerevisiae*. This suggests that the conformational rearrangements required for catalytic activation in human spliceosomes are more complex than those in yeast. The B\* complex catalyzes the first step of splicing to generate a cleaved 5' exon and intron-3' exon lariat intermediates, generating the spliceosomal C complex. Additional RNP rearrangements convert the C complex into the C\* complex, which then catalyzes the second step of splicing, and ligation of the 5' and 3' exons forms the mRNA and releases the intron as a lariat.

Although yeast and human spliceosomes share a common set of evolutionarily-conserved core spliceosomal proteins, the human spliceosomal proteins typically contain additional regions that are often unstructured (Fabrizio et al., 2009; Korneta and Bujnicki, 2012). Spliceosomes in humans also contain many additional proteins not present in *S. cerevisiae* (Agafonov et al., 2011; Fabrizio et al., 2009; Jurica and Moore, 2003; Wahl et al., 2009). The protein composition of the purified spliceosome is highly dynamic, with about 30 proteins released during the B-to-B<sup>act</sup> transition, including the U4/U6-specific proteins, the LSm proteins, and the so-called B-specific proteins (Agafonov et al., 2011; Bessonov et al., 2010). At least 35 proteins are recruited or stably integrated during B<sup>act</sup> formation in higher eukaryotes, including proteins of the PRP19/CDC5L complex and the so-called PRP19-related proteins (Makarov et al., 2002). The yeast equivalent of the human PRP19/CDC5L complex, the NTC, is critical for spliceosome activation, with NTC proteins stabilizing the RNA network of the spliceosome during activation as well as at later stages (Bertram et al., 2017a; Chan et al., 2003; Galej et al., 2016; Hang et al., 2015; Rauhut et al., 2016; Wan et al., 2016; Yan et al., 2016). Several additional B<sup>act</sup> proteins are required during spliceosome activation; in humans, some



of these essential proteins are not present in the yeast B<sup>act</sup> complex, such as most peptidyl-prolyl isomerases.

Recently, several structures of spliceosomal complexes from yeast and human have been determined by single-particle cryo-electron microscopy (cryo-EM) at resolutions that allow atomic model building: tri-snRNP (Agafonov et al., 2016; Nguyen et al., 2015, 2016), B-complex (Bertram et al., 2017b; Plaschka et al., 2017), B<sup>act</sup>-complex (Rauhut et al., 2016; Yan et al., 2016), C-complex (Galej et al., 2016; Wan et al., 2016), C\*-complex (Bertram et al., 2017a; Yan et al., 2017), ILS-Complex (Hang et al., 2015; Wan et al., 2017), and the P-complex (Bai et al., 2017; Liu et al., 2017; Wilkinson et al., 2017). The reported resolution levels are typically not uniform but rather vary significantly over the entire spliceosome. The best-resolved structural parts of the spliceosome in all available structures consist mainly of U5 snRNP components and the well-defined and stable catalytic U2/U6 RNP core comprising RNA and protein components. In contrast, regions that are more peripheral are usually determined at lower-resolution levels. This reflects the highly dynamic nature of the spliceosome that is required for the many remodeling steps during spliceosome assembly, in which a large number of components need to adapt their binding strength and their binding interactions to the spliceosome to enable the required rearrangements. For instance, BRR2 is relocated by ~250 Å from its binding site on the tri-snRNP to its site in the spliceosome at a later state during catalytic activation (Bertram et al., 2017b; Rauhut et al., 2016). A similar rearrangement can be seen for the U2 core, which moves by 175 Å during the B<sup>act</sup>-to-C transition (Bertram et al., 2017a; Galej et al., 2016; Wan et al., 2016). Such rearrangements require a very dynamic network of interactions between numerous spliceosome components and a coordinated dynamic behavior of these components during spliceosome assembly. Most importantly, the observed motions must be maintained under strictest control to guarantee productive spliceosome formation and functionality; nonetheless, until now, how the spliceosome controls these tasks remains entirely unknown. Indeed, it has been a major technical challenge to investigate this particularly dynamic behavior for 3D structure determination, as mobile structural components are usually poorly defined in the calculated 3D structure. Some highly mobile parts are even entirely absent from 3D density maps, and roughly a third of the total mass of spliceosome complexes is not even visible in the refined density maps.

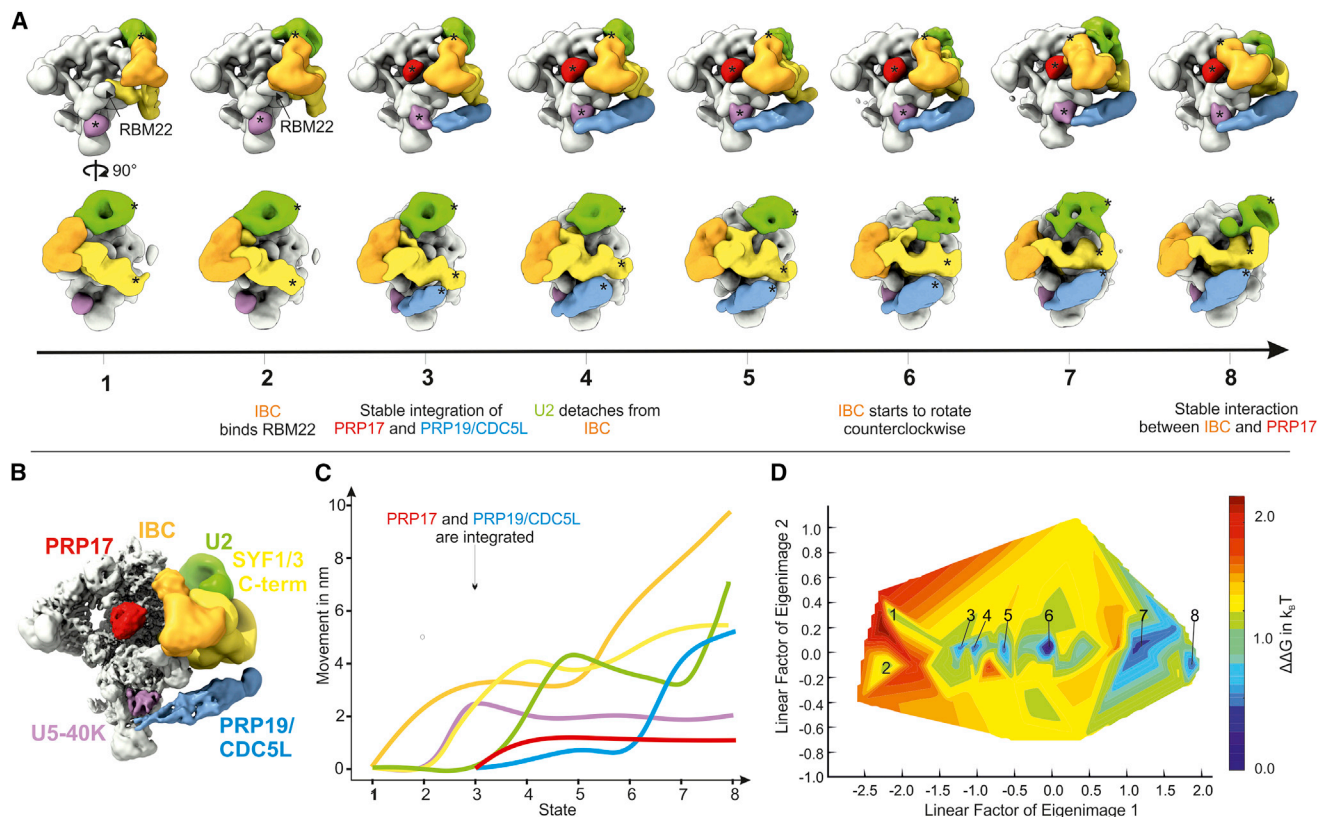
This extensive conformational variability of the spliceosome is almost certain to be key to its functionality. Therefore, we initially determined the structure of the human B<sup>act</sup> spliceosome at high resolution (Figure S1). Subsequently, we addressed the dynamic structural changes in a quantitative manner, by combining an iterative 3D classification approach with a newly developed method based on three-dimensional principal component analysis (3D PCA) (Haselbach et al., 2017). After extensive 3D classification, we obtained >100 structures of the spliceosome, which become ordered according to their conformational similarities by applying PCA. We were thereby able to identify the motion patterns of flexible components in a conformational landscape and to describe the key modes of motion of the mobile parts of the spliceosome, which almost completely cover a conformational continuum that consists of numerous major and minor confor-

mational states. Counting the particle numbers for each of the observed conformations even allows the conformational landscape to be converted into an energy landscape (according to Boltzmann) that quantitatively describes the low-energy barriers between the dynamic conformational states sampled by the spliceosome. All conformations of the spliceosome become ordered with respect to conformational similarities by applying PCA, suggesting that progression in the spliceosome assembly direction happens by going from left to right on the energy landscape along the first eigenvector. It is important to emphasize, however, that this does not imply that the spliceosome strictly undergoes these conformational transitions in an ordered manner even though they become increasingly similar to later spliceosomal states in the same direction. Importantly, the low energy barriers between the states renders the assembly pathway stochastic in nature. This implies that a unique assembly pathway does not exist as spliceosome formation can occur by “sampling” any major and minor state along the energy landscape. Critically, PCA analysis of these distinct structures revealed that productive progress during active spliceosome formation requires numerous protein and RNA interactions to be accurately fine-tuned, and often in highly coordinated manner, to enable the stable structural accommodation of functionally relevant protein components. Only the effective interplay of the spliceosome compositional properties and the ability of sampling motions permit the stable binding and accommodation of the protein components required for formation of a catalytically functional spliceosome. Our data thus reveal both how and why the overall conformation flexibility of the spliceosome is required for productive formation of a catalytically active complex.

## RESULTS

### Overall Conformational Dynamics of the B<sup>act</sup> Complex

Human B<sup>act</sup> spliceosome was affinity purified from a splicing reaction in HeLa nuclear splicing extracts with a minimized pre-mRNA substrate (containing a single 5' exon and an intron with a BS and 10 nucleotides [nts] downstream of the BS, but no 3' ss or 3' exon). This construct allows the B<sup>act</sup> complex to be assembled but not catalytically activated (due to its short polypyrimidine tract) (Bessonov et al., 2010) (see Figure S4). Purified human B<sup>act</sup> complex was then analyzed by cryo-EM to give a large dataset of 2.2 million individual particle images, from which we determined a high-resolution (3.4 Å) 3D structure of the human B<sup>act</sup> spliceosomal complex. Additionally, extensive computational sorting of images allowed us to determine numerous conformational states that are present simultaneously in the same sample. As expected, the spliceosomal core could be determined at high resolution throughout all eight major conformational states (Figures S1D, S5, and S6). In stark contrast, the peripheral parts vary strongly in their conformation and have less resolved 3D structures, depending on their level of mobility. It is important to note that the calculated structures are the result of the computational sorting of all images in the dataset, and therefore all those structures are simultaneously present in the same sample. This is an advancement over standard EM image processing strategies, in which only a selected subset of images



**Figure 1. Conformational Dynamics of the Human B<sup>act</sup> Spliceosome**

The energy landscape of the hB<sup>act</sup> spliceosome, describing the full range of major and minor conformational states detected, shown as a 2D plot along the two major modes of motion.

(A) Eight different conformational states of the hB<sup>act</sup> spliceosome were obtained by computational sorting of images. The states 1–8 were sorted for conformational similarity by PCA. Although core structures were determined at resolution levels of 3.6–4.5 Å, low-pass filtered versions of the unmasked full maps are shown here for improved visualization of conformational differences and to highlight the flexible parts of the spliceosome. The segmented subcomplexes undergoing the most significant global conformational changes are described in (B).

(B) An overview of the entire complex in state 8 with the same segmentation as in (A). The map was sharpened according to its local resolution.

(C) Quantification of the movements of the individual segments shown in (A) reveals a large motion parameter for components, which varies between 1–10 nm. The reference position to measure the relative movements is marked with an asterisk.

(D) The landscape describes the energy differences between all the conformational states found by computational classification. The coordinates describe the main conformational differences along the two most significant eigenvectors derived from a PCA of all 3D volumes. The energies were calculated according to the Boltzmann equation using the known number of individual particles in each volume. While there are distinct energy minima visible for each state (number labels), the overall energy landscape is flat and does not show energy barriers higher than 3 k<sub>B</sub>T, which can easily be overcome by the thermal noise power of the surrounding medium. No statistical significant populations of particles were observed in the white regions of the energy landscape indicating higher energy barriers.

contributes to the high-resolution structure. In contrast, by analyzing the entire dataset of particle images, we were able to gain more detailed insight into the variable parts of the macromolecular complex as well as into the conformational variability that is covered in solution. The observed conformational changes fit well with the considerable variation in protein binding and known integration of protein components into the spliceosome. The variable structural domains with the largest degree of conformational flexibility consist mostly of the PRP19/CDC5L complex (comprising four copies of PRP19, CDC5L, SPF27, PLRG1, CWC15, and CTNNBL1), the intron-binding complex (IBC, comprising SYF1, AQR, PPIE, ISY1, and ZNF830), the U2 core RNP (comprising U2 snRNA, U2 Sm,

U2-A', U2-B''), large parts of the U2 SF3A, and the proteins PRP17, SYF3, PPIL1, PPIL2, and U5-40K (Figures 1A–1C).

Although the overall energy landscape is relatively flat (Figure 1D), a few conformational states of the spliceosome are more populated at 4°C, which allowed us to determine higher-resolution structures for these populations. We determined eight structures of thermodynamically more stable conformational states (major states) that provided sufficient image statistics for high-resolution structure determination of the spliceosomal core. Notably, while the U2/U5/U6/pre-mRNA RNP core structure was highly similar for all, the composition and conformation of their peripheral regions differed significantly within the distinct major states (Figure 1A and Movie S1).

Our 3D PCA analysis of the entire conformational landscape further revealed numerous additional minor states that are less frequently sampled in solution (and therefore cannot be easily determined at high resolution), which nonetheless contribute to the energy landscape (shown as a 2D plot along the major modes of motion) (Figure 1D). This energy landscape can be considered as a plausible overall assembly pathway (from left to right) on which continuous conformational changes lead to an increased stabilization and accommodation of proteins that are known to be important for later splicing steps. We can thus define the forward assembly pathway from  $B^{\text{act}}$  to the catalytic active  $B^*$  spliceosome by comparing the eight major states with the conformation of the later (catalytically active)  $C^*$  spliceosome (Bertram et al., 2017a): along the path from state 1 to state 8, the  $B^{\text{act}}$  spliceosome adopts conformations that are increasingly similar to that of the  $C^*$  complex. Thus, the productive conformational changes required for  $B^*$  spliceosome formation can be interpreted as the spliceosome moving from left to right (states 1–8) on the energy landscape (Figure 1D). Boltzmann interpolation reveals energy barriers between the various states of maximum 3  $k_B T$ , revealing that the transition from one state to the other can be overcome energetically with the available thermal energy. In other words, despite the very large conformational differences between the observed states, none of the motions strictly requires additional chemical energy in terms of ATP/GTP hydrolysis (which are indeed absent from our preparation of purified spliceosomes trapped in vitreous ice) to convert from one state to another. It should be emphasized that the distinct spliceosome states we observed cover a very large conformational space, involving translational movements of protein components of several nm and rotations of up to 40 degrees (Figure 1). As the spliceosomes were chemically stabilized by the GraFix method (Kastner et al., 2008), one could assume a potential bias toward a rougher energy landscape if the crosslinking procedure were to act as a kinetic trap. Because of the fragile nature of non-crosslinked spliceosomes, this cannot be directly tested for the spliceosome; nonetheless, we have not observed such an effect when comparing GraFix-crosslinked with non-crosslinked complexes for other complexes, such as the 26S human proteasome (Haselbach et al., 2017) and the 70S E. coli ribosome (Kastner et al., 2008). Further, one would expect to obtain an even flatter energy landscape when working at a physiological temperature of 37°C as compared to that at 4°C; indeed, this is what we observed for the 70S ribosome (Fischer et al., 2010).

Whereas the ordering of states from left to right is a direct consequence of the PCA analysis and indeed corresponds to the spliceosome assembly direction, no conclusions can be drawn from this about the exact pathway of conformational changes taken by the spliceosome. Specifically, the very low energy barriers would allow the spliceosome to move in any direction along the energy landscape, with a notable propensity for adopting conformations in the states 1–8. As long as the spliceosome cannot reach any state beyond 8, it will be able to continuously sample all conformations. The subsequent step of splicing requires ATP by the RNA helicase PRP2 and would very likely render further spliceosome maturation beyond state 8 as irreversible. The equilibrium of conformational states in

our energy landscape would then be sufficient to provide spliceosomes that are in a conformation and assembly state required for progression of the spliceosome toward a catalytically active machine. We conclude that the spliceosome movements from one state to another reflect inherent conformational dynamics that are stochastic in nature but that overall lead to a productive assembly reaction.

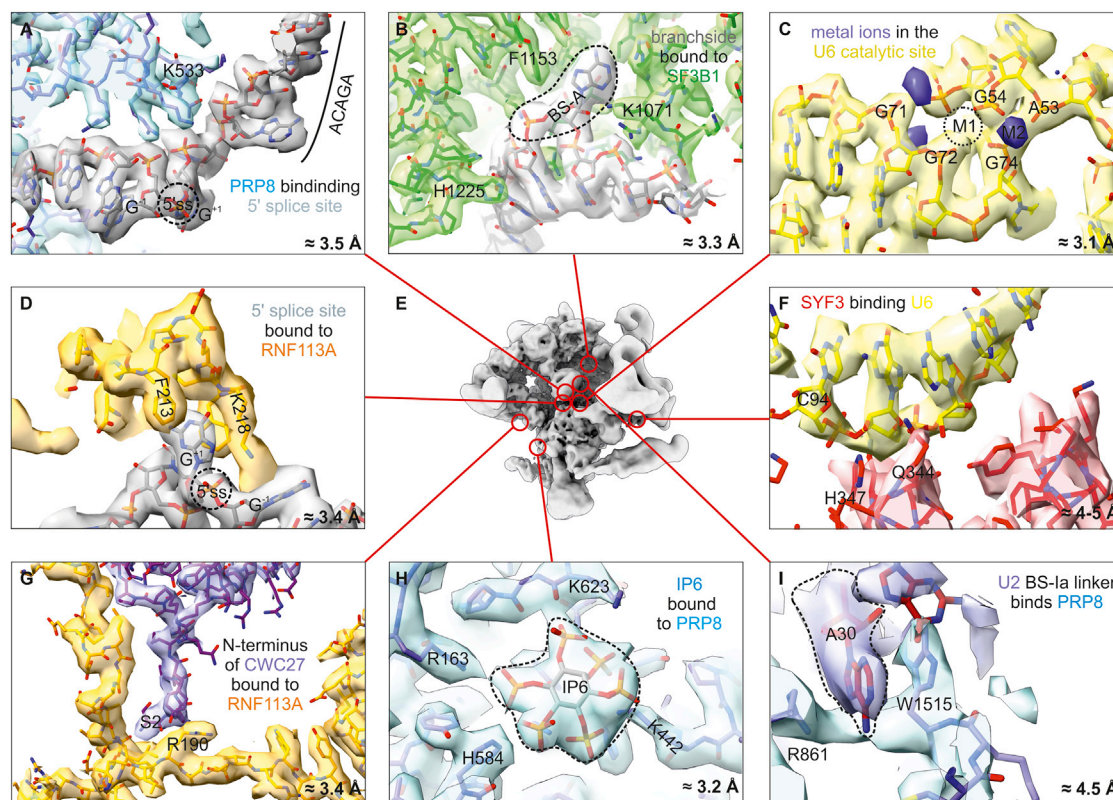
While our analysis shows that there is no unique assembly pathway for the spliceosome, the energy landscape, together with the structures for the eight major states, allowed us to extract more quantitative information about the dynamic behavior of the spliceosome and its requirements for a productive forward movement from states 1 to 8. Certain conformations that we observed required binding and stable accommodation of other protein components. For instance, the stable docking of the helical bundle of the PRP19/CDC5L complex to the core domain of the spliceosome is coupled to the tight binding of the PRP17 WD40 near the catalytic center; spliceosomes structures without visible PRP17 density do not reveal stable accommodation of the helical bundle. Further detailed analysis of the conformational trajectories revealed that the spliceosome undergoes an ordered remodeling by transferring the U2 core from an IBC- and SF3B-bound state to a SYF3-bound state, over a distance of 8 nm (Figures 1A and 1C). At the beginning of the main trajectory, AQR is only loosely bound to the complex, with U2 and the SYF1 N-terminal region as anchor points (state 1). It is then stabilized with a third interaction, followed by a slight reorientation of its domains (state 2). Proceeding along the trajectory, we observe that a lower energy barrier has to be overcome (Figure 1D), resulting in the appearance of a PRP19/CDC5L helical bundle and PRP17 in the structures in perfect unison (state 3). U2 then detaches from IBC and moves relative to the SYF1/3 complex (states 4 and 5). After the release of U2, AQR starts rotating counterclockwise toward PRP17, coupled to a movement of the SYF1/3 complex and the PRP19/CDC5L helical bundle (state 6). With this movement, U2 moves to new binding sites on IBC and SYF3, to finally bind the C-terminus of SYF1 and SYF3, while rearranging its internal conformation. Thus, the movement of U2 and the SYF complex is guided by multiple changes in the interaction patterns of several proteins along this pathway (Movie S1).

### Particularities of the Human $B^{\text{act}}$ Evident at 3.4 Å

To gain deeper insight into the architecture of the human  $B^{\text{act}}$  spliceosome, we combined the particles from the three central conformations (states 3–5) to yield a high-resolution core structure at 3.4 Å resolution, which is possible as these structures differ in mainly their periphery but not their core regions; to date, this is the highest resolution obtained for a spliceosome. Combining this high-resolution spliceosome structure with crosslinking mass spectrometric analysis of the sample allowed us to model the major parts of the functionally important regions of U2, U5, and U6 snRNAs, as well as the extended regions of the pre-mRNA around the 5' splice site and branch site (Tables S1, S2, and S3).

### The BS/U2 RNA Helix Is Sequestered by the U2 SF3B1 HEAT Domain

Major building blocks of the  $B^{\text{act}}$  spliceosome include the U2 snRNP SF3A and SF3B protein complexes, which contact the



**Figure 2. Structural Details about the Core of the Human  $B^{\text{act}}$**

(A) PRP8 binding to the ACAGA box of the U6 snRNA. PRP8 only recognizes the phosphate RNA backbone and is not involved in sequence recognition. (see also Figure S2).

(B) Binding of the branch site adenosine into a pocket formed by SF3B1.

(C) The catalytic side of U6 snRNA. One catalytic magnesium ion (M2) is coordinated by A53, G54, and A74, which is the same configuration as in the active catalytic site. However, the other catalytic magnesium ion (M1), which should be coordinated to G72, is not visible. Furthermore, two additional metal ions are visible; their function is still unknown. (see also Figure S2).

(D) The 5' ss of the pre-mRNA bound to RNF113A forms a pocket that recognizes the 5' ss adenosine. (see also Figure S3).

(E) Complete structure of the  $B^{\text{act}}$  spliceosome that serves as a reference for the location of structural details. Red circles indicate the positions of the more detailed views shown in panels (A–D) and (F–I).

(F) Binding site of the U2/U6 helix II on SYF3 that reveals significant flexibility over the whole trajectory.

(G) N terminus of the newly modeled protein CWC27. (see also Figure S3).

(H) A single inositol-hexakis-6-phosphate is bound to PRP8 close to the N-terminal region of RNF113A. (see also Figure S3).

(I) The U2 BS-Ia linker adenosine is recognized by two tryptophans of PRP8.

pre-mRNA intron at or near the BS, stabilizing the U2/BS helix. Of particular interest is the SF3B1 protein (within the SF3B complex), which contains a C-terminal domain of 20 HEAT repeats; within spliceosomes, SF3B1 can be crosslinked to the BS, to nucleotides upstream of the BS, and to the entire 3' end of the intron. A recent crystal structure of a protease-resistant human SF3B core complex revealed how the C-terminal HEAT domain of SF3B1 forms a stable intricate protein network with the intertwined three-propeller (WD40) cluster of SF3B3 (termed BPA, BPB, and BPC), the three-helix protein SF3B5, and the ZnF protein PHF5A (Cretu et al., 2016). This protein complex is located in the top domain of the human  $B^{\text{act}}$  ( $hB^{\text{act}}$ ) model, whereby the HEAT domain forms a more condensed ring-like structure (as compared to the more open toroidal form in the crystal structure) and touches the PRP8 RT/En domain, while BPA and BPC of SF3B3 contact the BRR2 helicase domain. The BS/U2 RNA helix

is clamped between the terminal HEAT repeats of the SF3B1 HEAT domain. The BS adenosine (BS-A) is flipped out of the BS/U2 helix and is located in a protein pocket built by the B helices of the HEAT repeats H15 to H17 and is capped by PHF5A. Indeed, at 3.3 Å local resolution in this region, the manner in which BS-A is contacted by multiple amino acid side chains is well resolved (Figure 2B).

#### **A Catalytic Magnesium Is Missing from the U2/U6 Core Region**

Similar to the yeast  $B^{\text{act}}$  spliceosome, the U2 and U6 snRNAs are rearranged in  $hB^{\text{act}}$  (as compared to the B complex), with the formation of the kinked stem of U6 ISL and the U2/U6 helices Ia and Ib as well as of the triple helix involving Hoogsteen base pair interactions between the U6 nts G46, A47, and U74 and the U6 nts A53–C55 of the catalytic triad (Fabrizio et al., 2009; Yean et al., 2000). At a local resolution of  $\sim 3.1$  Å in the heart of the catalytic

center, density for the catalytic metal ion M2 (presumably a  $Mg^{2+}$  ion) is clearly visible and is coordinated by the phosphates of the U6 nts A53, G54, and G74. In contrast, density for the catalytic metal ion M1, which should be coordinated by phosphates of nts G72 and U74, is either absent or has a significantly higher B factor and is thus not visible. Therefore, although the phosphate groups of the nucleotides at the U2/U6 catalytic center have adopted the appropriate positions to bind two catalytic metal ions, its present state would prevent it from being catalytically active. Intriguingly, a metal ion bound by the phosphate of G71 is located nearby. However, as this metal ion has also been observed at the same position in the catalytic center of catalytically active spliceosomes, such as the C and C\* complexes (which have both catalytic metal ions in place), it is likely to be a structural metal ion rather than the missing M1 ion (Figure 2C).

The 5'-terminal nt of U2 within the BS/U2 helix, and the 3'-terminal nt of U2 within the U2/U6 helix Ia, are spatially separated by about 20 Å and are bridged by a 4-nt U2 linker (nts 28 to 31). Interestingly, the U2 linker nucleotide A30, which is close to the U2/U6 helix Ia, is packed between the tryptophan residues 1515 and arginine 861 of PRP8 (Figure 2I), presumably stabilizing the U2 linker region. In turn, the U2/U6 helix II is bound by SYF3 (Figure 2F).

#### **RNF113A Connects Several Spatially Separated Parts**

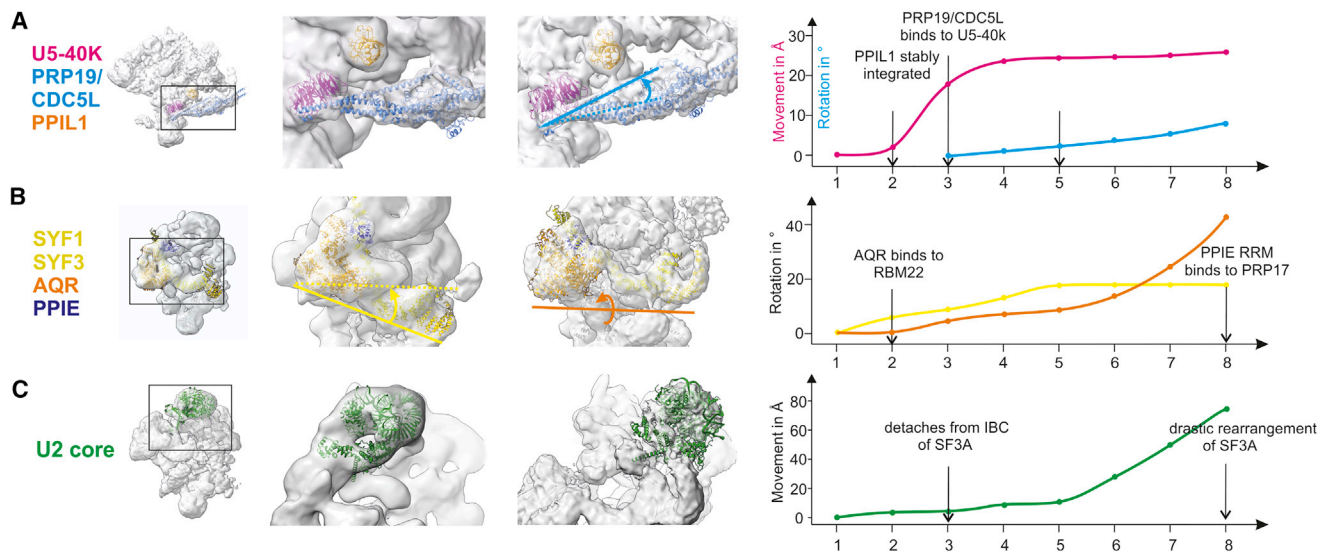
The 3'-terminal nts -2 to -4 of exon 1 are bound to a pocket in PRP8 and are base paired to nts 41 to 43 of U5 snRNA loop I, while nts +4, +5, and +6 are base paired to the first three nts of the U6 ACAGA box; this produces a kink in the 5' ss region that is further stabilized by interactions with PRP8 (Figures 2A and S2). The scissile bond at the 5' ss points toward the catalytic center but is spatially separated from the BS adenosine nucleophile by about 5 nm. Moreover, the 5' ss GU dinucleotides at the 5' end of the intron are in close contact with proteins. Specifically, the first 5'-terminal G is sandwiched between F213, F219, and K218 within the N-terminal zinc finger (ZnF) domain of RNF113A (Figure 2D). Moreover, the N-terminal end of SF3A2 is also positioned close to the GU dinucleotide, where it appears to interact with the U+2 residue. As these proteins not only spatially separate the 5' ss from the catalytic center but also hinder access of the BS adenosine to the 5' ss, they must be rearranged to liberate the 5' ss for its final docking into the catalytic center; using crosslinking data, we have now identified long stretches of RNF113A that are involved in this rearrangement (Figure S3). Specifically, a stretch of approximately 10 amino acid of RNF113A, directly N-terminal to the ZnF domain, passes below the Prp8-En domain in about 20 Å distance parallel to the 5' exon-binding channel. This then interacts with an inositol hexakisphosphate ( $IP_6$ ) molecule, which is bound by numerous residues of the PRP8 En domain (Figures 2H and S3D) and interacts with SLU7 in a later stage of the spliceosome (Fica et al., 2017). Further N-terminal, it contacts the MA3 domain of CWC22 and the PRP8 RH domain and then wraps around the PRP8 En domain, before closely interacting with the CWC27 PPIase domain (Figures 2G and S3A). Finally, the N-terminal stretch of RNF113A latches onto the BRR2 RNA helicase domain (Figure S3C). The RNF113A C-terminal RING domain is attached to the WD40 B domain of U2 SF3B3 at the top of the  $hB^{act}$  structure. The linker region that connects the

RNF113A RING and ZnF domains passes along a protein domain that is situated on top of the SF3B1 HEAT domain (repeats H8 to H10; Figure S3B) and is absent from the yeast  $B^{act}$ . However, due to a lower number of crosslinks in that region and to a significant structural variability of the domain throughout the trajectory, we were not able to identify this density.

#### **Detailed Analysis of the Conformational Changes in States 1 to 8**

As we could identify most densities using our high-resolution structure, we were able to analyze the conformational trajectory of the peripheral regions in more detail. Somewhat unexpectedly, we found that the helical bundle of the PRP19/CDC5L complex is not absent from the initial  $B^{act}$  state; rather, it is flexibly attached, connected only via CDC5L and IBC to the rest of the complex. A dramatic translational movement of U5-40K of about 25 Å creates a new binding surface for the PRP19/CDC5L helical bundle, thereby leading to its stable integration. In the further trajectory, the helical bundle moves upward with IBC, leading to the binding of PPIL1 (Figure 3A). Notably, IBC is at the heart of the movement, rearranging the entire peripheral region around the intron. It performs two independent rotations at the same time: (i) while AQR rotates counterclockwise toward the newly integrated PRP17 WD40, the C-termini of the two SYF proteins move upward around an axis that is almost perpendicular to the first axis (Figure 3B); and (ii) the IBC rotations go along with the unbinding of the U2 core region from IBC. The region then starts to sample several binding sites until it reaches the C-terminus of SYF3, which is more than 80 Å away from its original location (Figure 3C).

All of these mechanics must involve several binding and unbinding events at several positions in the complex. Of special interest are the binding partners of IBC and the U2 core, as they perform the most dramatic rearrangements. They are connected to the spliceosomal core by six protein bridges (B1 to B6) that can open and close at different stages of the trajectory (Figure 4A). Specifically, bridge B1 is formed during the transition from stage 1 to stage 2 and is located between the C-terminal RRM of RBM22 and the AQR's beta barrel (Figure 4B). Bridge B2 interconnects the IBC with the rest of the complex and is formed where SYF1 and SYF3 cross (Figure 4C). While this interaction remains connected throughout all states, it relocates together with the rest of IBC on its trajectory, leading to bending of the U2/U6 helix II and the displacement of the PPIase domain of PPIL2. Another bridge (B3) is formed that connects AQR with the rest of the complex via the RRMs of PPIE and RRM2 of SF3B4 (Figure 4D). This interaction is disrupted during the transition from state 7 to state 8 and is replaced by a new bridge between PPIE and the PRP17-WD40 domain. The connection between IBC and the core of U2 (B4) is formed initially by SYF1, ISY1, U2-A', and SF3A. This interaction becomes weaker during the transition between state 3 and state 4 (Figure 4E). After this interaction is completely disrupted, the U2 core remains connected to the core of the spliceosome via one bridge (B5) between the U2-B''RRM2 and the C-terminal domain of SF3A3 and the U2 RNA stem IIa/IIb (B6). While B6 remains throughout the trajectory, B5 breaks during the transition from stage 7 to stage 8 (Figures 4F and 4G). Note that the bridges describe



**Figure 3. Conformational Changes**

The amounts of conformational changes observed for the most mobile structural parts of the  $B^{act}$  spliceosome (states 1–8) is depicted. The left panel shows the overall  $B^{act}$  structure as a reference, with the marked regions of interest in various colors. The two panels to the right show the extreme conformations. All protein positions have been determined by rigid body fitting (see Table S1). Whereas the initial positioning was guided by crosslinking mass spectrometry (see Table S2), the overall resolution of the maps was sufficient for unambiguous placement. The right panel depicts the quantitative analysis of the motion parameter revealing translational and rotational structural rearrangements of 1–8 nm and up to 40 degree for the most mobile parts.

(A) A detailed flexibility analysis of the PRP19/CDC5L complex reveals only weak initial interactions with IBC. The helical bundle of the PRP19/CDC5L complex must first bind to the U5-40k protein and then to PPIL1 before it becomes stably integrated into the spliceosome. This stable integration is directly coupled to a repositioning of the U5-40k protein by more than 20 Å. The final step of integration of the helical bundle requires an additional upward rotation by more than 12°, which allows a direct connection to PPIL1 to be established.

(B) The binding of AQR to RBM22 leads to two rotational motions of the SYF proteins around two perpendicular axes. The first rotation (yellow) moves the C-terminal ends of SYF1 and SYF3 upward, enabling their contacts with U2. The rotation around the second axis brings AQR and PPIE closer to PRP17, also allowing new contacts to be established.

(C) As a consequence of the movements described in (B), the U2 core region moves toward the C-terminus of SYF1, while binding to different sites on IBC along the trajectory.

only the major visible interactions; however, numerous minor (yet perhaps also highly relevant) variations in binding interactions of spliceosomal components can also be observed during the entire trajectory. Even the interaction surfaces of bridges 1–6 are relatively small, which may explain why many of the different connections can transiently be changed and modified. As mentioned above, there are no significant energy barriers, and these changes—despite their very dramatic nature—do not require ATP/GTP hydrolysis to occur (which is also absent from the sample). The available thermal energy would thus be entirely sufficient to drive these conformational rearrangements. Notably, it is the total architecture of the network of interactions that constructively adds up to organize the dynamics of the spliceosome and to permit to the clockwork of visible conformational changes along the splicing trajectory.

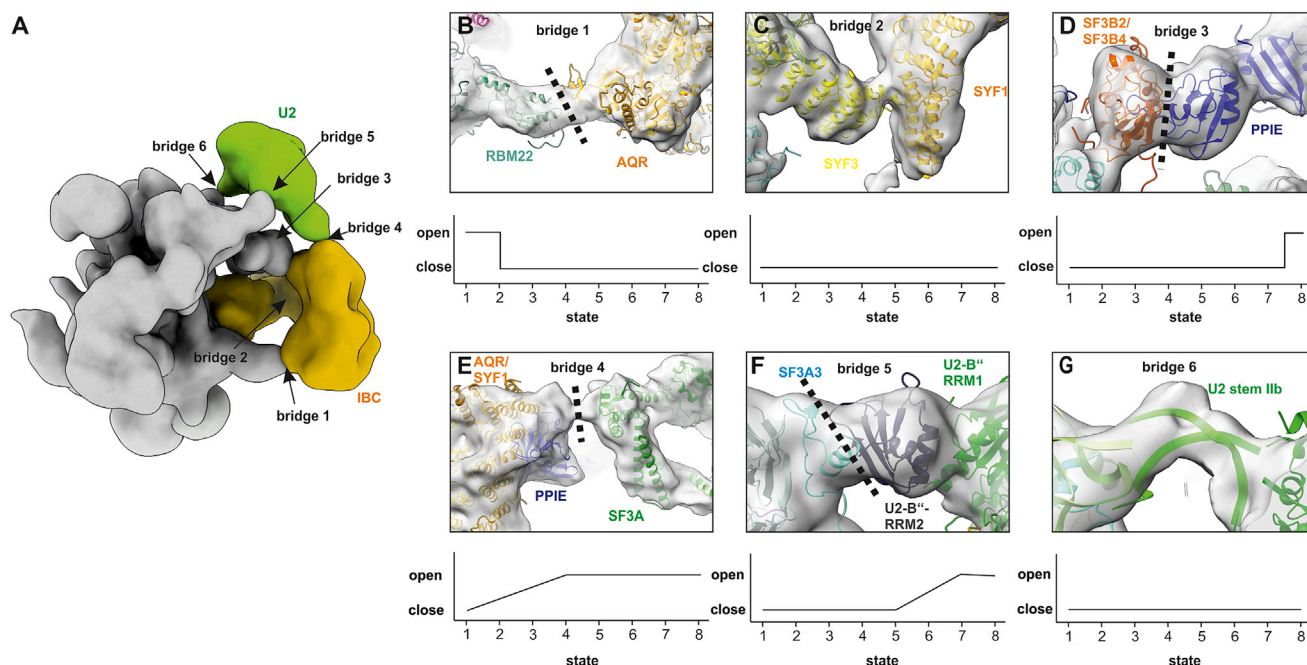
### Conformational Switch near the Catalytic Site

In addition to the larger rigid body domain movements of proteins in the periphery, a restructuring of protein domains can be seen in the core regions that were determined at higher resolution. The SKIP protein does not contain a stably folded domain but rather threads through larger parts of the complex. Two of its regions—amino acids (aa) 95–114, just N-terminal of the PPIL1 binding site, and aa 215–226—are also flexible in early states

and become simultaneously stabilized in states 4 and 5. Specifically, aa 95–114 bind to the HAT repeats 3 and 4 of SYF3, and the N-terminal aa 215–226, to HAT repeats 1 and 2 (Figure 5B). This is accompanied by a closer contact of the SYF3 HAT1 (Figure 2F) with U6 RNA and a slight rotation of the U2/U6 helix II. At the same time,  $\alpha$ -helix<sup>111–128</sup> of RBM22, which is situated close to the RBM22 ZnF domain, also becomes stabilized and interacts with SKIP, close to U6-C60 (Figure 5B). Indeed, due to their coordinated stabilization, SKIP, RBM22, and SYF3 create a pocket that tightly binds U6-C60 from state 4 onward. In addition to the U6-C60 binding, neighboring U6 nucleotides have an interface with the (previously flexible) SKIP protein, causing the U6-C60 nucleotide even to flip in another direction, which is likely to stabilize the structure of this part of U6 RNA close to the U6 bulge loop (Figure 5). Taken together, these results indicate that the highly inter-connected SKIP protein probably plays a major role in coordinating the dynamics and final stabilization of multiple flexible protein domains during the last stages of  $B^{act}$  complex maturation.

### DISCUSSION

Based on massive data from cryo-EM imaging of the human  $B^{act}$  spliceosome, we have quantitatively analyzed the motion



**Figure 4. Dynamic Bridges Between the Different Mobile Elements and the Core**

The main bridges that connect the mobile elements (IBC and U2 core) to the rest of the complex are shown. A detailed view of each bridge and the potential opening site(s) created, as well as a graph indicating the opening state, is shown. All parts were unambiguously placed according to their structural features and as guided by crosslinking mass spectrometry (see [Tables S1](#) and [S2](#)).

(A) A low-pass filtered overall structure is shown as reference, with the location of the bridges indicated.

(B) Bridge 1 is formed between AQR and RBM22 early on in the transition between states 1 and 2 and remains closed for the rest of the trajectory.

(C) Bridge 2 is formed between SYF1 and SYF3 and is closed during the entire trajectory. Only the angle between the two proteins changes slightly.

(D) Bridge 3 connects IBC via PPIE to the core of the spliceosomal complex via SF3B2/SF3B4. Even though IBC rotates continuously, this bridge disrupts only at a later stage (state 8), at which point a new connection between PPIE and PRP17 is formed.

(E) Bridge 4 connects IBC via PPIE and AQR/SYF3, and the U2 core via SF3A. This connection is maintained only at an early stage during the trajectory. After disruption, the U2 core and AQR move away in different directions from the actual bridging point.

(F) Bridge 5 connects the U2 core via U2-B'' with the core of the complex via SF3A3. This connection breaks later in the trajectory; however, it starts to reorient already directly after the U2 core detaches from PPIE/AQR (E).

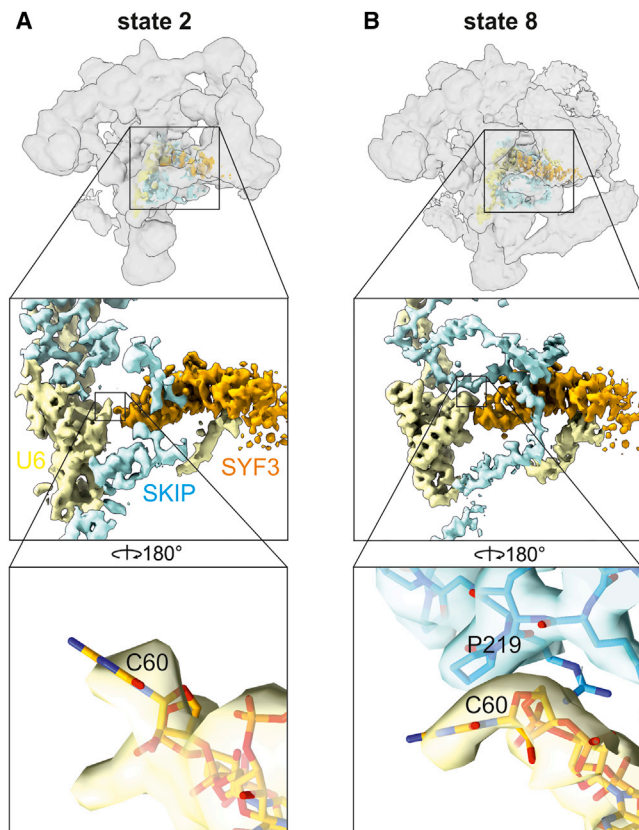
(G) Bridge 6 connects the U2 snRNP core with the core of the spliceosome via U2 snRNA. Although this bridge stays connected all the time (due to its covalent nature), structural rearrangements in its topology seem to occur.

patterns of the spliceosome. This has allowed us to visualize for the first time the functionally important remodeling steps of the hB<sup>act</sup> spliceosome and has provided us with deep insight about how such movements can be the driving force behind spliceosome assembly and functionality. This is a critical first step toward understanding how the spliceosome works in terms of organizing these dynamic features at a structural level.

The spliceosomal core structure is its most stable structural part (i.e., it has the least amount of freedom of motion, and is consistently present), allowing it to be determined at high resolution in all currently available structures of the spliceosome. In stark contrast, the next peripheral layer of splicing components to bind to the spliceosome can significantly change in their spatial positioning as well as in their composition during the spliceosome assembly pathway. This results in an increased mobility of this structural layer, which is critical to decipher as it contributes to the functional activation, and most likely also regulation, of the spliceosome. Combining our extensive 3D classification with 3D PCA provided the technical framework to

allow us to analyze these dynamics quantitatively, by determining the order of all B<sup>act</sup> spliceosome conformational changes along the two major modes of motion found as eigenvectors in the PCA analysis.

Despite these advances, we still cannot visualize a significant part of the B<sup>act</sup> complex. This leads us to propose that there is yet another layer of spliceosomal components that is organized in a structurally distinct manner. Such a layer may be organized even more transiently, with a higher degree of flexibility and/or mobility, as the mobile parts described here. In this additional layer, splicing-relevant proteins might be pre-organized by weak interaction forces, allowing them to be stably integrated into the spliceosome only at a later state of the assembly when the conformation of the spliceosome favors such a stable integration. Our mobility analyses of flexible spliceosomal parts highlight that spliceosome remodeling requires a certain conformational flexibility that is driven by the available thermal energy. Thus, the very flat energy landscape of the different states makes it energetically cheap to convert any state into another, as no



**Figure 5. Larger Overall Conformational Changes are Coupled to a Flipped Nucleotide Near the Catalytic Center**

The structural order close to the catalytic center of the spliceosome is increased at later stages of the trajectory.

(A) The zoomed areas highlight the interaction network of U6 snRNA and the proteins SYF3 and SKIP in state 2. Notably, cytosine 60 of U6 snRNA is now in a “flipped out” conformation. All three components interact only weakly and are not yet very well defined. SKIP has large unstructured regions that are not visible in the density.

(B) In state 8, the overall structural rearrangements of SYF3 lead to a translational separation from U6 snRNA by 5 Å, and a rotational separation by 4 Å. As a consequence, new binding surfaces are formed that lead to structural stabilization of the unstructured region of SKIP in this conformation. The flipped position of the U6 snRNA C60 is stabilized by P219 of SKIP.

more than  $3k_B T$  is required. This allows the spliceosome to “sample” a wide array of conformations, which are required to productively move it along the trajectory required to assemble a catalytically active spliceosome. In other words, each time a spliceosome component binds, the conformational space that the entire complex can adopt will be affected. As the spliceosome samples new conformations, it can create new binding interfaces that allow proteins to bind and become more strongly integrated into the spliceosome. This is a very stochastic approach that relies on the perfect orchestration of all spliceosomal components acting together productively in a complex network of interactions. For example, stable integration of PRP17 is a clear trigger for conformational changes in the spliceosome: only once PRP17 binds stably, the U5 U5-40K protein

changes its position significantly, which allows IBC to become stably integrated. The subsequent conformational changes of the spliceosome that bring the entire spliceosome closer to being a catalytically active species can only happen as a consequence of the preceding dynamic steps. This leads not only to massive rearrangements in U2 positioning but also to minor changes in the RNA at the catalytic center, resulting in the flip in orientation of a single nucleotide. We therefore conclude that the extremely large conformational changes evident in the periphery of the spliceosome have direct consequences on the formation of its catalytic center.

The process of sampling different conformations by the spliceosome along its assembly pathway is likely to be essential for both its regulation and functionality; this is supported by: (i) an interplay of dynamic overall motions of peripheral components, and (ii) any resulting changes of conformational freedom caused by binding interactions or any other interference of the total interaction network within the spliceosome. Our current analysis already allows us to detect numerous interaction patterns that change during the observed order of events from states 1 to 8. It will be very interesting to understand, at a structural level, how the spliceosome maintains its total interaction energies at an almost constant level, which makes it possible to dynamically adapt to all of these widely varying conformations.

Along the energy landscape of the  $B^{\text{act}}$  spliceosome, we found eight relatively stable intermediate structures, which we describe as major conformational states. However, we also observed a much higher number of minor states that the spliceosome can sample along the observed trajectory in the  $B^{\text{act}}$  spliceosome formation. It will now be important to study the flexibility patterns of the other, functionally well-defined assembly intermediates of the spliceosome, to obtain the full picture about how the spliceosome assembles, how it dynamically adapts, and how catalysis finally takes place.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Preparation of splicing extracts
  - *In vitro* splicing and MS2 affinity selection of  $B^{\text{act}}$  complex
  - 2D gel electrophoresis and mass spectrometry
  - Crosslinking of the  $B^{\text{act}}$  complex and crosslink identification
  - Cryo-EM sample preparation
  - Cryo-EM data acquisition
  - Image processing
  - Conformational landscape analysis
  - 3D volume refinement
  - Model building
- **DATA AND SOFTWARE AVAILABILITY**

## SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, three tables, and one movie and can be found with this article online at <https://doi.org/10.1016/j.cell.2018.01.010>.

## ACKNOWLEDGMENTS

We thank Thomas Conrad, Hossein Kohansal, Ulrich Steuerwald, and Winfried Lehndekel for expert technical support during sample preparation. Molecular graphics and analyses were performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco.

## AUTHOR CONTRIBUTIONS

I.K. and D.E.A. purified and analyzed the complexes biochemically. D.H. performed cryo-EM data acquisition (with some contribution from I.K.), image analyses, statistical analyses, and biophysical analyses. D.H., K.H., and B.K. modeled the density with the help of B.G. O.D. performed crosslinking mass spectrometry experiments and analysis under the supervision of H.U. H.S. supervised the cryo-EM work. H.S., R.L., B.K., and D.H. wrote the text with input from all authors.

## DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: October 31, 2017

Revised: December 23, 2017

Accepted: January 5, 2018

Published: January 17, 2018

## REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Agafonov, D.E., Deckert, J., Wolf, E., Odenwalder, P., Bessonov, S., Will, C.L., Urlaub, H., and Lührmann, R. (2011). Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis. *Methods Mol. Cell. Biol.* **31**, 2667–2682.
- Agafonov, D.E., Kastner, B., Dybkov, O., Hofele, R.V., Liu, W.-T., Urlaub, H., Lührmann, R., and Stark, H. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351**, 1416–1420.
- Bai, R., Yan, C., Wan, R., Lei, J., and Shi, Y. (2017). Structure of the Post-catalytic Spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**, 1589–1598.
- Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017a). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323.
- Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017b). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* **170**, 701–713.
- Bessonov, S., Anokhina, M., Krasauskas, A., Golas, M.M., Sander, B., Will, C.L., Urlaub, H., Stark, H., and Lührmann, R. (2010). Characterization of purified human Bact spliceosomal complexes reveals compositional and morphological changes during spliceosome activation and first step catalysis. *RNA* **16**, 2384–2403.
- Buchan, D.W.A., Minneci, F., Nugent, T.C.O., Bryson, K., and Jones, D.T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res.* **41**, W349–57.
- Chan, S.-P., Kao, D.-I., Tsai, W.-Y., and Cheng, S.-C. (2003). The Prp19p-Associated Complex in Spliceosome Activation. *Science* **302**, 279–282.
- Cretu, C., Schmitzová, J., Ponce-Salatierra, A., Dybkov, O., De Laurentiis, E.I., Sharma, K., Will, C.L., Urlaub, H., Lührmann, R., and Pena, V. (2016). Molecular Architecture of SF3b and Structural Consequences of Its Cancer-Related Mutations. *Mol. Cell* **64**, 307–319.
- De, I., Bessonov, S., Hofele, R., dos Santos, K., Will, C.L., Urlaub, H., Lührmann, R., and Pena, V. (2015). The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat. Struct. Mol. Biol.* **22**, 138–144.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.M. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132.
- Fabrizio, P., Dannenberg, J., Dube, P., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2009). The Evolutionarily Conserved Core Design of the Catalytic Activation Step of the Yeast Spliceosome. *Mol. Cell* **36**, 593–608.
- Fica, S.M., Tuttle, N., Novak, T., Li, N.-S., Lu, J., Koodathingal, P., Dai, Q., Staley, J.P., and Piccirilli, J.A. (2013). RNA catalyzes nuclear pre-mRNA splicing. *Nature* **503**, 229–234.
- Fica, S.M., Mefford, M.A., Piccirilli, J.A., and Staley, J.P. (2014). Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21**, 464–471.
- Fica, S.M., Oubridge, C., Galej, W.P., Wilkinson, M.E., Bai, X.-C., Newman, A.J., and Nagai, K. (2017). Structure of a spliceosome remodelled for exon ligation. *Nature* **542**, 377–380.
- Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* **466**, 329–333.
- Galej, W.P., Wilkinson, M.E., Fica, S.M., Oubridge, C., Newman, A.J., and Nagai, K. (2016). Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201.
- Goddard, T.D., Huang, C.C., and Ferrin, T.E. (2007). Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287.
- Guex, N., and Peitsch, M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723.
- Hang, J., Wan, R., Yan, C., and Shi, Y. (2015). Structural basis of pre-mRNA splicing. *Science* **349**, 1191–1198.
- Haselbach, D., Schrader, J., Lambrecht, F., Henneberg, F., Chari, A., and Stark, H. (2017). Long-range allosteric regulation of the human 26S proteasome by 20S proteasome-targeting cancer drugs. *Nat. Commun.* **8**, 15578.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell* **12**, 5–14.
- Kastner, B., Fischer, N., Golas, M.M., Sander, B., Dube, P., Boehringer, D., Hartmuth, K., Deckert, J., Hauer, F., Wolf, E., et al. (2008). GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55.
- Kim, S.H., and Lin, R.J. (1996). Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.* **16**, 6810–6819.
- Korneta, I., and Bujnicki, J.M. (2012). Intrinsic Disorder in the Human Spliceosomal Proteome. *PLoS Comput. Biol.* **8**, e1002641.
- Kuwasako, K., He, F., Inoue, M., Tanaka, A., Sugano, S., Güntert, P., Muto, Y., and Yokoyama, S. (2006). Solution Structures of the SURP Domains and the Subunit-Assembly Mechanism within the Splicing Factor SF3a Complex in 17S U2 snRNP. *Structure* **14**, 1677–1689.
- Leitner, A., Walzthoeni, T., and Aebersold, R. (2014). Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.* **9**, 120–137.

- Lin, P.C., and Xu, R.M. (2012). Structure and assembly of the SF3a splicing factor complex of U2 snRNP. *EMBO J.* **31**, 1579–1590.
- Liu, S., Li, X., Zhang, L., Jiang, J., Hill, R.C., Cui, Y., Hansen, K.C., Zhou, Z.H., and Zhao, R. (2017). Structure of the yeast spliceosomal postcatalytic P complex. *Science* **358**, 1278–1283.
- Macías, S., Bragulat, M., Tardiff, D.F., and Vilardell, J. (2008). L30 Binds the Nascent RPL30 Transcript to Repress U2 snRNP Recruitment. *Mol. Cell* **30**, 732–742.
- Makarov, E.M., Makarov, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Lührmann, R. (2002). Small Nuclear Ribonucleoprotein Remodeling During Catalytic Activation of the Spliceosome. *Science* **298**, 2205–2208.
- Nguyen, T.H.D., Galej, W.P., Bai, X., Savva, C.G., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52.
- Nguyen, T.H.D., Galej, W.P., Bai, X., Oubridge, C., Newman, A.J., Scheres, S.H.W., and Nagai, K. (2016). Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
- Plaschka, C., Lin, P.-C., and Nagai, K. (2017). Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621.
- 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.
- Schellenberg, M.J., Edwards, R.A., Ritchie, D.B., Kent, O.A., Golas, M.M., Stark, H., Lührmann, R., Glover, J.N.M., and MacMillan, A.M. (2006). Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. USA* **103**, 1266–1271.
- Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46.
- Ulrich, A., and Wahl, M.C. (2014). Structure and evolution of the spliceosomal peptidyl-prolyl cis-trans isomerase Cwc27. *Acta Crystallogr. D Biol. Crystallogr* **70**, 3110–3123.
- van Roon, A.-M.A.M., Oubridge, C., Obayashi, E., Sposito, B., Newman, A.J., Seraphin, B., and Nagai, K. (2017). Crystal structure of U2 snRNP SF3b components: Hsh49p in complex with Cus1p binding domain. *RNA* **23**, 968–981.
- Wahl, M.C., Will, C.L., and Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell* **136**, 701–718.
- Wan, R., Yan, C., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904.
- Wan, R., Yan, C., Bai, R., Lei, J., and Shi, Y. (2017). Structure of an Intron Lariat Spliceosome from *Saccharomyces cerevisiae*. *Cell* **171**, 1–13.
- Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., and Andrew, J. (2017). Post-catalytic spliceosome structure reveals mechanism of 3' splice site selection. *Science* **358**, 1–12.
- Xu, C., Zhang, J., Huang, X., Sun, J., Xu, Y., Tang, Y., Wu, J., Shi, Y., Huang, Q., and Zhang, Q. (2006). Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *J. Biol. Chem.* **281**, 15900–15908.
- Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2016). Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911.
- Yan, C., Wan, R., Bai, R., Huang, G., and Shi, Y. (2017). Structure of a yeast step II catalytically activated spliceosome. *Science* **355**, 149–155.
- Yang, B., Wu, Y., Zhu, M., Fan, S., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y., Chen, H.-F., et al. (2012). Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906.
- Yean, S.-L., Wuenschell, G., Termini, J., and Lin, R.-J. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* **408**, 881–884.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12.
- Zhang, X., Yan, C., Hang, J., Finci, L.I., Lei, J., and Shi, Y. (2017). An Atomic Structure of the Human Spliceosome. *Cell* **169**, 918–929.e14.
- Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332.

## STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE   | SOURCE   | IDENTIFIER  |
|---|--|---|
| Bacterial and Virus Strains   |  |   |
| Rosetta 2(DE3) Competent Cells  | EMD Millipore                                      | Cat# 71397  |
| Deposited Data  |  |   |
| Hsh49(RRM1)-Cus1(290-368)   | <a href="#">van Roon et al., 2017</a>              | PDB: 5LSL   |
| PPIL1-Skip  | <a href="#">Xu et al., 2006</a>                    | PDB: 1XWN   |
| PPIL1-Skip  | (structural genomics data, unpublished)            | PDB: 2K7N   |
| SF3a  | <a href="#">Lin and Xu, 2012</a>                   | PDB: 4DGW   |
| SF3A1-SURP1   | <a href="#">Kuwasako et al., 2006</a>              | PDB: 2DT6   |
| SF3A1-SURP2   |  | PDB: 2DT7   |
| PPIL2-RING  | (structural genomics data, unpublished)            | PDB: 1ZKC   |
| SF3B6(p14)-SF3B1(377-415)   | <a href="#">Schellenberg et al., 2006</a>          | PDB: 2F9D   |
| CWC27   | <a href="#">Ulrich and Wahl, 2014</a>              | PDB: 4R3E   |
| AQR   | <a href="#">De et al., 2015</a>                    | PDB: 4PJ3   |
| yBact complex   | <a href="#">Rauhut et al., 2016</a>                | PDB: 5LQW   |
| yBact complex   | <a href="#">Yan et al., 2016</a>                   | PDB: 5GM6   |
| yB complex  | <a href="#">Plaschka et al., 2017</a>              | PDB: 5NRL   |
| yB complex  | <a href="#">Bertram et al., 2017b</a>              | PDB: 5O9Z   |
| hC* complex   | <a href="#">Bertram et al., 2017a</a>              | PDB: 5MQF   |
| hC* complex   | <a href="#">Zhang et al., 2017</a>                 | PDB: 5XJC   |
| SF3B  | <a href="#">Cretu et al., 2016</a>                 | PDB: 5IEF   |
| SpliProt3D (database containing structural models of all human spliceosomal proteins) | <a href="#">Korneta and Bujnicki, 2012</a>         | <a href="http://iimcb.genesilico.pl/SpliProt3D/home/">http://iimcb.genesilico.pl/SpliProt3D/home/</a> |
| Human B <sup>act</sup> core map   | This study   | EMDB: 4255  |
| Human B <sup>act</sup> core model   | This study   | PDB: 6FF4   |
| Human B <sup>act</sup> state 1 map  | This study   | EMDB: 4233  |
| Human B <sup>act</sup> state 2 map  | This study   | EMDB: 4234  |
| Human B <sup>act</sup> state 3 map  | This study   | EMDB: 4235  |
| Human B <sup>act</sup> state 4 map  | This study   | EMDB: 4236  |
| Human B <sup>act</sup> state 5 map  | This study   | EMDB: 4237  |
| Human B <sup>act</sup> state 6 map  | This study   | EMDB: 4238  |
| Human B <sup>act</sup> state 7 map  | This study   | EMDB: 4239  |
| Human B <sup>act</sup> state 8 map  | This study   | EMDB: 4240  |
| Human B <sup>act</sup> state 8 complete model   | This study   | PDB: 6FF7   |
| Human B <sup>act</sup> state 1 core map   | This study   | EMDB: 4247  |
| Human B <sup>act</sup> state 2 core map   | This study   | EMDB: 4248  |
| Human B <sup>act</sup> state 3 core map   | This study   | EMDB: 4249  |
| Human B <sup>act</sup> state 4 core map   | This study   | EMDB: 4250  |
| Human B <sup>act</sup> state 5 core map   | This study   | EMDB: 4251  |
| Human B <sup>act</sup> state 6 core map   | This study   | EMDB: 4252  |
| Human B <sup>act</sup> state 7 core map   | This study   | EMDB: 4253  |
| Human B <sup>act</sup> state 8 core map   | This study   | EMDB: 4254  |
| Experimental Models: Cell Lines   |  |   |
| Human: HeLa S3 cells  | Helmholtz Center for Infection Research, Brunswick | N/A   |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE         | SOURCE                             | IDENTIFIER  |
|-----------------------------|------------------------------------|---|
| Recombinant DNA             |                                    |   |
| Plasmid: pMBP-MS2           | Macias et al., 2008                | Addgene plasmid # 65104   |
| Plasmid: pSP6-PM5-10        | Bessonov et al., 2010              | N/A   |
| Software and Algorithms     |                                    |   |
| Coot v0.8.8                 | Emsley and Cowtan, 2004            | <a href="https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/">https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/</a>       |
| COW beta                    | Luetlich et al. (unpublished data) | <a href="http://www.cow-em.de">www.cow-em.de</a>  |
| EMAN 2.1                    | Tang et al., 2007                  | <a href="http://blake.bcm.edu/emanwiki/EMAN2/">http://blake.bcm.edu/emanwiki/EMAN2/</a>   |
| Gautomatch                  | Dr. Kai Zhang                      | <a href="https://www.mrc-lmb.cam.ac.uk/kzhang/">https://www.mrc-lmb.cam.ac.uk/kzhang/</a>                                       |
| Gctf                        | Zhang, 2016                        | <a href="https://www.mrc-lmb.cam.ac.uk/kzhang/">https://www.mrc-lmb.cam.ac.uk/kzhang/</a>                                       |
| MATLAB 2014b                | MathWorks                          | <a href="https://de.mathworks.com/products/matlab.html">https://de.mathworks.com/products/matlab.html</a>                       |
| MotionCor2                  | Zheng et al., 2017                 | <a href="http://www.msg.ucsf.edu/em/software/index.html">http://www.msg.ucsf.edu/em/software/index.html</a>                     |
| PHENIX suite v1.12-2829-000 | Adams et al., 2010                 | <a href="https://www.phenix-online.org">https://www.phenix-online.org</a>   |
| pLink v.1.23                | Yang et al., 2012                  | <a href="http://pfind.ict.ac.cn/software/pLink/">http://pfind.ict.ac.cn/software/pLink/</a>                                     |
| PyMOL                       | Schrödinger LCC                    | <a href="https://www.pymol.org/2/">https://www.pymol.org/2/</a>   |
| RELION v2.0.2               | Scheres, 2012                      | <a href="http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page">http://www2.mrc-lmb.cam.ac.uk/relion/index.php/Main_Page</a> |
| SWISS-MODEL suite           | Guex and Peitsch, 1997             | <a href="https://spdbv.vital-it.ch/">https://spdbv.vital-it.ch/</a>   |
| UCSF Chimera v.1.11.2       | Pettersen et al., 2004             | <a href="http://www.cgl.ucsf.edu/chimera/">http://www.cgl.ucsf.edu/chimera/</a>   |
| UCSF ChimeraX beta          | Goddard et al., 2007               | <a href="https://www.rbvi.ucsf.edu/chimerax/">https://www.rbvi.ucsf.edu/chimerax/</a>   |

**CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to, and will be fulfilled by the Lead Contact: Holger Stark ([holger.stark@mpibpc.mpg.de](mailto:holger.stark@mpibpc.mpg.de)).

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

For *in vitro* splicing and purification of spliceosomes, nuclear extract from HeLa S3 cells was used. HeLa cells were grown in DMEM/F12 (1:1) medium supplemented with 5% NCS serum in a 30l fermenter (Applikon Biotek) at a density of  $6.5 \times 10^6$  cells/ml and harvested using a Sorvall BIOS 16 centrifuge.

The MBP-MS2 fusion protein was expressed in the *Escherichia coli* strain Rosetta 2 (DE3) (Novagen), which was grown in 2YT medium at 37°C.

**METHOD DETAILS****Preparation of splicing extracts**

To prepare nuclear extracts active in splicing (Dignam et al., 1983), HeLa cells were grown in DMEM/F12 (1:1) medium supplemented with 5% NCS serum in a 30l fermenter (Applikon Biotek) to a density of  $6.5 \times 10^6$  cells/ml and harvested by centrifuging for 10 min at 2000 rpm in a  $8 \times 2000$  mL BIOS rotor (Thermo Scientific). Cells were washed twice with ice cold PBS buffer (pH 7.4) and resuspended in 1.25 volumes of MC buffer [10 mM HEPES-KOH, pH 7.6, 10 mM KOAc, 0.5 mM Mg(OAc)<sub>2</sub>, 0.5 mM DTT] supplemented with 2 tablets of protease inhibitor cocktail (Roche) per 50 mL of the buffer. They were then incubated for 5 min on ice and lysed with 18 strokes of a Dounce homogenizer at 4°C. Nuclei were pelleted for 5 min at 10000 rpm in a F14-14x50cy rotor (Thermo Scientific). After removing the supernatant, 1.3 volumes of Roeder C buffer [25% (v/v) glycerol, 20 mM HEPES-KOH, pH 7.6, 0.2 mM EDTA pH 8.0, 420 mM NaCl] supplemented with 0.5 mM DTT and 0.5 mM PMSF were added per gram of nuclei. The latter were then lysed with 20 strokes of a Dounce homogenizer. The mixture was stirred slowly for 40 min at 4°C, followed by centrifugation for 30 min at 12300 rpm in a F14-14x50 rotor (Thermo Scientific). The supernatant was recovered and was immediately used for B<sup>act</sup> complex assembly as described below without dialysis or freezing.

### **In vitro splicing and MS2 affinity selection of B<sup>act</sup> complex**

Uniformly [<sup>32</sup>P]-labeled, m<sup>7</sup>G(5')ppp(5')G-capped PM5-10 pre-mRNA (Bessonov et al., 2010) was synthesized *in vitro* by SP6 runoff transcription. HeLa S3 cells, obtained from GBF, Brunswick (currently Helmholtz Center for Infection Research, Brunswick), tested negative for mycoplasma. HeLa nuclear extract was prepared essentially as previously described (Dignam et al., 1983), but without the final dialysis step. Spliceosomal complexes were isolated by MS2 affinity selection. The PM5-10 pre-mRNA containing three MS2 aptamers at its 5' end was incubated with a 10-fold molar excess of MBP-MS2 fusion protein and then added to a splicing reaction. The reaction contained 5 nM of <sup>32</sup>P-labeled pre-mRNA and 20% HeLa nuclear extract in buffer containing 3 mM MgCl<sub>2</sub>, 50 mM NaCl, 4 mM HEPES-KOH pH 7.9, 15 mM MES-NaOH pH 6.4, 2 mM ATP, and 20 mM creatine phosphate, and was incubated at 30°C for 160 min. A 50-fold molar excess of DNA oligonucleotides (M6: 5'-GGCGGTCTCGTC-3', and M12: 5'-CTCGTCGGCAGC-3') complementary to nucleotides -7 to -24 (relative to the 5' ss) of the PM5-10 substrate were added. The reaction was incubated at 30°C for an additional 20 min. Aggregates were removed by centrifugation, and the supernatant was loaded onto an MBP Trap HP column (GE Healthcare) after addition of 5 mM HEPES-KOH pH 7.9. The column was washed with G-150 buffer (20 mM HEPES-KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 150 mM NaCl), and complexes were eluted with G-150 buffer containing 1 mM maltose. Eluted complexes were loaded onto a 36 mL linear 5%–20% (w/v) sucrose gradient containing G-150 buffer and centrifuged at 24000 rpm for 13 hr at 4°C in a Surespin 630 (Thermo Scientific) rotor. Fractions of 1.1 mL were harvested through a capillary from the bottom using a peristaltic pump. The distribution of <sup>32</sup>P-labeled PM5-10 RNA across the gradient was determined by Cherenkov counting. Fractions were analyzed by denaturing 4%–12% NuPAGE (Life Technologies) followed by autoradiography. Five peak fractions containing B<sup>act</sup> complex were pooled, concentrated by Amicon 50kD cut-off unit, diluted with G-150 buffer to 1.5 mL to decrease the sucrose concentration to below 3% (w/v), and re-loaded on the same gradients with glutaraldehyde as fixative (Kastner et al., 2008). For biochemical sample validation, the same procedure was performed but without fixation in the second gradient. Gradients were fractionated into 200 µL fractions. RNA and protein compositions of purified complexes were determined by denaturing 1D PAGE and 2D gel electrophoresis.

### **2D gel electrophoresis and mass spectrometry**

Two-dimensional gel electrophoresis of affinity-purified spliceosomal complexes was performed as described previously (Agafonov et al., 2011) using 7% acrylamide mono 2D gels for proteins above 40 kDa, and 15% acrylamide mono 2D gels for proteins below 40 kDa. For mass spectrometry, Coomassie-stained protein spots were cut out of the 2D gel, and proteins were digested in-gel with trypsin and then extracted. The extracted peptides were analyzed in a liquid-chromatography coupled electrospray ionization mass spectrometer (LTQ Orbitrap XL) under standard conditions. Proteins were identified by searching fragment spectra against the human UniProt database using Mascot as a search engine.

### **Crosslinking of the B<sup>act</sup> complex and crosslink identification**

MS2 affinity-purified spliceosomal B<sup>act</sup> complexes were crosslinked with 150 µM BS3 for 30 min at 25 °C after the first gradient centrifugation and subsequently quenched with Tris-HCl pH 8.0. Approximately 25 pmol of B<sup>act</sup> complexes were trypsin-digested in-solution and analyzed essentially as described before (Leitner et al., 2014). After trypsin digestion, peptides were reverse-phase extracted and fractionated on a Superdex Peptide PC3.2/30 column (GE Healthcare). 50 µL fractions corresponding to an elution volume of 1.2–1.8 ml were analyzed in triplicate using Thermo Scientific Q Exactive HF and Orbitrap Fusion Lumos Tribrid mass spectrometers. Protein-protein crosslinks were identified by a pLink 1.23 search engine allowing for three missed cleavages (<http://pfind.ict.ac.cn/software/pLink>) and filtered at 1% FDR as recommended by the developer (Yang et al., 2012). For simplicity, the crosslink score is reported as a negative value of the common logarithm of the original pLink score, i.e., score = -log<sub>10</sub>(pLink Score). The crosslinks observed with at least three crosslink-spectrum matches are shown in Table S2. For model building, a maximum distance of 30 Å between the C $\alpha$  atoms of the crosslinked lysines was allowed.

### **Cryo-EM sample preparation**

Three peak fractions of the fixative gradient containing B<sup>act</sup> complexes were quenched with 100 mM aspartate, concentrated, and buffer-exchanged in an Amicon 50 kDa cut-off unit, yielding 20–50 pmol of B<sup>act</sup> in 250 µL. Sample quality was assessed by negative-staining EM. For cryo-EM, a self-made continuous carbon film (approximately 3 nm thick) was floated on the sample for 1 min at 4°C and then attached to a QUANTIFOIL R 3.5/1 grid (Quantifoil, Jena, Germany) without glow discharge. Grids were blotted for 2 s using the blotting sensor and freeze-plunged in a Leica EM GP (Leica, Wetzlar, Germany) at 75% humidity and 4°C.

### **Cryo-EM data acquisition**

Grids were imaged in a Titan Krios (FEI, Eindhoven, the Netherlands) equipped with a Cs-corrector (CEOS, Heidelberg, Germany) on a Falcon III detector in linear mode. The microscope's internal TEM scripting was used for automated imaging. After manually choosing 3600 carbon holes, their focus z height was determined and saved with a defocus offset between 800 and 4000 nm. The microscope was kept in exposure mode during imaging without further tracking and focusing steps that would require a change in the illumination conditions. Due to the minimized aberrations through the usage of a Cs corrector, a spot scanning mode could be used without the introduction of significant coma. This means for every carbon hole, nine images were taken from different non-overlapping areas using image/beam shift. The microscope was operated in nanoprobe mode with a beam diameter

of 800 nm at a nominal magnification of  $\times 110000$ , corresponding to 1.16 Å per pixel in the final image. The total electron dose (40 electrons per Å<sup>2</sup>) was fractionated into 20 frames. The first frame was discarded, as it had inhomogeneous illumination due to the camera shutter. A total of 32000 micrographs were collected, using about 5000 micrographs per 24 h and about 1 week of image acquisition time.

### Image processing

Individual image frames were aligned and weighted according to electron dose using MotionCor2 (Zheng et al., 2017) to reduce the effects of drift and charging. The CTF of the micrographs was determined using Gctf (Zhang, 2016). All micrographs and power spectra were assessed visually using the cow suite (manuscript in preparation). Micrographs with severe flaws, such as large amounts of contamination, strong drift, devitrification, or lacking Thon rings better than 5 Å, were discarded (8000 micrographs). About 3.3 million particles were selected from the remaining micrographs in a template-free manner using Gautomatch. Subsequently, several iterations of 2D classifications in RELION and COW were performed, using subsets with maximum 200000 particles. Particles belonging to classes that clearly showed ice contamination, charging, drift, or false-positive images (such as carbon edges) (about 1.1 million), or that did not show clear molecular views, were discarded. After this, the best class averages were used to generate an initial 3D model using EMAN2 (Tang et al., 2007).

### Conformational landscape analysis

The initial 3D model was used as a reference in a 3D classification with RELION (Scheres, 2012). For this, the dataset was split randomly into subsets of 200000, and each set was classified into 20 classes without any masking. Classes were visually inspected to verify that they show the overall shape of the B<sup>act</sup>. No further particles were removed from this analysis. All obtained 3D volumes were refined individually, filtered to a resolution of 20 Å, normalized, and then aligned in UCSF Chimera (Goddard et al., 2007) to the core structure. These aligned 3D volumes were subjected to 3D PCA in the COW software suite. The result of this analysis is a coordinate transformation given by:

$$X_i = \bar{X} + \sum_{j=1}^n a_{ij} e_j$$

where every volume  $X_i$  is described as linear combination of the average volume  $\bar{X}$ , and the eigenvectors  $e_j$  are multiplied with the linear factor  $a_{ij}$ . Additionally, the eigenvectors are sorted within the PCA procedure according to the significance of the variance they describe. For initial sorting, only the first eigenvector was used, resulting in a linear description of the most significant conformational change. The trajectory given by  $a_{i,1}$  was visually inspected, and the volumes with their corresponding particles were sorted according to their linear factors into 8 subsets. The whole analysis including the 3D classification was repeated twice with the sorted particle subsets.

From the final iteration, the linear factors toward the first and the second eigenvector were used to calculate a conformational landscape. Therefore, each class is represented as 3D coordinate with linear factor  $a_{i,1}$  and  $a_{i,2}$  as x and y coordinates and the particle number in the class as z coordinate. Using

$$\Delta\Delta G = k_B T \ln \frac{p_i}{p_0},$$

with T being the absolute temperature and  $k_B$  the Boltzmann constant, particle numbers  $p_i$  relative to the particle number in the most populated state were converted to changes of free energy  $\Delta\Delta G$ . An energy landscape was calculated using the scattered interpolator function in MATLAB.

### 3D volume refinement

All 8 classes were refined individually first without a mask and then later with a mask around the spliceosome core regions. Several attempts were made to combine data from different states to obtain a higher-resolution structure for the spliceosomal core. The combination of the states 3, 4, and 5 provided the highest-resolution structure for this.

### Model building

An initial rigid body fit of all visible proteins was made using models from different sources, which were docked into the density guided by crosslinks and the yeast B<sup>act</sup> structure (Rauhut et al., 2016) using UCSF Chimera (Goddard et al., 2007). Additional aid for regions, which had to be modeled at least partly *de novo*, was obtained using the secondary structure prediction server PSIPRED (Buchan et al., 2013). An initial rigid body refinement was performed using real space refinement in PHENIX (Adams et al., 2010) and subsequent manual modeling in Coot (Emsley and Cowtan, 2004). Secondary structure restraints were then generated using *phenix.ksdssp* for proteins. All secondary structure restraints were visually inspected, and additional restraints for RNA and ligands were added manually. Several iterative rounds of real space refinement in PHENIX and manual modeling in Coot were then performed, whereby the last PHENIX refinements included ADP refinement to calculate B factors.

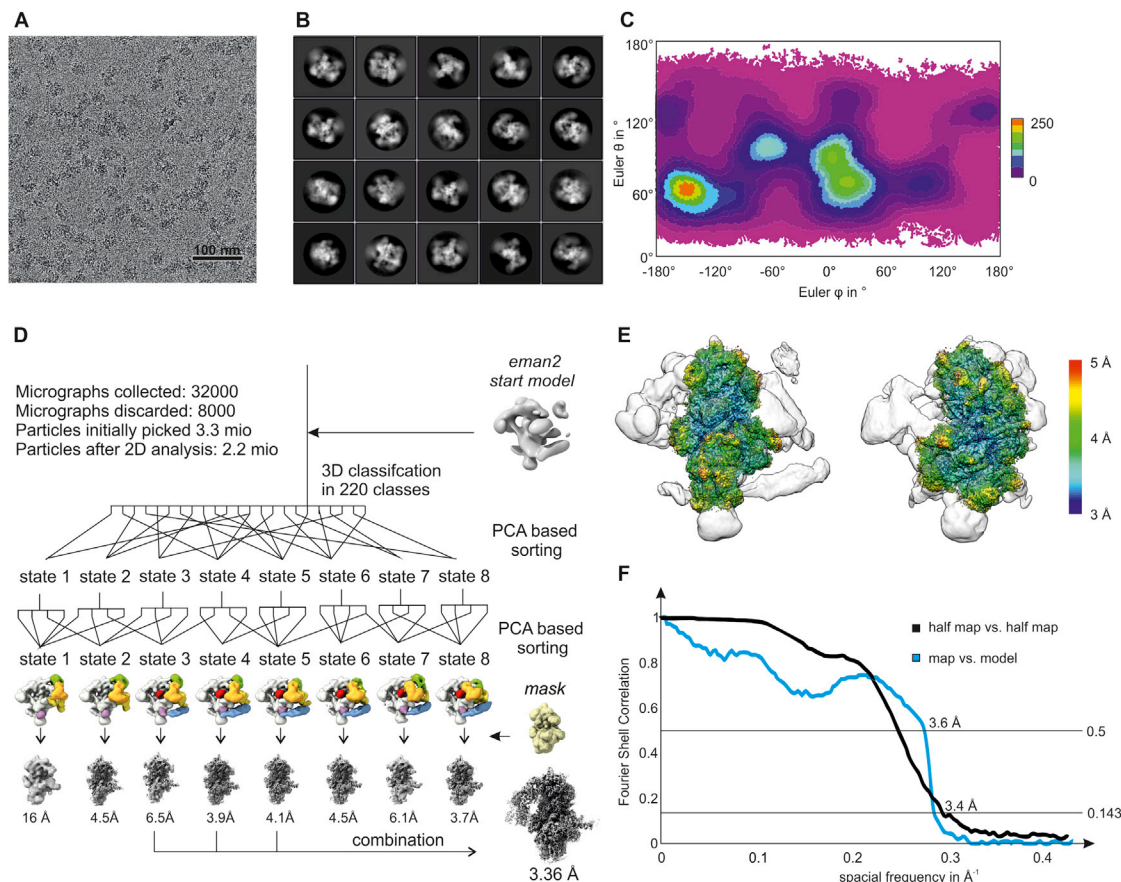
To account for local resolution differences in the EM density map, density fit analysis was used as a guideline to define the level of structural details interpreted in the final model (Table S1). Model quality was assessed using MolProbity within PHENIX.

To account for local resolution differences in the EM density map, density fit analysis was used as a guideline to define the level of structural details interpreted in the final model (Table S1).

#### DATA AND SOFTWARE AVAILABILITY

The COW software (developed in the Stark laboratory) is currently in an open beta test phase and can be downloaded from [www.cow-em.de](http://www.cow-em.de).

EM density maps of the complete complexes and the masked cores have been deposited in the EMDB under accession numbers EMD-4233, 4234, 4235, 4236, 4237, 4238, 4239, 4240 and EMD-4247, 4248, 4249, 4250, 4251, 4252, 4253, 4254, respectively. Raw micrographs have been uploaded to the EMPIAR database. The high-resolution map and its modeled atomic coordinates have been deposited in the EMDB and Protein Data Bank under the following accession numbers, EMD-4255 and PDB-6FF4. The composite model including all visible densities was deposited as alanine trace under the identifier PDB-6FF7.



**Figure S1. Cryo-EM Analysis, Related to STAR Methods and Figure 1**

(A) Typical raw electron micrograph of spliceosomes embedded in vitreous ice.

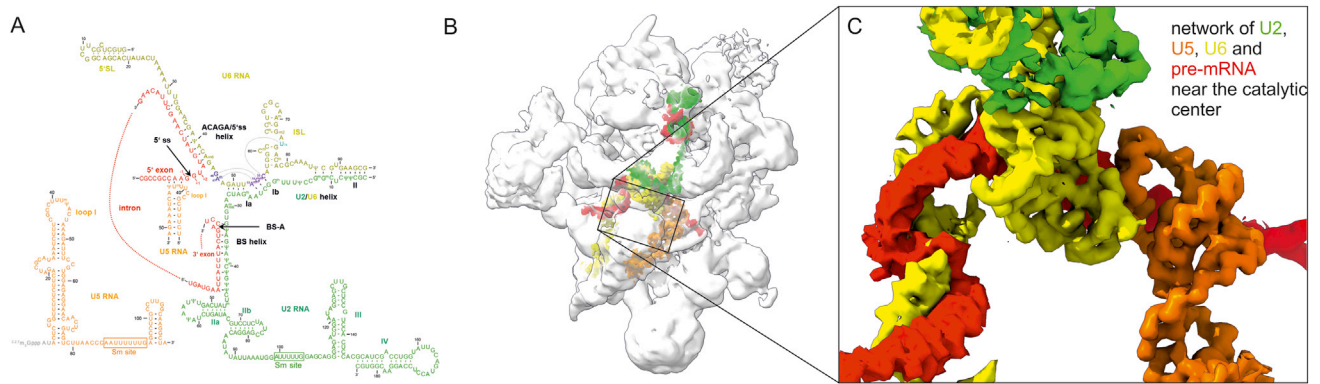
(B) Representative 2D class averages.

(C) The Euler angle distribution of the final 3.4 Å structure.

(D) Schematics of the particle sorting scheme. All micrographs were inspected visually, and micrographs showing optical artifacts or large contaminations were discarded. All picked particles were subjected to 2D classification, and particle classes showing particle views were subjected to 3D analysis. An initial model was generated from the best 2D class averages using eman2. With this, all particles were 3D classified in packages of 100,000 in 10 classes each. All classes were aligned in UCSF Chimera and subjected to PCA analysis. Consequently, the classes were sorted into eight states according to their linear factor toward the first eigenvector. Subsequently, all particles belonging to a certain state were combined and reclassified in packages of 100,000 in 10 classes and again sorted using PCA. In this stage, between 3%–13% of particles changed to the neighboring state. After sorting, all particles belonging to the same state were refined without using a mask. Afterward, a mask around core region was applied, and the structure was re-refined. Finally, as classes 3, 4, and 5 showed a very similar conformation in the core region, they were combined and refined to a 3.4 Å structure.

(E) Local resolution map of the spliceosomal core structure. The core structure shows a resolution range between 3 and 5 Å. The map of the complete particle shown as translucent overlay was determined at resolutions between 10 Å and 25 Å.

(F) Fourier shell correlation curves of two independently refined subsets of particles (map-map) as well as the final refined map and the model (map-model) are shown. The map-map curve indicates a resolution of 3.4 Å at a threshold of 0.143. The map model curve shows a resolution of 3.6 Å using the 0.5 threshold.

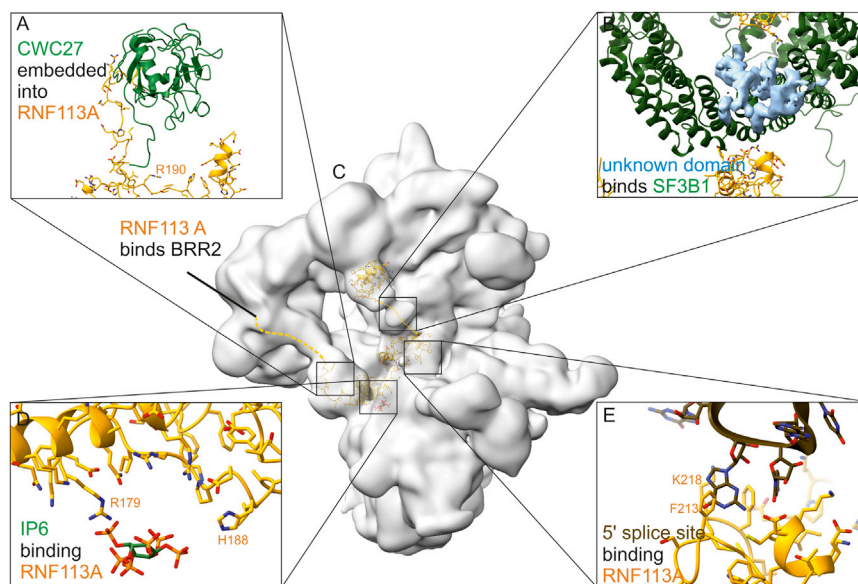


**Figure S2. RNA Network, Related to Figure 2**

(A) Schematic representation of the RNA network of the human B<sup>act</sup>.

(B) Density representation of the RNA network of the human B<sup>act</sup>.

(C) The pre-mRNA is bound through its interactions with U6 and U5 RNA.



**Figure S3. Path of RNF113A Through B<sup>act</sup>, Related to Figure 2**

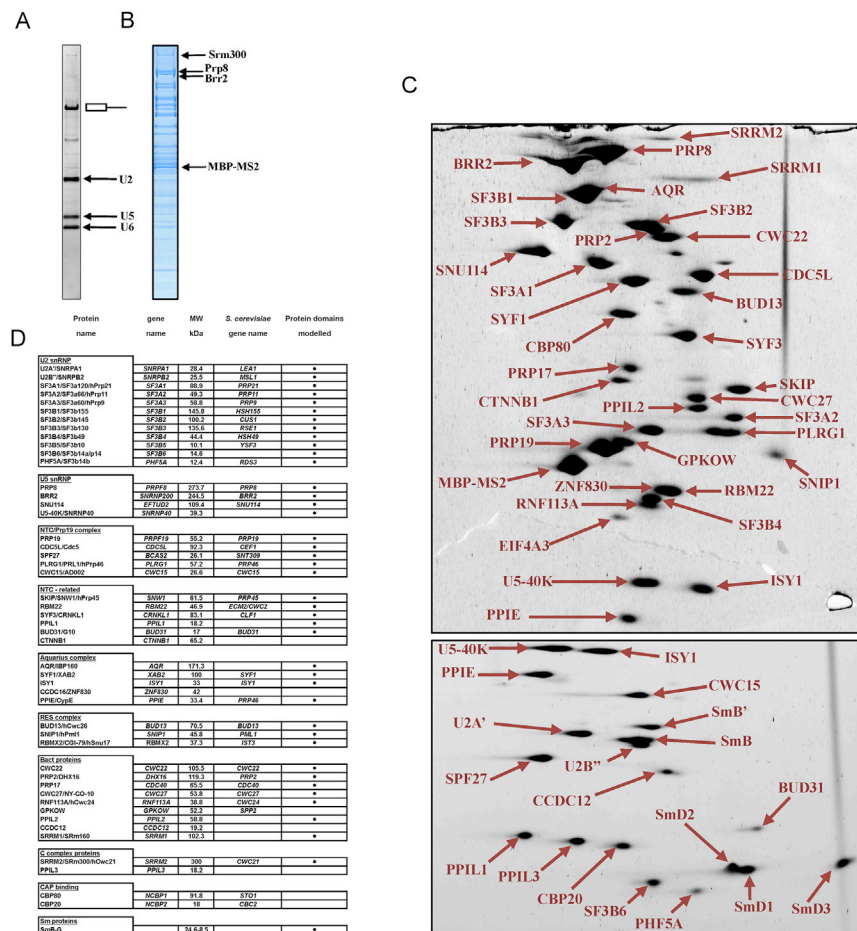
(A) RNF113A binds the N terminus of CWC27.

(B) An unidentified domain is visible that binds SF3B1.

(C) Path of RNF113A is shown as in the overall complex. Unmodeled parts are shown as a dashed line, based on crosslinking mass spectrometry evidence.

(D) RNF113A binds the PRP8-bound IP6.

(E) RNF113A binds the 5' ss.



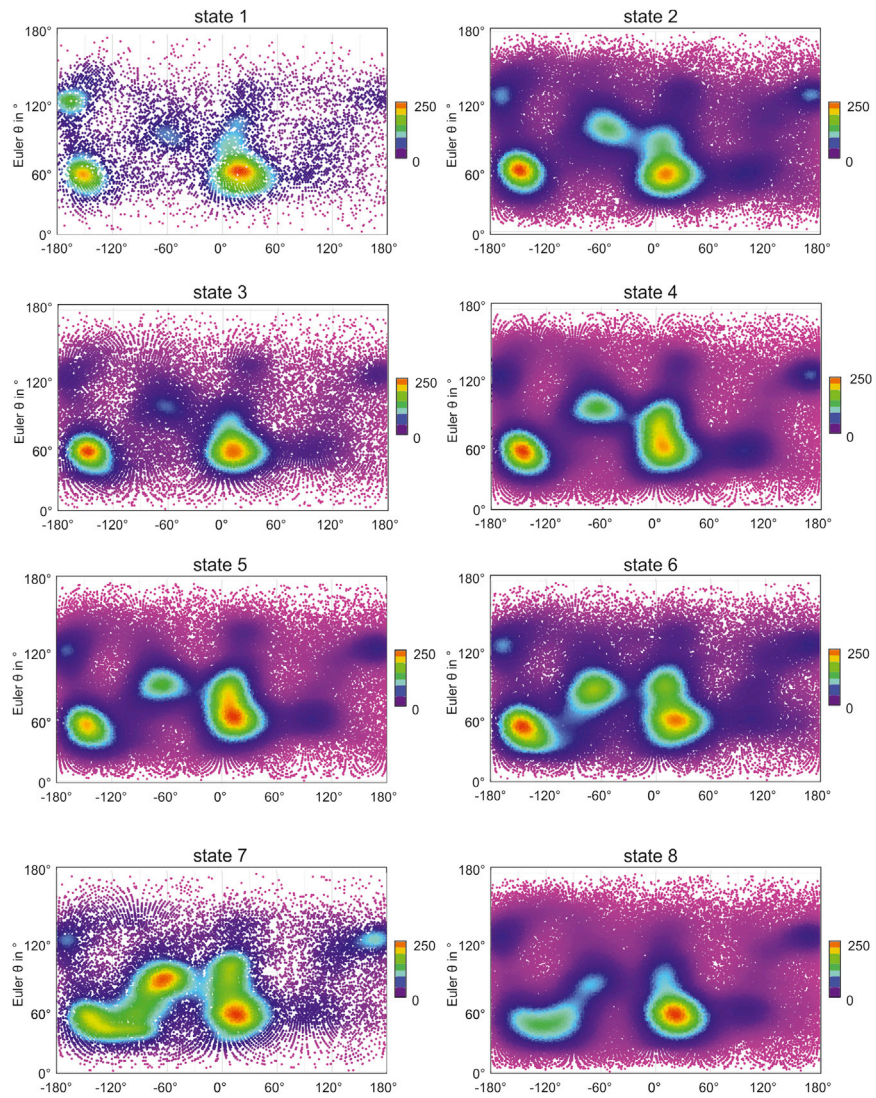
**Figure S4. Biochemical and Mass Spectrometric Analysis of the Sample, Related to STAR Methods**

(A) RNA gel of the sample. The U2, U5, and U6 snRNAs, as well as the unspliced pre-mRNA (box pictogram), are clearly visible. Due to the absence of U4 RNA, the complex shows a post-B state of the spliceosome.

(B) SDS-PAGE gel of the sample. Although a few proteins are clearly identifiable, a 2D analysis (shown in C) was necessary to distinguish the individual components.

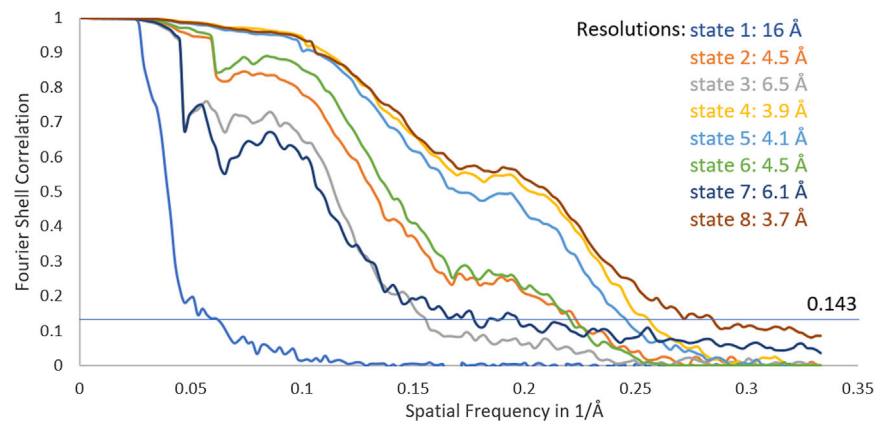
(C) 2D gel electrophoresis of the preparation. All proteins are clearly distinguishable, and B<sup>act</sup>-specific proteins are visible.

(D) A list of all proteins identified by mass spectrometry in the sample. Proteins were marked as modeled if at least parts of the respective protein could be identified in the final structure.



**Figure S5. Angular Distribution of All Models, Related to STAR Methods**

The Euler angle distribution of all 8 models did not reveal significant differences in angular orientation of particles that contributed to the structures. This indicates that there is no bias based on different variations in projection orientation during the classification procedure.



**Figure S6. FSC Curves, Related to STAR Methods**

The FSC curves for all 8 models are given.