

***High-resolution structure determination of  
human spliceosome complexes by cryo-EM***

Dissertation

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program

*Biomolecules: Structure - Function - Dynamics*

of the Georg-August University School of Science (GAUSS)

submitted by

**Karl Bertram**

from Cologne

Göttingen, November 2018

## **Thesis Committee**

**Prof. Dr. Holger Stark**

Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen

**Prof. Dr. Ralf Ficner**

Department of Molecular Structural Biology, Georg-August-University, Göttingen

**Prof. Dr. Henning Urlaub**

Bioanalytical Mass Spectrometry, Max Planck Institute for Biophysical Chemistry, Göttingen

## **Members of the Examination Board**

**Referee: Prof. Dr. Holger Stark**

Department of Structural Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen

**2<sup>nd</sup> Referee: Prof. Dr. Reinhard Lührmann**

Department of Cellular Biochemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

## **Further members of the Examination Board**

**Prof. Dr. Ralf Ficner**

Department of Molecular Structural Biology, Georg-August-University, Göttingen

**Prof. Dr. Detlef Doenecke**

Department of Molecular Biology, Georg-August-University, Göttingen

**Dr. Juliane Liepe**

Quantitative and Systems Biology, Max Planck Institute for Biophysical Chemistry, Göttingen

Date of oral examination: 20. December 2018



*To my dear mother, who introduced me to the world of science  
but could not live long enough to see the completion of this work.*



## **Acknowledgments:**

I would like to express my sincerest gratitude to several people who influenced, guided and/or advised me during the time I prepared this work.

Initially, I would like to cordially thank Holger Stark as the head of our group. Not only for offering me the chance to pursue my research ambitions in his laboratory but also for receiving his honest support whenever it was needed. Upon coming to Göttingen I was able to enter a lively group of welcoming colleagues that was on top scientifically equipped with everything imaginable. I can always trust in his supporting advice, which is most certainly benefitting in one or the other way. I truly appreciate it, thank you Holger!

Secondly, I deliberately like to thank Reinhard Lührmann as the leader of the department for Cellular Biochemistry for his collaboration and consistently great advice. He arranged for our cooperative projects to thrive from the spliceosomal side and was, and still is, always open for fruitful discussions that can easily last until nightfall.

The members of my thesis committee, Holger Stark, Ralf Ficner and Henning Urlaub then provided for the additional framework of scientifically or otherwise valuable advice that I was happy to benefit from. Thank you very much for your support.

On that way, I also express my honest gratitude to Wen-ti Liu and David Haselbach as the supervisors of my early work. Both of them introduced me to the practical world of cryo-EM and their motivation for helpful discussions and to pass on their knowledge in the busy environment of the lab was unmatched. I am truly grateful.

Besides these two, the members of our Office 113, Lukas Schulte, Fabian Henneberg and Kashish Singh became true friends over the years and accompanied me during the countless hours in the lab with their (unique) humour, good attitudes and helpful advice at innumerable occasions. It was a great pleasure working and enjoying the time with you guys!

Furthermore, I will thank the remaining members of our group, in particular Erik Schliep, Sabrina Fiedler, Georg Bunzel, Uwe Lücken and Dietmar Riedel but also Cole Townsend and Zhenwei Zhang as my former master students for providing a work atmosphere that I was always happy to return to in the mornings. I likewise enjoyed working together with the members of Reinhard Lührmann's department, such as Berthold Kaster, Dmitry Agafonov, Olexandr Dybkov, Ulrich Steuerwald and Klaus Hartmuth, who especially at the beginning of my work took great time and efforts to introduce me to the spliceosome and broadened my horizon with countless of interesting discussions. Thank you very much.

Finally, it is hard to express how truly much I thank my family, my parents Anja and Roland who unforgettably at all times encouraged, backed and enabled me to completely engage in whatever I was excited to do, no matter the difficulties, my sisters Josefine and Franziska, my partner Katharina and friends who have all together supported and guided me in my ambition of moving to Göttingen in order to do my doctoral studies over here. Even though all of them were fully aware what it meant, that I could inevitably spend less time with them by doing so. It was unimaginable knowing all of you at my side. Thank you!



# Content

<b>List of figures</b> .....	<b>ix</b>
<b>List of abbreviations</b> .....	<b>xi</b>
<b>1 Abstract</b> .....	<b>1</b>
<b>2 Introduction</b> .....	<b>2</b>
<b>2.1 Molecular machines</b> .....	<b>2</b>
2.1.1 The rise of molecular machines.....	2
2.1.2 The machinery of the Cell .....	2
2.1.3 Types, sizes and building blocks.....	3
2.1.4 <i>Macromolecular machines</i> .....	6
<b>2.2 Electron cryo-microscopy (cryo-EM)</b> .....	<b>8</b>
2.2.1 How to visualize atomic structures? .....	8
2.2.2 General features of the electron microscope.....	9
2.2.3 Optical aberrations and resolution.....	14
2.2.4 Sample preparation and imaging under cryo-conditions .....	17
<b>2.3 Image and data processing</b> .....	<b>18</b>
2.3.1 Micrograph quality control and pre-processing .....	18
2.3.2 Aberration estimation, particle picking and extraction.....	19
2.3.3 Particle classification and averaging.....	20
2.3.4 Angular reconstitution and 3D reconstruction .....	20
2.3.5 (Atomic) model building and interpretation.....	23
<b>2.4 Pre-mRNA splicing</b> .....	<b>25</b>
2.4.1 Basic splicing mechanics .....	27
2.4.2 General components of the spliceosome.....	29
2.4.3 Spliceosome assembly and function.....	32
<b>2.5 Aim of the study</b> .....	<b>38</b>
<b>3 Publications</b> .....	<b>39</b>
<b>3.1 Cryo-EM structure of a human spliceosome activated for step 2 of splicing</b> .....	<b>41</b>
<b>3.2 Cryo-EM structure of a pre-catalytic human spliceosome primed for activation</b> ..	<b>60</b>

<b>4 Discussion</b> .....	<b>86</b>
<b>4.1 The Pre-catalytic human spliceosome</b> .....	<b>86</b>
4.1.1 U2 snRNP components .....	87
4.1.2 Dramatic restructurings after tri-snRNP integration .....	87
4.1.3 Functional insights into the role of B-specific proteins.....	89
4.1.4 Differences in human and yeast spliceosome activation .....	90
<b>4.2 The catalytically activated human spliceosome</b> .....	<b>91</b>
4.2.1 Molecular architecture and structural conservation between species.....	91
4.2.2 Catalytic centre.....	92
4.2.3 Large-scale structural rearrangements modulate splicing activity .....	93
4.2.4 Extended ACAGA and BSH helices in the human system .....	94
<b>4.3 General aspects of splicing &amp; utilizing cryo-EM as a tool to reveal its mechanism</b>	<b>96</b>
4.3.1 Evolutionary conservation of RNA splicing .....	96
4.3.2 RNA helicases that modulate splicing activity from a distance.....	97
4.3.3 Completeness of the reconstructed spliceosome structures .....	98
4.3.4 Non-uniform resolution estimates in cryo-EM .....	98
4.3.5 Anisotropic resolution due to preferred particle orientations in Cryo-EM .....	99
<b>4.4 Perspectives</b> .....	<b>100</b>
<b>References</b> .....	<b>102</b>

## List of figures:

<b>Fig. 1</b> Modular composition of a macromolecular machine .....	7
<b>Fig. 2</b> General schematic of an electron cryo microscope (cryo-EM).....	10
<b>Fig. 3</b> Specific optical aberrations in the electron microscope.....	15
<b>Fig. 4</b> Simulated Projection sphere illustrating the relationship between 2D projections and their 3D origin.....	21
<b>Fig. 5</b> Typical metazoan pre-mRNA.....	27
<b>Fig. 6</b> Chemistry of pre-mRNA splicing .....	28
<b>Fig. 7</b> Sequence and predicted secondary structure of the five human spliceosomal snRNAs.....	30
<b>Fig. 8</b> Splicing cycle.....	33





## List of abbreviations:

2D	Two-dimensional
3D	Three-dimensional
aa	Amino acids
BS	Branch site
BS-A	Branch site adenosine
BSH	Branch-site helix
CB	Cajal bodies
CC	C-terminal helicase cassette (of the protein BRR2)
CCD	Charge coupled device
CMOS	Complementary metal oxide semiconductor
cryo-EM	Electron cryo-microscopy
C <sub>s</sub>	Spherical aberration coefficient
CTF	Contrast-transfer function
CX-MS	Crosslinking coupled to mass spectrometry
DQE	Detective quantum efficiency
FRC	Fourier Ring Correlation
FSC	Fourier Shell Correlation
GII intron	Group II self-splicing intron
HAT	Helical half a TPR (domain)
IBC	Intron-binding complex
ISL	Internal stem loop
LMNG	Lauryl Maltose Neopentyl Glycol
MSA	Multivariate statistical analysis
NC	N-terminal helicase cassette (of the protein BRR2)
nm	Nanometer
NTC	(Prp) nineteen complex
PCA	Principle component analysis
pre-mRNA	Precursor messenger RNA
PRP	Precursor RNA Processing (protein family)
RES	Retention and splicing (complex)
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RRM	RNA recognition motive
rRNA	Ribosomal RNA
SNR	Signal-to-noise ratio

snRNP	Small nuclear ribonucleoprotein
SS	Splice sites
STEM	Scanning transmission electron microscopy
TEM	Transmission electron microscopy
TMG	5' trimethylguanosine
TPR	Tetratricopeptide repeat
U-snRNA	Uridylic-acid-rich small nuclear ribonucleic acid
WD-40	Tryptophan-aspartic acid repeat domain
Znf	Zinc finger
µm	Micrometres

# 1 Abstract

The eukaryotic splicing of precursors to mRNA is facilitated by a highly dynamic, multi-megadalton macromolecular machine termed the spliceosome. The underlying chemical reaction features the excision of an intron, which is followed by the re-ligation of two exons with single nucleotide precision. The spliceosome therefore actively participates in the flow of genetic information. How catalysis is mechanistically achieved and why the dynamic nature of the molecular machine is essential for its function was poorly understood in the past.

This work presents the first high-resolution structures of human spliceosomes in the pre-catalytic and the catalytically activated phase of assembly. Elucidated by cryo-EM, the molecular architectures of the B and C\* complex reveal significant insights into the mechanism of catalytic activation and general activity. The pre-catalytic B complex thereby shows a distinctive spatial separation of the reactive pre-mRNA BS-A and 5'SS elements during spliceosomal assembly. Mechanistically, the structure sheds light on the tremendous restructuring events that take place upon the integration of the tri-snRNP into the pre-spliceosomal A complex. B-specific proteins like PRP38, SNU23, MFAP1 or SMU1 specifically stabilize the B complex configuration and prevent premature activation by contacting the important U6 snRNA ACAGA box helix and RNA helicase BRR2. Intriguingly, a detailed comparison between the yeast and human pre-catalytic spliceosome structures unexpectedly reveals a potentially different catalytic activation pathway in higher and lower eukaryotes.

The molecular architecture of the C\* complex highlights a profound conservation of the catalytic core of the assembly between species once the spliceosome is catalytically activated. Unexpected differences nonetheless exist between the structural organization of yeast and human spliceosomes: for example, the ACAGA box and BSH helices are characteristically extended in the human spliceosome, which potentially compensates for the degenerate appearance of the corresponding signal sequences in the pre-mRNA of higher eukaryotes. In addition, metazoan-specific proteins such as RBM22 or IBP160 (Aquarius) can be localized and likely assist in modulating the splicing activity by interacting with the pre-mRNA and proximal protein factors. Large-scale remodelling events of the remaining U2 snRNP components are furthermore found to convey their functionally essential dynamic trajectories onto the much smaller entities at the catalytic core of the C\* complex. For example, the BSH that spatially occupies the catalytic centre in the post-step 1 C complex is repositioned accordingly in the pre-step 2 C\* complex.

Besides clarifying the molecular architecture of the spliceosome itself, the results presented in this work contribute towards a better understanding of the involved assembly pathways and the mechanism of catalysis. The substantial differences between yeast and human spliceosomes during catalytic activation and in the handling of pre-mRNA stabilization within catalytically activated spliceosomes may furthermore add to the evolutionary understanding of RNA splicing.

## 2 Introduction

### 2.1 Molecular machines

Molecular machines are the functional entities that carry out all work within the cell. Comparable to their macroscopic counterparts, molecular machines process a defined set of raw materials in a characteristic way in order to obtain a specific product. In a molecular, microscopic context this means that molecular machines usually catalyse and control a well-defined chemical reaction by appropriately positioning the individual reactants at their specifically organized catalytic centre. The following sections will introduce the evolutionary development of molecular machines in life, their catalytic capabilities and their general, often modular appearance.

#### 2.1.1 The rise of molecular machines

Energy is employed in all kingdoms of life to chemically transform matter from the direct surroundings into structurally better-defined assemblies of functional relevance – a key principle of the survival and function of every self-sustaining organism. While thermal energy, in the natural, molecular context, is available at all times and directly resulting in Brownian motion, other energy sources – such as those stored in chemical bonds or electromagnetic fields – may be utilized whenever available or needed. In life, the process of evolution, fostered by billions of years of trial and error types of experiments, eventually found some of the most elegant solutions to the problem of how to combine a limited number of chemical elements and the energy available at a certain area in order to form small (organic) molecules, but equally well multi-million atom conglomerates of specific function.

Viewed as a combinatorial problem where a certain amount of atomic building blocks are available in order to be rearranged by limited quantities and types of energy until a “functional” solution is found, it turns out that the time required for this process to be successful is directly correlated to the total amount of sampling possible. Therefore, while the actual origin of life is still under debate, it becomes inevitable that small organic molecules with little structural complexity must have initially made the beginning. By adding time, the right environment and a further myriad rounds of recombination, ribonucleic acids (RNA) eventually assembled and likely laid the foundation for the very first molecular machines (Joyce, 1989; Yarus, 2018). Beginning with the most basic catalytically active entities of di- or mononucleotides, some of which are still utilized as essential cofactors in today’s modern life enzyme chemistry (White, 1976; Yarus, 2011), more complex structures were derived, paving the way for genetic reproduction and every other cellular function.

#### 2.1.2 The machinery of the Cell

Since the rate of an uncatalyzed chemical reaction is often found to be too slow to foster a biological process, kinetics need to be accelerated through more efficient catalysis. While evolution must have initially solved a seemingly infinite number of trials in order to construct the first molecular machine that e.g. catalyses a certain chemical reaction, the evolved spatial

arrangement of atoms in that enzyme then consequently allowed for the directed catalysis of e.g. the very same reaction with much improved specificity and rate. Viewed in the light of this *Gedankenexperiment*, efficient (macro) molecules that positively influence the rate of a specific (bio) chemical reaction are absolutely essential for maintaining homeostasis in any living organism. In fact, the metabolism and reproductive capability of any cell is compellingly dependent on the function of molecular machines, which often directly execute or are involved in enzymatic functionality.

While the exact number of different molecular complexes in the cell is not known, in particular those comprising the proteome, it is estimated that the human genome encodes anywhere from several tens of thousands to millions of functional entities (Ponomarenko et al., 2016). Out of this pool, enzymes make up a significant fraction and are commonly classified according to their functions into six subsets. *Oxidoreductases* catalyse redox reactions; *Transferases*, *Hydrolases* and *Lyases* specifically break and potentially reform a different chemical bond; *Isomerases* selectively isomerise a certain molecule; and finally *Ligases*, which specifically create bonds between two previously unconnected molecules (Cornish-Bowden, 2014). All of these molecular machines therefore execute a specific catalytic function. The structural and biochemical characterization of molecular machines during the last century has yielded detailed insights into the function, requirements, and order of the enzymatic processes that catalyse the chemistry of life.

### 2.1.3 Types, sizes and building blocks

The function and organization of molecular machines often appears to resemble that of the much larger mechanical devices, such as those that are man-made from metal or other materials. In this light, considering a *macroscopic* machine, it will always have been designed to serve a dedicated purpose. In a generalized manner this purpose then aims at facilitating a certain process as efficiently as possible; for example, cutting a piece of wood with a circular saw or providing rotational motion using the torque vectoring components of a fuel-powered engine. Every device of this kind will consist of a functional area, e.g. the saw blade in the example of the circular saw and some surrounding scaffolding material, such as the housing and handles of the depicted tool. Furthermore, every machine is constructed from a certain material, or a mixture of multiple; may it be wood, metal or plastic.

Enzymes, or *molecular* machines, are equally well composed from a limited set of structural entities: essentially the 20 amino acids and/or certain nucleic acids. Within the microscopic world, instead of materials like wood or metal, these molecular building blocks directly assemble the entire framework of any molecular machine: the mechanical scaffold or “housing” as well as the catalytically active site of the enzyme. Importantly, the catalytic centre of any enzyme can only be functional if the molecular scaffold enclosing it is found to be in its predestined spatial conformation. Hence, the catalytic properties of an enzyme are directly dependent on the structural entities that frame the catalytically active area. The latter may then catalyse a specific chemical reaction by arranging a defined set of reactants in a spatially-precise manner. Individual reactive groups, ions or molecules that are commonly present in close proximity can then interact with a selectively-bound reactant in order to perform the desired chemical reaction. This carefully-controlled molecular environment ensures a high selectivity for and control over the reaction educts and products, while precisely targeting the catalytic activity to the reactants of choice. Briefly illustrating this chemical precision it is interesting to

note that enzymes even routinely perform enantioselective catalysis (Jaeger and Eggert, 2004); a process that is generally hard to achieve in non-biocatalytic chemistry.

To elucidate some of the general features resulting in the great diversity of enzymatic functionality, the following subsections will give a short introduction to the commonly observed molecular frameworks of enzymatic complexes, their sizes, and a frequently recognized structural modularity.

### **2.1.3.1 Molecular frameworks**

While the moieties in the active centre of a molecular machine actually facilitate the chemistry of an enzymatic reaction, the molecular framework surrounding this area is equally important in maintaining and regulating the catalytic activity of any enzyme. In nature, enzymatic complexes are therefore generally composed of either nucleic acids entirely, a hybrid structure of nucleic acids and protein components, or exclusively proteins.

As mentioned before, biocatalysts relying on nucleic acid chemistry entirely are thought to be the earliest representatives of their kind. Termed *ribozymes* (*ribonucleic acid enzymes*), they specifically catalyse phosphate group transfers and peptide bond formation reactions (Fedor and Williamson, 2005). Some ribozymes thereby completely rely on nucleic acids as molecular scaffolds and catalytic moieties, emphasising the importance of ribonucleic acids as chemically diverse building blocks within molecular machine architecture. Enzymes of this family commonly catalyse phosphodiester bond cleavage and transfer by an  $S_N2$ -type-in-line reaction mechanism, a type of reaction that is crucially important in nucleic acid processing and splicing (Long and Uhlenbeck, 1993; Scott and Klug, 1996). One such example are self-cleaving hairpin ribozymes, which catalyse a specific cleavage reaction utilizing the 2' hydroxyl group of a nucleic acid's ribose as the nucleophile. The process thus generates a cleaved, free 5' hydroxyl terminus at one nucleobase and a cyclic 2',3' phosphate termini at the other end of the cleavage site (Ferré-D'Amaré et al., 1998). Strikingly, the RNA moieties involved were found perform this type of catalysis entirely based on nucleic acid chemistry alone, in particular without the involvement of water, metals or other cofactors during the reaction (Nesbitt et al., 1997). A second class of ribozymes that exclusively utilize ribonucleic acids as their molecular framework commonly feature divalent metal cations and water molecules as cofactors to perform self-splicing, instead of self-cleavage (Fedor and Williamson, 2005). The catalysed reaction is reversible and specifically removes a certain intronic sequence from a preliminary RNA transcript. During catalysis, cofactors like water molecules and two  $Mg^{2+}$  ions are involved in first cleaving a phosphodiester bond at a particular position then initiating the religation to a different, appropriately positioned downstream RNA 3' hydroxyl group in the reverse manner (Steitz and Steitz, 1993). The Group I and II self-splicing introns are particularly important representatives of this class of enzymes, of which the corresponding crystal structures significantly aided in explaining their catalytic mechanism (Golden et al., 1998; Kruger et al., 1982; Toor et al., 2008). Considering the rigidity and catalytic activity of these complexes it is noteworthy that divalent metal cations likewise often associate with nucleic acid-based molecular frameworks in a structurally supportive manner. Most relevantly, these interactions are promoted between hard, divalent cations like  $Mg^{2+}$  and the various oxygens of the nucleobases and their phosphates. Furthermore, metal cations are frequently integrated at different regions of the molecular framework that are characteristic for the function of the ribozyme. Efficiently neutralizing negative charges (e.g. those of the phosphate backbone), these

cations thus enable a closer and structurally more rigid packing of RNA-based structures (Draper, 2004; Draper et al., 2005).

Further extending the defined but limited chemistry of nucleic acid moieties alone, molecular machines of greater catalytic potential developed. The incorporation of protein domains into the previously RNA-only molecular framework therefore resulted in a significant hybrid class of enzymes termed ribonucleoproteins (RNP), catalysing reactions of greater variability. While the actual catalytic centre of these complexes is still comprised of RNA only, some or many of the structural domains enveloping the active site actually consist of amino acids. Evolutionarily, a plausible hypothesis states that ribosomes were the first representatives of this hybrid RNP class of enzymes (Root-Bernstein and Root-Bernstein, 2015; 2016). Composed of significantly more RNA than amino acid residues by weight (Londei et al., 1983), its active centre promotes catalysis through structurally defined interactions between nucleobases and their ribose moieties, water molecules and the amino acid educts and products (Leung et al., 2011). The peptide transfer required for protein synthesis is then achieved in a two-step tetrahedral intermediate mechanism (Hiller et al., 2011). The spliceosome family represents another crucially important member of the hybrid molecular framework RNP class of enzymes. While spliceosome dependent precursor messenger RNA (pre-mRNA) splicing is catalysed through a two-metal mechanism by an RNA-derived catalytic centre, similar to that of the Group II self-splicing intron ribozyme described above, 80-90% of the molecular machine is composed of a protein scaffold, rather than of RNA itself (Fica et al., 2013; Will and Lührmann, 2011). Proteins, not RNA, therefore control many of the structural and regulatory aspects of the highly regulated and precise pre-mRNA splicing process. Other examples of functionally relevant RNP derived enzyme families are furthermore given by telomerase and RNase P, both either catalysing the cleavage or ligation of (poly-)RNA molecules, respectively (Marquez et al., 2006; Nguyen et al., 2018). The addition of protein factors to RNA based catalytic networks thus significantly enhanced their regulatory and functional capabilities. While the catalytic potential of the above mentioned enzymes still seems to be limited to peptide bond formation and RNA based hydrolysis or ligation reactions, some of the most fundamental chemistry in life, including pre-mRNA splicing and protein synthesis, is actually facilitated by this class of highly complex molecular machines.

Walking down the timeline of evolution, metabolic processes in biology nonetheless eventually required an even more diverse chemistry than that achievable by an entirely RNA based catalysis only. Extensive three-dimensional protein structures seemed to fulfil this task and, compared to their RNA based counterparts, consequently offered a structurally more versatile backbone as well as a chemically highly variable pool of side chain moieties that ergo then participate in three dimensional structure formation and/or catalysis (Narlikar and Herschlag, 1997). While peptide entirely derived enzymatic complexes initially may have been functional solely utilizing eight or nine disparate amino acids, evolution soon favoured the incorporation of more than 20 different types of side chain moieties during protein synthesis (Müller et al., 2013). In consequence, the latter then allowed for a great wealth of different, chemically versatile molecular machines to emerge. Considering the innumerable catalytic mechanisms employed by protein-derived enzymes, however, a detailed explanation of their mechanisms would be far beyond the scope of this introduction. Nevertheless, the structural framework composing protein-derived molecular machines is as important in order to enable their catalytic capabilities and proper function as it is to the RNA derived examples described above.

### 2.1.3.2 *Functional domains and their size*

In analogy to a macroscopic machine, molecular machines are often composed of more than a single functional unit. The amino acids that assemble these individual units are commonly arranged in a specific spatial arrangement, which is characterized by primary, secondary and tertiary structural elements. Extending this classification, so called protein domains may be defined in order to describe a functional unit of compact, mostly rigidly-folded primary and secondary structure elements that are assembled into a recognizable, characteristically shaped domain of tertiary structure (Richardson, 1981). Similar motifs likewise exist in RNA-derived structures, where hairpin- or stem-loop based structures form stable, generally independent molecular units in three dimensional space (Svoboda and Di Cara, 2006). These structurally well-defined domains, which often retain their specific function also in an isolated context, may then serve as the smallest unit of functional building blocks available to nature in order to construct any kind of molecular machine in a modular manner.

Given the wealth of different domain structures that evolved over time, many of them appear to be well conserved throughout species and are consequently utilized in a large number of molecular complexes. Some of these domains, such as the *Tetratricopeptide Repeat* (TPR) or HEAT folds often stack up to form alpha-helical solenoid structures that may selectively influence the molecular framework of a complex (Blatch and Lässle, 1999; Kobe et al., 1999). Others, like e.g. *RNA Recognition Motives* (RRM), *zinc fingers* (Znf) or tryptophan-aspartic acid repeats (WD-40), facilitate the stable association of nucleic acids with hybrid molecular complexes, or other tasks of structural maintenance (Bandziulis et al., 1989; Neer et al., 1994; Pabo et al., 2001). Still others provide examples for domain folds that may independently exert specific catalytic functions. While the latter is sometimes required for the overall functionality of a (macro-) molecular machine, the catalytic activity itself may also get lost or become dormant in different variants of an otherwise homologous domain. Hence, the structurally well-conserved domain then exerts a different but not catalytic function. Prime examples for these cases are the RNase H or RecA (-like) domains, which are found to perform catalytic activity in many complexes while seemingly carrying out different, non-enzymatic functions in others (Davies et al., 1991; Schellenberg et al., 2013; Schmitt et al., 2018).

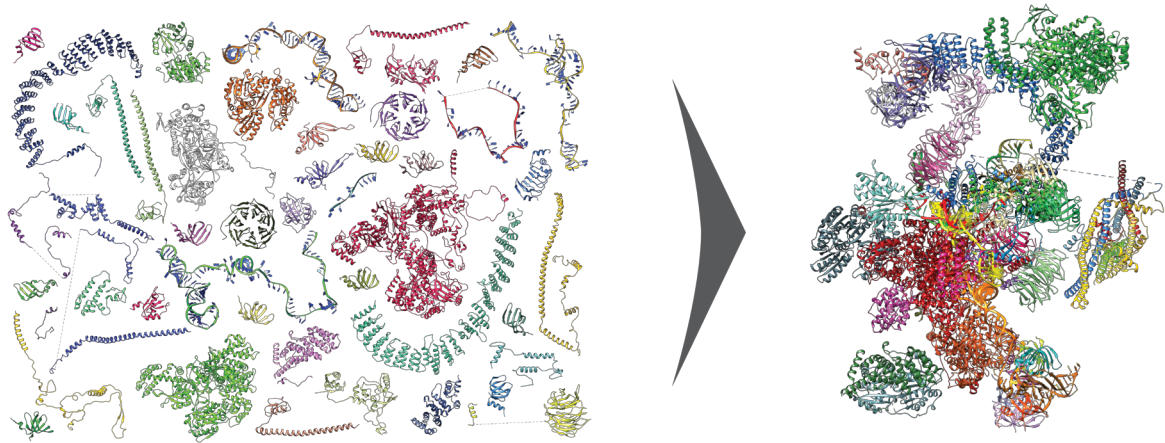
Even though the specific function of the various domains may be exceptionally diverse, the most frequently observed size or molecular weight is not. Commonly, the recognized size of a protein domain was rather determined to be around 125-150 amino acids (aa) in size, or roughly ~ 14kDa in molecular weight (Berman et al., 1994). Exceptions obviously exist, where particular domain frameworks were found to be not much larger than 36 amino acids in some cases or almost 700 residues long in others (Jones et al., 1998). Nevertheless, the commonly observed complexity of the average domain structure seems to be limited by folding kinetics, theoretically yielding the maximum free energy of unfolding for entities of 100 amino acids in lengths (Xu and Nussinov, 1998). Thus likely making it the optimal size for a protein domain.

### 2.1.4 *Macromolecular machines*

While all molecular machines are ultimately composed out of one or multiple domains, not all of the resulting enzymes are similar in size. Even though most functional entities in cellular homeostasis are represented by a single domain structure (Berman et al., 1994), some are conglomerates of larger or very large proportions. In fact, significant examples like the proteasome, ribosome or spliceosome structures form molecular machines of exceptional



dimensions, sometimes containing 80 or more distinct units of protein or RNA, of which most again harbour various functional domains in a single chain (Agafonov et al., 2011; Fischer et al., 2010; Unno et al., 2002). These conglomerates of enzymatic functionality are thus termed *macromolecular machines*, a denomination also commonly used in the literature (Barford and Hopfner, 2016; Nogales and Scheres, 2015; Saibil, 2012).



**Fig. 1 Modular composition of a macromolecular machine.** Depicted is a pseudo-atomic model of the human C\* complex spliceosome (Bertram et al., 2017b). The proteins and RNA moieties that constitute the fully assembled complex (on the right) are shown individually on the left-hand side in a comic-like ribbon-style representation. Individual domains of the same protein are not necessarily differently coloured.

As implied before, macromolecular machines are assembled on a modular basis. Hence, the structure and function of various, potentially hundreds of different domains are combined within one complex (Fig. 1). This modularity is found to be a widely adopted feature in nature, as it practically enables evolution to mix and match previously developed molecular assemblies in order to form novel complexes of a potentially entirely different function, wherever needed. In the process, however, certain domains of otherwise great homology frequently undergo a change in function upon utilization in a complex of a different kind. Often, this change is furthermore accompanied by a slight variation in primary amino acid sequence and/or structure. For example, a specific domain's catalytic functionality that is essential to one macromolecular complex can be transformed into a merely structurally supportive task in another (compare domain functionality described in section 2.1.3.2 and as another example: EF-G or SNU114, a GTPase that is catalytically functional in the ribosomal context but not necessarily in that of the spliceosome (Carlson et al., 2017; Fica et al., 2017)).

The structural arrangement of the chemical moieties that assemble each domain and any macromolecular complex is therefore directly related to its function and is thus of exceptional importance. Hence, structural analysis of large macromolecular machines provides valuable insights into the architecture and spatial composition of the complex in question. Comprehending a molecular machine's three-dimensional structure at high resolution may therefore ultimately reveal its mechanism of action – in direct analogy to knowing the detailed blueprints of a man made, macroscopic machine.

## 2.2 Electron cryo-microscopy (cryo-EM)

In order to analyse the spatial organization of such structures at sufficient detail, various techniques are available and commonly utilized. This section will briefly elucidate the prospects of using microscopic magnification to visualize macromolecular structures at up to atomic resolutions.

### 2.2.1 How to visualize atomic structures?

The human eye is generally not capable of resolving structural detail that is smaller than a couple of hundred micrometres ( $\mu\text{m}$ ) in size. Auxiliary tools are therefore required to increase the resolution that is effectively achievable. The most commonly utilized approach is thus to simply magnify the projected image of a certain object. Consequentially, the latter procedure isotopically increases the distance between specific points of the object in the final image. In order to achieve such proportional magnification, various kinds of microscopy techniques have been developed during the last centuries. Beginning in the 17<sup>th</sup> century, arrays of glass lenses staggered in optical microscopes were successfully used to significantly magnify small objects for the first time. Utilizing photons in the visible light spectrum it thus became possible to depict and identify previously invisible microscopic features up to a certain resolution. The term *resolution* thereby describes the minimal distance at which two points (at specimen level) may still be distinguished as such (e.g. in the magnified image). Unfortunately, the maximal attainable resolution on the optical axis in any kind of microscopy that visualizes illuminated, passive objects is limited by the physical relationships that are described by the *Abbe diffraction limit*. The latter thus describes the resolution limiting correlation between the illuminating wavelength and the optical properties of a microscope. More precisely, the principle states that two separated objects in the projected image may only be resolved as such if the distance between them is greater or equal to the wavelength divided by twice the numerical aperture of the lens, even when using perfect instrumentation. Hence, as conventional light microscopic lenses only achieve numerical aperture values of around 1 they can not resolve structural features separated by  $\sim 200$  nm or less, as the shortest wavelength of visible light (in the blue spectrum) is approximately 400 nm.

Picking up on the initial proposal to use microscopy in order to resolve and analyse the atomic details of a (macro) molecular machine it becomes obvious that a diffraction-limited light microscopic system is physically not capable of achieving this task. As an example, to visualize two organically bonded carbon atoms that are typically situated  $\sim 0.12 - 0.15$  nm apart, an illuminating wavelength of at least 0.3 nm would be required. Glass lenses, however, cannot effectively manipulate electromagnetic radiation at wavelengths smaller than  $\sim 200$  nm. Thus, preventing their utilization in conventional high-resolution microscopy.

As the wavelengths of the utilized radiation critically influences the maximum attainable resolution, different means of illumination are required to overcome these limitations. In fact, electrons in motion adopt wave-like properties and can be precisely manipulated by arrays of electromagnetic lenses, even if accelerated to wavelengths far smaller than those required to satisfy the Abbe diffraction limit for structures of molecular size. Electron microscopy can therefore be used to study the structure and function of macromolecular complexes.

## 2.2.2 General features of the electron microscope

The electron microscope generally offers the potential to utilize accelerated electrons in order to probe a specifically prepared, often very thin sample of matter. Initially constructed by Ernst Ruska and Max Knoll in 1931 at the Fritz-Haber-Institute in Berlin, the technology ever since matured to become a powerful tool in determining molecular structures at up to atomic resolution (Renaud et al., 2018). Nowadays commonly utilized in the life and material sciences, modern electron microscopes are capable of resolving a samples structural detail through various modes of operation. Transmission Electron Microscopy (TEM) classically represents the most utilized method, which illuminates a thin specimen in a bright- or dark field mode to project a magnified image of the sample onto a detector. Scanning (transmission) electron microscopy (S(T)EM), in contrast, features a focused beam of electrons that is scanning over or though the sample point-by-point while other, less frequently applied technics may additionally analyse secondary radiation (e.g. X-rays) generated by the primary, incident beam of electrons. The spectroscopic analysis of the radiation that is modified or emitted by the sample is likewise possible.

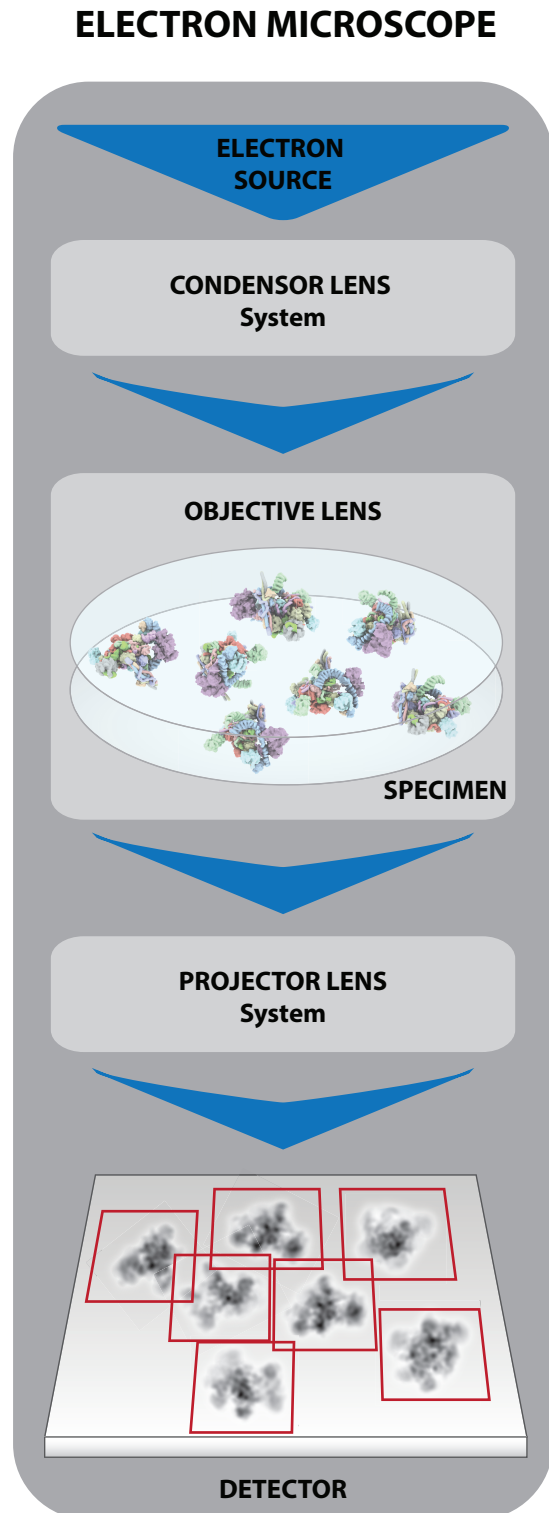
Since TEM is the main method that is used within the scope of this work, the following elaboration will briefly describe and introduce the function and capabilities of a transmission electron microscope. As the basic optical principals of image formation are very similar in light or electron microscopes, their setup is somewhat comparable, while important differences exist.

### 2.2.2.1 Setup

Beginning with the source of illumination, electron microscopes classically feature one of three different kinds of electron sources. The electron emitting part of the assembly is furthermore frequently described as the cathode of a microscope. Historically, heated tungsten filaments or lanthanum hexaboride ( $\text{LaB}_6$ ) crystals are used to provide the electrons required for imaging by emission into the *liner tube* of the microscope column. More recent devices, however, instead feature an ultra-thin tipped, zirconium oxide-coated tungsten crystal that is operated as field- or *Schottky* emitter in order to extract a temporally and spatially coherent beam of electrons at desirably high currents (Swanson and Schwind, 2009). Extracted by a positive potential difference between the tip of the emitter and an anode, emitted electrons pass through the *Wehnelt* aperture into an accelerator stack of conductive plates. Since the wavelength of an electron is directly correlated to its momentum (hence, its mass and velocity), as initially described by the French physicist *de Broglie*, acceleration is required to obtain electron radiation of the desired wavelength and high energy. The latter is practically achieved in the microscope by exposing the emitted beam of electrons to an increasingly positive electric field that is created by a potential difference applied perpendicular to the path of flight. Acceleration voltages of around 300 kV are currently utilized in high-end equipment to generate electron radiation of approximately 2 pm in wavelength. The accelerated electrons are then ready to be used for imaging, in analogy to the photons that are emitted from a light source in an optical microscope.

In order to achieve the imaging of a sample, accelerated electrons have to be refracted by electromagnetic coils and lenses. As electrons carry a negative charge, not only an externally applied electric field (such as that utilized during acceleration) but also a magnetic field effectively influences the propagation of an electron through space by the exertion of *Lorentz force*. A magnetic field created in close proximity thus affects the trajectory of any electron wave. In fact, electron-optical elements provide a specifically shaped magnetic field by sending an electrical current through coils of wire. Assemblies built to induce a spherical magnetic field therefore ultimately behave as an electromagnetic lens (Scherzer, 1936), in functional analogy to their optical counterparts made from glass or other transparent materials. Even though the exact mechanism of action is different from that observed for a glass lens element, electromagnetic lenses eventually possess the capability to deflect or condense incident electron radiation. As the refractive power of an electromagnetic lens is directly dependent on the strength of its magnetic field, both can be adjusted by changing the current that is sent through the coils of wire within the element. Therefore, in contrast to glass lenses, electromagnetic lenses offer the possibility to be adjusted in refractive strengths, a feature that is readily used in any kind of electron microscopic application.

Continuing the path through the microscope column, the accelerated electron beam is channelled into an array of two or three electromagnetic condenser lenses, thus preparing the beam to illuminate the sample. The interplay between these optical elements – which are operated at adjustable refractive powers and various exchangeable apertures – then allows for a precise shaping of the beam to finely tune the illuminated area and the angle of incidence. Furthermore, the coherence and brightness of the incident electron illumination (“light”) are readily adjustable, as the combination of a fixed diameter aperture and the adjustable refractive strengths of certain lenses within the condenser lens array



**Fig. 2 General schematic of an electron cryo microscope (cryo-EM).** The illuminating electrons are extracted and accelerated in the tip of the instrument at the electron source. Passed through and shaped by the condenser lens system, the radiation then interacts with the sample that is inserted into the objective lens. The image information is then passed on and recorded at the detector.

may be used to physically block fractions of the electrons that are initially emitted by the source. Once optimized in shape, dose and coherency characteristics, the beam enters the upper part of the objective lens that is physically mounted just below the condenser lens system.

The objective lens itself then images the sample. Divided into upper and lower portions, the specimen, which is commonly applied to a thin metal grid of 2-3 mm in diameter (section 2.2.4), is then inserted in-between these two parts of the objective lens. The so-called upper and lower pole pieces of the lens then precisely shape the magnetic field to channel the beam of electrons through the pole piece gap and the sample. As the exiting wave now contains the spatial information of the imaged sample area, the electron-optical features of the objective lens need to be particularly well-adjusted and monitored to prevent aberrations in the final, recorded projection.

The virtual image of the specimen, which is now physically carried in the scattered electron wave, may then be magnified in a setup relatable to that of an optical microscope. Hence, another array of intermediate lenses magnifies the virtual image through manipulating the convergence and opening angles of the beam. Importantly, in this process the user may freely choose the desired degree of magnification without the exchange of any hardware, as electromagnetic lenses are adjustable in refractive power. The actual magnification capabilities of an electron microscope therefore often reach from less than fifty to several million times magnification. Once brought to the desired magnification the emergent beam passes through a final projector lens that eventually adapts the geometry of the beam in order to properly expose a downstream detector or fluorescent screen.

As electrons interact strongly with any kind of matter, including the gas molecules in the ambient air, a high vacuum is required at all times when utilizing electron radiation in a microscope setup. Vacuum pumps are thus mounted in several stages onto the microscope column and operate to remove as many contaminant molecules from the liner tube atmosphere as possible. Practically, turbo molecular and ion-getter pumps create local vacuums of up to  $1 \times 10^{-9}$  mbar.

### **2.2.2.2 Contrast formation and transfer**

The formation and transfer of contrast is essential for the imaging and interpretation of any kind of data obtained with a microscope. The term contrast thereby describes the magnitude of difference in intensity of a certain signal with respect to that originating from the immediate surroundings. A high contrast in imaging therefore increases the likelihood of discriminating an actual feature from other features or noise in the direct vicinity and is thus essential to draw conclusions from the recorded data.

In TEM, two physically different principals of contrast formation generally contribute towards the overall image contrast that is finally recorded on the detector. First, while illuminating a somewhat transparent sample, some of the illuminating radiation is either absorbed in various ways or deflected at high scattering angles far away from the optical axis by certain, mostly dense features within the specimen. Since some of the illuminating radiation has now been removed from the imaging process, the corresponding positions in the final image appear to be depicted darker in general (of lower intensity), with respect to their more intensely illuminated surroundings (Lenz, 1971). Therefore, amplitude contrast is formed.

A second, more significant way of contrast formation in electron microscopy is described by phase contrast. In simple terms, parts of the initially plane, incoming electron wave interact with

a sample of *weak phase objects* while passing through it. In the process, certain fractions of that wave are deflected or elastically scattered away from the main trajectory of the illuminating wave at a specific angle by the electrostatic potential of the sample (Oxley et al., 2017). Since the observed scattering angles in this scenario are usually small (i.e., several degrees) (Wade, 1992), the scattered parts of the electron wave may not escape from, but are instead refracted back into, the optical system by the objective lens. The respective wave components are thus still available for imaging and will consequently also reach the detector. The distance travelled (i.e., path length) of these refracted parts of the wave, however, has now changed, therefore introducing a phase shift between the scattered and unscattered parts of the formerly plane, incident wave (Reimer, 2013). Mathematically, the appearance of these waves can be modelled by specific wave functions that in particular describe the wavelength (e.g. the energy of the accelerated electron), amplitude and phase of any (composite) wave at a given position. When traversing down the microscope, scattered and unscattered parts of the wave thus interfere at characteristic positions or planes in the instrument. After passing the objective lens, for example, the composite wave function shows strong amplitude maxima in the back focal plane of that lens, which ultimately represent the diffraction pattern of the originally imaged specimen in reciprocal space. Traveling further, the patterns of interference are altered again in order for the composite wave function to eventually reproduce the density distribution of the original specimen at the image plane of the lens. Since the process of image formation can be considered a linear process in first approximation (Lenz, 1971), the density information contained in the composite wave function may additionally be magnified in a linear fashion by exploiting the basic principles of optical magnification using lens arrays. When finally reaching the detector, which is mounted in a plane that is approximately conjugate to the image plane of the objective lens, the optical system of the microscope is aligned in a way to promote interference of all parts of the wave that reach this plane. If a shift in phase is now present in some parts of that wave it will positively or negatively interfere with the unaltered parts of the originally incident wave. Hence, intensity minima or maxima are introduced at various spatially defined positions on the detector. Effectively, this interference of phase-shifted waves thus translates a practically undetectable phase shift into fluctuations of amplitude, which can then be identified on the detector as changes in exposure, hence, the read out intensity. The actual appearance of these intensities, however, is substantially modulated by the scattering angles, certain aberrations and the refractive power of the objective lens (i.e.. via defocus). In particular, spherical aberration and defocus thereby represent optical phenomena that add substantial phase shifts to the beam and thus significantly modulate the intensities that are eventually detected. By means that are explained in much greater detail here (Lenz, 1971), contrast is only efficiently transferred for features that scatter the incoming electron wave at specific angles, given a fixed focal length (strength of the objective lens) and constant spherical aberration. The signal that eventually reaches the detector is thus not entirely representative of the complete spatial relationships of the originally imaged object but is rather filtered in a complex manner, which can be precisely described by the (phase) *contrast transfer function* (CTF) (Erickson and Klug, 1970). In consequence, some features of a specific *spatial frequency* (as the CTF is evaluated in reciprocal space) generate a strong positive contrast (i.e., are depicted brighter), while others are imaged with negative (darker) or no contrast at all (hence, are invisible). As the contrast modulations described above are strongly (de)focus dependent, micrographs are often deliberately recorded in an underfocused fashion to increase the contrast for features of low spatial frequency. Finally, the detector records the resulting variations in intensity near a plane conjugate to the image

plane, where, more accurately, the probability of detecting an electron is proportional to the squared amplitude of the composite wave function at this position.

Since in a typical life science TEM application most interactions between the incident electron wave and the sample material result in small-angle scattering events, phase contrast generally contributes more than 90% of the overall contrast in the recorded image (Penczek, 2010a). In order to successfully conserve and reproduce the original spatial relationship of the features in the sample and the projected image through phase contrast, a highly sophisticated and tuned optical system is required. Perturbations of the latter due to (sometimes inevitable) flaws in hardware design or from external sources thus significantly reduce the quality and resolution of any image projected through it (section 2.2.3). Hence, efforts are being made to correct these aberrations.

### **2.2.2.3 Image detection**

Once the electron wave has interacted with the sample and thus transfers its spatial information, an efficient way of detection is required in order to record that information. As the refracted and unrefracted wave fronts interfere at the surface of the detection device, intensity minima and maxima are formed, resulting in amplitude and phase contrast (see section 2.2.2.2). These intensity perturbations are then recorded at a defined spatial position on the detector and thus, after readout, may be correlated to each other during image processing.

In conjunction with the recorded signal, however, every detector inevitably adds additional, signal-independent intensities to the final image, which are commonly classified as noise. Even though noise will accumulate in the final image from different sources, the amount that is added by the detector itself at different spatial frequencies is of great importance. A critical quality measure of the readout signal is therefore given by the detector's *detective quantum efficiency* (DQE) (Dainty et al., 1976). The latter is defined by the square of the ratio of the output signal-to-noise (SNR) to that of the input SNR (McMullan et al., 2014). Consequently, the DQE describes how well a detector is capable of recovering the properties of an original input signal with respect to that which is eventually read out from the device after recording. Hence, high DQE values close to one represent an almost perfect signal recovery, while values approaching zero indicate that a signal can hardly be distinguished from noise.

Photographic film was historically used for the purpose of image recording not only in photography or light microscopy but also in electron microscopic applications. Electrons, like photons, darken a film material through specific chemical interactions with its substrate. Areas where the illuminating waves constructively interfere are thus blackened more than areas where less, no, or destructive interference is observed. Importantly, the darkening of the film material is linearly correlated to the amount of interference observed at the corresponding position. Therefore, the amplitude of the composite wave is recorded. While photographic film provides good imaging characteristics and a high DQE in many recording scenarios, its physical handling is exceptionally tedious. As the film material needs to reside within the high vacuum environment of the microscope for imaging, an undesirable breach of that vacuum, which potentially introduces moisture or other contaminants into the column, is inevitable once exposed or new film material has to be recovered or inserted into the microscope. Every film plate of every micrograph that was recorded then has to be separately developed and digitized for analysis in downstream image processing applications. With the advent of other detection devices the use of film thus was replaced with faster, more convenient, and higher-throughput methods of data collection.

Modern image detection of electron microscopic signals is therefore most commonly achieved by the use of semiconductor-based detectors. Charge coupled device (CCD) chips, which are excited by photons produced through electrons hitting a thin scintillation material above the sensitive detection layer, were formerly used for this purpose, as a direct digital readout of the recorded signal provides a great advantage in efficiency over the manual development of photographic film. CCDs, however, are not hard enough to endure direct electron irradiation at the doses required in electron microscopy. Therefore, improved complementary metal oxide semiconductor (CMOS) detector designs eventually allowed the construction of a direct electron detector (McMullan et al., 2016). The latter then allowed for a removal of the scintillation material and a much thinner design of the active detection layer (i.e., “back-thinning”) to provided a significantly improved DQE compared to all other detection methods available to this date. In combination with a fast readout rate it thus became feasible to operate the detectors in several recording modes, allowing to further improve the quality of the (raw) data. Single electron counting and the possibility of reading out many frames (hundreds to thousands) per second improves the spatial accuracy of the recorded signals, thus significantly contributing towards the widespread adoption and success of electron (cryo-) microscopy techniques today (Grigorieff, 2013; Li et al., 2013).

### **2.2.3 Optical aberrations and resolution**

The most fundamental aim of microscopy is to extract and reproduce the exact spatial relationships of an object of interest within a projected, often magnified image of that object. The precise focusing of rays originating from one point at the object plane to another point on the image plane near the detector is therefore of absolute importance. In electron microscopy, any real optical system introduces aberrations that reduce the imaging quality compared to an ideal, hypothetical system. As in light microscopy, certain aberrations are inevitably introduced by the physical characteristics of an optical element while others are caused by variations in production quality or the physical alignment of one unit with respect to another. In practice, various types of aberrations interfere destructively with the imaging process, ultimately limiting the maximal attainable resolution.

Besides their wave-like properties, electrons also possess the characteristics of a particle that may be accelerated, deflected or otherwise manipulated by certain fields. Electron trajectories can therefore be calculated and visualized by applying the laws of relativistic mechanics. While a detailed mathematical description is beyond the scope of this introduction (but may be found here (Reimer, 2013)), a recapitulatory explanation of the most significant aberrations in electron microscopy and their general effects on image quality is given below.

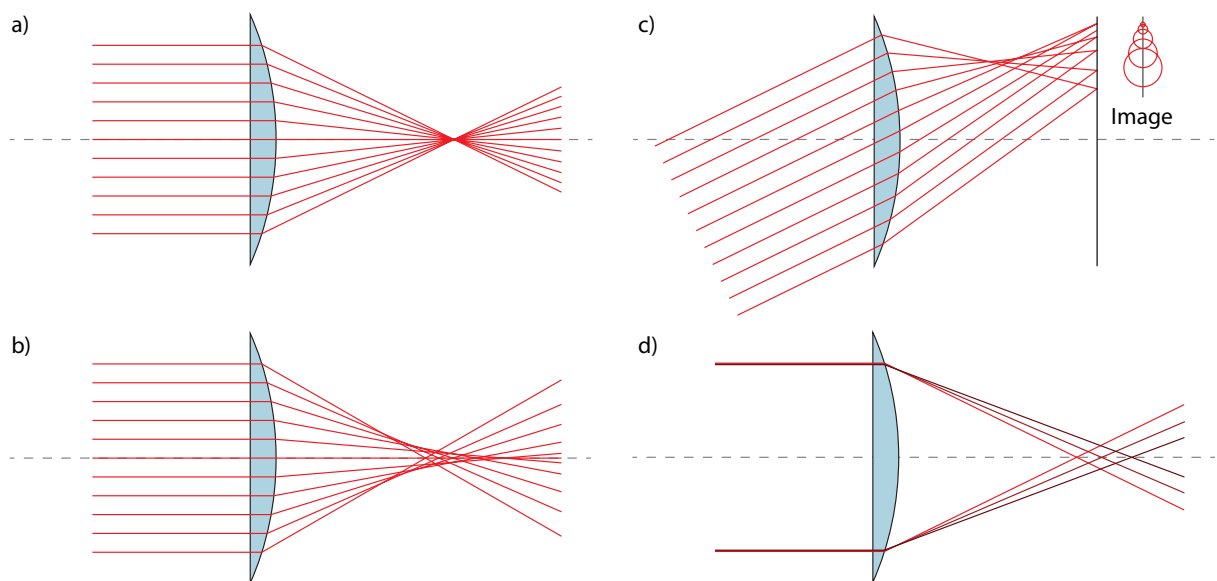
#### **2.2.3.1 Optical aberrations**

Electron microscopical aberrations can be classified into isotropic and anisotropic effects of various order, resulting in on- or off-axis errors (Reimer, 2013). Due to the practical importance of retrieving high-resolution image information in single particle cryo-EM, this section briefly describes the effects of spherical aberration, axial astigmatism, distortion, coma and chromatic aberration.

Spherical aberration belongs to the class of isotropic, on-axis aberrations and represents one of the most significant imaging defects in any kind of microscopy. As electron lenses induce a spherical magnetic field around circular coils of wire, which gradually decays with an increasing



distance to these wires, rays passing through that field closer to the coil experience a stronger deflective force, hence, more refractive power than rays passing through the centre of the lens (Fig. 3b). Where a similar behaviour can be observed in spherical optical lenses (although due to a different physical effect), corresponding rays originating from one point above the lens are consequently focused not in one spot, but on different planes behind the optical element. Therefore, a specific feature of the original object is not projected as a sharp point in the image plane but effectively blurred out by the effective broadening of focal points. The magnitude of spherical aberration is described by the spherical aberration coefficient ( $C_s$ ) and decreases with increasing strengths of the lens (Reimer, 2013). Hence, stronger lenses in the intermediate or projection system experience less spherical aberration effects than a generally weaker objective lens. Typical values of the  $C_s$  for an objective lens lie between  $C_s = 0.5 - 3$  mm. As in light microscopy, the deteriorating effects of spherical aberration can be reduced by introducing additional optical elements; that is, a  $C_s$  corrector (Hosokawa et al., 2003).



**Fig. 3 Specific optical aberrations in the electron microscope.** Illustrated is the optical path of a parallel beam of electrons (red) that is incident from the left side of the schematic and refracted by a stylized converging electron lens (blue). The dashed line in grey depicts the optical axis. (a) A perfect, imaginary lens that focuses all incoming parallel radiation at the indicated focal point. No aberrations are visible. (b) The effect of spherical aberration onto the focussing of parallel radiation. The parts of the beam that are passing the lens at a more central position are generally refracted less strongly. (c) The effect of coma that is induced by parallel rays that enter the lens at an angle with respect to its optical axis. An originally point-shaped feature in the imaged object is characteristically distorted towards a specific side at the image plane. (d) Chromatic aberration that is induced by a varying refractive index of the lens with respect to the energy of the incident radiation. High energy electrons are depicted by the dark red line, hence, are refracted less; low energy electrons by the bright red line, which thus are refracted more strongly.

A further significant on-axis aberration related to spherical aberration is axial astigmatism. Caused by imperfections in the rotational symmetry of the magnetic field of a lens, astigmatism likewise results in the misfocusing of a point on the object plane in the image plane. Where spherical aberration is induced by a radial variance in magnetic field strength, astigmatism generally describes an elliptical or otherwise non-spherical variance in that field. Rays passing the lens in sagittal and meridional direction therefore focus at different points around the image plane. In first-order (or two-fold) astigmatism (three- and higher fold variants also exist) the resulting distance between minimal and maximal focal height is commonly given in nm and thus

represents a coefficient for the magnitude of the astigmatism observed in the image. Since astigmatism, such as that introduced by the objective lens, affects the entire field of view in a predictable manner, it is also described as the angular dependency of defocus (THON, 1971). The first-order astigmatism of a lens may be corrected by stigmatism of the beam utilizing a weak quadrupole lens (Reimer, 2013) or computationally during image processing, as mentioned in section 2.3.2.

Distortion represents an off-axial aberration that causes anisotropic magnification. Non-linear forms of distortion for example cause the effect that points further away from the optical axis are displaced in a pincushion, barrel or other fashion. Effectively, describing the first two examples, imaged lines that do not pass through the optical axis are thus bent outwards or inwards, respectively. Linear distortion generates an additional form of anisotropic magnification that results in a proportional, often defocus-dependent change in magnification over different areas of a micrograph (Rubinstein et al., 2015). Distortion thus anisotropically varies the spatial correlation of features in the projected image with respect to those in the original image. It may be corrected optically by tuning specific lens characteristics or computationally, as initially introduced by Crowther & Sleytr (Crowther and Sleytr, 1977).

Having contained the above mentioned imaging defects, coma is one of the most critical aberrations for high-resolution imaging in electron microscopy (Ishizuka, 1994). As generally induced by rays that enter a lens at an angle with respect to its optical axis, refracted rays are then focussed at different positions in the image plane depending on the position where they initially entered the magnetic field of the lens (Fig. 3c). Coma thus results in the imaging of an off-axis point in the object plane as circles of increasing radius in the projected image. Appearance wise, these circles are additionally shifted away from their expected Gaussian image point at a distance correlating to their radius (Reimer, 2013). This effect therefore creates an image reminiscent of a comet tail. In contrast to the effects of e.g. two-fold astigmatism, coma cannot easily be eliminated during image processing and therefore needs to be corrected optically, before the exposure, by a particular coma-free alignment procedure. As initially developed by Zemlin and colleagues in 1978, the procedure consists of aligning the beam exactly onto the optical axis of the imaging lens (Zemlin et al., 1978), thus effectively reducing the effects of axial coma.

Finally, chromatic aberration represents another important aberration in high-resolution phase contrast microscopy. As generally observed in all kinds of microscopy, chromatic aberration is inevitably induced by the energy spread of the illuminating radiation and not by imperfections of the lens. In electron microscopy, the focal length of a magnetic lens is directly proportional to the energy of the incident radiation and the strength of the magnetic field (or the current) that is sent through the coils (Reimer, 2013). Therefore, electrons with different energies (thus possessing a different wavelength) are refracted at disparate focal lengths by the same lens (Fig. 3d). Chromatic aberration is thus best minimized by the use of a coherent source of illumination, such as a (cold) field-emitter and/or the use of a monochromator (Kimoto, 2014; Tsuno, 2011).

### **2.2.3.2 Optical resolution**

As introduced in section 2.2.1, the point-to-point resolution in electron microscopy is generally not diffraction-limited. Even though electrons with short wavelengths (typically in the range of a few pm) are utilized as the source of illumination, and satisfactory numerical apertures would theoretically satisfy the Abbe diffraction limit for higher resolutions, the optical quality of electromagnetic lenses is significantly worse than that of their light optical counterparts

(Penczek, 2010b). Aberrations, therefore, prevent the electron optical system from reaching the theoretically possible specifications. Hence, the maximal attainable resolution in electron microscopy is not diffraction-, but rather aberration-limited.

In practice, any kind of aberration directly interferes with the spatially proportional reproduction of an objects features in the imaged projections. During imaging (in real space), the original, unaltered image function of the specimen is thus always convolved with that of the sum of all aberrations; hence, a specific point spread function. Consequently, the information recorded in the final image does not represent the exact spatial relationships of the original specimen but is rather blurred out by the effects of various aberrations that are added cumulatively. A major objective in improving the point-to-point resolution in electron microscopy is thus to contain and reduce the aberrations introduced during the imaging process.

### 2.2.4 Sample preparation and imaging under cryo-conditions

Biomolecules that are to be imaged within an electron microscope inevitably need to be exposed to high levels of radiation as well as to the ultra-high vacuum of the microscope column (typically up to  $10^{-9}$  mbar). Attempts of probing the sample in a liquid environment are therefore condemned to fail, as evaporation would dry out the specimen while contaminating the vacuum. Setting aside the possibility to stain and dry the sample (e.g. utilizing heavy atom metal salts), a method that was classically used but is limited in resolution due to the grain size of the stain, the challenge to image unstained biological specimen was eventually solved by Dubochet and Alasdair McDowell by embedding the sample into a thin layer of glycerol or non-crystalline (hence, vitreous) ice (Dubochet et al., 1985; 1988). A procedure that is routinely used in today's high-resolution electron microscopy once represented a cornerstone in the development of the method. The electron microscopic probing of frozen samples was thus named electron cryo-microscopy (cryo-EM) and its inventors were awarded the Nobel Prize in Chemistry in 2017.

While embedded in a thin layer of ice, proteins or any other biomolecules remain in their natural buffer, fully hydrated and physically fixed at a static position – desirable attributes not necessarily offered by all methods of structural biology. The formation of ice crystals, however, needs to be prevented as they would strongly diffract the incident electron beam and overshadow the actual signal that originates from the specimen. Once a cooling rate of more than  $10^5$  K s<sup>-1</sup> is reached, ion enriched water tends to freeze into an amorphous, glass like solid in a process termed *vitrification* (Brüggeller and Mayer, 1980). For thin samples, like those typically featured in a cryo-EM application, rapid cooling is performed by applying the sample to a thin metal grid as mechanical support and then plunging the entire preparation into a liquefied gas, such as ethane cooled to approximately 93 K (-180 °C) (Dobro et al., 2010). In order to generate thin layers of ice, which is a prerequisite for reaching sufficiently high cooling rates and electron permeability later in the microscope, excess water or buffer is usually blotted away from the grid just before plunge freezing.

Another positive effect of probing biological samples with electron radiation at low temperatures (e.g. at 77 K for liquid nitrogen) is the decreased sensitivity to radiation damage (Bammes et al., 2010; Stark et al., 1996). While any crude biological specimen at room temperature would immediately vaporize when exposed to an electron dose that is required for imaging, the same sample does stay intact long enough in a low temperature frozen state. Nevertheless, in particular when compared to the doses that are commonly utilized in electron

microscopy of relatively radiation insensitive inorganic sample materials (e.g., several thousand electrons per  $\text{\AA}^2$  per second), cryo-EM of biological specimen must be performed under much lower dose conditions (e.g., up to 50 electrons per  $\text{\AA}^2$  per second). Hence, the name *low dose cryo-EM* was established. Unfortunately, the low dose imaging of these specimen generates images of critically low signal to noise ratio. Extensive image processing is therefore required in order to enhance the signal and eventually reconstruct a high-resolution three-dimensional model of the originally imaged (macro-) molecular complex of interest.

### 2.3 Image and data processing

Any image or micrograph recorded by TEM eventually provides the two-dimensional (2D) projections of a three-dimensional (3D) object of interest, e.g. a large macromolecular complex particle (Saibil, 2012). In the process, the entire density distribution of the 3D object is projected along one specific (projection) axis into a single 2D image (Penczek, 2010c). The relative angular relationship between this axis and an arbitrary reference axis in a coordinate system is given by the *projection angle*, which is determined by the angular relationship between the particle's orientation on the grid and that of the incident, illuminating radiation. As the latter is commonly not tilted during image acquisition, the angle at which a 3D particle's appearance is projected onto a detector is solely dependent on the orientation of that particle on the grid. To finally reconstruct a 3D model of the initially imaged particle, its projections need to be recorded at various projection angles (Penczek, 2010c). In cryo-EM, the latter is readily achieved by applying and freezing a highly purified sample of the molecular complex of interest to an EM-grid (Dobro et al., 2010). As the particle surfaces ideally interact with the grid substrate in an unspecific manner, a random distribution of particle orientations is ideally obtained. Therefore, once imaged, projections of single particles at (ideally) random angles are recorded on every micrograph and may consequently be utilized for 3D reconstruction. Since many images of (the same kind of) single particle are recorded and utilized, the procedure is commonly described as *single particle cryo-EM*.

In order to reconstruct and analyse the three-dimensional density map of a macromolecular complex from the 2D projections that were recorded in an electron microscope, extensive image processing is required. The following sections will thus briefly introduce the most essential procedures and finally complete the explanation of cryo-EM as a method for high-resolution single particle 3D structure determination.

#### 2.3.1 Micrograph quality control and pre-processing

Every set of raw micrographs has to be pre-processed and analysed accordingly once recorded by the microscope. As modern detectors allow for a dose fractionation through the consecutive recording of multiple frames at very short exposure times (section 2.2.2.3), each exposure of the same frameset needs to be aligned with respect to the others. The latter is commonly achieved by a selection of software packages and effectively reduces the effects of beam-induced motion blurring or charging, that would otherwise diminish image quality (McLeod et al., 2017). The resulting framesets of the aligned images are then averaged and optionally dose weighted in order to improve the signal to noise ratio of the final micrograph representation, while at the same time optimizing the representation of high spatial frequency information in the data (Zheng et al., 2017). Once processed, the original stacks of raw-micrograph images that are

generally of a large file-size may now be deleted. The pre-processed, averaged and dose-weighted newly synthesized micrographs are further processed.

The quality of each individual micrograph may nevertheless significantly vary throughout the dataset. Even though areas on a grid that are to be recorded are usually carefully selected in advance, local contaminations are still imaged occasionally and thus need to be removed from the dataset at later stages. In particular variances in ice thickness, particle aggregation or charging effects represent instances where the projections of single particles become uninterpretable and are thus unusable for subsequent reconstruction procedures. Since automated routines are currently not available, micrographs have to be visually inspected and consequently sorted according to the general quality requirements introduced above using certain software (unpublished work, Stark lab).

Additionally, the so-called *power spectrum* of a micrograph may be calculated at this time; an operation where the real space image information is mathematically transformed to reciprocal, or *Fourier* space. The resulting *Thon-rings* may then be manually evaluated and characterized to discard recordings that show crystalline ice, drift (mechanical movement of the sample or sample holder), charging, or a very weak signal to noise ratio in the image (Cong and Ludtke, 2010; THON, 1971). Approximately 50-90% of the originally recorded micrographs are usually evaluated as acceptable and thus passed on to the next step in image processing.

### 2.3.2 Aberration estimation, particle picking and extraction

As described in section 2.2.3, certain aberrations may be detected and corrected during image processing. Additionally, as introduced in section 2.2.2.2, the signal recorded on each micrograph is mostly phase contrast derived and thus always modulated by a CTF. For high-resolution 3D reconstruction purposes, however, an eventually homogeneous signal distribution over all spatial frequencies is of superior importance (Zou et al., 2011). The power spectra of the pre-processed micrographs are therefore utilized to fit an approximation of the respective CTF for every micrograph. As the CTF is directly dependent on the focal height of the recording, fitting the function to the real data also reveals its respective defocus parameters. In addition to the amplitude and direction of the 2-fold astigmatism that is often present in the data, the CTF zero crossings are precisely mapped in the same procedure (Vulovic et al., 2010). This analysis, usually automatically performed by software, thus allows the application of basic image processing procedures such as phase flipping or astigmatism correction (Zhang, 2016; Zhou et al., 1996). Specifically the flipping of phases that would otherwise have resulted in a negative image contrast composes new images of the particle projections that consequently feature an all-positive image contrast. The above described procedures are thus absolutely essential to eventually retrieve the high spatial frequency information from any phase contrast transfer function modulated image (Cong and Ludtke, 2010).

Once the CTF and 2-fold astigmatism parameters are determined, the position of all single particle projections on the micrograph must be determined; a process commonly referred to as *particle picking*. The resulting coordinates are consequently used to extract the imaged particles from the micrographs, ideally discarding artefacts and blank spaces from further processing. Even though at present day this task is commonly performed by software (Woolford et al., 2007), the precise picking of particle coordinates is often complex due to the diverse shape and appearance of various macromolecular complexes (Cong and Ludtke, 2010) as well as variations in image contrast e.g. as a result of a varying ice thickness. As soon as the particle coordinates

are obtained, every projection is cropped from its original micrograph at a box size that is approximately 1.5 - 2x the diameter of the particle. All further basic image-processing procedures are then performed on these stacks of particle images only.

### 2.3.3 Particle classification and averaging

As explained in section 2.2.4, the low dose imaging that is commonly applied during single particle image acquisition inevitably results in a noisy representation of the recorded image. One of the most critical challenges in (single particle) cryo-EM image processing is therefore to improve the generally low signal-to-noise ratio (SNR) in the recorded image, which is inevitable when using low dose exposure techniques. The first successful attempts to master this task experimentally were initially presented by Henderson and colleagues, who effectively averaged the signal of millions of Bacteriorhodopsin particles arranged in a 2D crystal lattice during imaging to generate micrographs of satisfactory signal-to-noise ratio (Henderson et al., 1990; Unwin and Henderson, 1975). In single particle applications, however, the data recorded is noisy and not directly interpretable to atomic detail. Nevertheless, the signal of single particle projections may be equally well improved with respect to the ambient noise by averaging with other projections of the same kind during image processing (Henderson, 1995). As the projections of single particles are usually recorded at random projection angles and not in a crystalline lattice, however, a careful alignment and classification of all images is obligatory before averaging by summation (Frank, 1975). After alignment and classification, the same kind of projections thus show the same features at the same position while the noise in the image is generally distributed randomly. Therefore, after averaging, the signal of the particle features increases proportionally with the number of images utilized, while the signal that originates from random noise cancels out in a similarly proportional manner.

Many algorithmic approaches have been developed in order to align and classify sets of noise contaminated single particle projections (Marabini and Carazo, 1996; Schatz and van Heel, 1990; van Heel et al., 1992). Some of which, in particular those that are based on multivariate statistical analysis (MSA) approaches (Schatz and van Heel, 1992), are still utilized today. While iterative rounds of cross correlation based multiple reference alignments (Saxton and Frank, 1976) and principle component analysis (PCA) based classification allow for a precise segregation of single particle projections into classes of similar appearance, methods that are based on *maximum likelihood* estimations and *Bayesian statistics* are gaining importance in the field and are nowadays used on a routine basis (Scheres, 2012a; Sigworth, 1998). Finally, 2D class averages of sufficiently improved SNR and best possible structural homogeneity are required to proceed to angular reconstitution and 3D reconstruction steps.

### 2.3.4 Angular reconstitution and 3D reconstruction

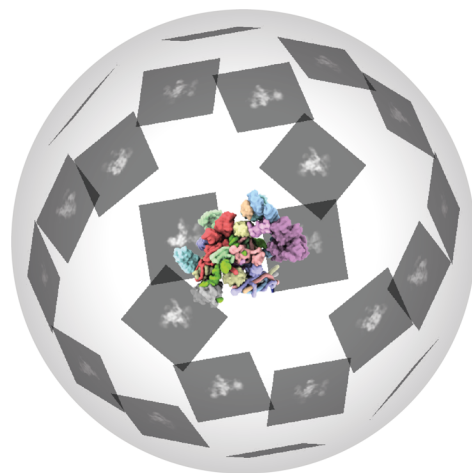
Even though 2D projections of a macromolecular complex are typically recorded in a single particle cryo-EM type of experiment, a 3D reconstruction of the sample is generally desired. Methods to routinely achieve this task were initially applied to electron micrographs of single particle projections in the 1970s, even though the mathematical foundation was developed long before (Crowther and Klug, 1975; Frank, 1981). In order to reconstruct a 3D volume from its 2D projections, the latter need to be back projected while spatially arranged at their corresponding, original projection angles (Fig. 4). In effect, a virtual 3D model of the originally imaged object is created at the centre of an imaginary projection sphere, which can then be analysed at various

threshold levels that again represent the density distribution of the original object. Even though the above described procedure is well characterized in theory and efficiently computed in reciprocal space, its application to real electron microscopy data is not trivial. The most challenging task in this regard is represented by the (re-) assignment of the projection angles to the arbitrarily recorded 2D images. Although the entire 3D density distribution is conserved within a 2D projection of a 3D particle, the relative angular relationship of that projection to the originating particle is lost. The knowledge of the exact projection angles, however, is critical for the success of any back projection attempt during 3D reconstruction (Penczek, 2010c).

Depending on the prior knowledge regarding the structure of an imaged sample, various methods of angular reconstitution have been developed in the past. One of the most direct ways to achieve the latter, if applicable, is to utilize the intrinsic symmetry of a macromolecular complex to reconstruct the relative angular relationship of its asymmetric units. As for the first time presented by David DeRosier and Aron Klug in 1968, they were able to reconstruct a 3D model of a Bacteriophage T4 tail from a single micrograph through symmetry analysis of its diffraction pattern and subsequent Fourier synthesis (de Rosier and Klug, 1968). Even though the resulting physical model was of poor resolution and eventually carved from wood, they were nevertheless able to prove the concept of back projection as an incredibly valuable tool in 3D reconstruction from electron microscopic images. Although the appearance of symmetry greatly reduces the complexity of angular reconstitution, far from all (macro) molecular complexes actually exhibit a periodic structure. Other methods were therefore developed to reconstitute the angular relationship of non-symmetric, single particle projections.

Another successful approach that considers multiple pairs of projections recorded from the *same* particle is still utilized today. To image these projection pairs, the sample is recorded multiple times (at least twice) at various tilt angles during a tomography type of experiment. Termed *random conical tilt*, the method relies on correlating sets of particles that are deliberately imaged on a common tilt axis, with all others. If projections are consequently related amongst themselves by a rotation around the common tilt axis, only one more angle is required to be determined in order to describe the entire set of angular relationships. In this manner, Frank and Radermacher demonstrated the angular reconstitution and 3D reconstruction of a 50S Ribosomal subunit in 1986 (Radermacher et al., 1986; 1987).

Preceding these developments, a method that is still commonly utilized today does not require additional experimental procedures to correlate the projection axis of randomly oriented, single particle recordings. Once, however, again established for structures of high symmetry, a procedure that correlates pairs of “common lines” in the Fourier domain of single particle



**Fig. 4 Simulated Projection sphere illustrating the relationship between 2D projections and their 3D origin.** Various 2D projections of the same 3D particle that is oriented in the centre of the sphere are illustrated. The 2D projections are thereby arranged on the arbitrarily sized projection sphere according to their correctly determined projection angle. A (simulated) projection of the central 3D particle would thus exactly yield the depicted 2D projections and vice versa. Hence, a back-projection of the 2D projections at their corresponding angle would in return yield the original 3D model in the centre.

projections was utilized (Crowther et al., 1970). The reciprocal space based method is based on the *projection-slice theorem* (Garces et al., 2011) and allows the correlation between features in a projection and the corresponding object of one additional dimension. Hence, for example a 2D projection and its 3D object of origin. Marin van Heel refined and modified this procedure to utilize real space common lines in *sinograms* to reconstitute the angular relationship of any projection from the same 3D object (van Heel, 1987). Effectively, both variants work to a satisfactory extent if the input 2D projections 1) all belong to the same 3D object and 2) are of sufficient SNR. These prerequisites thus render it essential that the sample in question is 1) of high purity, hence, all projections come from the same 3D object and 2) of sufficient SNR, hence, the experimentally recorded projections have been pre-processed accordingly (section 2.3.3).

Once 3D information concerning the particle of interest is available, e.g. through initial attempts of angular reconstitution, a crystal structure or any other model that is compositionally related, computed projections of this existing model can then be utilized as an alignment reference for the recorded data. Since the reference projections are computed, not empirically recorded, their angular relationship is known. All particles that are aligned to a projection of this kind may therefore be used at once in a 3D reconstruction with an angular relationship that originates from and is equal to that of the original reference projection (Penczek et al., 1994). The method is therefore named *projection matching* and reviewed in more detail here (Elmlund and Elmlund, 2015). As an existing model is used as a reference, the procedure may likewise be utilized to iteratively refine a 3D structure against the data through consecutive rounds of alignment, 3D reconstruction and re-projection of the reconstructed volume. This procedure thus allows an iterative improvement of the back-projected 3D volume, hence, its re-projections, and consequently results in a more precise alignment of the recorded data to the computed re-projections of higher resolution during the next iteration. Maximum-likelihood based methods of classification in combination with projection matching procedures therefore allow the refinement of cryo-EM data to highest resolutions (Scheres, 2012b).

### **2.3.4.1 Resolution after 3D reconstruction**

Objective measures for the assessment of resolution are of great importance in evaluating the informative value of any reconstructed 3D volume. As introduced in section 2.2.3.1, the optical resolution of the microscope does not necessarily equal the resolution of the 3D volume after 3D reconstruction. Instead, the point-to-point resolution theoretically obtainable in an ideal exposure of a 2D micrograph is further degraded by the low SNR in a typical cryo-EM recording, alignment- and aberration correction errors, and additional optical aberrations introduced by recording procedures that are optimised for speed (e.g. beam tilt induced coma) (Penczek, 2010b). The resolution assessment of cryo-EM data after image processing therefore rather concentrates on the consistency of the result, which is generally an averaged image in the 2D classification scenario and a density map after 3D reconstruction. While a wealth of procedures have been developed over the years, the *Fourier Ring Correlation* (FRC) in 2D (Heel and Stöfler-Meilicke, 1985; Saxton and Baumeister, 1982) or *Fourier Shell Correlation* (FSC) in 3D (Harauz and van Heel, 1986) are most commonly utilized in the field today. The FSC thereby represents a 1D function of the modulus of spatial frequency, where its values are computed correlation coefficients between two similar but independently reconstructed 3D volumes. Consequently, values of the FSC curve close to one represent a strong similarity between the two 3D volumes at a given spatial frequency while values approaching zero suggest a random, thus no correlation. A specific, somewhat arbitrary cut-off value (e.g. 0.5 or 0.143) is then used to determine the



resolution (Rosenthal and Henderson, 2003). Defined dataset refinement procedures and FSC cut-off values eventually represent the “gold-standard” of resolution determination (Henderson et al., 2012). The obtained numerical value thereby commonly represents spatial frequency (e.g.  $1/\text{\AA}$  in units), while an inverse of this number is used to describe the apparent resolution in units of real space distance, e.g.  $\text{\AA}$ . Resolution estimations of this kind are additionally utilized during refinement procedures in order to prevent overfitting and to guide the process in general (Scheres, 2012b). 3D maps may therefore be filtered and post-processed at their respective maximum resolutions without the risk of an over- or underestimation of the information content that is actually present in the data.

Once the resolution of a reconstructed 3D model is evaluated, its numerical value is expected to represent a consistent and reproducible indication for the information content of the map. In general, it is accepted that resolutions worse than  $30 \text{\AA}$  solely describe the overall shape of a macromolecular complex in rough detail. In the range of  $30 - 10 \text{\AA}$  it then becomes feasible to fit and identify rigidly folded whole protein structures or their domains. A map with a  $10 - 4 \text{\AA}$  resolution range is expected to clearly reveal the shape of alpha-helices and beta-sheets, eventually separating the later and allowing an identification of the alpha-helix pitch at resolutions close to  $4 \text{\AA}$  (Baker et al., 2012). Maps better than  $4 \text{\AA}$  in resolution are finally expected to reliably reveal densities for individual side-chain moieties, where resolutions of  $3, 2 \text{\AA}$  or even lower values continuously increase the certainty of their fit, up to exposing true atomic detail at around  $1 \text{\AA}$  (Popp et al., 2018).

### 2.3.5 (Atomic) model building and interpretation

Reconstructed 3D volumes that are refined to resolutions better than  $\sim 10 \text{\AA}$  are commonly fitted with a (pseudo-) atomic model, while actual atomic modelling usually becomes feasible at resolutions better than  $4 \text{\AA}$ . The resulting model then approximately defines the position of ideally all atoms in the structure. Depending on the quality of the map, this model may finally be used to determine and describe the molecular architecture of a large macromolecular complex, the conformation and interaction sites of protein or nucleic acid domains or even, at sufficiently high resolutions, the chemistry involved in the function of a molecular machine. The evaluation of an atomic model thereby allows for a more straightforward and comprehensive interpretation of the data that is obtained by a typical single particle cryo-EM experiment. The construction, refinement and validation steps that are required for the generation of such a model are not trivial and an in-depths description of the involved processes would be beyond the scope of this introduction. Nevertheless, to provide a general overview of the steps that are necessary to model atomic coordinates, the most important operations are shortly elaborated below. More detailed reviews may be found here (Brown et al., 2015; Davis et al., 2007; Jensen, 2010; Terwilliger et al., 2012).

#### 2.3.5.1 Docking of known structures

As mentioned above, (pseudo-) atomic models are typically built for EM density maps of approximately  $10 \text{\AA}$  in resolution or better. The modelling process usually follows a well-established procedure that initially takes potentially available, prior structural and biochemical information into account. In particular the atomic coordinates of single (sub-) domains or other structural models that are deposited within a database are often of good use. Since the biochemical composition of an investigated macromolecular complex is usually known,

modelling attempts may be limited to exclusively those structural domains that are expected to be a member of the probed structure. As macromolecular complexes are often assembled from various subunits or domains of previously known proteins, structural models of the latter were often solved in advance by means of X-ray crystallography or NMR studies. Utilizing these models as rigid-bodies, exhaustive translational and rotational searches can be performed in order to match their appearance with a part of the electron density map (Rossmann et al., 2005). Several software packages are available to perform this task semi-automatically (Siebert and Navaza, 2009; Topf et al., 2008), even though manual adjustments by a trained user are almost inevitable to obtain meaningful results. Furthermore, additional information from biochemical studies or restraints obtained by crosslinking coupled to mass spectrometry (CX-MS) analysis (Schmidt and Urlaub, 2017) may be imposed onto the structural model. In particular CX-MS type of experiments often yield experimentally validated structural restraints that are tremendously valuable in most real applications (Dybkov et al., 2018; Rhode et al., 2003; Schmidt and Urlaub, 2017). In this way, small protein domains with a rather generic overall shape may also be reliably fitted into densities at areas that exhibit a less favourable local resolution. Should no coordinate model be available for a specific domain structure of interest, several methods are available for *de-novo* construction.

### **2.3.5.2 *De-novo* modelling of unknown structures**

The information content that is extractable from any electron density map is directly correlated to its resolution. Successful *de-novo* modelling attempts are thus strongly dependent on the (local) resolution of the map, where areas of higher resolution are usually modelled with greater confidence than those of lower resolution. The *de-novo* modelling procedure is therefore largely dependent on the actual features that can be identified in the map. A resolution of 10-6 Å initially allows to fit alpha-helices into characteristic strings of density that are readily observed (DiMaio and Chiu, 2016). However, since the specific density elements that represent side-chain moieties are not recognizable at these resolutions, it is obligatory to determine the register and general position of each helical element through prior information (Baker et al., 2010). The latter may be obtained by tracing a sequence that is leading away from an already positioned, adjacent part of the same domain or by considering additional restraints that are revealed by crosslinking studies. A reliable and justifiable *de-novo* placement of the structure is otherwise impossible.

For structures that are to be modelled into maps of a higher resolution (e.g. better than 4 Å), a true *de-novo* modelling approach that actually sequences the density can be feasible; even though practically the latter often emerges to be a challenging task. Ambiguous side-chain densities like those of equally sized amino acids (e.g. aspartate and asparagine, phenylalanine and tyrosine, serine and alanine), or other small amino acids often impede the sequence identification from maps that are between 3 and 4 Å in resolution. The initial modelling of longer strands of amino acids is thus required in order to identify the register of a given domain fragment. Once that is identified, the corresponding model may be extended in either direction, directly tracing the density while imposing the primary sequence of the known protein (Baker et al., 2010). Software like *Coot* or *Chimera* thereby provide the graphical user interface for this work and other generally supportive algorithms for model building (Emsley and Cowtan, 2004; Pettersen et al., 2004). Software that automatically approximates the sequence of a given density fragment is also available (Wang et al., 2015), providing the potential to automate the above described procedures.

### 2.3.5.3 Refinement and validation of atomic coordinates

Once a (pseudo) atomic model is built for the interpretable parts of a structure, its geometry and chemical plausibility generally need to be refined against the experimental data. The procedure thus aims at maximizing the fit of the model to the experimentally derived electron density map while minimizing steric violations and clashes. The latter is primarily achieved by imposing prior knowledge concerning the general chemistry of proteins onto the molecules in question (Brown et al., 2015). Multiple software packages are available to perform this task, utilizing various methodical approaches that were often originally developed to refine X-ray crystallography data in a comparable manner. The algorithmic solutions currently available operate with varying success, while the global energy minimization routines offered by software suites like *REFMAC* or *Phenix* are most commonly utilized in the field (Adams et al., 2010; Murshudov et al., 2011). A careful manual supervision of the often automated atomic coordinate refinement procedures is nevertheless recommended to prevent unrealistic steric clashes or other general misfits of the model. To geometrically validate the refinement procedure, software like *MOLprobity* was developed (Davis et al., 2007), which independently analyses the coordinate model and consequently denominates errors, clash scores or chemically implausible structures.

As soon as a model is finally optimized in its chemical plausibility and fit to the map, it may be analysed in detail; also without the simultaneous examination of the map it was derived from. The visualisation options provided by 3D rendering software like Chimera (Pettersen et al., 2004) or *PyMol* (DeLano Scientific LLC, USA) then allow to study the relationship between the structure and function of a macromolecular complex in great detail.

## 2.4 Pre-mRNA splicing

DNA harbours the blueprints for all proteins and most molecular entities that constitute life. The flow of genetic information thereby typically conveys these instructions from the chromosomal DNA via single-stranded molecules of messenger RNA (mRNA) to the protein synthesizing machinery in the cytosol. Within eukaryotic systems, however, the genetically encoded information is usually segregated by noncoding intervening sequences (*introns*) (Gilbert, 1978). Eukaryotic introns can span regions of several ten- to ten thousands of nucleotides (Gotoh, 2018). While the origin of these intronic sequences is still under vigorous debate (Irimia and Roy, 2014; Mattick, 1994; Rogozin et al., 2012), they often interrupt functionally important areas on the DNA and thus likewise on the premature RNA transcript thereof. Hence, the genetic information on a gene cannot be immediately passed on to the protein synthesizing machinery without further processing. As the cellular DNA is usually not modified for this purpose, the corresponding intron sequences have to be excised with great precision from the premature RNA transcript by RNA splicing.

The splicing of RNA in general and that of pre-mRNA in particular thus describes an ancient biochemical process in life that is facilitated by a chemically rather simple reaction. In essence, the procedure results in the sequential removal of a certain intron sequence from a consecutive strand of (pre-m) RNA by the hydrolysis and subsequent relegation of the flanking *exon* sequences; hence, a splicing reaction. The precise removal of potentially very large introns through splicing therefore provides a mechanism to restore or even establish the intended function of a gene product. At the same time, the process is capable of modifying the information

content or structural appearance of a raw transcript on its way from a DNA encoded gene to a functional protein or other RNA derived entity (Will and Lührmann, 2011).

Splicing occurs in all kingdoms of life (Yoshihisa, 2014), where the chemically similar outcome is catalysed via various routes and mechanisms. In Archaea and Eukaryotes the reaction is commonly facilitated through one of the major three pathways: Group I & II self-splicing introns (Saldanha et al., 1993) and the spliceosome dependent removal of introns from a strand of regular pre-mRNA (Will and Lührmann, 2011). Due to the severity of problems arising from frame shifted or otherwise misprocessed splicing products, the chemical reaction, albeit simple in nature, must be highly regulated and catalysed with exceptional fidelity. Group I and II self-splicing introns achieve the latter by autonomously folding into a fully functional ribozyme that is endogenously primed to splice out a defined part of its own sequence (Golden et al., 1998; Lambowitz and Zimmerly, 2011; Marcia and Pyle, 2012). The specifically arranged molecular framework of the premature RNA transcript thereby directly forms the catalytically active enzyme structure by itself, predominantly lowering or eliminating the necessity for additional protein factors to aid during catalysis. Due to the necessarily complex nature of the raw transcript, however, which self-evidently also needs to encode the blueprint of the catalytically active splicing machinery, only a very limited number of RNAs within the cellular context are processed in this manner.

Pre-mRNA transcripts, as a significantly more diverse family of splicing substrates within the eukaryotic cell, are instead targeted by a more intricate, *trans*-acting macromolecular machine termed the *spliceosome*. Functionally, its complex molecular architecture ensures a precise pre-mRNA target- and splice site selection as well as the subsequent catalysis of the splicing reaction itself. The spliceosome dependent splicing of pre-mRNA (for simplicity, henceforth referred to as “splicing”) is usually facilitated in the nucleus concomitantly with transcription (Tilgner et al., 2012). As introduced above, it represents one of the most significant pathways to process, influence and manipulate genetic information on its way from a gene to a functional protein. Alternative splicing, as commonly observed in higher eukaryotes, furthermore allows for a specific modification and diversification of the protein products derived from a single gene on the DNA. Depending on the cellular context it thus enables one gene to encode for various proteins; a feature that is particularly essential for the rise and development of multicellular organism like ourselves (Roy et al., 2013).

The spliceosome-dependent splicing of pre-mRNA is therefore a process of exceptional relevance for the development and homeostasis of eukaryotic life. The mechanism and function of the particularly dynamic molecular machine was consequently investigated for decades through primarily classic biochemical research. Even though important discoveries were made during this time, the understanding of the spliceosome’s precise structural architecture and mechanism of action was at all times limited by the relatively poor spatial resolution of the biochemical experiments. In fact, only high-resolution models of the entire macromolecular machine obtained by means of structural probing, i.e. through electron microscopy, offer the chance to significantly increase the comprehension of the structural relationships of the splicing process.

While no high-resolution structures of a catalytically competent spliceosomal complex were available upon the start of this project, numerous have been published within the last three years. In order to offer the best possible overview of the current state of research I will therefore have to anticipate some of my own results within this introduction, which were in parts also

published in the meantime by our competitors. Generally, the field of splicing and spliceosome research has made a tremendous leap forward during the last years, as the numerous recently unveiled spliceosome structures enabled an in-depth understanding of the involved molecular mechanisms for the first time.

## 2.4.1 Basic splicing mechanics

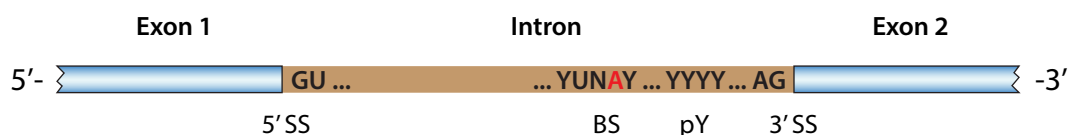
Comparable to the well-conserved chemistry of intracellular RNA polymerization, a ubiquitous two-step reaction pathway facilitates the splicing of RNA. In this regard, the initial cleavage of the premature poly RNA molecule is executed at a defined position during step 1 of the reaction, whereas step 2 subsequently completes the removal of the intron while facilitating the relegation of the adjoining exon sequences at the same time. Focussing on the process of major spliceosome-dependent pre-mRNA splicing in humans, the following sections will briefly elucidate the requisite sequence features of the RNA substrate as well as the basic chemistry of the reaction.

### 2.4.1.1 Sequence encoded features of the pre-mRNA

To accurately splice any kind of RNA transcript a precise definition of the corresponding intron and exon boundaries is crucial. Pre-mRNA therefore contains several well-conserved, endogenous signal sequences that direct catalytic spliceosome activity exclusively to the intended splice sites.

As a continuous strand of RNA is cleaved twice in order to remove an enclosed intron part of the assembly, two dedicated splice sites (SS) are required. These SS are usually characterised by a defined sequence of nucleotides, which are directly encoded within the substrate RNA molecule. Since pre-mRNA exons usually code for a particular amino acid sequence, SS defining nucleotides are instead located within the non-translated intron region of the molecule (Fig. 5). Depending on whether recognized by a U2-type (major) or U12-type (minor) spliceosome, the corresponding sequence snippets are highly conserved throughout the genome and across species (Sheth et al., 2006; Turunen et al., 2012).

In general, tracing the pre-mRNA molecule in 5' to 3' direction, the interface between the 3' end of the exon (i.e., exon 1) and the 5' end of the intron is typically referred to as the 5' splice site (5' SS). Across species and for the major spliceosome, this SS is ubiquitously defined by a GU or GC doublet as the first nucleotides of the intron, followed by a slightly more degenerate purine sequence of 3-4 bases in metazoans (Konarska, 1998; Sheth et al., 2006).



**Fig. 5 Typical metazoan pre-mRNA.** Exon 1 and Exon 2 are depicted in blue, the intron is depicted in brown. The 5' and 3' splice sites (SS), the branch site (BS) and polypyrimidine tract (pY) are indicated accordingly. The letters residing within the illustration of the intron represent single RNA nucleotide bases, where Y represents any pyrimidine. The branch site adenosine (BS-A) nucleotide is marked in red.

In a similar manner, the 3' splice site (3' SS) also features an equally well conserved AG doublet of nucleotides across all species, representing the two most terminal 3' nucleotides of any intron

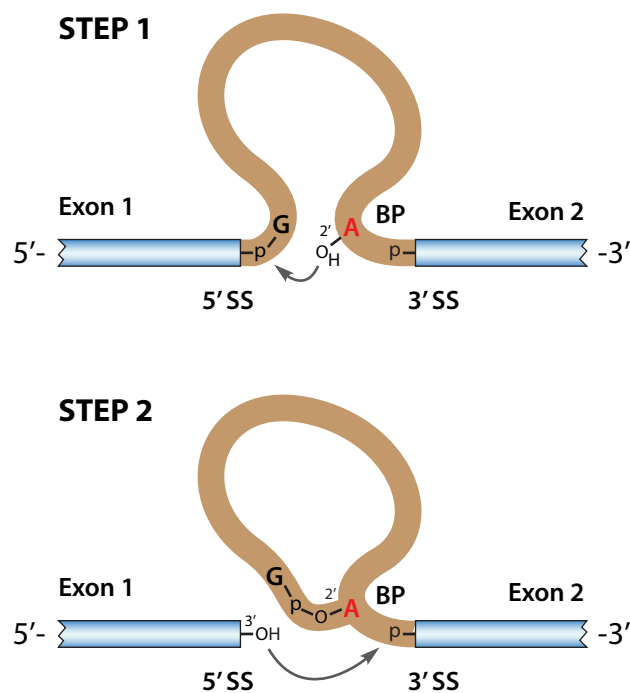
(Sheth et al., 2006). In metazoans, the 3' SS is furthermore often preceded by a characteristic poly-pyrimidine tract (pY) (Coolidge et al., 1997).

During the first step of splicing (also see section 2.4.1.2) a characteristic intron lariat structure is formed at the branch site (BS), involving a specific adenosine nucleotide (hence, the *Branch Site Adenosine* (BS-A)). Comparable to the importance of an accurate SS selection, the specific BS-A has to be recognized and utilized in a similarly careful manner by the spliceosome and is thus flanked by an additional set of conserved nucleotides in every intron element. While the BS region in *S. cerevisiae* (henceforth referred to as yeast) is remarkably conserved (i.e., UACUAAC; the BS-A nucleotide is underlined), the related sequence in metazoans only features a conserved uracil and the BS-A itself (i.e. yUnAy in the human case, where *y* represents a greater likelihood for a pyrimidine and *n* any nucleotide) (Gao et al., 2008). The branch site sequences of higher eukaryotes are thus more degenerate and proper BS selection is amongst others further supported by other nucleotides like those of the pY tract (Gao et al., 2008).

### 2.4.1.2 Chemical basis of splicing

RNA nucleotides naturally polymerise through the formation of a specifically conserved bond between an oxygen atom of the phosphoester that is attached to the 5' carbon of the +1 nucleotide ribose molecule and the hydroxyl oxygen bonded to the 3' carbon of the previous -1 nucleotide ribose. In order to facilitate step 1 of splicing, this bond has to be broken and substituted by an  $S_N2$ -type transesterification reaction. Initiating the reaction, the 2' hydroxyl oxygen of the conserved BS-A is arranged in a geometrically favourable way that allows for a nucleophilic attack on the phosphoester of the last intron nucleotide adjacent to the 5' splice site (5' SS) (Fica et al., 2013). While a free 3' hydroxyl group of the most terminal exon 1 nucleotide remains, the 5' end of the intron is now covalently bonded to the 2' hydroxyl oxygen of the BS-A. Hence, a characteristic *intron lariat* structure is formed (Fig. 6).

Step 2 of splicing accordingly requires the corresponding 3' splice site (3' SS) to be positioned adjacently to the step 1-derived, free 3' hydroxyl group at the terminus of exon 1. Once again arranged in an appropriate geometry, a second  $S_N2$ -type transesterification reaction can occur, resulting in the attack of the aforementioned 3'



**Fig. 6 Chemistry of pre-mRNA splicing.** Exon 1 and Exon 2 are depicted in blue, the intron is depicted in brown. The 5' and 3' splice sites (SS) and the branch site (BS) is indicated accordingly. Step 1 of the reaction is initiated by an attack of the characteristic 2'OH from the BS-A nucleotide on the phosphate of the 5' SS residing guanosine nucleotide, forming the characteristic intron lariat structure. Step 2 is then facilitated by an attack of the exposed 3'OH group of Exon 1 on the corresponding phosphate Exon 2. Upon completion of the reaction both exons are spliced together and the intron lariat structure is removed from the assembly.

hydroxyl oxygen from exon 1 onto the phosphoester of the first exon 2 nucleotide. As the intron attached 3' hydroxyl group is eventually substituted by the corresponding exon 1 hydroxyl group, the intron lariat is effectively spliced out of the (pre-) mRNA and removed from the assembly. Concurrently, both exons are re-ligated due to the substitutive nature of the reaction (Will and Lührmann, 2011).

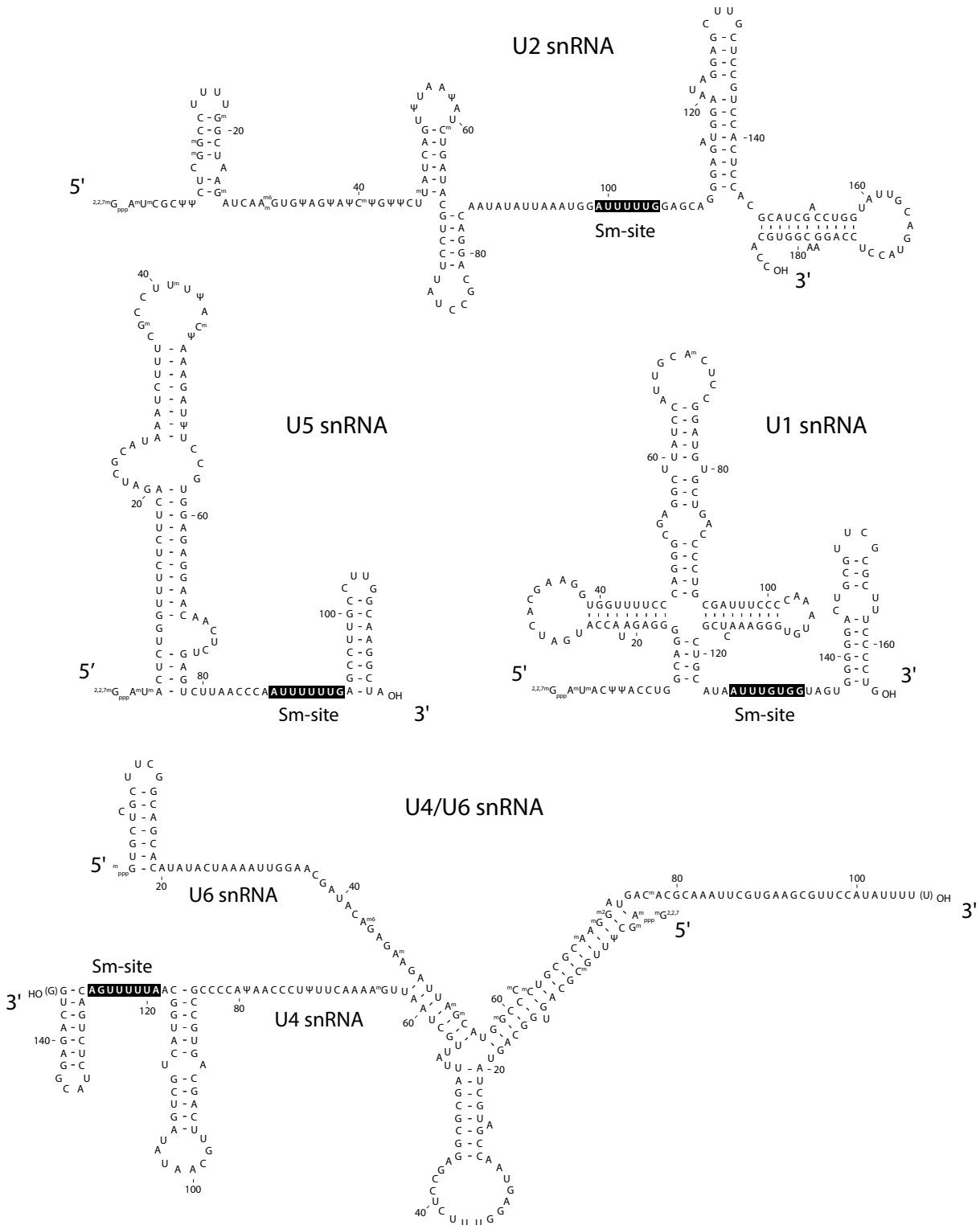
As typically observed for most chemical reaction mechanisms,  $S_N2$ -type reactions likewise require a precise steric configuration of the involved reactants in order to achieve the highest rate. Within the catalytic centre of the spliceosome, RNA moieties and metal cations (e.g.,  $Mg^{2+}$ ) directly interact with the pre-mRNA substrate and aid in appropriately positioning the latter for catalysis (Fica et al., 2013). RNA, not proteins, therefore catalyse the actual splicing reaction, rendering the spliceosome a ribozyme.

## 2.4.2 General components of the spliceosome

Comprised of RNA and protein components alike, the spliceosome represents one of the largest macromolecular machines of the cell, reaching several megadaltons in mass at certain assembly stages. In order to comprehend its catalytic mechanism and regulation, a more in-depth appreciation of the properties and characteristics regarding the most integral building blocks is essential. A brief introduction is thus provided below.

### 2.4.2.1 RNA

Each major spliceosomal complex is eventually utilizing five different *uridylic-acid-rich small nuclear ribonucleic acids* (U-snRNA), namely the U1, U2, U4/U6 and U5 snRNAs (Hodnett and Busch, 1968; Will and Lührmann, 2006). Human U-snRNAs (Fig. 7) are rather short (106-187 nt) in length and transcribed in the nucleus predominantly by RNA polymerase II (Will and Lührmann, 2001). U1-U5 snRNAs feature a common 5' trimethylguanosine (TMG) cap (Guthrie and Patterson, 1988) and are able to bind seven auxiliary Sm-proteins (the *Sm-core*) in a heptameric ring-like manner via a conserved Sm-binding site, which is situated towards the 3' end of the strand (Branlant et al., 1982; Urlaub et al., 2001). The Sm-core itself is assembled in the cytoplasm after the newly synthesized snRNAs have been exported from the nucleus. Promoted by the binding of a complete set of Sm proteins, further RNA processing steps like TMG cap hypermethylation (Mattaj, 1986) or modification and trimming at the 3' end of specific snRNAs are performed in the cytosol (Seipelt et al., 1999). The Sm-core bound snRNAs are then re-imported into the nucleus for further processing and assembly into a functional spliceosome (Fischer et al., 1993). In contrast, the U6 snRNA is transcribed by RNA polymerase III and does not contain a sequence encoded Sm-binding site (Dahlberg and Lund, 1991). Instead, it interacts with seven *Sm-like* (LSm) proteins, which also form a ring-like structure in a similar geometry and position compared to that of their Sm counterparts. Due to all U6 snRNA assembly and processing stages taking place in the nucleus, nuclear export is neither required nor performed during its biogenesis (Achsel et al., 1999).



**Fig. 7** Sequence and predicted secondary structure of the five human spliceosomal snRNAs. The nucleotide sequence is depicted in the labelled direction. A base pairing of nucleotides is indicated by a dash, the conserved Sm-binding sites by white writing on black background. The chemical modification of certain nucleotides is denoted as follows: Ψ = pseudouridine; m(number) = methylation (at the correspondingly numbered position of the nucleobase), no number indicates a 2'-O-methylation; ppp = tri-phosphate.

Importantly, defined snRNA regions fold into characteristically shaped 3D stem-loops and helices, some of which are only transiently formed. The resulting RNA-RNA interactions are



often highly conserved across species and essential for proper spliceosome functionality. The corresponding structures are either formed intramolecularly, as in the case of several stem-loops of U2 snRNA (Perriman and Ares, 2007), or intermolecularly, as observed during the association of the U2 or U4 and U6 snRNAs (Sun and Manley, 1995; Will and Lührmann, 2001). Modulating these interactions, certain bases experience posttranscriptional modifications equal to those that were likewise reported for their ribosomal RNA (rRNA) counterparts (Penzo and Montanaro, 2018). Site-specific pseudouridylation or 2'-O-methylation is performed at conserved regions of the U1, U2, U4 and U6 snRNA (Karijolic and Yu, 2010; Reddy and Busch, 1983). An intricate network of RNA-RNA interactions thus further extends the structural and chemical abilities of the spliceosomal molecular framework.

#### **2.4.2.2 Protein**

Even though the splicing reaction itself is catalysed by RNA, proteins, by mass, make up the largest portion of the molecular machine. In fact, more than 200 different proteins are reported to associate with the human spliceosome, while the assemblies of lower eukaryotes such as yeast are less complex in molecular composition (Agafonov et al., 2011). The spliceosome is thus an exceptionally protein-rich *ribonucleoprotein* (RNP) machine, in particular when compared to other large ribozymes like the ribosome.

Spliceosomal proteins assist in integrating the tremendously complex snRNA network; guiding and sequestering pre-mRNA substrates; performing RNA helicase, binding and translocation activities; and participating in spliceosome regulation and localisation (section 2.4.3). Proteins are essential for successful, spliceosome-derived pre-mRNA splicing (Will and Lührmann, 2011), although the catalysis of some artificial phosphoester linkages by snRNAs alone was described in the literature (Valadkhan and Manley, 2001).

Historically, some of the most important *precursor RNA processing* (PRP) alleles were identified in yeast, as demonstrated by large temperature-sensitive mutation screens during the late 1980s and 1990s (Moore et al., 1993; Noble and Guthrie, 1996). By searching for defects in, or the accumulation of, splicing products or intermediates, previously unknown genes and protein families that often harboured Zn-finger, RNA-helicase or other sequence motives known to be associated with protein-RNA interactions were characterized. A great homology between the splicing factors of different species (Fabrizio et al., 2009; Will and Lührmann, 2011) eventually led to the assumption that splicing is a particularly conserved process in life. The latter led to the assumption that the yeast spliceosome is generally assembled by a core set of essential protein components, whereas that of higher eukaryotes, metazoans or humans must have acquired additional, potentially regulatory factors during evolution (Fabrizio et al., 2009). Besides those factors that fulfil merely structural tasks, eight highly conserved DEXD/H-box ATPases or RNA helicases were identified to associate and act in the spliceosome (Cordin and Beggs, 2013). More of the most important proteins that are functionally involved in spliceosome assembly and operation are specifically introduced in section 2.4.3.

#### **2.4.2.3 Small nuclear ribonucleoproteins (snRNP)**

Given the wealth of different molecular components that need to be assembled in order to form a functional spliceosome, a pre-organization of those components *in vivo* can simply be assumed for kinetic reasons. Indeed, spliceosome formation is enabled by the merger of four *small nuclear ribonucleoprotein* (snRNP) particles, which were first described by Lerner and Steitz (Lerner et al., 1980; Lerner and Steitz, 1979) and eventually found to add many but not all of the

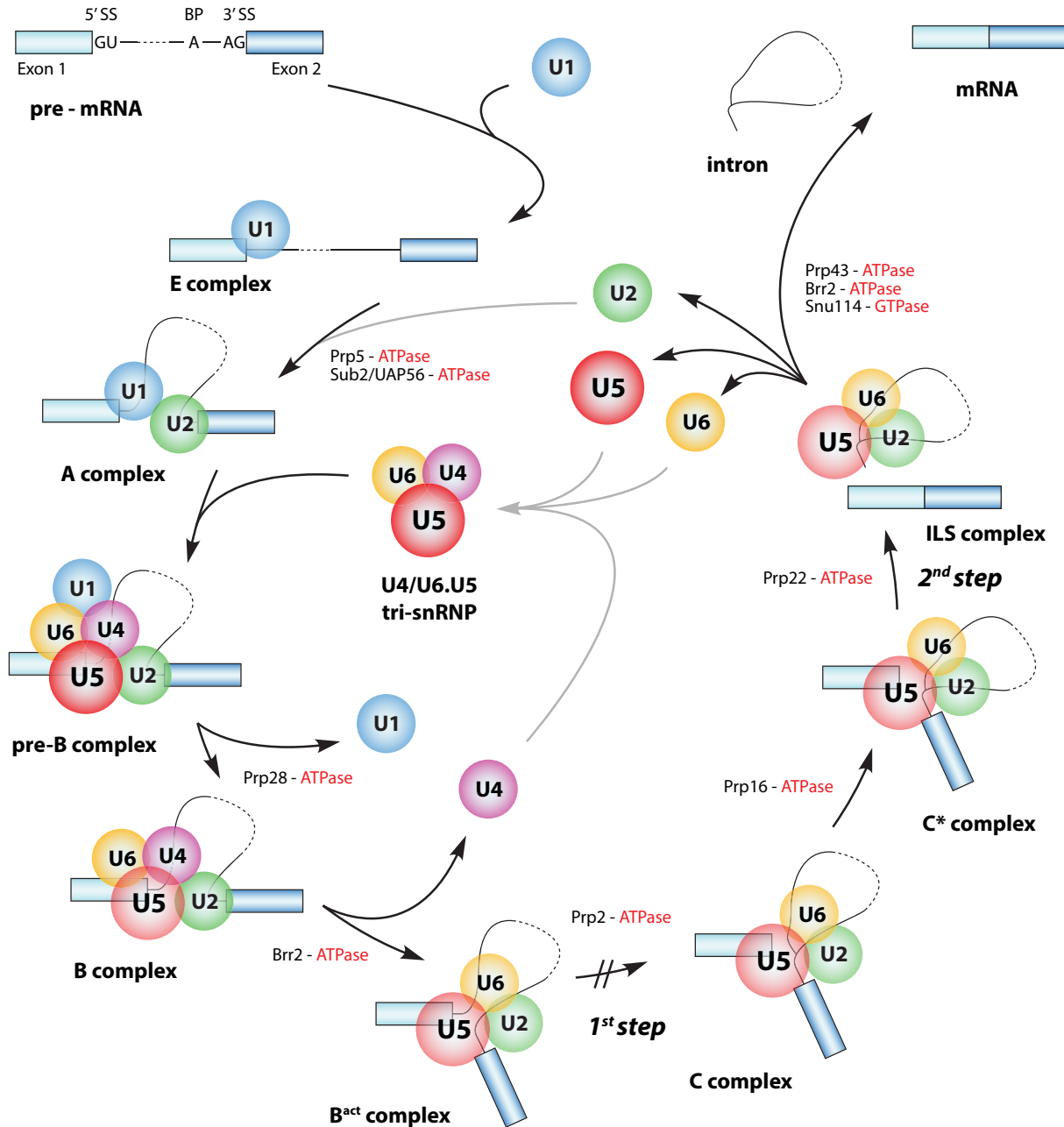
essential molecular components to the functional complex. In general, each snRNP is comprised of one or two snRNA(s), as well as ~ 10 – 30 associated protein factors (Kambach et al., 1999). The aforementioned Sm-proteins that are associated with the snRNAs are thereby essential for functional snRNP formation (Ségault et al., 1995). Since the U4 and U6 snRNAs are typically associated with each other through intermolecular base pairing, they form a characteristic *di*-snRNP assembly in combination with proteins like Prp 3, 4 and 31 (Wahl and Lührmann, 2015). Named according to the incorporated snRNA, U1, U2, U4/U6 and U5 snRNPs are commonly distinguished regarding the major spliceosome (Moore et al., 1993).

Once formed, snRNPs (e.g., U1 and U2 snRNP) are then directly recruited by the pre-mRNA, or interact to merge into larger assemblies like the *tri*-snRNP (U4/6 and U5 snRNP) before being integrated into a competent, pre-catalytic spliceosome complex (section 2.4.3). Besides the snRNPs, certain proteins were additionally found to co-purify and thus likely pre-assemble into larger building blocks, which then associate with the emerging spliceosome. In particular the intron-binding complex (IBC) (De et al., 2015), the RES (retention and splicing) complex (Dziembowski et al., 2004) and most importantly the PRP19/CDC5L complex (*nineteen complex* in yeast, NTC) (Makarova et al., 2004) belong into this group. The remarkable remodelling events leading to catalytic activation and progression are further described below.

### 2.4.3 Spliceosome assembly and function

One of the most distinctive characteristics of the spliceosome is the dynamic nature of its assembly and catalytic function (Will and Lührmann, 2011). Spanning more than 40 years of research, a tremendous amount of intra- and intermolecular interactions have been identified between many of the involved molecular entities, which transiently strengthen and/or dissociate during specific stages of catalysis. The spliceosome thus not only appears to be a particularly large macromolecular machine, but also one that alters its appearance, size and molecular composition during each round of splicing in an unprecedented manner.

Consequentially, multiple spliceosomal complexes exist in parallel within the cell at different functional states. Building up or transitioning between these states, specific proteins or snRNAs are added (stabilized) or removed (destabilized) from the complex in a particularly conserved fashion (Will and Lührmann, 2011). The spliceosome thus assembles and functions in a stepwise manner, beginning at the first interaction of the U1 snRNP with the 5' SS and ending with the disintegration of the molecular machine after splicing is completed with step 2 of the reaction. Once fully assembled, a defined set of core components is stably associated around the pre-mRNA substrate at the centre of the molecular machine. The latter then enables peripheral factors to join, leave and remodel this core at specific stages of the reaction. In general, spliceosome formation can be divided into three major stages: First, the *pre-spliceosomal* assembly at which the 5' SS and BS is determined; second, the *pre-catalytic* complex formation states, which are primed but not yet competent for catalysis; and finally the third, *catalytically activated* stage of spliceosomal assembly that performs the intron excision and re-ligation of the exons.



**Fig. 8 Splicing cycle.** The snRNPs of the spliceosome are represented by coloured spheres and labelled accordingly. The directionality of the reaction or the corresponding assembly / disassembly pathway is marked by black arrows. Grey arrows indicate a presumed pathway of utilisation. The various helicases involved in spliceosome remodelling and function are indicated at their corresponding position of activation. For further information, see text.

The rearrangements facilitated during the catalytic cycle are striking and have been subject to many years of extensive investigation. Summarizing these findings, most of the well-known spliceosomal intermediates are depicted within the *splicing cycle* (Fig. 8). The wealth of biochemical and structural data regarding whole spliceosomal complexes and their constituting components now enables an in-depths analysis and comparison of discrete assembly and activation states within the splicing cycle. Even though this cycle, as depicted above, was constantly expanded throughout the years, it is by far not complete and further intermediate, even more transient structural states of the spliceosome are likely to emerge in the future. Nevertheless, particularly when analysing the knowledge that was acquired through the

structural elucidation of spliceosomal components, it is nowadays possible to gain a good overview on basic spliceosome mechanics. The following sections will accordingly briefly guide through the mechanics of spliceosomal pre-mRNA splicing and further introduce significant proteins and conserved RNA structures. Even though yeast spliceosomal complexes, so far, were found to be largely homologous in structure and function, specific differences exist. As this work elucidates the structure and function of human spliceosomal complexes, the explanations below will largely focus on the human system. Nevertheless, wherever necessary details specific to yeast may sometimes add to a better understanding and are thus occasionally included.

### **2.4.3.1 Substrate recognition and 5' splice site definition**

The splicing of pre-mRNA is classically initiated by the recognition of the substrate through its 5' SS and BS. Even though the corresponding nucleotide sequences are later on handed over to other molecular components of the assembled spliceosome, the fidelity of the entire mechanism is determined at this earliest step of assembly. Components of the U1 snRNP are responsible for the recognition and binding of the 5' SS. Nucleotides at the 5' tail of the U1 snRNA thereby directly base pair with the correspondingly conserved nucleotide sequence on the pre-mRNA (Stark et al., 2001). Supporting this interaction, proteins like the U1-70k and U1-C additionally contact the splice site in metazoans and aid in its stable integration into the U1 snRNP (reviewed by (Will and Lührmann, 2006)). Once integrated into the snRNP, the fate of the 5' SS is determined, creating *Complex E*.

Next, the recognition of the BS is facilitated by pre-organized factors in the U2 snRNP. Again, the U2 snRNA serves as molecular probe that identifies and binds the conserved branch site sequence downstream of the 5' SS on the pre-mRNA. Forming a characteristic RNA double helix, the *branch-site helix* (BSH), the central BS-A nucleotide is bulged out to subsequently serve as the nucleophile during the initiation of step 1 of splicing (Wahl et al., 2009). Stabilizing this interaction, the U2 snRNP-associated protein heterodimer U2AF65/35 contacts the pre-mRNA pY tract downstream of the BS. Proteins of the SF3A and SF3B families additionally interact with the intron 6 to 26 nucleotides upstream of the BS-A (Gozani et al., 1996), stably sequestering the BS-A nucleotide midway. Upon productive binding of the U2 snRNP particle to the pre-mRNA, the snRNP is then remodelled by the DEAD-box RNA helicase PRP5 (Liang and Cheng, 2015) in order to associate with an up- or downstream situated E complex. The resulting pre-spliceosome is named the *A complex*.

### **2.4.3.2 Pre-catalytic spliceosome assembly**

Once the A complex is formed by integration of the pre-mRNA and the U1/U2 snRNPs into the pre-spliceosome, further components are required to build up the catalytic centre. Factors of the pre-formed U4/U6.U5 *tri-snRNP* are essential for this purpose and thus have to be incorporated into the complex. The U4/U6.U5 tri-snRNP itself is comprised of its three snRNAs in addition to more than 25 distinct proteins, some of which are most essential to the spliceosomal function and assembly (Behrens and Lührmann, 1991). In particular the U5 snRNP derived proteins PRP8 and BRR2 as well as the U6 snRNA, which eventually form the catalytic core of the spliceosome, play exceptional roles in the subsequent splicing procedure. PRP8 thereby provides the general scaffold for spliceosome assembly and BRR2 represents a major player in spliceosome activation (see below). Specific nucleotides of the U6 snRNA eventually facilitate catalysis (Madhani, 2013; Madhani and Guthrie, 1992). Furthermore, a conserved stretch of U4 and U6 snRNA nucleotides is tightly engaged in a duplex helical structure within the complex

(Agafonov et al., 2016; Bringmann et al., 1984; Hashimoto and Steitz, 1984). The latter structural element is thus particularly characteristic for the pre-catalytic state of the assembly. As indicated by biochemical and structural studies, most of the U6 snRNA nucleotides involved in this duplex helix, however, are eventually required to participate in the formation of the catalytic centre of the spliceosome (Agafonov et al., 2016; Fica et al., 2013). The latter thus exemplarily illustrates the requirement for substantial rearrangements of the various (pre-) spliceosome particles during the process of catalytic activation.

In order to proceed with the assembly process, the pre-formed U4/U6.U5 *tri-snRNP* loosely associates with the A complex in order to form the pre-catalytic *pre-B complex*. The merger is initiated by the formation of a short helix between the 5' and 3' ends of U2 and U6 snRNA (U2/U6 helix II), respectively, functionally fusing both complexes together (Boesler et al., 2016). The resulting spliceosome now contains all snRNPs but is still in need of substantial rearrangements in order to prepare for catalytic activity.

Just prior to catalytic activation, RNA helicase PRP28 initiates the transfer of the pre-mRNA 5' SS from the U1 snRNA to the similarly highly conserved U6 snRNA *ACAGA box* sequence (Boesler et al., 2016; Staley and Guthrie, 1999). In consequence, U1 snRNA loses contact with the pre-mRNA and the U1 snRNP is significantly destabilized, ultimately leading to the formation of the pre-catalytic *B complex*. At the same time, nucleotides of the characteristic U5 snRNA loop 1 were identified to interact with those of the most 3'-terminal end of the 5' exon ("exon 1") (Sontheimer and Steitz, 1993), indicating a remodelling event that leads to an incorporation of the adjacent 5' SS into the U5 snRNP within the B complex. A recently published structure of the yeast B complex obtained by electron cryo-microscopy in fact reveals that the pre-mRNA/U6 snRNA comprised ACAGA-box helix rests particularly close to the U5 snRNA loop 1, even though virtually no parts of the 5' exon itself are visualized within this structure (Plaschka et al., 2017). Furthermore, the yeast structure suggests that U5 snRNP associated Ski2-like helicase Brr2 has moved more than 200 Å away from its original position in the human tri-snRNP. In the yeast B complex, Brr2 instead contacts its U4 snRNA substrate (Laggerbauer et al., 1998), primed to act on the above-mentioned U4/U6 duplex helical structure.

Catalytic activation and the formation of the active centre is eventually achieved by further substantial rearrangements, ultimately initiated by the BRR2-catalysed unwinding of the U4/U6 duplex (Cordin and Beggs, 2013). By doing so, most proteins belonging to the U4/U6 di-snRNP, B complex-specific proteins and the U4 snRNA are substantially destabilized and removed from the assembly (Fabrizio et al., 2009). Where in humans about 20 proteins leave the stable association with the pre-catalytic spliceosome, more than 25 are added in return, particularly those that belong to the RES, IBC and NTC complexes (Will and Lührmann, 2011). The resulting, still pre-catalytic but now activated spliceosome is called *Complex B<sup>act</sup>*.

The rearrangements that facilitate the B-to-B<sup>act</sup> complex transition are astounding and can nowadays be followed by a comparison of the recently published structures of the human (Haselbach et al., 2018; Zhang et al., 2018) and yeast (Rauhut et al., 2016; Yan et al., 2016) B<sup>act</sup> complexes. Briefly summarizing its state of assembly, the complex is almost fully primed for subsequent catalytic activation. The catalytic centre was thus formed by an intricate network of U2 and U6 snRNA, amongst others involving the important U2/U6 snRNA Helix I and the U6 snRNA internal stem-loop (ISL) structure, which are snugly embedded into basic pockets of PRP8. Nevertheless, one Mg<sup>2+</sup> ion that is essential for catalysis (Fica et al., 2013) has not yet bound to the catalytic U6 snRNA moieties. At the same time, exon 1 resides tightly enclosed by

PRP8 near the 5' SS. Hence, its important *RT/EN-domain* (PRP8<sup>RT/En</sup>) now adopts a *closed conformation*. Additionally, the 3'-terminal end of exon 1 is held in place by U5 snRNA loop 1. The BSH and its branched out BS-A nucleotide, albeit still sequestered by various SF3B proteins, is at the same time brought in closer to the 5' SS in preparation for step 1 of splicing. Several proteins from the PRP19 complex furthermore contact and thus potentially stabilize the catalytic RNA network and PRP8, amongst others for example the TPR-containing SYF3, CWC15, PRLG1, SKIP or CDC5L.

### **2.4.3.3 Catalytic activation and step 1 of splicing**

The final progression in the catalytic activation of the spliceosome is eventually dependent on the action of RNA helicase PRP2. Specifically binding the intron 25-30 nucleotides downstream of the BS (Liu and Cheng, 2012) it is proposed to initiate the transition from the B<sup>act</sup> complex to the *B\* complex*. The latter then finally represents the fully catalytically activated form of the spliceosome that catalyses step 1 of the reaction. Even though no structural and little biochemical data is available describing this probably very transient complex, it seems likely that SF3A and SF3B proteins, which remained in contact with the BSH since its selection by the U2 snRNP, are removed in order to align the BS-A at the catalytic centre. Indeed, various studies show that PRP2 is involved in this process from a distance (Lardelli et al., 2010; Ohrt et al., 2012).

Once step 1 of splicing is catalysed by the B\* complex, the catalytically activated spliceosome progresses to a post step 1 state, termed the *C complex*. As for the structures of the B<sup>act</sup> complex, yeast and human C complex assemblies were recently visualized by cryo-EM (Galej et al., 2016; Wan et al., 2016a; Zhan et al., 2018). Analysing these structures, it becomes apparent that indeed the previously SF3 protein sequestered BSH is now released from the U2 proteins and integrated at the heart of the spliceosomal core. Contacted by the step 1 splicing factors CCDC49, CCDC94 and ISY1, the BSH is now situated in close proximity to the catalytic centre. Within the C complex, the most 3' terminal nucleotide of exon 1 thus only rests several angstroms away from the newly formed intron lariat structure attached to the BS-A, close to the former 5' SS. In order to facilitate step 2 of the reaction, the BSH-intron lariat structure, including a wealth of proximal protein factors, again need to be significantly rearranged. In contrast to the restructuring events that are observed in the pre-catalytic spliceosomes, however, the remodelling of the catalytically activated complexes is generally more restricted to the reactive RNA components (e.g. the BSH, exon 2, etc.) as well as very peripheral proteins. The catalytic core of the assembly – which is comprised of the proteins PRP8, SNU114, the U5 SM-core, PRL1 and the U5 and U6 snRNAs – thus generally retains its overall structure while forming the single catalytic centre.

A progression in the catalytic cycle then eventually requires the detachment and removal the BSH, including the attached intron lariat structure, from the catalytic site at the core of the spliceosome. In fact, the RNA helicase PRP16 is reported to lead to the destabilization of CCDC49 and CCDC94 as well as a remodelling of ISY1 (Horowitz, 2011), which allow the BSH and the associated proteins to reposition approximately 20 Å away from their previous location at the catalytic centre. Being primed for step 2 of the reaction, the spliceosome has now reached the *C\* complex* configuration. As several structures in yeast (Fica et al., 2017; Yan et al., 2017), and recently also in humans (Zhang et al., 2017), indicate, the complex is in fact primed for exon ligation, even though exon 2 could not yet be visualized in any published model of the C\* complex state. Besides the remodelling events introduced above and a wealth of others that are

not discussed here in detail, the characteristically ring-shaped WD40 domain of PRP17 is translocated more than 70 Å between the C and C\* complex states. In the latter, it now tightly interacts with the BSH region of the assembly, stabilizing its position and therefore the pre-step 2 configuration of the complex.

#### **2.4.3.4 Step 2 of splicing and spliceosome disassembly**

The C\* complex appears primed to complete the splicing reaction by finally excising the intron lariat structure and ligating both exons. How exactly this process is achieved in the human system, however, remains poorly understood. Biochemically, the action of RNA helicase PRP22 is reported to be involved in exon ligation and pre-mRNA release (Fourmann et al., 2013), as it is already present at the periphery of the C\* complex. Unsurprisingly, two recently published cryo-EM studies of a yeast post-step 2 spliceosome, the *P complex*, revealed an overall configuration that is substantially comparable to that of the C\* complex (Liu et al., 2017; Wilkinson et al., 2017). As consistently reported, the latter studies show that the yeast P complex utilizes the conserved and readily positioned 5' SS and BS-A nucleotides in order to form non-Watson-Crick base pairs with the 3' SS adjacent to exon 2. Once positioned, the free 3' OH of exon 1 may then attack the correspondingly positioned phosphate of exon 2, ultimately completing the splicing reaction by exon ligation and intron removal. Yeast Prp22 is then found to act on exon 2, 15 – 21 nucleotides away from the site of ligation, potentially aiding in the removal of the spliced mRNA from the complex. The exact situation in the human system, however, remains to be elucidated, as particularly alternative splicing events may require a more highly sophisticated regulation of exon ligation during step 2 of the reaction.

Once splicing is completed, as mentioned above, PRP22 is likely involved in removing the spliced mRNA from the spliceosome. The remaining *intron-lariat complex* (ILS complex) is then thought to be dissembled in order to recycle the involved splicing factors for subsequent rounds of splicing (Will and Lührmann, 2011). Even though no structural model of a human or *S. cerevisiae* ILS complex are available to date, the corresponding structure of an *S. pombe* spliceosome was reported in the past (Yan et al., 2015). Besides the missing mRNA, its appearance is largely comparable to that of the yeast P complexes.

The final disassembly and recycling procedures of the spliceosome are structurally poorly understood. Nevertheless, the DEAH-box helicase PRP43 and its cofactors NTR1 and NTR2 play an important role (Fourmann et al., 2017). Acting on the intron lariat RNA structure near the U2-BSH (Fourmann et al., 2016), the ILS complex is segregated from the intron lariat and then dismantled into smaller snRNPs or protein assemblies (Arenas and Abelson, 1997; Fourmann et al., 2013; 2017). Single snRNPs are then targeted to and recycled within *Cajal bodies* (CB) (Schaffert et al., 2004; Staněk et al., 2003). A subsequent release of the rearranged or restructured snRNPs may then result in their participation in further rounds of splicing.

### 2.5 Aim of the study

The spliceosome-dependent splicing of precursors to mature mRNA is an astonishingly complex process that takes place in the nucleus of the eukaryotic cell. Many years of biochemical studies, structural probing and the elucidation of individual spliceosomal components by means of X-ray crystallography or NMR have provided countless hints and a rough understanding of the involved mechanisms. It was thus shown that intricate networks of dynamically adjusted interactions between the individual components of the spliceosome regulate and facilitate the splicing reaction in a highly coordinated manner. Even though much could be learned in the past about the separate factors that constitute the molecular machine and its biochemical behaviour, the spliceosome, assembled from up to hundreds of individual components, is only functional as a whole. Examining the complex structure at high resolution in its entirety, not separated into individual components, is thus a promising approach to reveal even more of the essential intermolecular interactions and work towards a better understanding of the dynamic nature and functional mechanics of the spliceosome. Since biochemical studies are unlikely to reach the spatial resolution that is required for this purpose, the aim of this study is to elucidate the high-resolution structure and function of complete human spliceosomal complexes.

Utilizing cryo-EM as a tool, this study targets the recording and analysis of entirely functional spliceosomes at high resolution, not only individual parts of the assembly. Focussing on the human pre-catalytic B complex as well as the catalytically activated, pre-step 2 C\* complex, we hope to contribute towards the development of a more detailed mechanistic model of pre-mRNA splicing.



## 3 Publications

Two publications elaborating on the structure and function of a human spliceosomal B and C\* complex, respectively, represent the major body of results in this work. Both publications are attached as they were originally published in the following section.



### **3.1 Cryo-EM structure of a human spliceosome activated for step 2 of splicing**

**Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R.**

**Cryo-EM structure of a human spliceosome activated for step 2 of splicing**

2017

Nature *542*, pp. 318–323

# Cryo-EM structure of a human spliceosome activated for step 2 of splicing

Karl Bertram<sup>1\*</sup>, Dmitry E. Agafonov<sup>2\*</sup>, Wen-Ti Liu<sup>1</sup>, Olexandr Dybkov<sup>2</sup>, Cindy L. Will<sup>2</sup>, Klaus Hartmuth<sup>2</sup>, Henning Urlaub<sup>3,4</sup>, Berthold Kastner<sup>2</sup>, Holger Stark<sup>1</sup> & Reinhard Lührmann<sup>2</sup>

**Spliceosome rearrangements facilitated by RNA helicase PRP16 before catalytic step two of splicing are poorly understood. Here we report a 3D cryo-electron microscopy structure of the human spliceosomal C complex stalled directly after PRP16 action (C<sup>\*</sup>). The architecture of the catalytic U2–U6 ribonucleoprotein (RNP) core of the human C<sup>\*</sup> spliceosome is very similar to that of the yeast pre-Prp16 C complex. However, in C<sup>\*</sup> the branched intron region is separated from the catalytic centre by approximately 20 Å, and its position close to the U6 small nuclear RNA ACAGA box is stabilized by interactions with the PRP8 RNase H-like and PRP17 WD40 domains. RNA helicase PRP22 is located about 100 Å from the catalytic centre, suggesting that it destabilizes the spliced mRNA after step two from a distance. Comparison of the structure of the yeast C and human C<sup>\*</sup> complexes reveals numerous RNP rearrangements that are likely to be facilitated by PRP16, including a large-scale movement of the U2 small nuclear RNP.**

The spliceosome catalyses pre-mRNA splicing and is assembled by recruitment of U1 and U2 small nuclear (sn)RNPs to the 5' splice site and branch site, respectively, of the pre-mRNA intron, followed by integration of the preformed U4/U6.U5 tri-snRNP<sup>1</sup>. This generates the spliceosomal B complex, and after extensive RNP rearrangements, the pre-catalytic B<sup>act</sup> complex is formed (Extended Data Fig. 1a). The latter is converted into the catalytically active B<sup>\*</sup> complex by RNA helicase PRP2 (refs 2,3). During activation, a catalytic RNA–RNA network that resembles the catalytic core of group II self-splicing introns<sup>4–6</sup>, as well as the spliceosome's active site, are established. The B<sup>\*</sup> complex catalyses step one of splicing, in which the branch site adenosine (BS-A) carries out a nucleophilic attack at the 5' splice site. This yields the cleaved 5' exon and the lariat–3' exon, in which the 5' end of the intron is covalently attached to the BS-A, forming a branched intron structure. Concomitantly the spliceosomal C complex is formed. The branched intron region must be displaced from the catalytic centre of the spliceosome after step one to allow juxtapositioning of the step-two reactants, the 3'-OH of the 5' exon (the nucleophile for step two) and the 3' splice site. This remodelling is catalysed by RNA helicase PRP16 and leads to the C<sup>\*</sup> complex<sup>7,8</sup>. The latter catalyses step two, during which mRNA and the spliced-out intron are generated. Efficient catalysis of step two requires additional proteins, including SLU7, PRP18 and PRP22 (refs 9–13). However, the precise nature and dynamics of many spliceosome structural rearrangements, especially those occurring just before or during step two, remain unclear.

Cryo-electron microscopy (cryo-EM) of a post-step-two intron–lariat spliceosome (ILS) from *Schizosaccharomyces pombe* provided the first molecular insight into the architecture of the spliceosomal catalytic RNP core<sup>5,14</sup>. Cryo-EM structures revealing the molecular architecture of the *Saccharomyces cerevisiae* and human U4/U6.U5 tri-snRNPs, and more recently of the *S. cerevisiae* B<sup>act</sup> and C spliceosomal complexes, have also been reported<sup>15–22</sup>. However, to our knowledge, no high-resolution structures of human spliceosomal complexes are currently available. Here we report a 3D cryo-EM structure of a human spliceosomal C complex stalled after PRP16 action but before catalytic step two, which

together with protein crosslinking, allows us to determine the spatial organization of its protein and RNA components.

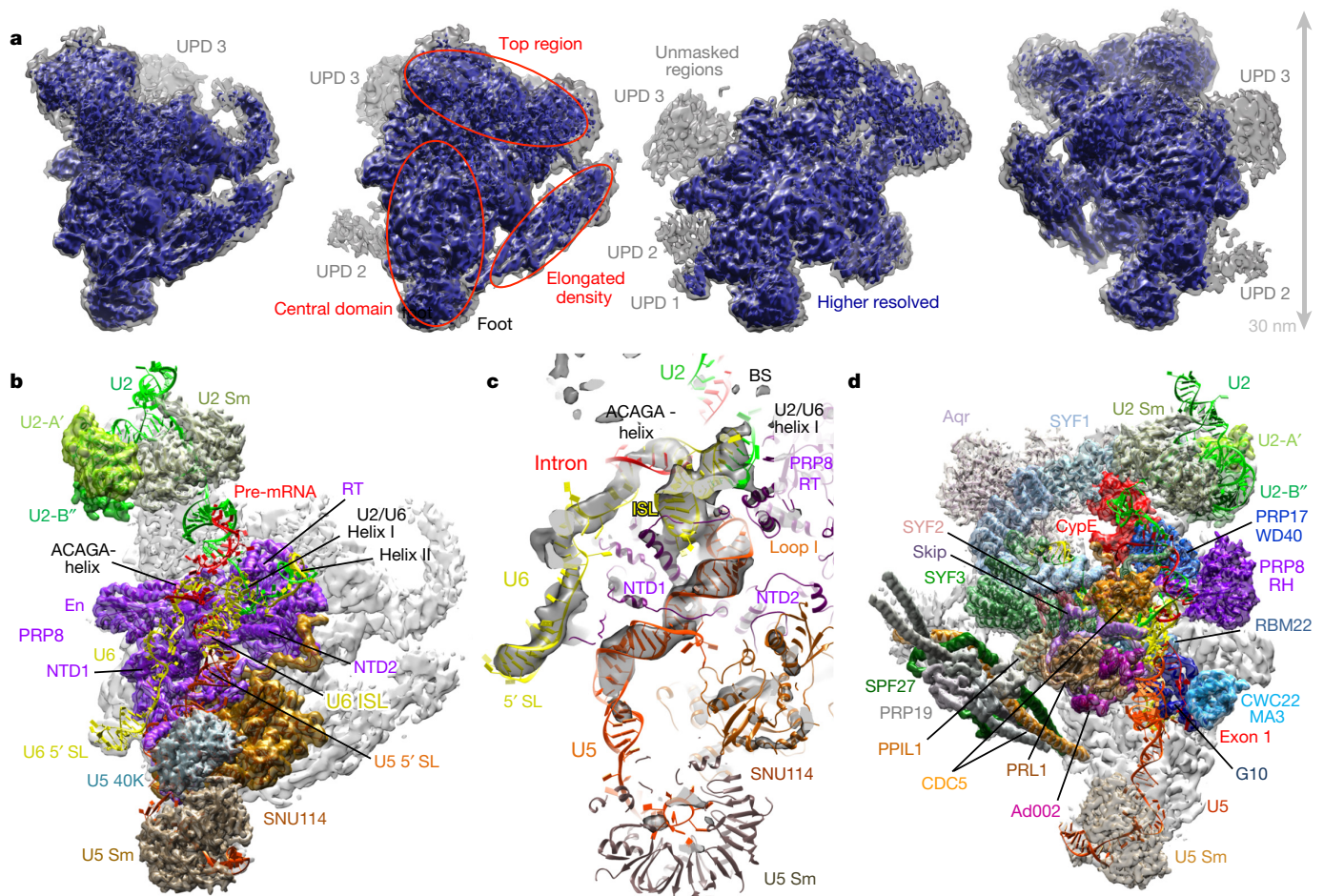
## Overview of the C<sup>\*</sup> spliceosome structure

By lowering the pH of the *in vitro* splicing reaction, we were able to affinity-purify human spliceosomal complexes that were stalled before step two of splicing but after the action of PRP16 (denoted C<sup>\*</sup>) (Extended Data Fig. 1), and shown by chase experiments to be functional (Extended Data Fig. 1). The 3D structure of the human C<sup>\*</sup> complex was determined by cryo-EM, with an overall resolution of 5.9 Å and approximately 4.5 Å in areas of the catalytic RNP core (Extended Data Fig. 2). A pseudo-atomic model was built for the more stable part of C<sup>\*</sup>, where the resolution sufficed for clear identification of structured protein domains and double-stranded RNA elements, allowing us to fit known X-ray structures or homology models of structured regions of C<sup>\*</sup> components into the EM density map (Fig. 1, Supplementary Table 1, Supplementary Video 1). Chemical protein crosslinking coupled to mass spectrometry (CX-MS) (Supplementary Table 2), validated the locations of large proteins and facilitated the docking of smaller ones.

A comparison of the overall structure of the human C<sup>\*</sup> complex with that of the *S. cerevisiae* B<sup>act</sup> or C complex, and *S. pombe* ILS, reveals that the central domains of all four complexes possess similar size (approximately 20 × 12 nm) and morphology<sup>14,19–22</sup>. Like in the yeast spliceosomes, the lower central domain of C<sup>\*</sup> contains U5 snRNP components, that is, the U5 snRNA, PRP8, SNU114, and U5-40K proteins, and the U5 Sm core (Fig. 1b, c). Furthermore, the catalytic U2–U6 RNA structural element of the spliceosome (Fig. 2), together with the U6 ACAGA box/intron helix, is docked in a similar manner into the active site region of PRP8 (ref. 23) (Extended Data Fig. 3). The positions of the U6 5' stem-loop (close to the U5 Sm core) and of U2/U6 helix II (at the upper part of the main body) are also conserved, probably owing to their stabilization by similar sets of proteins (Extended Data Figs 3 and 4). For example, the human G10 protein interacts like its yeast counterpart (Bud31) with U6 snRNA close to the U6 5' stem-loop and anchors it to the PRP8 NTD1 domain (Extended Data Fig. 4).

<sup>1</sup>Department of Structural Dynamics, MPI for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. <sup>2</sup>Department of Cellular Biochemistry, MPI for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. <sup>3</sup>Bioanalytical Mass Spectrometry, MPI for Biophysical Chemistry, Am Fassberg 11, D-37077 Göttingen, Germany. <sup>4</sup>Bioanalytics Group, Institute for Clinical Chemistry, University Medical Centre Göttingen, Robert-Koch-Straße 40, D-37075 Göttingen, Germany.

\*These authors contributed equally to this work.



**Figure 1 | 3D cryo-EM structure of the human C\* complex and overview of the molecular organization of its proteins and RNAs. a**, Different views of the C\* EM density map (rotated around the vertical axis). Better resolved densities are shown in blue. Masked regions not seen in the 5.9 Å structure are grey. UPD 1–3, unmasked peripheral domains

visible in the unmasked C\* EM density map. **b**, Location of U5 and U2 snRNP proteins, and the spliceosomal RNAs. **c**, Regions of the highest density within the central domain, and fit of RNA and protein domains. **d**, Location of the NineTeen complex (NTC) and NTC-related proteins (rotated approximately 180° relative to **b**). Aqr, aquarius.

While individual homologues of the yeast Cwc2 and Ecm2 proteins are absent in humans, the human RBM22 protein shares homology with both yeast Cwc2 and Ecm2, suggesting it is a fused version of both proteins<sup>24</sup>. Indeed, some regions of human RBM22 that are homologous to regions of the *S. cerevisiae* Cwc2 and Ecm2 proteins, are located in similar positions near the U6 internal stem loop (ISL) in the human C\* complex and the *S. cerevisiae* C complex (Extended Data Fig. 4). Thus, spliceosomal protein domains conserved in humans and yeast perform a similar function (that is, stabilizing critical snRNA structures), but are organized in a different manner in higher eukaryotes.

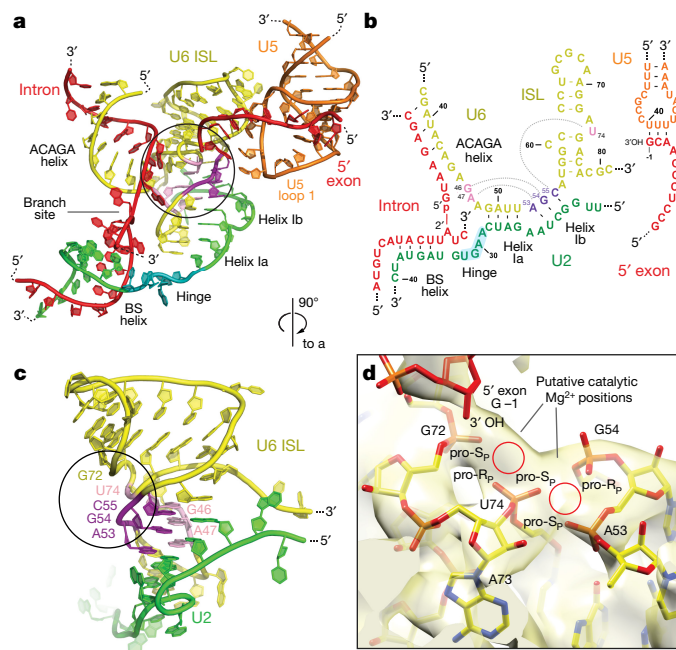
Like their orthologues in the yeast spliceosomes, the human HAT<sup>25</sup> (half a TPR) proteins SYF1 (also known as XAB2) and SYF3 (also known as CRNKL1) form long, curved  $\alpha$ -helical solenoids that cross one another close to U2/U6 helix II (Fig. 1, Extended Data Fig. 5a, b). Moreover,  $\alpha$ -helical elements of human SYF2 (GCIP p29) share an interface with SYF3 and U6 snRNA upstream of U2/U6 helix II (Extended Data Fig. 5). The N-terminal HAT repeats of SYF1 share a large interface with the aquarius protein, located at an equivalent position at the top of the C\* and ILS complexes, and extend to the nearby U2 Sm core domain, with its attached U2-A' and U2-B'' proteins (Fig. 1, Extended Data Fig. 5d, e). SPF27, the C-terminal part of CDC5 and four copies of PRP19 form a helical bundle that fits into the elongated density element at the side of the C\* main body, and is at an equivalent position in the *S. pombe* ILS (Fig. 1, Extended Data Fig. 5a).

### Structure of U2–U6 catalytic RNP core

Consistent with previous biochemical studies<sup>26</sup>, the catalytic centre of the human C\* complex contains a three-helical junction and is highly similar to the catalytic centre of the *S. cerevisiae* C complex. Indeed, the structure of the catalytic U2–U6 RNA of the latter fits well, with some adjustments, into an RNA-shaped element in our C\* model (Figs 1c and 2, Extended Data Fig. 6a, b). The approximately 4.5 Å resolution in this part of the C\* complex allows the unambiguous placement of the loop and kinked stem of the U6 ISL, the two U2/U6 helices Ia and Ib (which are stacked on each other), and the sharp turn separating the ACAGA box helix of U6 snRNA and U2/U6 helix Ia (Fig. 2, Extended Data Fig. 6a, b). The topography of U6 nucleotides U74 and G72, and of A53 and G54, is such that they could potentially coordinate Mg<sup>2+</sup> ions 1 and 2 (which we cannot unambiguously discern at 4.5 Å resolution), respectively. This is consistent with biochemical studies showing metal ion coordination of the corresponding nucleotides in yeast U6 (ref. 27). The similarity between the C and C\* catalytic centres is also underscored by the conservation of protein–RNA interactions in the catalytic RNP core (Extended Data Figs 3b–f and 5b).

Following catalytic step one, the cleaved 5' exon remains tightly bound to the spliceosome. In C\*, the 3'-terminal nucleotides of the 5' exon, including the 3'-OH of the exon (the step two nucleophile), are located close to the catalytic centre, with the 3' hydroxyl close to the putative position of the Mg<sup>2+</sup> ion 1 (Fig. 2, Extended Data Fig. 6f). Exon nucleotides –1 to –3 are base-paired with U5 loop





**Figure 2 | Structure of catalytic core RNA elements in the C\* complex.**

**a, c,** Structure of C\* catalytic RNA network. Black circle, catalytic centre; U6 catalytic triad and catalytic triplex nucleotides, purple and pink, respectively. **b,** Base pairing interactions in the spliceosome's RNA core. Dotted lines, proposed tertiary interactions<sup>4</sup>. **d,** Position of metal-coordinating phosphate groups of U6 snRNA. Fit of nucleotides proposed to coordinate Mg<sup>2+</sup> into the EM density low-pass filtered to 4.5 Å. The metal-coordinating, non-bridging pro-S<sub>p</sub> and pro-R<sub>p</sub> oxygens of the phosphate groups are indicated.

nucleotides U40 to U42 (Extended Data Fig. 6f), consistent with previous studies<sup>28,29</sup>. We can trace an additional nine nucleotides of the exon RNA, which thread through a narrow tunnel between the linker region of the PRP8 reverse transcriptase/endonuclease (RT/En) domain and the PRP8 NTD1 domain, and then pass along the rearranged hairpin loop from the linker regions of PRP8 (the so-called switch loop) and the MA3 domain of CWC22 (Extended Data Fig. 7a, c). The 20 N-terminal amino acids of SRm300 are juxtaposed with the PRP8 switch loop (Extended Data Fig. 7a, c), similar to its probable yeast counterpart, Cwc21, in the C complex<sup>21,22</sup>. Thus, this conserved N-terminal region of SRm300, together with the switch loop, probably plays an evolutionarily conserved role in stabilizing the 5' exon during splicing catalysis.

CWC22 straddles the exon binding channel, with its MA3 domain bound to the PRP8 RT/En domain and its N-terminal MIF4G domain attached to domain 1 of SNU114, as well as to the helicase eIF4AIII (Extended Data Fig. 7). eIF4AIII is located close to the likely position of nucleotides -20 to -25 of the 5' exon (Extended Data Fig. 7b) and is part of the exon junction complex (EJC) that is deposited at a later stage 20–24 nucleotides upstream of the exon–exon junction of spliced mRNAs in higher eukaryotes<sup>30</sup>. While the other EJC proteins are not stably recruited to the C\* complex, our structure nonetheless provides the first direct evidence for the location of EJC proteins in human spliceosomes and how they potentially interact with the exon RNA.

### Repositioning of branched intron region

Directly after step one of splicing, the branched intron structure—that is, the U2/branch site helix in which the BS-A is covalently linked to the 5' end of the intron—remains at the catalytic centre of the C complex, in part stabilized by the Cwc25, Yju2 and Isy1 proteins<sup>21,22</sup>. Consistent with our purified C\* complex representing a post-PRP16 stage, the branched intron structure is no longer positioned in the catalytic centre, but rather the phosphodiester bond between the BS-A and 5' end of the intron is located approximately 20 Å away (Fig. 3a, b).

This displacement distance is highly conserved between group IIB introns and the spliceosome<sup>6</sup>. In the C\* EM density map, a helical density element extends from the U2/branch site helix, indicating formation of an extended, distorted helix between the U2 snRNA and nine intron nucleotides upstream of the branch site (that is, the extended U2/branch site helix) (Fig. 3b). The latter is also formed in the yeast B<sup>act</sup> and C complexes and ILS<sup>14,20,22</sup>, and thus formation of this helix is a general feature, at least after spliceosome activation.

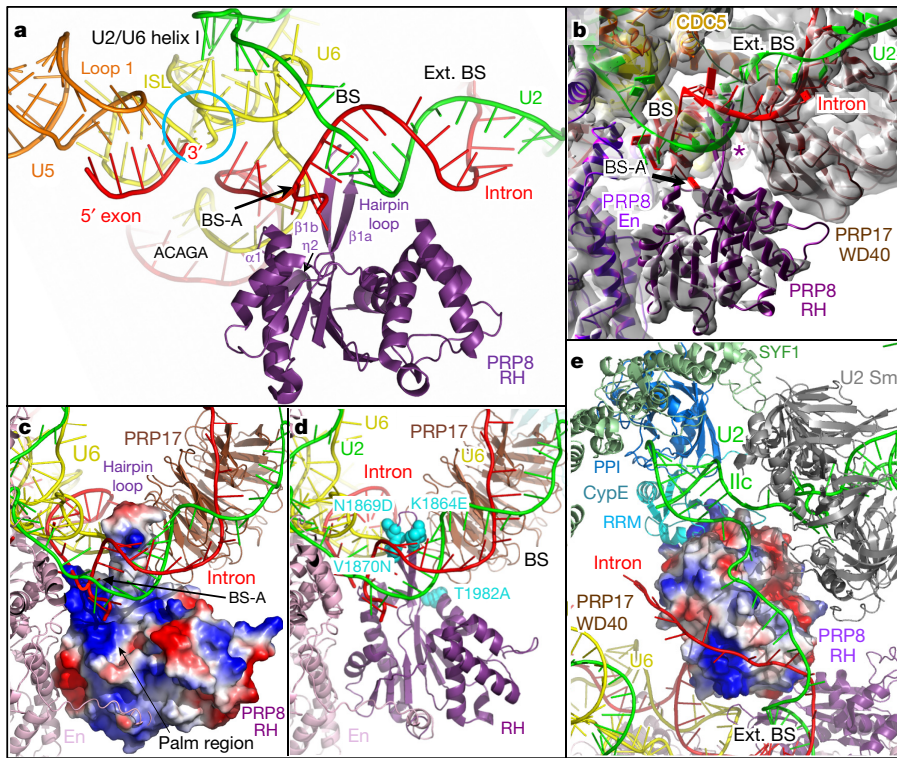
### PRP8 and PRP17 stabilize the branch site position

The PRP8 RNase H-like (RH) domain is located at different positions in the B<sup>act</sup>, C, C\* and ILS complexes (Extended Data Fig. 8). In C\*, the RH domain interacts via its N-terminal  $\alpha$ -helical region with extended  $\alpha$ -helices close to the tip of the En domain of PRP8 (Fig. 3c). At the same time, the RH domain's  $\beta$ -hairpin loop is inserted between the groove of the U2/branch site helix and the U6 ACAGA box/intron helix, contacting nucleotides of the U2 snRNA and the backbone of the U6 ACAGA sequence (Fig. 3a). The branched intron structure rests on a basic region of the RH domain (Fig. 3c). Thus the RH domain appears to play an important part in stabilizing the conformation of the branched intron structure in C\*. Contact between the RH domain and the branched intron region also provides a potential structural basis for the suppression of BS-A mutations by RH domain mutants<sup>31–33</sup> (Fig. 3d). Like the RH domain, the PRP8  $\alpha$ -finger is also found in different positions in the B<sup>act</sup>, C and C\* complexes. In C\* the  $\alpha$ -finger is located in the catalytic centre, suggesting that it may help to position the 3' splice site for step two or, alternatively, that it might have to be repositioned to allow step two catalysis (Extended Data Fig. 8a, b).

In the C\* complex, the WD40 domain of the step-two factor PRP17 lies at a strategically important position between the U2 Sm core and PRP8 RH domain, close to the catalytic centre (Fig. 3c, e). The WD40 domain contacts C-terminal helices and the  $\beta$ -hairpin loop of the PRP8 RH domain (Fig. 3c, d), and the N-terminal region of CDC5 (Extended Data Fig. 9a, b). The PRP17 WD40 domain also contacts the extended U2/branch site helix and U2 nucleotides that connect the extended U2/branch site helix with the putative U2 helix IIc (Fig. 3e, Extended Data Fig. 9c, d). Consistent with its close interaction with U2 snRNA, mutations in the latter are lethal in the absence of Prp17 (ref. 34). As both the RH domain of PRP8 and the U2 Sm core domain undergo PRP16-mediated remodelling (see below), the Prp17 WD40 domain may help stabilize the second-step conformation of the spliceosome, explaining its function as a second-step splicing factor<sup>35</sup>.

### Remodelling during the catalytic phase

Although the structure of the central domain of the spliceosome does not change substantially during the catalytic steps of splicing, the U2 snRNP undergoes large-scale movements. In the C complex, the U2 Sm core, together with its attached Lea1 and Msl1 (human U2-A' and U2-B'') proteins, has rotated and moved approximately 175 Å from its previous position in the B<sup>act</sup> complex (Fig. 4a). This rearrangement requires destabilization of the U2 SF3a and SF3b proteins, which occurs during the Prp2-mediated catalytic activation of the B<sup>act</sup> complex<sup>3,36,37</sup>, and allows the U2/branch site helix to dock in the catalytic centre of the spliceosome. During PRP16-mediated transformation of the C complex into C\*, the branched intron region moves away from the active site and the U2 Sm core moves approximately 100 Å, accompanied by a rotational movement (Fig. 4b). The U2 snRNA contains four nucleotides (A29 to U32 in humans) located between the branch site interacting region and nucleotides that base pair with U6 to form U2/U6 helix Ia (Fig. 2b), which could act as a hinge and allow these movements. Proteins that potentially aid in the repositioning of the U2 Sm core include SYF1 and SYF3. The N-terminal HAT repeats of SYF1 contact the U2 core domain via the U2-A' protein both in C and C\*, and regions of both SYF1 and SYF3 undergo movements in a concerted manner with the U2 Sm core domain (Fig. 4c). The large-scale movements of the U2 Sm core domain may be driven by the repositioning

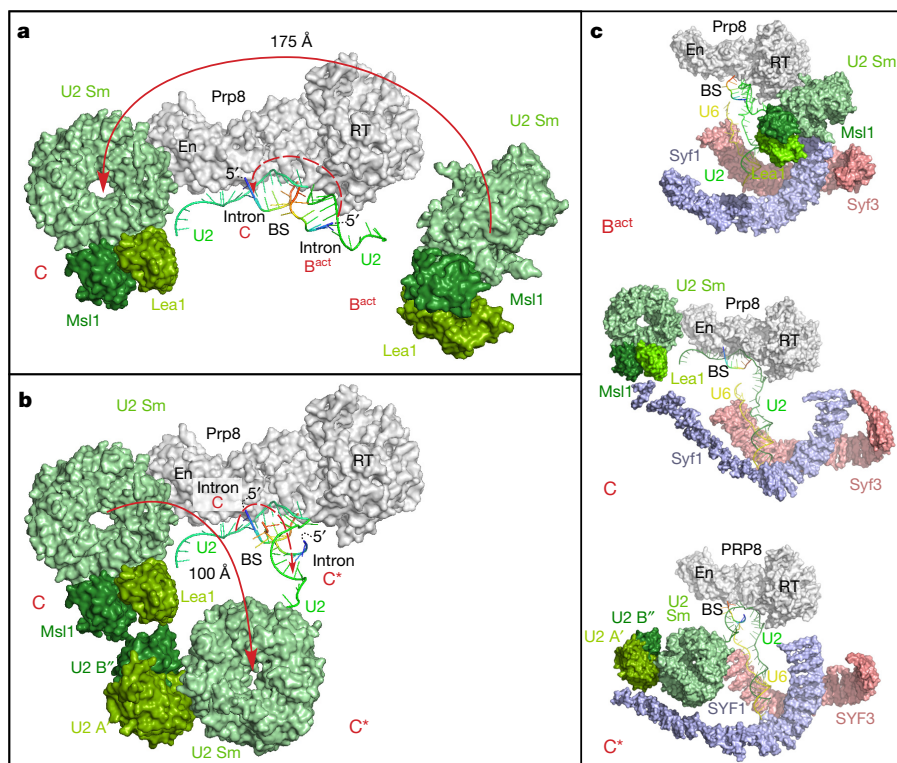


**Figure 3 | Location of the branched intron structure and its interaction with the PRP8 RH and PRP17 WD40 domains.** **a**, Position of branched intron structure and interaction of PRP8 RH with U2/branch site helix. Catalytic centre, blue circle. Ext. BS, extended U2/branch site helix. **b**, Fit of U2/branch site and extended helix. Asterisk,  $\beta$ -hairpin of RH domain. **c**, Electrostatic surface potential of PRP8 RH and interactions with U2/branch site helix and PRP17 WD40. **d**, RH domain residues (blue spheres, numbering according to *S. cerevisiae*) that when mutated suppress BS-A mutations<sup>31–33</sup>. **e**, Electrostatic surface potential of PRP17 WD40 and interactions with U2 snRNA and CypE.

of the U2/branch site helix, as the rotational movements of the latter closely mirror those of the U2 Sm core (Fig. 4a, b), with a similar distance between both elements observed at each stage (Fig. 4c).

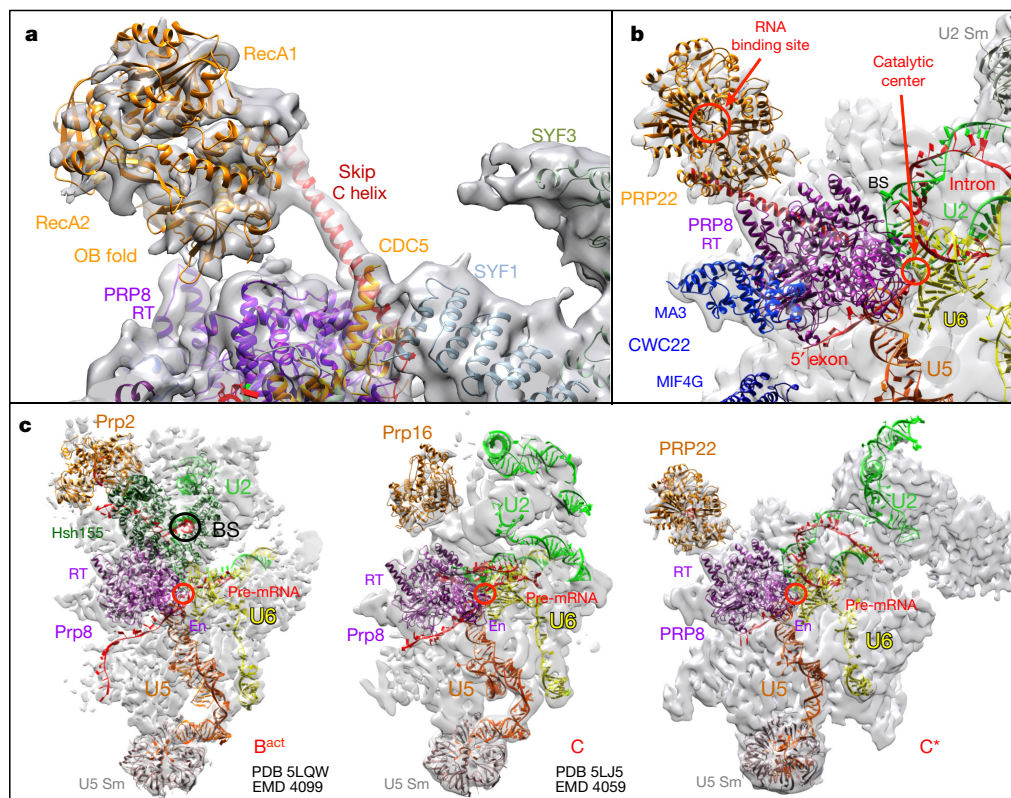
A comparison of the cryo-EM structure of the *S. cerevisiae* C complex (pre-Prp16 state) and the human C\* complex (post-Prp16 state) reveals that an unexpectedly large number of RNA and protein domains are restructured during this PRP16-mediated transition of the spliceosome (Extended Data Fig. 10), assuming that the observed structural differences are not species-specific differences. These include

repositioning of: (1) the extended branched intron structure; (2) the PRP17 WD40 domain; (3) the entire 3' domain of U2 snRNP; (4) SYF1 and SYF3; (5) The RH domain and  $\alpha$ -finger of PRP8; and (6) destabilization of Yju2 (CCDC130 in human) and Cwc25 (CWC25 in human) (refs 8, 38). The high interconnectivity of these restructured domains suggests that their remodelling by PRP16 occurs in a coordinated manner. The PRP16-mediated movement of the branched intron away from the catalytic centre might initially be triggered by PRP16 destabilizing or repositioning one or more proteins (for example,



**Figure 4 | Large-scale movements of the U2 snRNP Sm core domain during the catalytic phase of splicing.** **a**, **b**, Repositioning of the U2 Sm core domain and U2-A' and B'' proteins (*S. cerevisiae* Lea1 and Msl1, respectively) and U2/branch site helix (aligned relative to Prp8 En/RT domain) during the B<sup>act</sup> to C complex transition (**a**) and the C to C\* transition (**b**). Movement of U2 Sm core and U2/branch site helix, solid and dashed red lines, respectively. Intron, indicated by rainbow colouring. **c**, Position of U2 Sm core relative to SYF1 and SYF3 in the B<sup>act</sup>, C, and C\* complexes.





**Figure 5 | The RNA helicase PRP22 is located on the periphery of the human C\* complex, far from its target. a**, Fit of PRP22 in the unmasked, C\* EM density map. **b**, The RNA binding site of PRP22 is approximately 100 Å from the catalytic centre. This distance would accommodate

around 15 to 20 nucleotides of RNA and thus potentially allow PRP22 to interact with 15 nucleotides of the 3' exon near the exon–exon junction<sup>45</sup>. **c**, Position of PRP2, PRP16 and PRP22 in their respective spliceosomal complexes, aligned relative to the PRP8 RT/En domain.

Cwc25/CWC25 and Yju2/CCDC130), as opposed to acting directly on the branched RNA structure. The displacement of spliceosomal proteins probably creates new conformational space and enables the sampling of different RNP conformations within the spliceosome<sup>39</sup>. The resulting rearranged RNA/RNP conformation would then be stabilized by the formation of new protein–protein and protein–RNA interactions (for example, interaction of the branched intron with the PRP8 RH domain). A similar mechanistic principle can be envisioned for the RNP rearrangements in the spliceosome that are mediated by other spliceosomal helicases, such as PRP2 (ref. 19).

### Insights into the role of helicase PRP22

We do not find density to accommodate nucleotides containing the 3' splice site, suggesting that it is flexible. For efficient positioning of the 3' splice site in the catalytic centre, the second step factors SLU7, PRP18 and PRP22 are required (reviewed in ref. 40). While SLU7 and PRP18 are missing from our C\* complex, PRP22 is stoichiometrically present. The helicase domain of PRP22 fits into UPD3 of the C\* unmasked EM density (Fig. 1), where it is attached via its C-terminal oligonucleotide/oligosaccharide-binding (OB) fold domain to the long  $\alpha$ -helix of Skip and to the PRP8 RT domain (Fig. 5a), consistent with earlier biochemical and genetic work<sup>41–43</sup>. As PRP22 contributes to step two of splicing in an ATP-independent manner<sup>11</sup> and is located far from the C\* catalytic centre (Fig. 5b), it probably plays an indirect role in positioning the 3' splice site for catalysis of step two. Following catalytic step two, PRP22 facilitates the release of the spliced mRNA in an ATP-dependent manner<sup>44</sup>. To achieve this, PRP22 may have to undergo a conformational change in order to bind the spliced 3' exon, which is required for its activity<sup>45</sup>. As CWC22 is located on both sides of the exon-binding channel, its removal may be a prerequisite for release of the spliced mRNA from the latter. Indeed, during mRNA release from the post-step two spliceosome, there is nearly quantitative displacement

of CWC22 (ref. 46). PRP22 is bound close to the region of the PRP8 RT domain where the MA3 domain of CWC22 is located (Fig. 5b), and thus its action might potentially lead to displacement of CWC22. PRP2, PRP16 and PRP22, which bind transiently and in a sequential manner, are all located at the periphery of their respective spliceosomal complexes, at a distance from their presumed targets, which are located at or near the catalytic centre (Fig. 5c). Thus, all three helicases appear to act indirectly, at a distance, as recently proposed for PRP2 (ref. 19) and PRP16 (ref. 47).

A comparison of the cryo-EM structures of the human C\* and *S. pombe* ILS indicates that the rearrangements accompanying this transition are more subtle than those occurring during the C to C\* transition. However, determination of the structures of the human B<sup>act</sup>, C and ILS complexes will be needed to clarify whether structural changes proposed to occur during splicing catalysis are indeed evolutionarily conserved.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 18 November 2016; accepted 4 January 2017.**

**Published online 11 January 2017.**

1. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
2. Kim, S. H. & Lin, R. J. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol. Cell. Biol.* **16**, 6810–6819 (1996).
3. Warkocki, Z. *et al.* Reconstitution of both steps of *Saccharomyces cerevisiae* splicing with purified spliceosomal components. *Nat. Struct. Mol. Biol.* **16**, 1237–1243 (2009).
4. Fica, S. M., Mefford, M. A., Piccirilli, J. A. & Staley, J. P. Evidence for a group II intron-like catalytic triplex in the spliceosome. *Nat. Struct. Mol. Biol.* **21**, 464–471 (2014).



5. Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349**, 1191–1198 (2015).
6. Robart, A. R., Chan, R. T., Peters, J. K., Rajashankar, K. R. & Toor, N. Crystal structure of a eukaryotic group II intron lariat. *Nature* **514**, 193–197 (2014).
7. Schwer, B. & Guthrie, C. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature* **349**, 494–499 (1991).
8. Ohrt, T. *et al.* Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *RNA* **19**, 902–915 (2013).
9. Frank, D. & Guthrie, C. An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev.* **6**, 2112–2124 (1992).
10. Horowitz, D. S. & Abelson, J. A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of pre-mRNA splicing in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **13**, 2959–2970 (1993).
11. Schwer, B. & Gross, C. H. Prp22, a DEXH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J.* **17**, 2086–2094 (1998).
12. Chua, K. & Reed, R. Human step II splicing factor hSlu7 functions in restructuring the spliceosome between the catalytic steps of splicing. *Genes Dev.* **13**, 841–850 (1999).
13. Horowitz, D. S. & Krainer, A. R. A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev.* **11**, 139–151 (1997).
14. Yan, C. *et al.* Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).
15. Nguyen, T. H. *et al.* The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52 (2015).
16. Nguyen, T. H. *et al.* Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302 (2016).
17. Wan, R. *et al.* The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**, 466–475 (2016).
18. Agafonov, D. E. *et al.* Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* **351**, 1416–1420 (2016).
19. Rauhut, R. *et al.* Molecular architecture of the *Saccharomyces cerevisiae* activated spliceosome. *Science* **353**, 1399–1405 (2016).
20. Yan, C., Wan, R., Bai, R., Huang, G. & Shi, Y. Structure of a yeast activated spliceosome at 3.5 Å resolution. *Science* **353**, 904–911 (2016).
21. Wan, R., Yan, C., Bai, R., Huang, G. & Shi, Y. Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904 (2016).
22. Galej, W. P. *et al.* Cryo-EM structure of the spliceosome immediately after branching. *Nature* **537**, 197–201 (2016).
23. Galej, W. P., Oubridge, C., Newman, A. J. & Nagai, K. Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643 (2013).
24. Rasche, N. *et al.* Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J.* **31**, 1591–1604 (2012).
25. Preker, P. J. & Keller, W. The HAT helix, a repetitive motif implicated in RNA processing. *Trends Biochem. Sci.* **23**, 15–16 (1998).
26. Anokhina, M. *et al.* RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* **32**, 2804–2818 (2013).
27. Fica, S. M. *et al.* RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234 (2013).
28. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* **68**, 743–754 (1992).
29. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
30. Le Hir, H., Izaurralde, E., Maquat, L. E. & Moore, M. J. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19**, 6860–6869 (2000).
31. Query, C. C. & Konarska, M. M. Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants. *Mol. Cell* **14**, 343–354 (2004).
32. Siatecka, M., Reyes, J. L. & Konarska, M. M. Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev.* **13**, 1983–1993 (1999).
33. Collins, C. A. & Guthrie, C. Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. *Genes Dev.* **13**, 1970–1982 (1999).
34. Ben-Yehuda, S. *et al.* Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in *Saccharomyces cerevisiae*. *Genetics* **156**, 1503–1517 (2000).
35. Jones, M. H., Frank, D. N. & Guthrie, C. Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc. Natl Acad. Sci. USA* **92**, 9687–9691 (1995).
36. Lardelli, R. M., Thompson, J. X., Yates, J. R., III & Stevens, S. W. Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *RNA* **16**, 516–528 (2010).
37. Ohrt, T. *et al.* Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *RNA* **18**, 1244–1256 (2012).
38. Tseng, C. K., Liu, H. L. & Cheng, S. C. DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *RNA* **17**, 145–154 (2011).
39. Konarska, M. M. & Query, C. C. Insights into the mechanisms of splicing: more lessons from the ribosome. *Genes Dev.* **19**, 2255–2260 (2005).
40. Horowitz, D. S. The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* **3**, 331–350 (2012).
41. Schneider, S. & Schwer, B. Functional domains of the yeast splicing factor Prp22p. *J. Biol. Chem.* **276**, 21184–21191 (2001).
42. Albers, M., Diment, A., Muraru, M., Russell, C. S. & Beggs, J. D. Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *RNA* **9**, 138–150 (2003).
43. Gahura, O. *et al.* Prp45 affects Prp22 partition in spliceosomal complexes and splicing efficiency of non-consensus substrates. *J. Cell. Biochem.* **106**, 139–151 (2009).
44. Company, M., Arenas, J. & Abelson, J. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* **349**, 487–493 (1991).
45. Schwer, B. A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release. *Mol. Cell* **30**, 743–754 (2008).
46. Fourmann, J. B. *et al.* Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev.* **27**, 413–428 (2013).
47. Semlow, D. R., Blanco, M. R., Walter, N. G. & Staley, J. P. Spliceosomal DEAH-box ATPases remodel pre-mRNA to activate alternative splice sites. *Cell* **164**, 985–998 (2016).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank T. Conrad for HeLa cell production, H. Kohansal for preparing HeLa cell nuclear extract, and U. Steuerwald, W. Lendeckel, I. Öchsner, M. Raabe and U. Pleßmann for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft (SFB 860) to R.L., H.S. and H.U.

**Author Contributions** D.E.A. developed the purification strategy and characterized the C\* complex. O.D. and H.U. analysed protein–protein crosslinking data. K.B., W.-T.L. and H.S. prepared grids and collected EM data. B.K., K.B. and K.H. carried out model building and refinement. K.B., B.K., K.H., H.S. and R.L. analysed the structure. All authors were involved in data interpretation. The manuscript was written by R.L. and C.L.W., with input from all authors. R.L., H.S., B.K. and H.U. initiated and orchestrated the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.K. ([b.kastner@mpi-bpc.mpg.de](mailto:b.kastner@mpi-bpc.mpg.de)), H.S. ([hstark1@gwdg.de](mailto:hstark1@gwdg.de)) or R.L. ([reinhard.luehrmann@mpi-bpc.mpg.de](mailto:reinhard.luehrmann@mpi-bpc.mpg.de)).

**Reviewer Information** *Nature* thanks Y. Cheng, N. Toor and J. Valcarcel for their contribution to the peer review of this work.

## METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**In vitro splicing.** Uniformly [ $^{32}\text{P}$ ]-labelled,  $m^7\text{G}(5')\text{ppp}(5')\text{G}$ -capped MINX pre-mRNA was synthesized *in vitro* by T7 runoff transcription. HeLa S3 cells were obtained from GBF, Braunschweig (currently Helmholtz Zentrum für Infektionsforschung, Braunschweig) and tested negative for mycoplasma. HeLa nuclear extract was prepared essentially as previously described<sup>48</sup>, but without the final dialysis step. To isolate C\* complexes, splicing was performed with 5 nM of  $^{32}\text{P}$ -labelled pre-mRNA and 20% (v/v) HeLa nuclear extract, in buffer containing 3 mM  $\text{MgCl}_2$ , 50 mM NaCl, 4 mM HEPES-KOH pH 7.9, 12 mM MES-NaOH pH 6.4, 2 mM ATP and 20 mM creatine phosphate, and was incubated at 30 °C for different time periods. Samples were analysed on a denaturing 4–12% NuPAGE gel (Life Technologies) and pre-mRNA and splicing intermediates and products were visualized with a Typhoon phosphorimager (GE Healthcare).

**MS2 affinity selection of splicing complexes.** Spliceosomal complexes were isolated by MS2 affinity selection. In previous studies, affinity purified human C complexes were formed *in vitro* on mutated pre-mRNA substrates that are unable to undergo the second step of splicing<sup>49–51</sup>. Here, we used the MINX pre-mRNA substrate<sup>52</sup> that contains an intron flanked by a 5' and 3' exon, allowing a functional analysis of complexes that assemble on it. In brief, MINX pre-mRNA containing three MS2 aptamers at its 3' end RNA was incubated with a tenfold molar excess of MS2-MBP fusion protein and then added to a splicing reaction. After incubating at 30 °C for 5 h and centrifuging to remove aggregates, the reaction was loaded onto a 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  $\text{MgCl}_2$ , 150 mM NaCl) and complexes were eluted with G-150 buffer containing 1 mM maltose. Eluted complexes were loaded onto a 36 ml linear 10–30% (v/v) glycerol gradient containing G-150 buffer (20 mM HEPES-KOH pH 7.9, 1.5 mM  $\text{MgCl}_2$ , 150 mM NaCl), centrifuged at 23,000 r.p.m. for 15 h at 4 °C in a Surespin 630 (Thermo Scientific) rotor and fractions were harvested from the bottom. The distribution of  $^{32}\text{P}$ -labelled MINX RNA across the gradient was determined by Cherenkov counting. Fractions were analysed by denaturing 4–12% NuPAGE (Life Technologies) followed by autoradiography. Peak fractions containing the first step splicing intermediates were pooled, concentrated by centrifugation with an Amicon 50 kD cut-off unit, diluted to decrease the glycerol concentration and reloaded on the same gradients with glutaraldehyde as fixative<sup>53</sup>. For biochemical sample validation, the same procedure was performed but without fixation in the second gradient. The 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 in ref. 54, using a 7% acrylamide mono gel in the second dimension for analysis of proteins larger than 25 kDa, or 15% acrylamide for proteins smaller than 25 kDa. For mass spectrometry, coomassie-stained protein-spots were cut out of the 1D or 2D gels, and proteins were digested in-gel with trypsin and extracted. The extracted peptides were analysed in a liquid-chromatography coupled electrospray ionization mass spectrometer (LTQ Orbitrap XL) under standard conditions. Proteins were identified by searching fragment spectra against the NCBI non-redundant (nr) database using Mascot as a search engine.

**Chase of purified C\* complexes with nuclear extract.** Total IgGs against human PRP16 were purified as described previously<sup>55</sup>. Affinity-purified C\* complexes formed on  $^{32}\text{P}$ -labelled MINX-MS2 pre-mRNA were incubated with splicing buffer alone (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 3 mM  $\text{MgCl}_2$ , 2 mM ATP, 20 mM creatine phosphate) or additionally in the presence of 20% HeLa nuclear extract prepared according to ref. 48. For antibody inhibition experiments, the splicing mixture was pre-incubated with  $3 \mu\text{g} \mu\text{l}^{-1}$  of anti-PRP16 antibody at 30 °C for 15 min as described<sup>55</sup>. The splicing reaction was initiated by addition of  $^{32}\text{P}$ -labelled MINX-MS2 pre-mRNA or C\* complex assembled on  $^{32}\text{P}$ -labelled MINX-MS2 pre-mRNA. The reaction was incubated at 30 °C for 0–60 min. RNA was recovered, separated by SDS-PAGE, and visualized with a Typhoon phosphorimager (GE Healthcare).

**Crosslinking of the C\* complex and crosslink identification.** Following MS2 affinity selection and the first density gradient centrifugation step, purified spliceosomal complexes were crosslinked with 150  $\mu\text{M}$  BS3 for 40 min at 20 °C and purified further by a second density gradient centrifugation step. Approximately 25 pmol of C\* complexes were pelleted by ultracentrifugation and analysed essentially as described before<sup>56</sup> with the following modifications: precipitated material was dissolved in a solution containing 4 M urea and 50 mM ammonium bicarbonate, reduced with DTT, alkylated with iodoacetamide, diluted to 1 M

urea and digested with trypsin (1:20 w/w). Peptides were reverse-phase extracted and fractionated by gel filtration on a Superdex Peptide PC3.2/30 column (GE Healthcare). 50- $\mu\text{l}$  fractions corresponding to an elution volume of 1.2–1.8 ml were analysed in quadruplicate on a Thermo Scientific Q Exactive HF mass spectrometer. Protein–protein crosslinks were identified by pLink 1.23 search engine (<http://pfind.ict.ac.cn/software/pLink>) and filtered at FDR 1% according to the recommendations of the developer<sup>57</sup>. For simplicity, the crosslink score is represented as a negative value of the common logarithm of the original pLink score, that is  $\text{Score} = -\log_{10}(\text{pLink Score})$ . For model building, a maximum distance of 30 Å between the  $\text{C}\alpha$  atoms of the crosslinked lysines was allowed. Approximately 97% of all crosslink-assigned spectra correspond to crosslink distances of 30 Å or less.

**EM and image processing.** Purified spliceosomes, stabilized by GraFix<sup>53</sup>, were allowed to adsorb on a thin carbon film before negative staining or rapid plunge freezing into liquid ethane at 100% humidity and 4 °C. Micrographs of negatively stained particles were recorded in a CM200 electron microscope (FEI/Phillips, the Netherlands) at room temperature and approximately 50,000 particles were picked by hand. The latter were then used to *de novo* build a negative stain starting structure that was refined to around 25 Å by using 3D maximum-likelihood alignment and 3D classification<sup>58</sup>. The resulting model was subsequently used as an initial reference in cryo-particle image processing and classification. Cryo-images were recorded at  $-193 \text{ }^\circ\text{C}$  in a Titan Krios electron microscope (FEI Company, The Netherlands) on a Falcon II direct electron detector at a nominal 88,000 $\times$  magnification resulting in a calibrated pixel size of 1.59 Å on the specimen level. Seventeen frames were recorded for each micrograph with an average dose of 2.1  $\text{e}^-$  per frame per Å<sup>2</sup>. Motion correction and spatial frequency weighed frame summation was achieved using the unblur software suite<sup>59</sup> (<http://grigoriefflab.janelia.org/unblur>). Summed micrograph images were then evaluated based on CTF parameters and only those revealing isotropic Thon rings were used for particle picking and extraction. Using the particle picking software Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) and 40° projections of the negative stain model filtered to 40 Å as a reference, we extracted approximately 2.5 million particle images from the pre-sorted cryo-micrographs and applied several particle sorting steps at the 2D and 3D level. 2D multivariate statistics and classification was first applied to the non-aligned particle images and subsequently to the aligned particles. In each round, particles contributing to bad classes were excluded from further processing. The remaining  $\sim 1,708,000$  particles were split into seven equally sized groups and subsequently applied to seven separate rounds of 3D classification in RELION featuring six classes each. The  $\sim 393,000$  particles pooled from all satisfactory classes were then subjected to further rounds of 3D classification in RELION<sup>71</sup>. For the high resolution structure determination, the 136,534 particles finally contributing to the best 3D class were used for refinement revealing an 8.4 Å resolution structure (which we refer to as the unmasked EM density map). Roughly 20% of the spliceosome density was not clearly defined at this level of resolution. As these densities largely disappear during the higher-resolution structure calculations, we excluded them with a mask in the final rounds of the refinement. A soft mask with a cut-off of 7 voxels was used for the refinement and for the determination of resolution. We obtained the final map with a resolution of 5.9 Å as determined by Fourier shell correlation calculated from two independent data sets with a threshold of 0.143. A local resolution plot revealed that there are areas of higher resolution in the catalytic RNP core of the C\* complex that approach 4.5 Å. Some peripheral regions have somewhat lower resolution (Fig. 1 and Extended Data Fig. 2c). More dynamic areas or those with components with non-stoichiometric occupancies are not visible in the 5.9 Å structure.

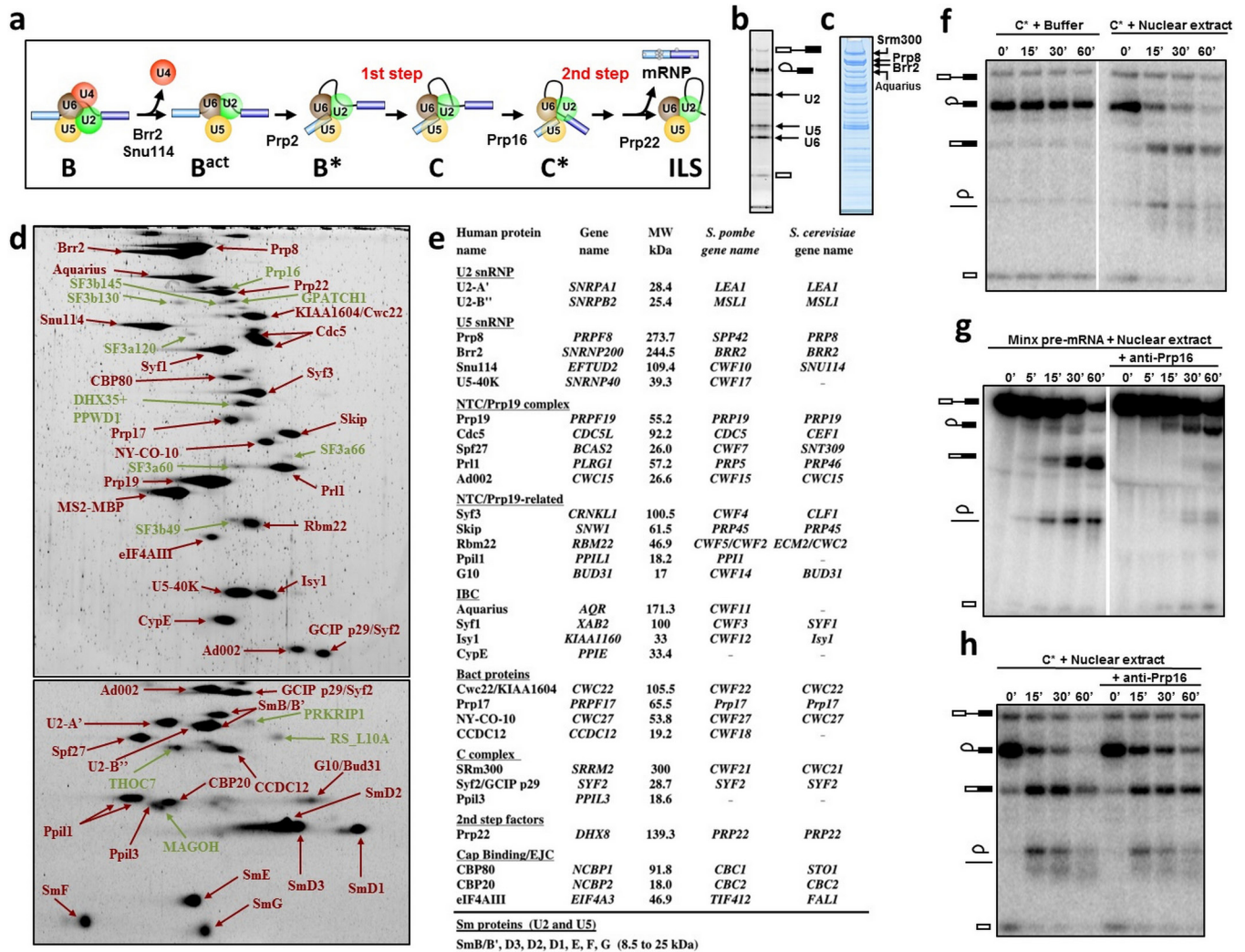
**Model fitting and building.** Available X-ray or homology models of proteins were fit into the EM density by Chimera<sup>60</sup>. Individual models of substructures (for example, domains or structural motifs) were further fitted as rigid bodies by Coot<sup>61</sup>. After visual inspection, the models were adjusted manually in the density, the disordered regions were removed and regions that were reorganized or were not present in the initial models (for example, loops and various elements of secondary structure) were built in Coot. Homology models of proteins were either obtained by using the respective functions of the SWISS-MODEL suite<sup>62</sup> or directly adapted from the SpliProt3D database<sup>63</sup>. Initial human snRNA and intron models were obtained by using the *S. cerevisiae* C complex spliceosomal RNA<sup>14,21</sup> as reference, and homology modelled according to the human sequence using the ModeRNA package<sup>64</sup> (<http://genesilico.pl/moderna/>). The snRNA fit was improved at rigid body level in Chimera and subsequently adjusted in Coot. The 5' stem loop of U6 and the U2/U6 helix II were generated by rigid-body fitting of idealized double-stranded RNA helices. Exon 1 was modelled *ab initio* into available density in Coot and the resulting model later verified by comparison to the *S. cerevisiae* C spliceosome model<sup>22</sup>. Once the entire coordinate model was built up, all proteins were truncated to poly-alanine level and a global minimization real space refinement was conducted against the 5.9 Å cryo-EM density using the real space refine

program from the PHENIX suite<sup>65</sup>. The RNA model was subsequently validated by the MolProbity server<sup>66</sup> and had an all atom clash score of less than 12 and no bad bond lengths or angles. Final visualization was carried out with Chimera and PyMOL (<http://www.pymol.org>). See also Supplementary Table 1 for protein and model building information for all modelled human C\* proteins.

**Data Availability.** All data generated or analysed during this study are included in this published article and its Supplementary Information files. The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMD-3547 (8.4 Å map) and EMD-3545 (5.9 Å map). The atomic model has been deposited in the Protein Data Bank under accession code 5MQF.

48. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
49. Bessonov, S., Anokhina, M., Will, C. L., Urlaub, H. & Lührmann, R. Isolation of an active step I spliceosome and composition of its RNP core. *Nature* **452**, 846–850 (2008).
50. Jurica, M. S., Sousa, D., Moore, M. J. & Grigorieff, N. Three-dimensional structure of C complex spliceosomes by electron microscopy. *Nat. Struct. Mol. Biol.* **11**, 265–269 (2004).
51. Golas, M. M. *et al.* 3D cryo-EM structure of an active step I spliceosome and localization of its catalytic core. *Mol. Cell* **40**, 927–938 (2010).
52. Zillmann, M., Zapp, M. L. & Berget, S. M. Gel electrophoretic isolation of splicing complexes containing U1 small nuclear ribonucleoprotein particles. *Mol. Cell. Biol.* **8**, 814–821 (1988).
53. Kastner, B. *et al.* GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat. Methods* **5**, 53–55 (2008).
54. Agafonov, D. E. *et al.* Semiquantitative proteomic analysis of the human spliceosome via a novel two-dimensional gel electrophoresis method. *Mol. Cell. Biol.* **31**, 2667–2682 (2011).
55. Ortlepp, D. *et al.* The mammalian homologue of Prp16p is overexpressed in a cell line tolerant to Leflunomide, a new immunoregulatory drug effective against rheumatoid arthritis. *RNA* **4**, 1007–1018 (1998).
56. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical crosslinking of protein complexes and identification of crosslinking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protocols* **9**, 120–137 (2014).
57. Yang, B. *et al.* Identification of crosslinked peptides from complex samples. *Nat. Methods* **9**, 904–906 (2012).
58. Sander, B. *et al.* Organization of core spliceosomal components U5 snRNA loop I and U4/U6 Di-snRNP within U4/U6.U5 Tri-snRNP as revealed by electron cryomicroscopy. *Mol. Cell* **24**, 267–278 (2006).
59. Campbell, M. G. *et al.* Movies of ice-embedded particles enhance resolution in electron cryo-microscopy. *Structure* **20**, 1823–1828 (2012).
60. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
61. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
62. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
63. Korneta, I., Magnus, M. & Bujnicki, J. M. Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* **40**, 7046–7065 (2012).
64. Rother, M., Rother, K., Puton, T. & Bujnicki, J. M. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res.* **39**, 4007–4022 (2011).
65. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
66. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).
67. Xu, C. *et al.* Solution structure of human peptidyl prolyl isomerase-like protein 1 and insights into its interaction with SKIP. *J. Biol. Chem.* **281**, 15900–15908 (2006).
68. Wang, X. *et al.* A large intrinsically disordered region in SKIP and its disorder-order transition induced by PPIL1 binding revealed by NMR. *J. Biol. Chem.* **285**, 4951–4963 (2010).
69. Stegmann, C. M., Lührmann, R. & Wahl, M. C. The crystal structure of PPIL1 bound to cyclosporine A suggests a binding mode for a linear epitope of the SKIP protein. *PLoS One* **5**, e10013 (2010).
70. Perriman, R. J. & Ares, M., Jr. Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev.* **21**, 811–820 (2007).
71. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).

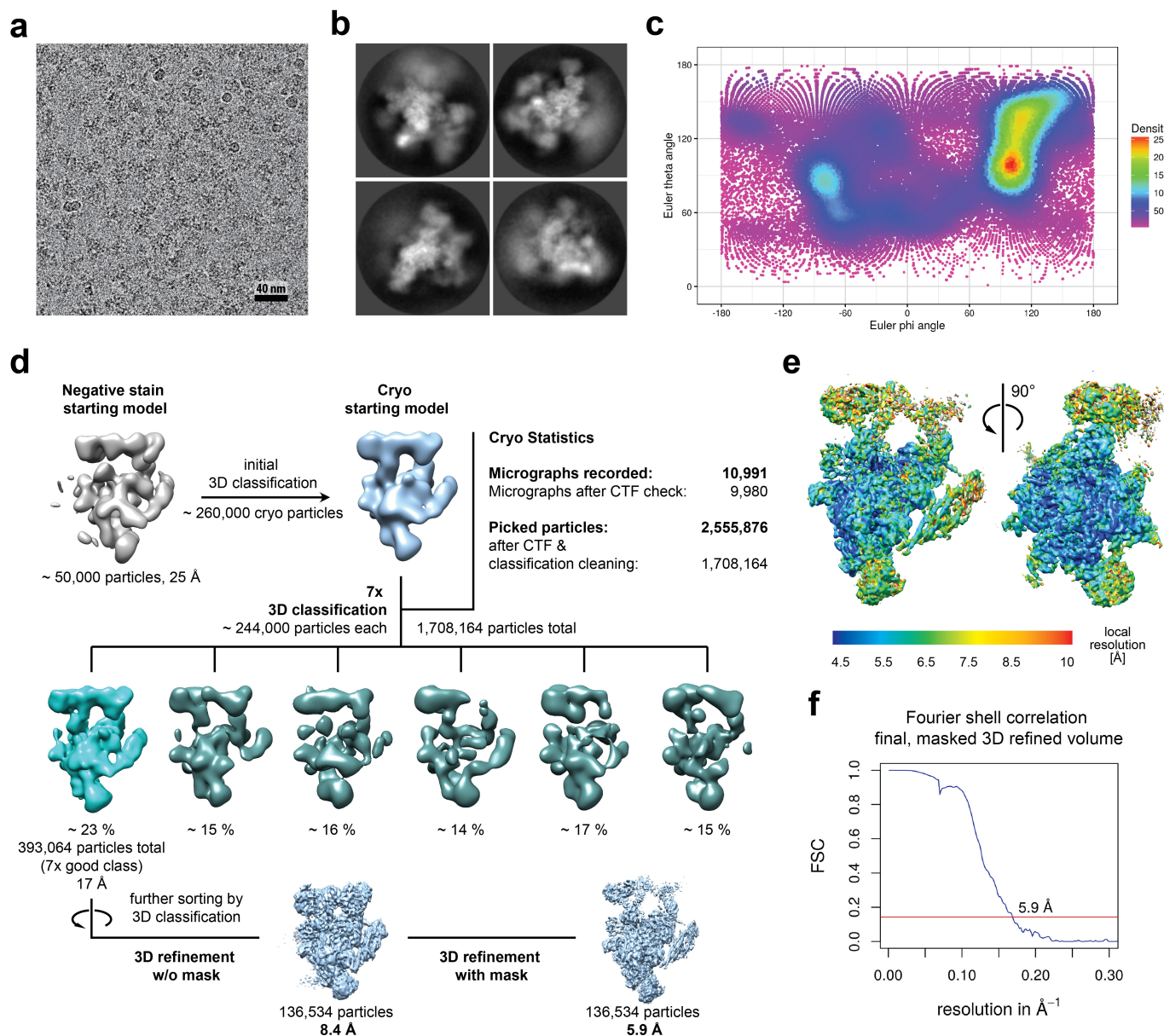




**Extended Data Figure 1 | Compositional and functional characterization of affinity-purified, human C\* complexes.**

**a**, Spliceosomal assembly pathway, starting with complex B. **b**, **c**, RNA (b) and protein (c) composition of MS2 affinity-purified, human C\* complexes. Only high molecular weight proteins are labelled. C\* complexes contained predominantly the first-step splicing intermediates (cleaved 5' exon and the lariat-3'-exon), and U2, U5 and U6 snRNA. **d**, **e**, Abundant C\* proteins (red) identified by 2D gel electrophoresis, summarized in **e**. Less abundant proteins, green. Upper and lower panel: proteins >25 kDa or <25 kDa, respectively. Proteins present in substoichiometric amounts in previous human C complex preparations (for example, SF3a and SF3b proteins)<sup>49,51</sup> were lacking or present only in very low amounts. SRm300 does not migrate well into the second dimension of the 2D gel, but is clearly observed on a 1D gel. The step-two factors SLU7 and PRP18 were not detected, which may explain why spliceosomes are stalled after step one at pH 6.4. All abundant proteins or domains/regions thereof were modelled into the C\* EM density

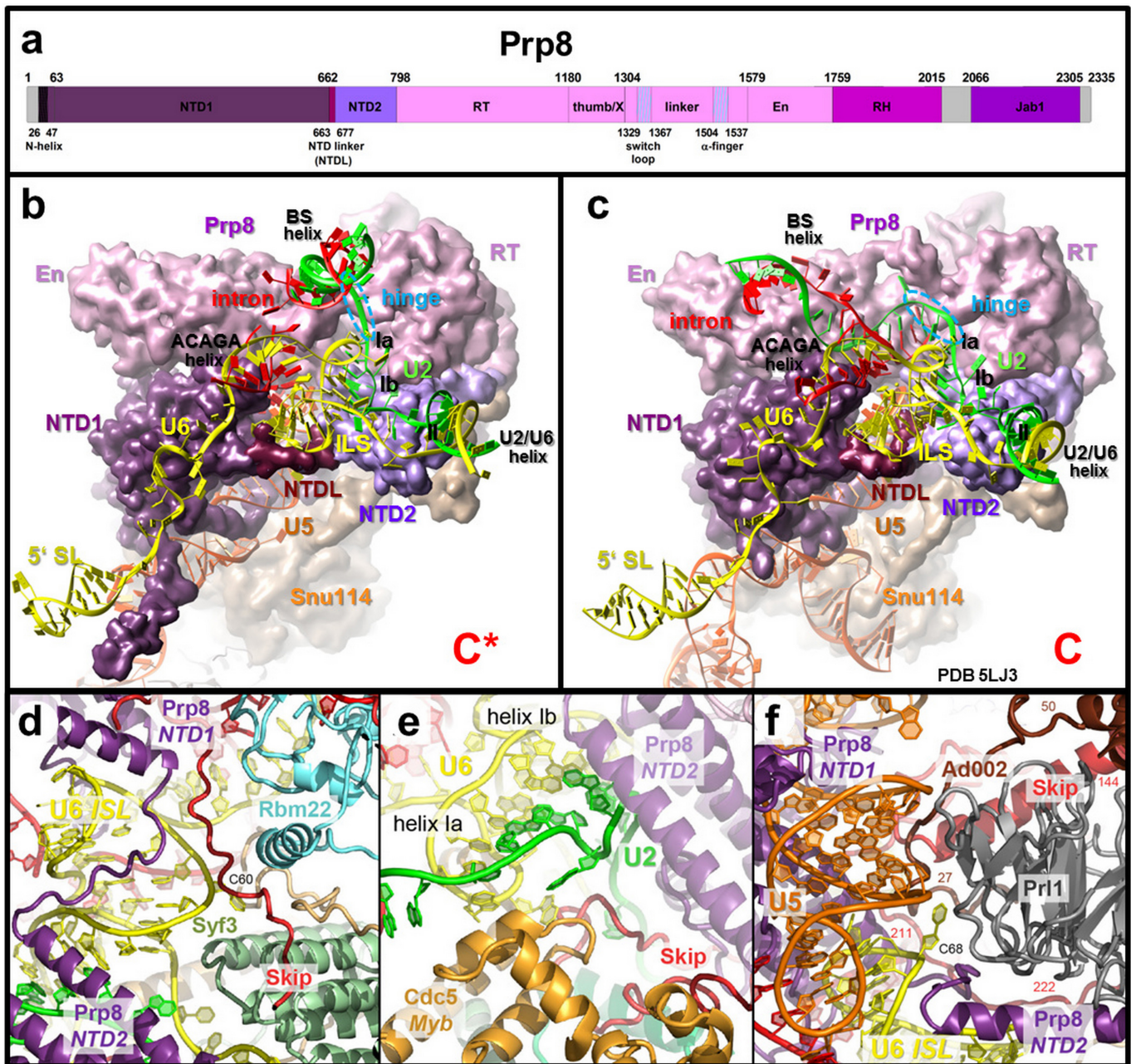
map, except BRR2, ISY1, NY-CO-10, CCD12, PPIL3, CBP80 and CBP20 (Supplementary Table 1). As BRR2 could not be located in the C\* EM density map, it is probably highly flexible in C\*. **f**, C\* complexes formed on <sup>32</sup>P-labelled MINX pre-mRNA at pH 6.4 were purified and incubated in HeLa nuclear extract at pH 7.9 or in buffer under splicing conditions. Splicing intermediates were efficiently chased into splicing products after 15 min in the presence of nuclear extract, but not buffer alone, indicating that our purified C\* complexes are functional and not dead-end complexes. **g**, **h**, MINX pre-mRNA (**g**) or affinity-purified C\* complexes (**h**) were incubated with HeLa nuclear extract with or without anti-PRP16 antibodies. Efficient catalysis of step two of splicing was observed if purified C\* complexes, but not pre-mRNA alone, were incubated in nuclear extract pre-incubated with PRP16 antibodies, indicating that the C\* complexes are indeed stalled after the action of PRP16. All 2D analyses and *in vitro* splicing experiments were performed at least twice in two independent experiments.



**Extended Data Figure 2 | Cryo-EM and image processing of the human C\* complex.** **a**, Typical cryo-EM raw image of human C\* spliceosomes recorded with a Titan Krios (FEI Company) electron microscope at a nominal magnification of 88,000 $\times$  with a Falcon II direct electron detector resulting in a pixel size of 1.59 Å per pixel. **b**, Several representative class averages showing different views of the C\* particle after 2D classification. **c**, Euler angle distribution of all particle images that contributed to the final 3D map. The coordinates describe the  $\phi$  and  $\theta$  angles. Size and colour of the plotted dots indicate the number of particles at any given Euler angle. Although several sets of particle orientations dominated, an almost complete angular coverage was obtained. **d**, Computational sorting scheme. Imaged micrographs were first evaluated and sorted according to the quality of their Thon rings in local power spectra. Roughly 2.5 million particle images were then selected from the remaining micrographs. In a second sorting step, particle images were again discarded based on the quality of Thon rings in classified, local power spectra. Following evaluations in Fourier space, particles were then excluded according to multiple rounds of 2D classifications. The remaining 1,708,164 particle images were split into seven subsets of approximately 244,000 particles,

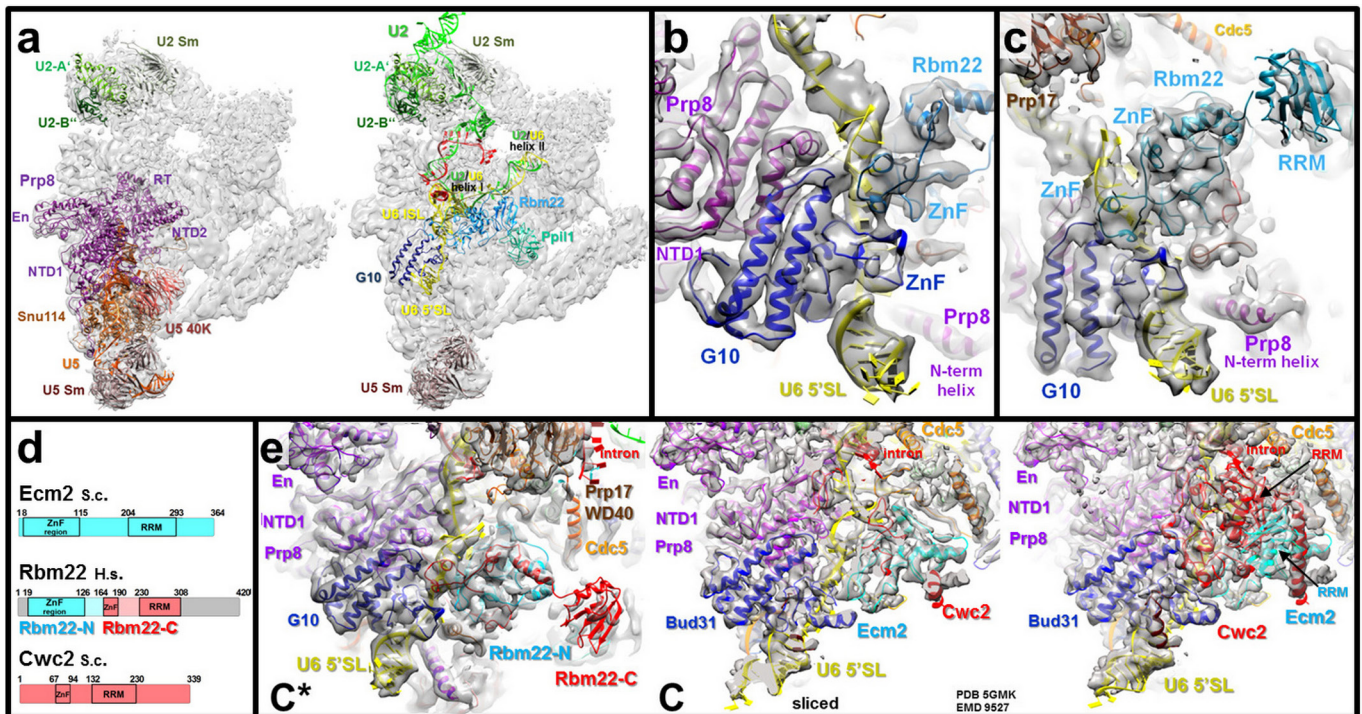
and each subset again separated into six classes by 3D classification in RELION. One of the results is shown as an example. Images contributing to the best defined spliceosome structure (around 23%) were then merged again into one particle subset. Further rounds of 3D classification in RELION led to a refinement of particle homogeneity within the final subset. The latter, now consisting of 136,534 particles was refined to a structure at 8.4 Å resolution without masking. The final structure at 5.9 Å resolution was obtained by applying a soft mask during the final steps of the refinement process. To evaluate details at the core of the C\* EM density, the unfiltered map obtained after the final 3D refinement was low-pass filtered to 4.5 Å and sharpened using a B-factor of  $-350$  in RELION. **e**, Local resolution plot reveals a resolution distribution from approximately 4.5 to 10 Å with some less well-defined parts at the periphery of the complex. Higher resolution regions (in blue, up to 4.5 Å resolution) were obtained for the centrally located catalytic core of the spliceosome. **f**, Fourier-shell correlation function of two independently refined half data sets indicates a global resolution of 5.9 Å for the masked C\* spliceosome comprising around 80% of the visible density of the whole spliceosome with respect to the unmasked density volume.





**Extended Data Figure 3 | Protein–RNA interactions in the catalytic RNP core of the C\* complex.** **a**, Domain structure of the human PRP8 protein. **b**, **c**, The catalytic U2–U6 RNA core is shown docked into the active-site pocket of PRP8 (space filling model) in the human C\* and *S. cerevisiae* C complex<sup>22</sup>. BS, branch site. **d**, Interaction of the N-terminal HAT repeats of SYF3 and the linker between PRP8 NTD1 and 2 with the U6 ISL in the C\* complex. **e**, Interaction of the CDC5 Myb domain with U2/U6 helix Ia. **f**, Interaction of WD40 domain of PRL1 with U5 stem Ic and the loop of the U6 ISL. Not only is the docking of the U2–U6 RNA

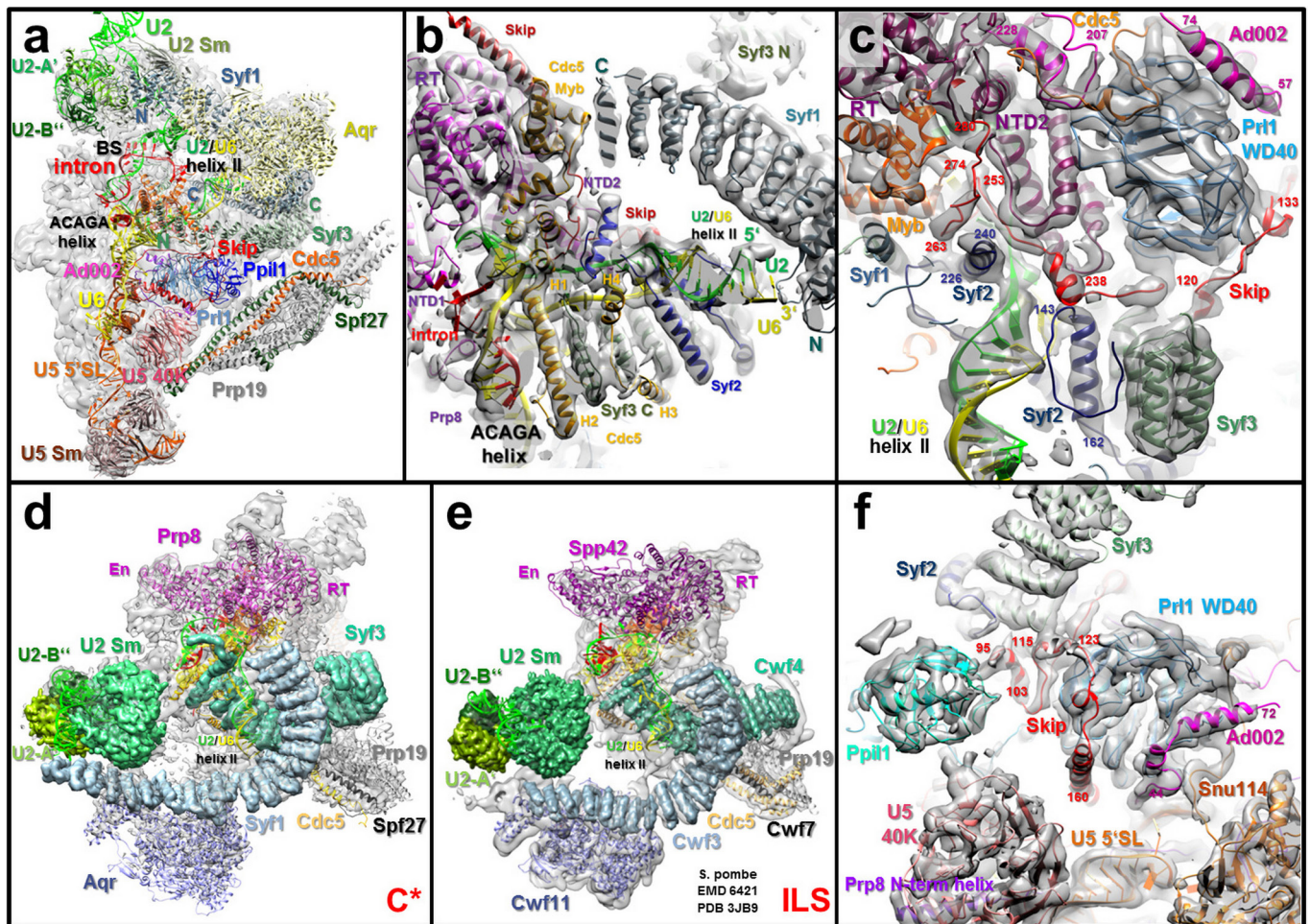
core within the active site pocket of PRP8 similar in C and C\*, but also the interaction of the linker between PRP8 NTD1 and 2 with the major groove of the U6 ISL, of SYF3 with the lower stem of the U6 ISL, and of the CDC5 Myb domain with the U2/U6 helix Ia. Moreover, the WD40 domain of PRL1 interacts with U5 stem Ic and the U6 ISL loop in both organisms<sup>14,21,22</sup> and  $\alpha$ -helix 2 of CDC5 runs along the groove of the U6 ACAGA box helix (Extended Data Fig. 5b). N-terminal parts of Skip and Ad002 (amino acid positions are indicated) also contact the U6 ISL.



**Extended Data Figure 4 | Interactions of spliceosomal proteins with the U6 snRNA in the C\* complex.** **a**, Overview of the location of various proteins and functionally important regions of the U6 and U2 snRNAs in the C\* complex. **b**, **c**, Close up of the fit of G10/BUD31, the PRP8 NTD1, and RBM22, as well as the U6 5' stem loop, into the EM density map of C\*. **d**, Comparison of the domain structure of human RBM22 with that of its apparent homologues in yeast, Cwc2 and Ecm2. **e**, Comparison of the organization of homologous domains between RBM22, and Ecm2 and Cwc2, in the human C\* and *S. cerevisiae* C complexes. The middle and right diagrams show different slices of the C complex EM density. Some regions of human RBM22 that are homologous to regions of the *S. cerevisiae* Cwc2 and Ecm2 proteins, are located near the U6 ISL in the

human C\* complex, similar to the situation in the *S. cerevisiae* C complex. For example, the N-terminal part of RBM22, comprising two zinc-finger domains that are homologous to those present in *S. cerevisiae* Ecm2, could be fit into a density element close to G10/BUD31, at a position similar to that in the *S. cerevisiae* C complex (and *S. pombe* ILS, not shown). The third zinc-finger domain of RBM22 (homologous to Cwc2) is located close to the 5' stem loop of U6 snRNA. The RRM domain of RBM22 (homologous to the RRM of Cwc2) fits into a less well-defined density element close to PPIL1, aquarius, and Skip, consistent with protein crosslinking data (Supplementary Table 2). Thus, the RRM of RBM22 is located at a different position compared to that of its likely yeast homologue (Cwc2/Cwf2) in the yeast spliceosomes.



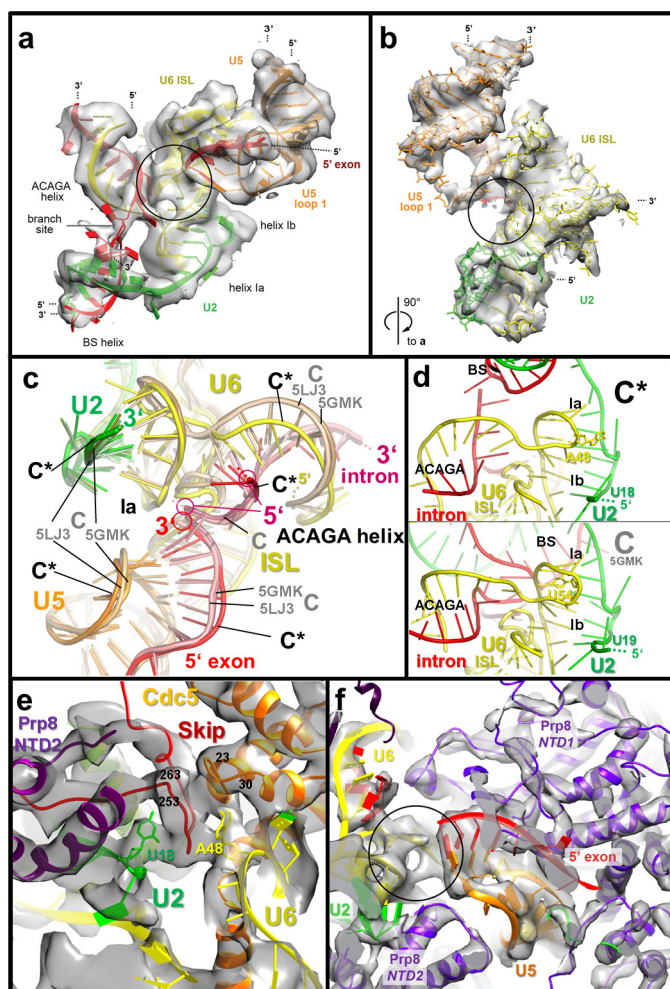


**Extended Data Figure 5 | Molecular organization of the NTC and NTC-related proteins is similar in the C\* complex and *S. pombe* ILS.**

**a**, Overview of the location of various proteins in the C\* complex. The N-terminal HAT repeats of SYF1 share a large interface with the aquarius (Aqr) protein. At the interface between SYF1 and aquarius there is a density element that accommodates cyclophilin E (CypE), a spliceosomal protein not found in yeast (Fig. 1). In C\*, the density accommodating aquarius is less well-defined, presumably owing to flexibility of this part of the complex. Crosslinks between ISY1 and aquarius indicate that part of ISY1 is located at the top of C\* (Supplementary Table 2). The meandering path of Skip (Prp45) and Ad002 (Cwc15) around the PRL1 WD40 domain is shown. N-terminal parts of Skip and Ad002 also contact the U6 ISL (see also Extended Data Fig. 3). The position of the PRP19 helical bundle (containing SPF27, the C-terminal part of CDC5 and four copies of PRP19) is stabilized by interactions of the C-terminal end of CDC5 with the WD40 domain of the U5-40K protein, and by interactions with the cyclophilin PPIL1. The WD40 domains of PRP19 are not visualized, presumably owing to their flexibility. N or C, N or C terminus. **b**, Magnified view showing the fit of parts of SYF1, SYF2 and

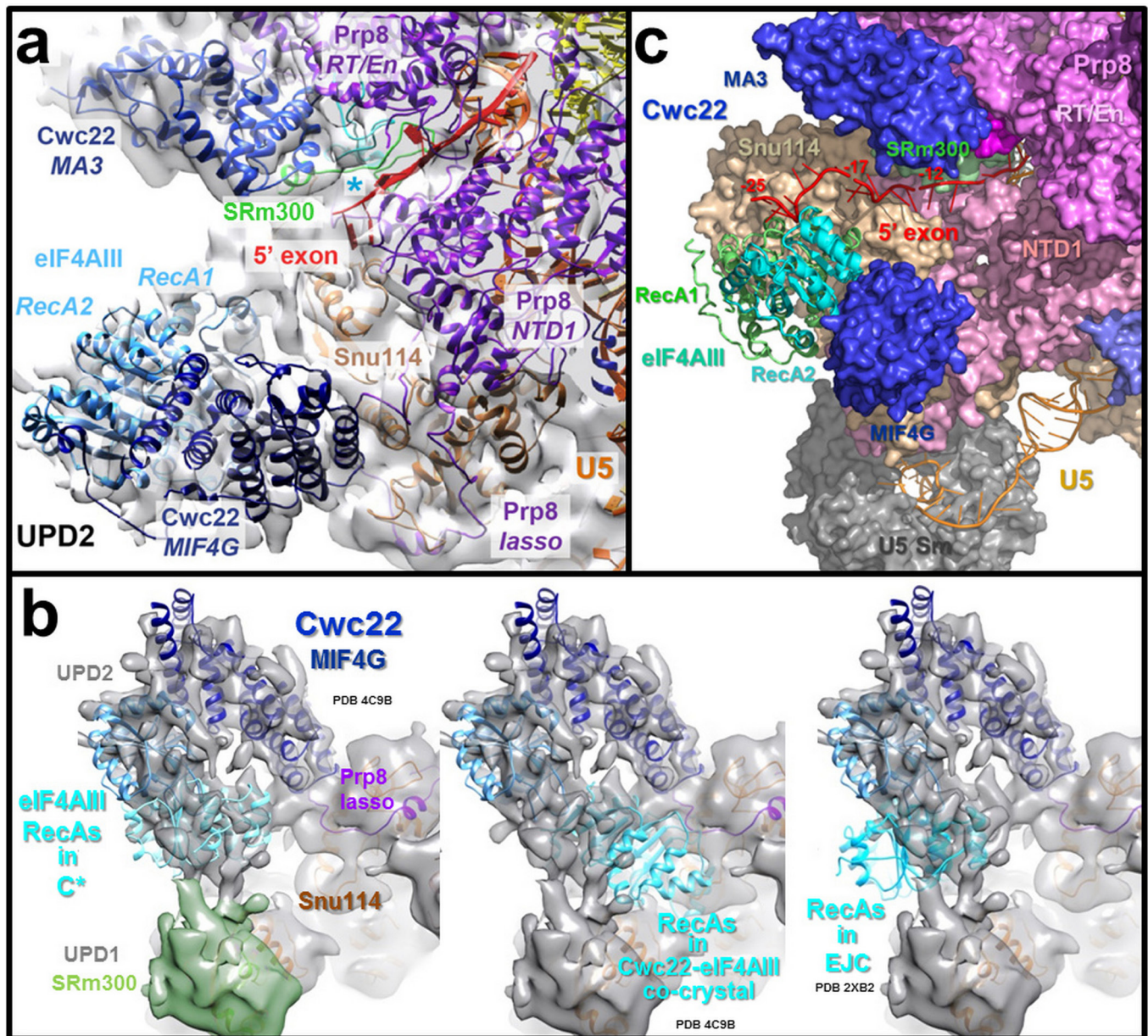
SYF3 and their neighbouring proteins into the C\* EM density map. In C\*, the C-terminal HAT repeats of SYF1 contact the N-terminal Myb domain of CDC5 (which in turn binds to the PRP8 RT domain). CDC5  $\alpha$ -helix 2 runs along the U6 ACAGA box/intron helix. **c**, Fit of Skip, PRL1, Ad002, SYF2 and their neighbouring proteins into the C\* EM density map. Major parts of Skip (amino acid numbers shown in red) bridge the WD40 domain of PRL1 (Prp46 in *S. cerevisiae*) and SYF3 N-terminal HAT repeats 2–5 with the PRP8 NTD2. SYF2 (amino acid numbers in blue) contacts U2/U6 helix II. The N-terminal part of Ad002 (amino acid numbers in pink) contacts the PRL1 WD40 domain (see also panel f). **d, e**, Comparison of the structure/organization of SYF1 and SYF3 (shown as space filling models), aquarius and other NTC and NTC-related proteins in the human C\* complex and the *S. pombe* ILS. The N-terminal region of CDC5, the N-terminal  $\alpha$ -helices of SYF3, WD40 domain of PRL1, major parts of Skip and Ad002, and PPIL1 (Ppi1 in *S. pombe*) have nearly the same structure and position as their yeast homologues (see also panels a–c). **f**, N-terminal amino acids of Skip interact with PPIL1, consistent with previous biochemical and structural studies<sup>67–69</sup>.





### Extended Data Figure 6 | 3D RNA network of the human C\* complex.

**a, b**, Fit of the catalytic core RNA into the 5.9 Å EM density map of C\* (**a**) or into the EM density map low-pass filtered to 4.5 Å (**b**). The topography of the RNA density is consistent with formation of the catalytic triplex, as found in the C complex and ILS. That is, the Watson–Crick faces of U6 nucleotides G46 and A47 are oriented towards the Hoogsteen faces of G54 and A53, respectively, and can potentially form two base triples. The bulged U74 of the U6 ISL would be in a position to stack on G46 and may form the third triple with U6 C55 (Fig. 2). Black circle, catalytic centre. **c**, Overall similarity of the catalytic RNA network in the yeast C versus human C\* complex. Superimposition of the core RNA elements from human C\* with those from the *S. cerevisiae* C complex. Dark-coloured RNAs are from C\*, whereas the lighter-coloured RNAs are from the C complex (PDB 5LJ3) and (PDB 5GMK). In the C\* complex, the U6 ACAGA box/intron helix is tilted slightly (compared to the C complex), such that the U6 snRNA moves 5–6 Å towards the core of the complex, which is probably due to the significantly different positions of the branched intron structures in the different spliceosomal complexes. **d**, The conformation of U6 nucleotide A48 (U54 in yeast U6) is significantly different in the yeast C and human C\* complex. Whether U6 A48 adopts this distinct conformation in other human spliceosomal complexes or whether it represents an intermediate state following PRP16 remodelling that subsequently adopts a conformation similar to U6 U54 in the yeast C complex before step two catalysis is unclear. **e**, U6 A48 is bound in a protein pocket comprised of amino acids from Skip and CDC5. Fit of U6 and U2 snRNA nucleotides, and amino acids of CDC5 and Skip into the C\* EM density map. **f**, Interactions of the U5 loop 1 and PRP8 with the 5' exon, and location of the 3' end of the 5' exon close to the catalytic centre (black circle).

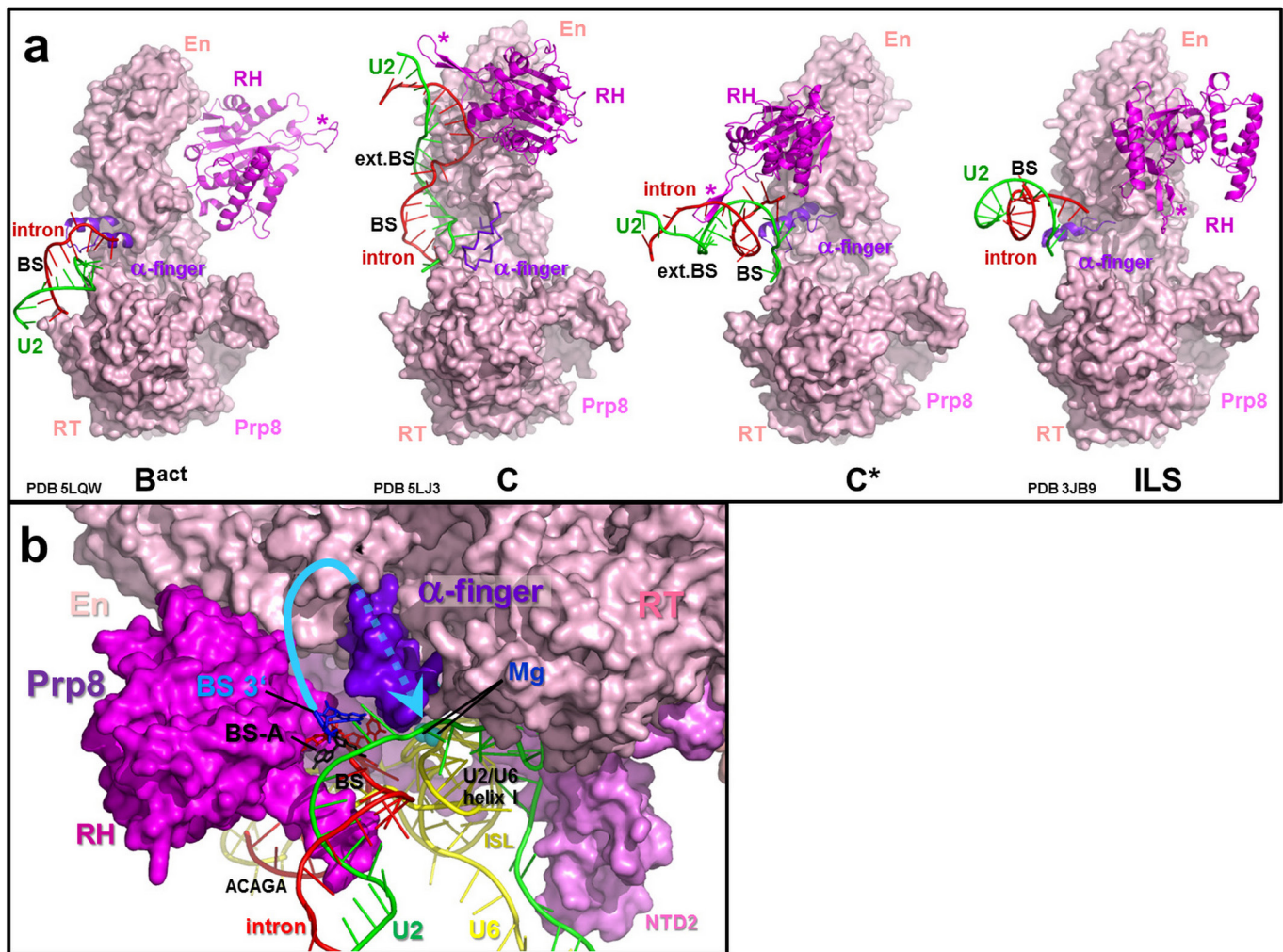


**Extended Data Figure 7 | Location and structure of CWC22 and its interaction partner, the EJC helicase eIF4AIII, in the C\* complex.**

**a**, EM density fit of various proteins forming the channel in which the 5' exon is bound. The MA3 domain of CWC22 binds the PRP8 RT/En domain and its N-terminal MIF4G domain is attached to domain 1 of SNU114. The MIF4G domain is only clearly visible in the unmasked EM density of the C\* complex, and maps to UPD2 (panel **a** and **b**). The PRP8 switch loop is shown in light blue. Close to the PRP8 switch loop, there is a density element into which the approximately 20 N-terminal amino acids of SRm300 can be accommodated in a manner similar to the arrangement of the homologous sequence of Cwc21 in the C complex of *S. cerevisiae* (indicated in green colour). **b**, EM density fit of the CWC22 MIF4G domain and the RecA domains of the eIF4AIII RNA helicase in the UPD2 density of the unmasked C\* model. Consistent with crosslinks (Supplementary Table 2), the RecA2 domain contacts the MIF4G domain. In the C\* complex, the relative orientation of the two RecA domains is such that they adopt a partially closed conformation, in contrast to their open (inactive) conformation in the MIF4G–eIF4AIII co-crystal structure (middle panel; PDB 4C9B) and their closed conformation in the crystal structure of the EJC complex, in complex with RNA (right most panel; PDB 2XB2). Shown is the poor fit of the RecA1 domain when in the open

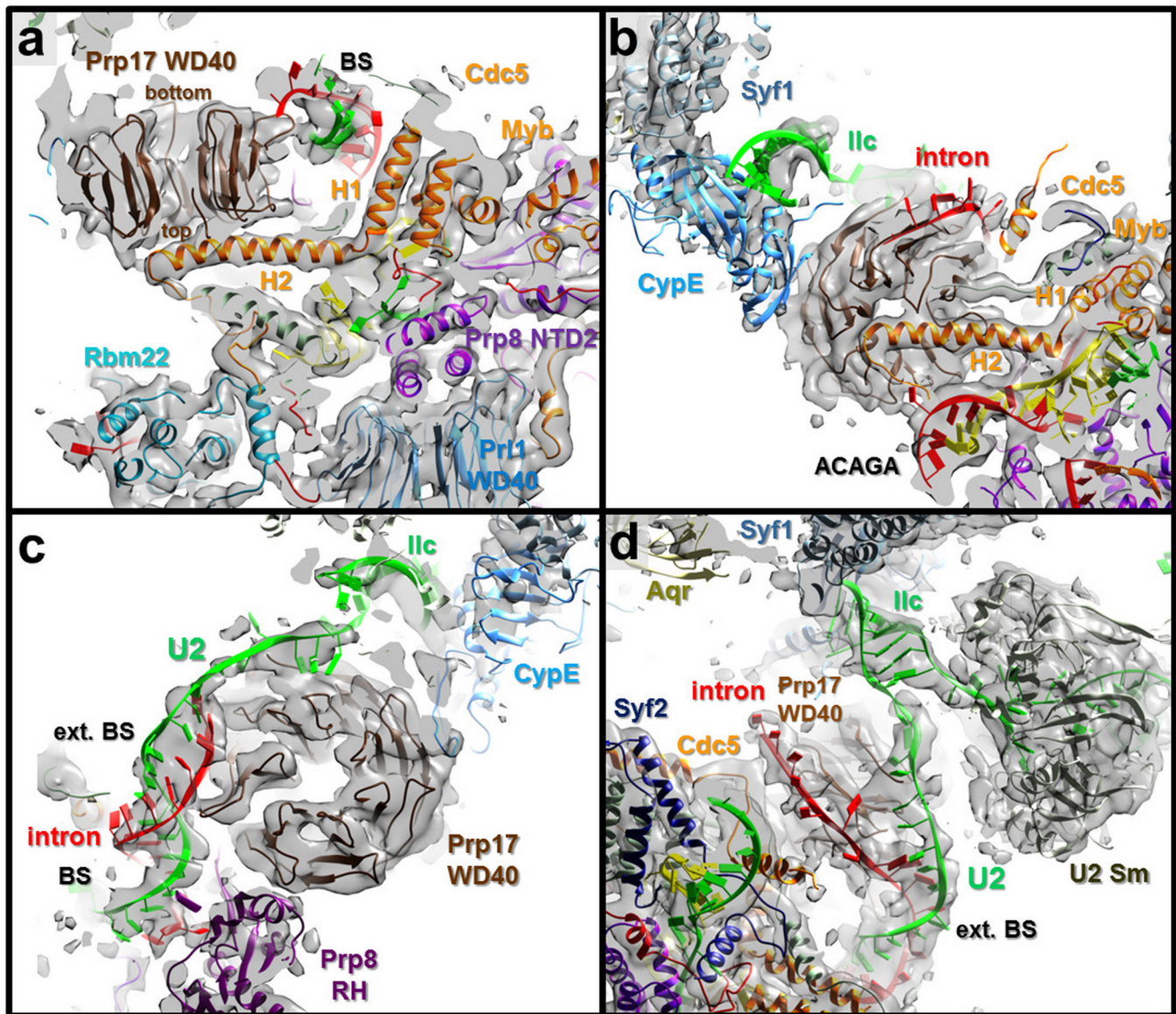
or completely closed conformation. Protein crosslinks between eIF4AIII and an N-terminal region of SRm300 indicate that part of this region is located in the neighbouring UPD1 density element of the unmasked C\* complex map, adjacent to domain 4 of SNU114 (which is not a part of UPD1), consistent with the location of two  $\alpha$ -helical elements of Cwc21 in an equivalent position in the *S. cerevisiae* C complex<sup>22</sup>. We cannot locate the RS domain of SRm300, probably because it is very flexible and located at the periphery of the C\* complex. **c**, Model of the position of an extended 5' exon RNA (red) relative to the RecA domains of eIF4AIII in the C\* complex. Extension of the 5' exon RNA to position –26 (relative to the 5' splice site) was performed by inserting nucleotides –26 to –18 based on the crystal structure of the isolated EJC bound to RNA and additionally modelling in exon nucleotides –17 to –13 (in an extended conformation), which we cannot clearly localize based on the EM density map. This suggests that eIF4AIII is already located at an optimal distance (that is, around 20 nucleotides upstream) from the 5' splice site (and thus also from the exon–exon junction that is formed after step two of splicing). Tight binding of eIF4AIII to the 5' exon probably occurs at a stage after formation of the C\* complex, based on the observation that the RecA domains are not completely closed at this stage.





**Extended Data Figure 8 | Variable position of the PRP8 RH domain and  $\alpha$ -finger in the  $B^{act}$ , C and  $C^*$  complexes.** **a**, Comparison of the position and structure of the PRP8 RH domain (magenta ribbon diagram) and  $\alpha$ -finger<sup>15</sup> (purple) relative to the PRP8 RT/En domain (pink space filling model) that is oriented in the same manner in  $B^{act}$ , C,  $C^*$  and the ILS. The position of the U2/branch site helix alone ( $B^{act}$  and ILS) or with the extended U2/branch site helix (as found in C and  $C^*$ ) is also shown. In the yeast C complex, the RH domain interacts with the 3' part of the U2 snRNP, Cwc25 and the extended region of the U2/branch site helix<sup>21,22</sup>, whereas in  $C^*$  the RH domain interacts with extended  $\alpha$  helices close to the tip of the En domain of PRP8 and its  $\beta$ -hairpin loop is inserted between the groove of the U2/branch site helix and the U6-ACAGA

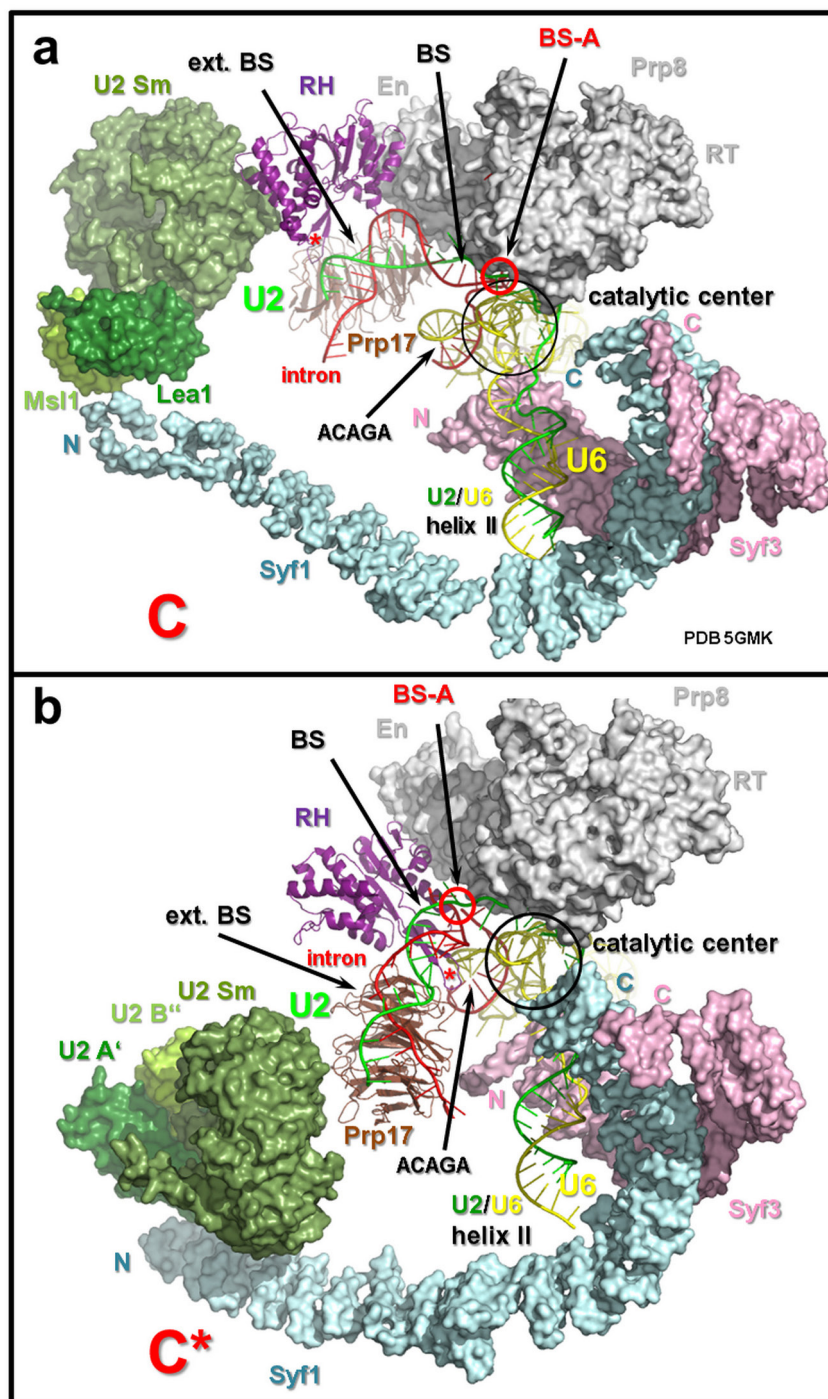
box/intron helix. Repositioning of the RH domain from C to  $C^*$  requires not only a translational, but also rotational movement that is coordinated with repositioning of the U2/branch site helix. Asterisk, position of the  $\beta$ -hairpin loop of the RH domain. **b**, Position of the PRP8  $\alpha$ -finger (shown in purple in the space filling model) in the human  $C^*$  complex. In the latter, the  $\alpha$ -finger is found close to the proposed position of the catalytic  $Mg^{2+}$  ions (turquoise spheres). The PRP8  $\alpha$ -finger could potentially aid in the placement of the 3' splice site for step two catalysis (indicated by a dashed blue line). Alternatively, it is conceivable that the  $\alpha$ -finger must be repositioned to allow proper positioning of the 3' splice site, which could be achieved by the interaction of a protein (for example, SLU7 or PRP18) or docking of the 3' splice site into the catalytic centre.



**Extended Data Figure 9 | Interaction of the WD40 domain of PRP17 with the extended U2/branch site helix and U2 nucleotides linking the U2/branch site helix and U2 Sm core.** **a, b**, Interaction of the PRP17 WD40 domain with the extended U2/branch site helix, CDC5 helix 2 and CypE. The fit of the aforementioned protein domains and RNA, plus additional neighbouring proteins, into the C\* EM density map is shown. Consistent with our protein crosslinks (Supplementary Table 2), loops on the top-side of the WD40 domain interact with the N-terminal part of  $\alpha$ -helix 2 from the N-terminal region of CDC5, which in turn contacts the U6 loop between the ACAGA box and the U2/U6 helix Ia. CypE contacts SYF1, the WD40 domain of PRP17 and the U2 3' terminal RNP domain. **c, d**, Fit of the extended U2/branch site helix and U2 nucleotides linking the latter with the U2 Sm core domain into the C\* EM density map. Intramolecular U2 snRNA helices IIa and IIc are mutually exclusive

structures<sup>70</sup>, and it was recently shown that U2 helix IIc is formed in the *S. cerevisiae* C complex<sup>21,22</sup>. In the human C\* complex, U2 nucleotides 88–95 can be fit in a helical conformation into a helical density element located directly adjacent to the U2 Sm core. This is consistent with the possibility that U2 intermolecular helix IIc is formed (at least partially) in the C\* complex, as helix IIc formation involves U2 nucleotides directly upstream of the the U2 Sm site. While the density for these U2 nucleotides is well-defined, EM density for complementary U2 nucleotides that are potentially involved in the formation of a helix are observed only at a lower threshold. This suggests that U2 helix IIc is not very stable in our C\* complex, and thus may be only partially formed. EM density that would accommodate U2 intra-molecular helix IIa or IIb is not detected in the C\* map, suggesting that they are unstable or not formed.





**Extended Data Figure 10 | Structural changes in the spliceosome that are probably mediated by the action of PRP16. a, b,** Comparison of the position/orientation of the PRP8 RH domain, U2 Sm core plus U2-A' (*S. cerevisiae* Lea1) and U2-B'' (*S. cerevisiae* Msl1), SYF1 and SYF3 HAT repeats, PRP17 WD40 domain and the catalytic RNA core, aligned relative to the PRP8 En/RT domain, in the *S. cerevisiae* C complex and the human C\* complex. Black circle, catalytic centre. Asterisk, position of the  $\beta$ -hairpin loop of the RH domain. Structural changes probably

facilitated by PRP16 include: (1) repositioning of the extended branched intron structure away from the catalytic centre together with a similar movement of the PRP17 WD40 domain; (2) a large scale movement of the entire 3' domain of U2 snRNP and rearrangement of SYF1 and SYF3; (3) translocation of the PRP8 RH domain by approximately 35 Å (bottom) to 60 Å (tip of  $\beta$ -hairpin loop) such that it interacts with the branched intron structure in the C\* complex; (4) rearrangement of the PRP8  $\alpha$ -finger; and (5) destabilization of Yju2/CCDC130 and Cwc25/CWC25.

### **3.2 Cryo-EM structure of a pre-catalytic human spliceosome primed for activation**

**Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H.**

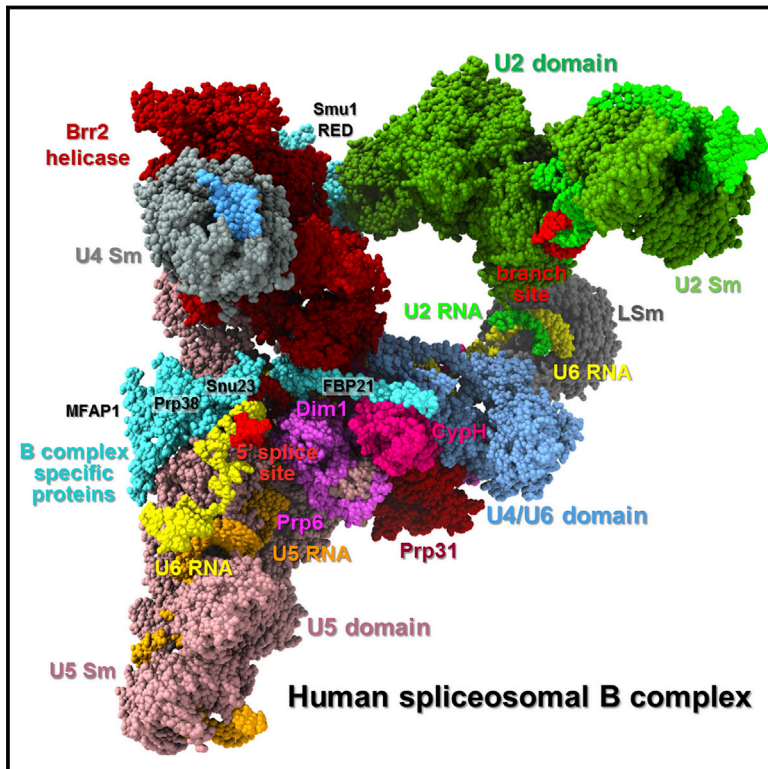
**Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation.**

2017

Cell 170, pp. 701–713.e711.

# Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation

## Graphical Abstract



## Authors

Karl Bertram, Dmitry E. Agafonov, Olexandr Dybkov, ..., Berthold Kastner, Reinhard Lührmann, Holger Stark

## Correspondence

b.kastner@mpibpc.mpg.de (B.K.),  
reinhard.luehrmann@mpibpc.mpg.de (R.L.),  
hstark1@gwdg.de (H.S.)

## In Brief

Structure of the human spliceosome getting ready for action.

## Highlights

- Molecular architecture of the human spliceosomal B complex elucidated by cryo-EM
- Dramatic restructuring of the human tri-snRNP following its integration into complex B
- Elucidation of the location and likely functions of B-specific proteins in complex B
- Marked differences in the early activation pathways of human and yeast spliceosomes



# Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation

Karl Bertram,<sup>1,5</sup> Dmitry E. Agafonov,<sup>2,5</sup> Olexandr Dybkov,<sup>2</sup> David Haselbach,<sup>1</sup> Majety N. Leelaram,<sup>2</sup> Cindy L. Will,<sup>2</sup> Henning Urlaub,<sup>3,4</sup> Berthold Kastner,<sup>2,\*</sup> Reinhard Lührmann,<sup>2,6,\*</sup> and Holger Stark<sup>1,\*</sup>

<sup>1</sup>Department of Structural Dynamics, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>2</sup>Department of Cellular Biochemistry, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>3</sup>Bioanalytical Mass Spectrometry, MPI for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany

<sup>4</sup>Bioanalytics Group, Institute for Clinical Chemistry, University Medical Center Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany

<sup>5</sup>These authors contributed equally

<sup>6</sup>Lead Contact

\*Correspondence: [b.kastner@mpibpc.mpg.de](mailto:b.kastner@mpibpc.mpg.de) (B.K.), [reinhard.luehrmann@mpibpc.mpg.de](mailto:reinhard.luehrmann@mpibpc.mpg.de) (R.L.), [hstark1@gwdg.de](mailto:hstark1@gwdg.de) (H.S.)  
<http://dx.doi.org/10.1016/j.cell.2017.07.011>

## SUMMARY

Little is known about the spliceosome's structure before its extensive remodeling into a catalytically active complex. Here, we report a 3D cryo-EM structure of a pre-catalytic human spliceosomal B complex. The U2 snRNP-containing head domain is connected to the B complex main body via three main bridges. U4/U6.U5 tri-snRNP proteins, which are located in the main body, undergo significant rearrangements during tri-snRNP integration into the B complex. These include formation of a partially closed Prp8 conformation that creates, together with Dim1, a 5' splice site (ss) binding pocket, displacement of Sad1, and rearrangement of Brr2 such that it contacts its U4/U6 substrate and is poised for the subsequent spliceosome activation step. The molecular organization of several B-specific proteins suggests that they are involved in negatively regulating Brr2, positioning the U6/5'ss helix, and stabilizing the B complex structure. Our results indicate significant differences between the early activation phase of human and yeast spliceosomes.

## INTRODUCTION

The spliceosome forms stepwise on its pre-mRNA substrate by sequential recruitment of snRNPs (small nuclear ribonucleoproteins) and numerous other proteins (Papasaïkas and Valcárcel, 2016; Wahl et al., 2009). During early spliceosome assembly, U1 and U2 snRNPs interact with the 5' splice site (ss) and branch site (BS), respectively, of the intron, forming the spliceosomal A complex. The pre-formed U4/U6.U5 tri-snRNP then joins the A complex, generating the pre-B complex in which the tri-snRNP is not yet stably bound (Boesler et al., 2016). After conformational changes, including Prp28 RNA helicase-mediated exchange of U1 with U6 at the 5'ss, the pre-catalytic B complex with stably associated tri-snRNP is formed (Boesler et al., 2016; Staley

and Guthrie, 1999). Complex B undergoes extensive compositional and conformational rearrangements, including dissociation of U1 and U4, yielding the activated B<sup>act</sup> complex. The latter is converted into a catalytically active spliceosome (designated B\*) that catalyzes step I of splicing, yielding the cleaved 5' exon and intron-3' exon lariat intermediates. At this stage, the spliceosomal C complex is generated and after additional rearrangements, the C\* complex catalyzes step II, resulting in ligation of the 5' and 3' exons and release of the spliced-out intron.

During spliceosome assembly and activation a dynamic RNA-RNA network involving snRNAs and the pre-mRNA is formed (Staley and Guthrie, 1998). U4 and U6 snRNA are extensively base paired in the tri-snRNP and B complex. During activation, the U4/U6 helices are unwound and a highly structured RNA network forms between the pre-mRNA and U2, U6, and U5 snRNAs, generating the spliceosome's catalytic RNA core (Fica et al., 2013; Staley and Guthrie, 1998). The U5 snRNP proteins Prp8 and RNA helicase Brr2 play central roles during catalytic activation. Prp8 is a major scaffolding protein that interacts with Brr2 and Snu114 (Wahl et al., 2009) and also forms a pocket that encompasses the catalytic RNA network of activated spliceosomes (Galej et al., 2014; Hang et al., 2015). Brr2 initiates spliceosome activation by unwinding the U4/U6 snRNA helices (Laggerbauer et al., 1998; Raghunathan and Guthrie, 1998). As Brr2 and its RNA substrate are present in the tri-snRNP and B complex, its activity must be regulated to ensure that U4/U6 unwinding first occurs during activation. In contrast to the situation in the *S. cerevisiae* tri-snRNP, Brr2 is located in human tri-snRNPs at a large distance from the U4/U6 helices, thereby preventing their premature unwinding (Agafonov et al., 2016). Whether Brr2 is repositioned and engages its substrate in the human B complex, and if so, how Brr2 is negatively regulated at this stage of spliceosome assembly, is unclear.

Yeast and humans share a common set of core spliceosomal proteins that are evolutionarily conserved, but human spliceosomes contain many additional proteins not present in *S. cerevisiae* (Fabrizio et al., 2009). For example, human B complexes contain a set of B-specific proteins, which include hSnu23, RED, Smu1, hMFAP1, FBP21, hPrp38, NPW38, and



NPW38BP (Agafonov et al., 2011). These proteins are conserved in higher eukaryotes, but only hSnu23, hPrp38, and hMFAP1 have homologs in *S. cerevisiae* (Agafonov et al., 2011; Ulrich and Wahl, 2017). B-specific proteins first bind during B complex formation and are missing or much less abundant in B<sup>act</sup> complexes (Agafonov et al., 2011). The function of most of these proteins remains unclear, but they are not required for tri-snRNP recruitment to the B complex (Boesler et al., 2016). Instead, they likely contribute to the activation process, as indicated for Prp38 (Schütze et al., 2016; Xie et al., 1998), via currently unknown mechanisms. Several B-specific proteins are implicated in alternative splicing decisions, but the mechanisms whereby they achieve this is unclear (Papasaikas et al., 2015; Spartz et al., 2004).

3D electron cryo-microscopy (cryo-EM) structures of the *S. cerevisiae* tri-snRNP, B<sup>act</sup>, C, and C\* complexes and *S. pombe* post-splicing ILS spliceosomes (Galej et al., 2016; Nguyen et al., 2016; Rauhut et al., 2016; Wan et al., 2016a, 2016b; Yan et al., 2016, 2017; Fica et al., 2017), and of the human tri-snRNP and C\* complex were recently reported (Agafonov et al., 2016; Bertram et al., 2017; Zhang et al., 2017). These studies revealed the spliceosome's molecular architecture during its activation and catalytic activity. However, relatively little information is available about the molecular organization of the spliceosome at early assembly stages before its activation. The structural organization of the human U4/U6.U5 tri-snRNP revealed by cryo-EM indicates that this major spliceosomal subunit is initially maintained in an inactive state. Thus major rearrangements in tri-snRNP components must occur to generate the functional centers required to prime the B complex for the subsequent activation step. To elucidate these structural changes, we determined the 3D cryo-EM structure of the human spliceosomal B complex at a core resolution of 4.5 Å and determined its spatial organization with the aid of protein crosslinking.

## RESULTS AND DISCUSSION

### Structure Determination and Model Building

By lowering the Mg<sup>2+</sup> concentration of the in vitro splicing reaction, we could affinity-purify human spliceosomes containing stoichiometric amounts of U2, U4, U5, and U6 snRNAs, but which were essentially devoid of U1 snRNA, indicating that they were stalled prior to Brr2-mediated activation and thus are spliceosomal B complexes (Figure S1). Consistent with this, highly abundant proteins in these complexes included U2 and tri-snRNP proteins (except RBM42, Prp28, and Sad1, which were absent or present in low amounts), U2AF, B-specific proteins, and pre-mRNA binding proteins, such as the CBPs. Except for moderate amounts of Skip (hPrp45) and RBM22, Prp19 complex proteins, or other proteins typically present in activated spliceosomes were absent or present in very low amounts (Figure S1D). Chase experiments with micrococcal nuclease treated extract showed that our purified B complexes are functional (Figure S1G). However, incubation of the latter with ATP, did not induce significant displacement of U4 snRNP (Figure S1F), indicating that Brr2 is still inhibited at this stage.

The 3D structure of the human B complex was determined by cryo-EM (Figure S2), revealing a globular head domain and a

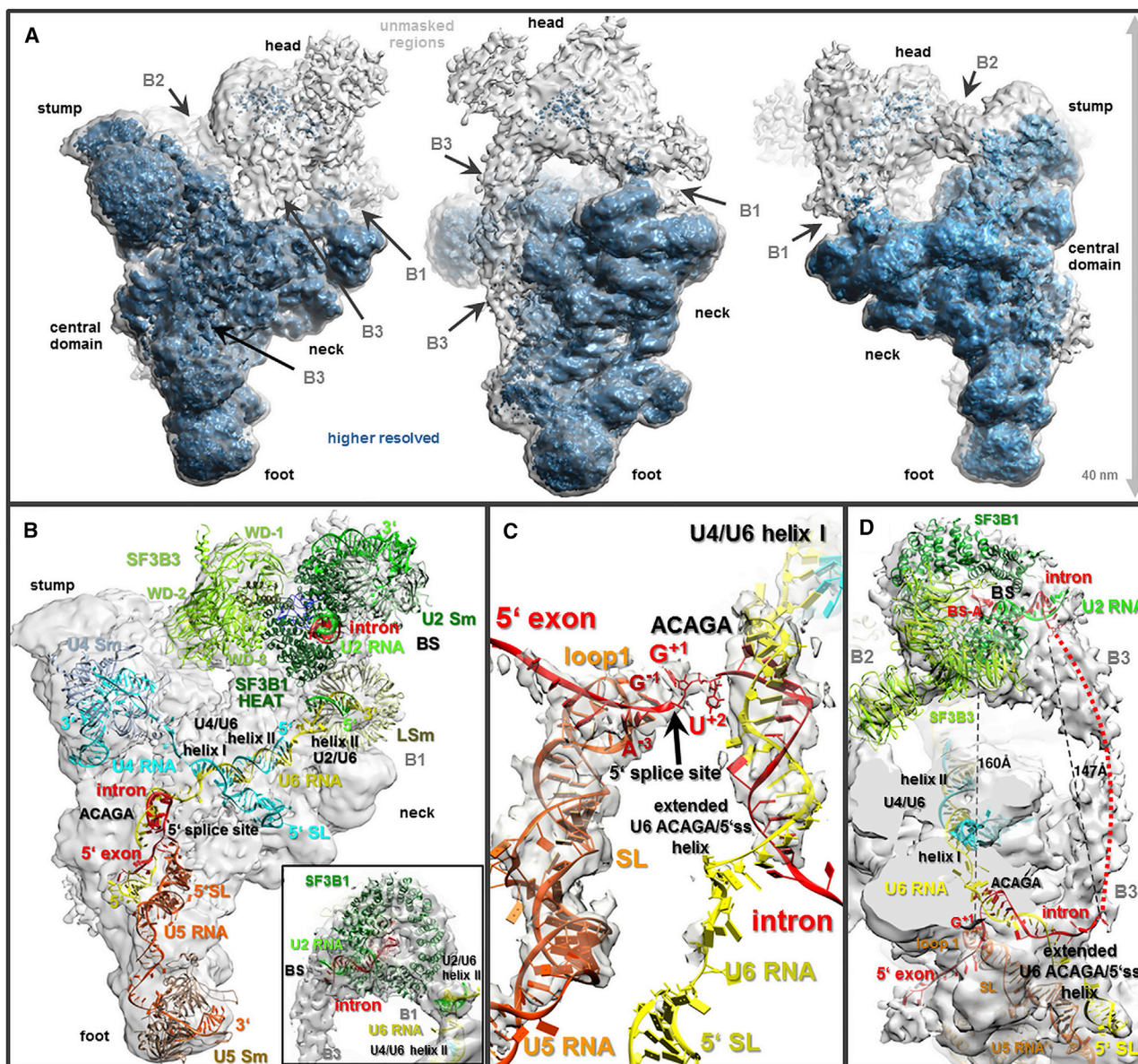
triangular body with a central domain with adjacent “foot,” “stump,” and “neck” domains (Figure 1; Movie S1). Three major densities (B1–B3) bridge the head to the neck (B1) or stump (B2), or run as an extended density element almost parallel to the central axis of the main body (B3). Most of the triangular domain is well-defined and its structure was determined at an overall resolution of 4.5 Å (Figures 1 and S2). However, the head and connecting bridges and some areas of the triangular body are more dynamic or contain components with substoichiometric occupancies and are thus less well resolved (Figures 1 and S2). By fitting known X-ray structures or homology models of structured regions of B complex components into the EM density map (summarized in Table S1), and aided by chemical protein crosslinking coupled with mass spectrometry (Table S2), we generated a pseudo-atomic model for the more stable parts of the B complex (Figure 1).

### U2 snRNP Is Located in the B Complex Head Domain

Several U2-SF3a and SF3b proteins contact the pre-mRNA at or near the BS in A, B and B<sup>act</sup> complexes, stabilizing the U2/BS helix (Gozani et al., 1996). In the *S. cerevisiae* B<sup>act</sup> complex the U2/BS helix is sequestered between the terminal HEAT repeats of the C-terminal HEAT domain of Hsh155 (SF3B1/SF3b155 in human) (Rauhut et al., 2016; Yan et al., 2016). Consistent with earlier immuno-EM studies (Wolf et al., 2009), we could fit the core of the SF3b complex together with the U2/BS helix, in a closed conformation as found in the B<sup>act</sup> complex (Rauhut et al., 2016; Yan et al., 2016), into the B complex head domain (Figure 1). The SF3B3 WD40 domains are oriented toward the stump and the circular SF3B1 HEAT domain is oriented toward the neck of the body, while the U2 Sm core structure is located close to bridge B3 (Figure 1). One end of the U2/BS helix is located at the interface between the head and B3 (Figure 1). Based on protein crosslinks, U2 SF3a proteins are also likely located in this region, while the largely intrinsically unfolded C-terminal region of U2 SF3A1 appears to extend from the head through B3 into the lower part of the main body, where it is crosslinked to Prp8 and U5-40K (Figure S1H).

### Organization of U4, U6, and U5 snRNA and the 5'ss Region of the Pre-mRNA

We were able to trace most nucleotides (nts) of U4, U6, and U5 snRNA in the B complex body (Figure 1). The three-way helical junction of the U4/U6 snRNA is located in the upper part close to the neck. The U4 Sm and U6 LSm protein rings are located in the stump or at the B1 bridge between U2 snRNP and the neck, respectively. B1 contains, among others, U2/U6 helix II, in which the 5' end of U2 and 3' end of U6 are base paired (Figures 1B and 3C). The major U5 stem-loop (SL) has a similar conformation in B and the tri-snRNP, with the U5 Sm core located in the foot of both (Figure 1B). However, in B, U5 loop 1 adopts a different conformation and loop 1 nucleotides U41 and U43 are base paired with 3' terminal nucleotides of the 5' exon, i.e., A-3 and G-1 (Figures 1C and S3A). An additional RNA helical element comprising the U6 ACAGA box base paired to several nucleotides near the 5' end of the intron is located ca 3–4 nm from the U5 loop 1/5' exon helix (Figures 1B and 1C), confirming that the U6 ACAGA/5'ss helix is formed in our



**Figure 1. Cryo-EM Reconstruction of the Human B Complex**

(A) Different views of the B complex EM density map (rotated around the vertical axis). Blue, better-resolved densities. Gray, masked regions not seen in the 4.5 Å structure. B1, B2, B3, and density bridges connecting the head with the triangular body.

(B) Location in the unmasked density of the U2 SF3b protein core and U2 Sm core in the head, and the U4 and U5 Sm cores, U6 LSm core, and major RNA elements in the main body. Inset, fit of the SF3B1 HEAT domain in complex with the U2/BS duplex.

(C) Fit of the U5 SL, 3' end of the 5' exon, 5'ss nucleotides, and extended U6 ACAGA/5'ss helix into the 4.5 Å EM density map.

(D) Path of B3 (in the unmasked density) from the U2/BS helix to the end of the extended U6 ACAGA/5'ss helix. Stippled red line, possible path of the intron. Distances between the branch adenosine and 5'ss, and between the U2/BS helix and extended U6/5'ss helix, are indicated.

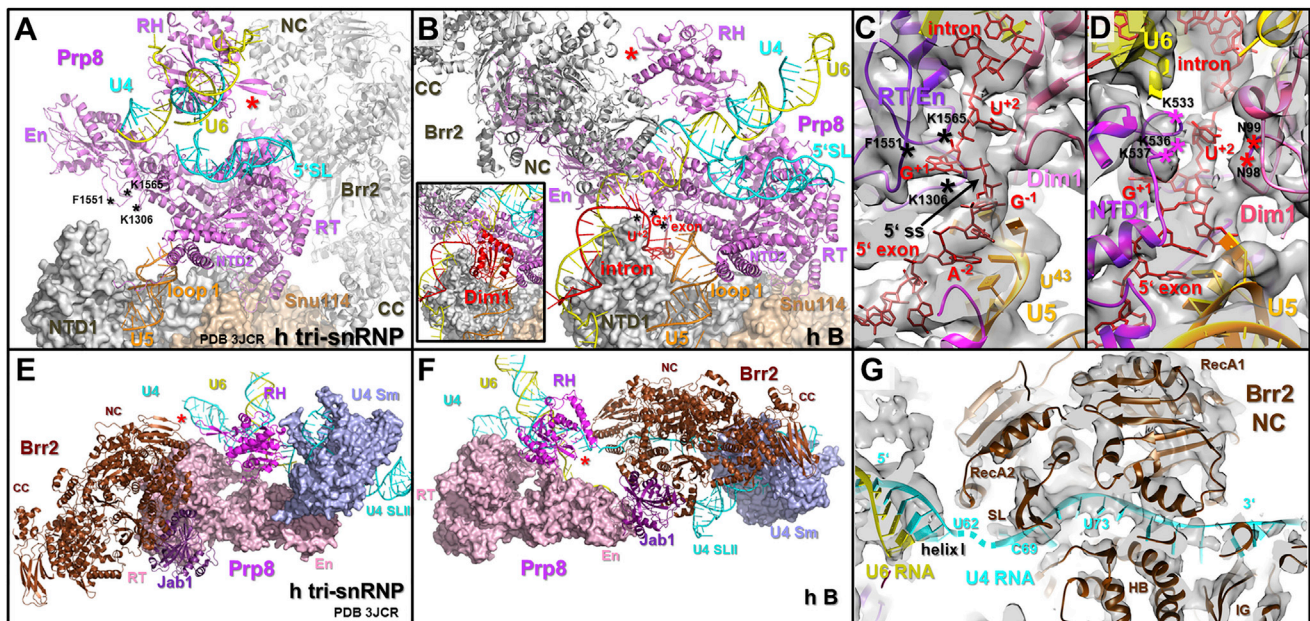
See also [Figures S1, S2, and S3](#), [Tables S1 and S2](#), and [Movie S1](#).

B complexes. The 5'ss GU nucleotides at the intron's 5' end (G+1 and U+2) are positioned between both helical elements in an extended conformation with the two bases pointing away from each other (Figure 1C), which differs significantly from their conformation in the activated spliceosome (Figure S3A).

Interestingly, U6 nucleotides 30 to 40, immediately upstream of the ACAGA box, also form base pairs (including non-canonical

ones) with additional intron nucleotides downstream of the 5'ss (Figure 1C). As an extended U6 ACAGA/5'ss helix is also present in the human C\* complex (Bertram et al., 2017; Zhang et al., 2017), but is not observed in yeast spliceosomes (Figure S3B–3D), it may be a structural feature of the spliceosome solely in higher eukaryotes. Such extended helical elements may help to stabilize short RNA helices, such as the human U6 ACAGA





**Figure 2. Substantial Prp8 and Brr2 Rearrangements and Formation of a 5'ss Binding Pocket in the B Complex**

(A and B) Open and partially closed conformation of Prp8 in the human tri-snRNP (A), and the B complex (B), respectively, aligned relative to the Prp8 NTD1 domain and Snu114. Black asterisks, amino acids in Prp8 linker loops close to G+1 of the 5'ss in the B complex. For clarity, the Dim1 protein is only shown in the inset of (B). Red asterisk, the  $\beta$  hairpin of the Prp8 RH domain. For a general overview of the position of U5 proteins in the B complex EM density map, see Figure S4A.

(C) Fit of the 5'ss nucleotides into the 4.5 Å EM density map of the B complex. G+1 is bound in a pocket comprised of amino acids of Prp8 linker loops (indicated by black asterisks). Base pairing of 5' exon nucleotides with U5 loop 1 is also shown.

(D) U+2 of the 5'ss is located in a protein pocket comprised of loops of both Prp8 NTD1 and of Dim1. Amino acids of Prp8 NTD1 and Dim1 close to U+2 in the B complex are indicated.

(E and F) Large-scale rotational movement of Brr2's helicase domain from the Prp8 RT domain in the tri-snRNP (E) to the Prp8 En domain in the B complex (F), aligned relative to the RT/En domain.

(G) Docking of Brr2's NC onto the central single-stranded region of U4 snRNA close to U4/U6 helix I. SL, separator loop. U4 nucleotides between U62 and C69 are flexible and are thus indicated by a stippled line.

See also Figures S3 and S4 and Tables S1 and S2.

box/5'ss helix. The extended bridge B3 (spanning ca 15 nm) connects the region where the U2/BS helix is located with the end of the extended U6 ACAGA/5'ss helix, suggesting B3 contains, in addition to protein, the middle part of the intron that connects these functionally important RNA helices (Figure 1D). The 5'ss and BS, which later must be juxtaposed for catalytic step 1 to occur, are physically separated by ca 16 nm in the B complex (Figure 1D). The 5'SL of U6 is located in a position similar to that observed in the yeast B<sup>act</sup> or human C\* complexes, close to the U5 Sm core (Figures 1B and S3). As the U6 5'SL and ACAGA box are likely located in the upper part of the isolated human tri-snRNP, they must be substantially repositioned upon tri-snRNP integration into the spliceosome.

#### Prp8 Adopts a Partially Closed Conformation and, Together with Dim1, Forms a 5'ss Binding Pocket

The B complex central domain contains the U5 Prp8, Snu114, U5-40K and Dim1 proteins (Figures 1A and S4A). Whereas the positions and structures of the latter three, and of the Prp8 N-terminal domain 1 (NTD1), are very similar in the human tri-snRNP and B complex, the position of the Prp8 RT/En domain is clearly different. In the tri-snRNP, Prp8 has an open conformation,

whereby the RT/En domain points upward and the tip of the En domain is separated from NTD1 by ca 5 nm (Figure 2A). In the B complex, the RT/En domain has rotated by ca 20° around its long axis compared to its position in the tri-snRNP, and moved downward, such that several loops emanating from the RT/En linker are juxtaposed with Prp8 NTD1 and Dim1 (Figures 2B and S4). While Prp8 adopts a partially closed conformation in the B complex, Prp8's active site pocket (formed by the RT/En, NTD1, and NTD2 domains), into which the catalytic U2/U6 RNA network docks during activation, is not completely closed, as found in catalytically activated spliceosomes (Figure S4B).

The partial rearrangement of Prp8 after tri-snRNP integration into the spliceosome, generates a protein pocket, comprised of residues of the Prp8 RT/En linker region, Prp8 NTD1 and Dim1, that binds the 5'ss GU dinucleotide in an extended conformation. That is, G+1 of the intron is sandwiched between loop regions (containing amino acids [aa] K1306 and F1551 to K1565, respectively) emanating from the Prp8 RT/En linker, while U+2 is contacted by a loop (aa 93–101) of Dim1, and by a helical region of Prp8 NTD1 (aa 532–537) (Figures 2B to 2D). Our structure thus suggests that the evolutionarily conserved Dim1 protein plays a direct, previously unknown role in 5'ss recognition in

the B complex. Recognition of the extended 5' ss by an intricate network of RNA-RNA and RNA-protein interactions, as shown here, provides the structural basis for sequestering the 5' ss at this pre-catalytic stage of splicing.

The Prp8 RH domain is located above the RT/En linker, in both the B complex and human tri-snRNP, indicating that it has undergone a similar downward shift as the RT/En domain. However, the RH domain has rotated by ca 180° about its central axis in B and thus its spatial orientation differs dramatically between the two complexes (Figures 2A and 2B). The 5' ss GU can be crosslinked to Prp8's RH domain at an early stage of spliceosome assembly (Reyes et al., 1999). As the RH domain is separated from the 5' ss by ca 6 nm in our B complex (Figure 2B), the proposed RH-5' ss interaction appears to take place at an earlier assembly step, presumably before disruption of the U1/5' ss interaction.

### Brr2 Is Dramatically Rearranged and Binds to Its RNA Substrate in the B Complex

In human tri-snRNPs, Brr2 and the Prp8 Jab1 domain that tightly binds to it, are located close to the RT end of the Prp8 RT/En domain, and the active N-terminal helicase cassette (NC) of Brr2 is located ca 10 nm away from the U4/U6 duplex and the U4 Sm core structure (Figures 2E and S4A). In contrast, in the B complex, Brr2 is positioned near Prp8's En domain, ca 20 nm away from its position in the tri-snRNP, and the Prp8 Jab1 domain now contacts the tip of Prp8's En domain (Figures 2F and S4A). This large-scale movement of Brr2 would require its rotation by ca 180° around the long axis of the tri-snRNP part of the B complex. Moreover, in the latter, the U4 Sm core is now located at the interface between Brr2's helicase cassettes, such that Brr2's NC is positioned between the U4 Sm core and U4/U6 helix I, which are connected via the central single-stranded region of U4 snRNA (Figures 2F, 2G, and S1). This region of U4 is required for U4/U6 duplex unwinding by Brr2 in vitro, indicating that it functions as a docking site for Brr2's NC (Mozaffari-Jovin et al., 2012). Consistent with this, the central single-stranded region of U4 runs across the two RecA domains of the NC, and RecA2 binds U4 nucleotides 69 to 73 downstream of U4/U6 helix I and is thus positioned close to helix I (Figure 2G). Thus in our B complex, Brr2 is bound to its RNA substrate and is poised to unwind the U4/U6 duplex and initiate the spliceosome activation process, but its helicase activity is still negatively regulated (Figure S1F). Interestingly, Brr2 does not appear to contact the double-stranded region of U4/U6 helix I (Figure 2G), which would be required for its ultimate unwinding. As described below, this contact is potentially prevented by the B-specific protein FBP21.

### Remodeling of the U4/U6 Di-snRNP and Prp6 during B Complex Formation

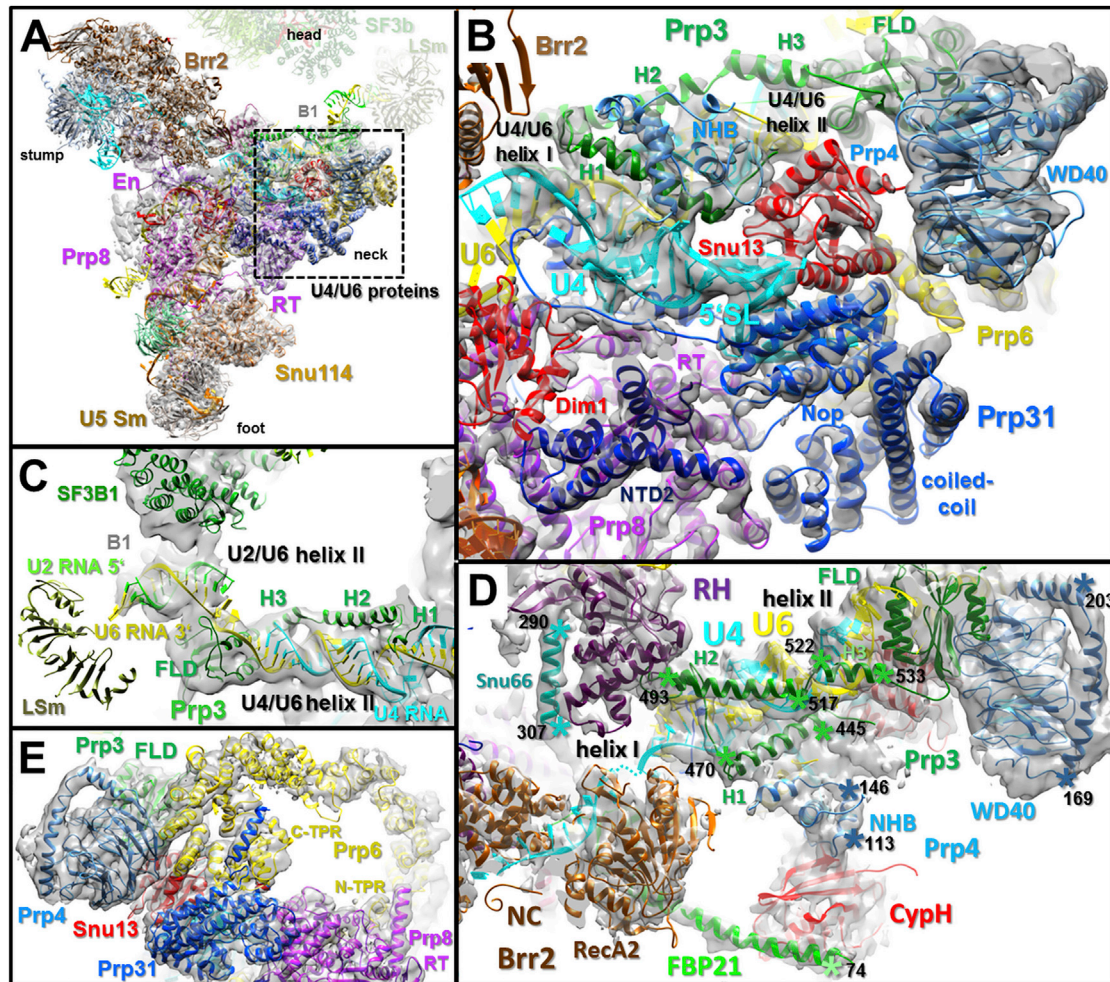
The major domains of the U4/U6 proteins interacting with the U4/U6 three-way junction are also located close to the B complex neck (Figure 3). Snu13, which directly interacts with the U4 k-turn motif, is located between the U4 5' SL and U4/U6 stem II, and also interacts with the C-terminal WD40 domain of Prp4 (Figure 3B). The Prp31 Nop domain interacts with Snu13 and the U4 k-turn motif in a manner similar to that observed in

the co-crystal structure of Snu13, Prp31 and the U4 5' SL (Liu et al., 2007). The Prp31 coiled-coil domain, however, has rotated ca 50° and is now closer to the Prp31 Nop domain in the B complex, and thus adopts a more compact structure (Figure 3B, see also below). The C-terminal part of Prp31 has an extended conformation and runs between the phosphodiester backbone of the U4 5' SL and Dim1 (Figure 3B).

Prp3 interacts closely with various parts of the U4/U6 duplex. Its C-terminal ferredoxin-like domain (FLD) binds to the 3' terminal single-stranded region of U6 (Liu et al., 2015) and likely stabilizes the neighboring U2/U6 helix II in bridge B1 (Figure 3C). Three  $\alpha$  helices (H1-H3), located N-terminal to Prp3's FLD, interact directly with U4/U6 helices I (H1 and H2) and II (H3) (Figure 3D). As both U4/U6 helices are subsequently unwound by Brr2, Prp3 may contribute to Brr2 regulation at this stage. Based on numerous protein-protein crosslinks, the N-terminal part of Prp3 runs across the Prp4 WD40 domain, and then passes along bridge B1, where it crosslinks with several proteins of the U6 LSm core and finally contacts the U2 SF3A1 and SF3B1 proteins in the head (Figure S5A). Thus, Prp3 may also play an important role in stabilizing this major connection between U2 and the tri-snRNP during early spliceosome assembly. Prp4 bridges several U4/U6 proteins with each other. For example, its C-terminal WD40 domain interacts with Snu13, Prp3's FLD and Prp6, while its N-terminal-most helical bundle (NHB) bridges cyclophilin H (CypH) to  $\alpha$  helix H1 of Prp3 (Figures 3B and 3D). Moreover, the N-terminal region of Prp4 additionally interacts with U2 SF3A proteins and Brr2's helicase domain, as indicated by multiple crosslinks (Figure S5B).

A comparison of the cryo-EM structures of the isolated tri-snRNP and B complex reveals that in addition to Prp8 and Brr2, several U4/U6 components and Prp6 also undergo significant rearrangements upon/after tri-snRNP integration into the spliceosome. For example, in the B complex, the U4 Sm and U6 LSm core structures have undergone large-scale translocations (Figures 2E, 2F, 4A, and 4B). Moreover, the entire complex of U4/U6 proteins, which includes Snu13, Prp4's WD40, Prp31's Nop, and Prp3's FLD domains, and the C-terminal TPR repeat domain of Prp6, has rotated counter-clockwise by ca 20-30° with respect to the Prp8 RT domain. Thus they, together with the U4/U6 three-way junction, are shifted closer to the Prp8 RT/En domain in the B complex (Figure 4B). The Prp6 C-terminal TPRs interact with Prp4, Snu13 and Prp3 in a very similar manner in both complexes, but they additionally contact Prp31's rearranged coiled-coil domain only in B (Figure 3E, and see below). After tri-snRNP integration into the spliceosome, a major rearrangement also occurs within the N-terminal-most TPR repeats of Prp6, which are more extended in the B complex and whose interaction with Prp8's RT/En and RH domains has changed substantially (Figures 4A and 4B). The rearranged Prp6-Prp8 interaction appears to be stabilized in the B complex by  $\alpha$ -helical regions of Snu66, which form a bridge stretching from Prp6's N-terminal TPRs to the Prp8 RT/En linker region, close to the switch loop (Figure 4D). The remaining part of Snu66, which is thought to be largely intrinsically unfolded, appears to be located close to Brr2's NC and CC, Prp8's linker, En, RH, and Jab1 domains, and Smu1, as indicated by protein-protein crosslinks (Tables S1 and S2).





**Figure 3. Location of U4/U6 Proteins and Their Interaction with the U4/U6 Three-Way Junction**

(A) Location of the U4/U6 proteins in the B complex neck region.

(B) Fit of the indicated U4/U6 protein domains/regions and U4/U6 three-way junction into the 4.5 Å EM density map of the B complex.

(C) Location of U2/U6 helix II, Prp3's FLD, and the U6 LS domain in or close to the B1 bridge at the interface between the head and neck of complex B.

(D) A helical domain (NHB) of Prp4, located N-terminal of its WD40 domain, links CypH via Prp3 helix H1 to the U4/U6 three-way junction. This network is connected via FBP21 to Brr2's NC and the U6 ACAGA/5'ss helix (see also Figure 5). Selected amino acids (black numbers) are indicated by asterisks.

(E) Fit of Prp6's C-terminal TPR repeats into the 4.5 Å EM density of the B complex and their interactions with the major domains of U4/U6 proteins.

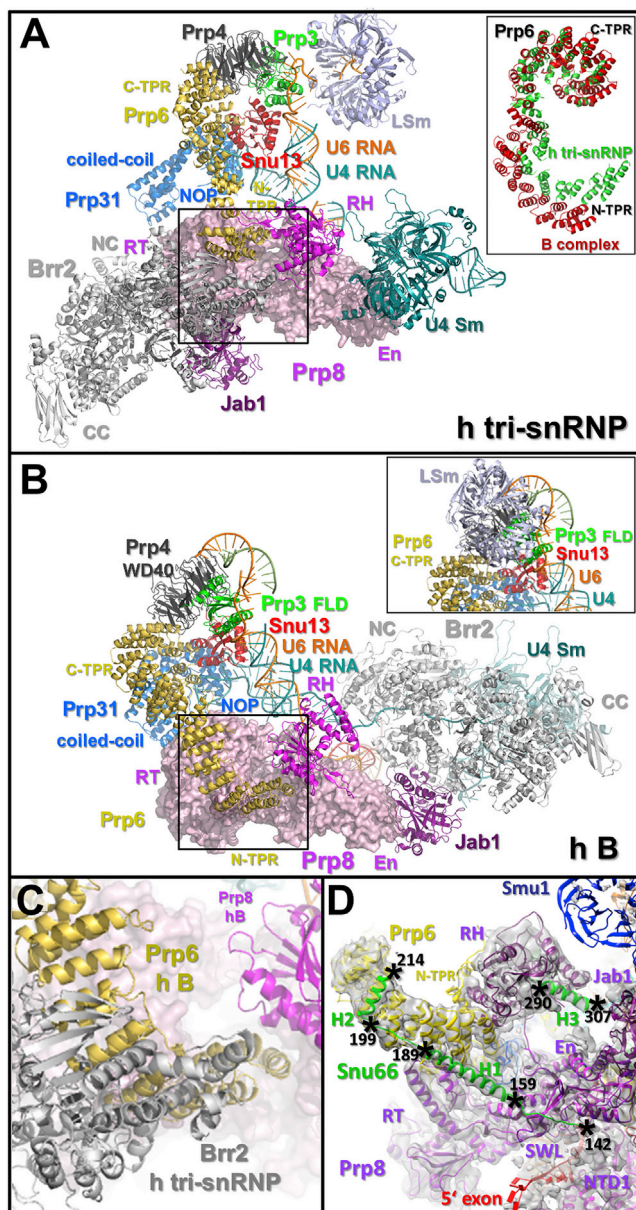
See also Figure S5 and Tables S1 and S2.

### Location and Potential Functions of the B-Specific Proteins

A number of B-specific proteins, whose domain organization is shown in Figure 5A, could be located in the B complex EM density map (Figure 5B), providing first insights into their potential functions. The B-specific protein Smu1 possesses an N-terminal, ca 180-aa-long, structured helical region (NTR), that contains LisH and CTLH domains, which is connected via a short linker region to a C-terminal WD40 domain (Ulrich et al., 2016a). In the B complex, Smu1's C-terminal WD40 domain is located at the interface between Brr2's CC and the WD40 domain 2 of SF3B3. Moreover, based on crosslinks, we could place Smu1's NTR in a density element that bridges the WD40 domains of Smu1 and SF3B3 (Figure 5C). Thus, Smu1 and SF3B3's WD40 domains are the major constituents of the B2

bridge, and Smu1 may help to stabilize the position of Brr2 after its rearrangement. Smu1's WD40 domain is positioned at the interface of Brr2's NC and CC (Figure 5D), and thus might influence the relative orientation of both helicase cassettes and thereby potentially modulate the helicase activity of the NC (Santos et al., 2012). The strategically important position of Smu1 in the B complex could explain why its mutation or deletion has profound effects on pre-mRNA splicing (Papasaiikas et al., 2015; Spartz et al., 2004).

Smu1 forms a heterodimer with the B-specific RED protein, stably binding a short region of the latter (i.e., aa 209–222) via its NTR (Ulrich et al., 2016a). RED is largely intrinsically unfolded and we could not locate its precise position in the B complex EM density map. However, in our purified B complex, crosslinks were detected between the Smu1 NTR and the short region of



**Figure 4. U4/U6 snRNP and Prp6 Remodeling during B Complex Formation**

(A and B) Organization of the indicated U5 and U4/U6 protein domains, as well as C-terminal TPR repeats of Prp6 in the human tri-snRNP (A) and the human B complex (B). Both complexes are aligned relative to the Prp8 RT/En domain. As they obscure the U4/U6 proteins, the LSm proteins were omitted from (B), but instead are shown in the inset of (B). Inset in (A), Superposition of Prp6's TPR domain in the tri-snRNP (green) and B complex (red), aligned relative to the Prp6 C-terminal TPR repeats.

(C) Superposition of the Brr2 helicase domain from the isolated human tri-snRNP onto the B complex structure, aligned relative to the RT/En domain, indicating a clash with the rearranged Prp6 TPR domain in the B complex in the absence of Brr2 rearrangement. Superimposed regions of the tri-snRNP and B complex are indicated by boxes in (A) and (B).

(D) N-terminal  $\alpha$  helices of Snu66 connect the Prp6 N-terminal TPR repeats with the Prp8 RT/En, RH, and NTD1 domains. SWL, switch loop within the Prp8 RT/En domain. Asterisks, selected amino acid positions of Snu66.

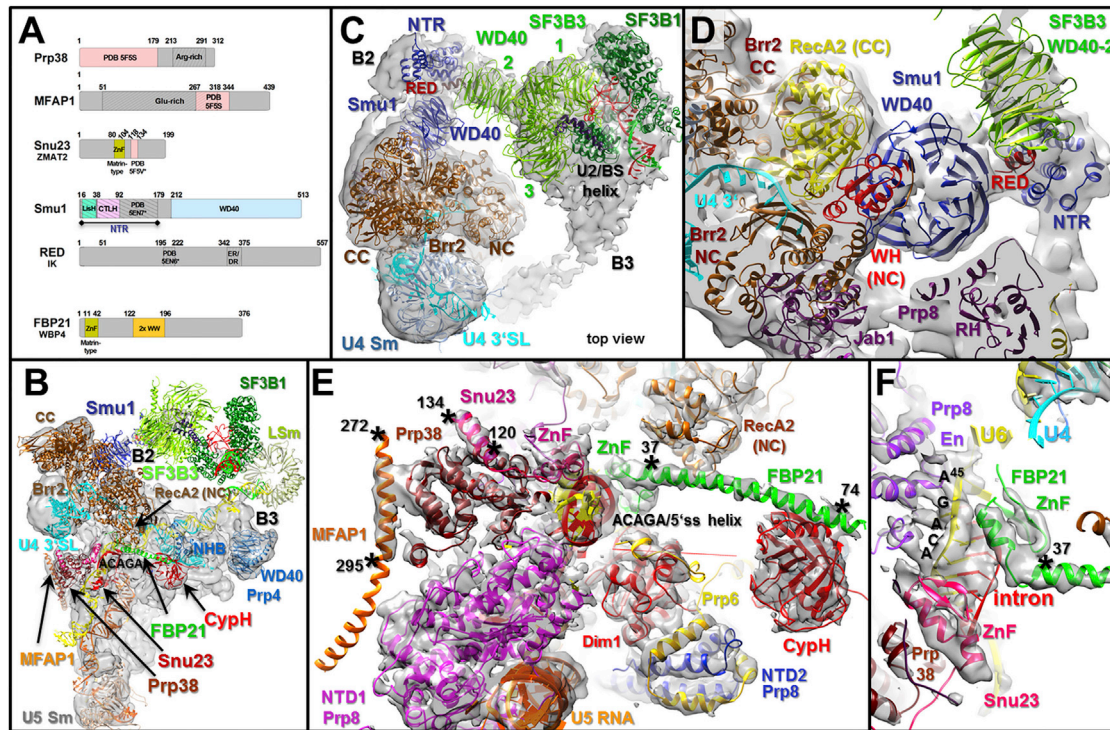
See also Tables S1 and S2.

RED that binds to it, and between the latter and Prp8's RH domain. Moreover, we observed numerous crosslinks between the N-terminal part of RED and several U2 proteins in the head, while its C-terminal region was crosslinked to the Prp8 RT/En domain and NTD1 (Figure S6A). Thus, like Smu1, RED appears to play a role in bridging U2 with U5 proteins in the B complex.

Several B-specific proteins contact the U6/5'ss helix and may aid in its repositioning within the spliceosome. The N-terminal domain (NTD) of Prp38 interacts simultaneously with several B complex proteins, and was crystallized in complex with short helices of MFAP1 and Snu23 (Ulrich et al., 2016b). The Prp38 NTD is located close to the U6/5'ss helix, and appears to interact with the U6 ACAGA sequence (Figures 5B and 5E). The  $\alpha$  helices of Snu23 and MFAP1 that bind Prp38 fit into neighboring EM densities, and the N-terminal Snu23 zinc finger (ZnF) also interacts with the upper part of the extended U6 ACAGA/5'ss helix (Figures 5E and 5F). Prp38 is required to convert B into a B<sup>act</sup> complex (Schütze et al., 2016; Xie et al., 1998), but little is known about the molecular mechanism whereby it contributes to the activation process. A comparison of the position of the U6 ACAGA/5'ss helix in B and B<sup>act</sup> complexes indicates that it must be repositioned during activation to ultimately juxtapose the 5'ss and U2/U6 catalytic center (Figure S3). As Prp38 and Snu23 contact the U6 ACAGA/5'ss helix, they may facilitate the repositioning of the latter during activation. In addition, or alternatively, they may help recruit proteins such as Cwc24 to the 5'ss during activation (Yan et al., 2016). MFAP1 appears to play a role in connecting the head and main body of the B complex, as indicated by protein crosslinks (Figure S6B).

Finally, in addition to Smu1, the B-specific protein FBP21 also contacts Brr2 and thus potentially aids in maintaining Brr2 in an inactive state in the B complex. The N-terminal region of FBP21 contains tandem WW domains (Huang et al., 2009), and is predicted to contain at its very N terminus a ZnF that appears to be followed by a long  $\alpha$ -helical element (Table S1). In the B complex, FBP21's N-terminal region indeed adopts a matrin-like ZnF structure that binds across the minor groove of the U6 ACAGA/5'ss helix (Figure 5F). The adjacent  $\alpha$  helix fits into a density element that runs from the U6/5'ss helix toward the RecA2 domain of Brr2's NC and the CypH protein (Figures 3D and 5E). CypH interacts with Prp4, which in turn communicates via Prp3 with U4/U6 helix I (Figure 3D). The position of FBP21's N-terminal  $\alpha$  helix between Brr2's NC (bound just upstream of U4/U6 helix I) and CypH, which indirectly binds U4/U6 helix I via Prp4 and then Prp3, suggests that it may act to maintain U4/U6 helix I and its associated proteins at a sufficient distance from Brr2's NC (i.e., hold it at bay), thereby preventing Brr2 contact with the double-stranded region of its substrate and potentially maintaining it in an inactive state. Our crosslinking results also suggest that the C-terminal region of FBP21 contacts Brr2's helicase domains and Prp4 (Figure S6C). Thus, a complex interaction network involving FBP21, Brr2, and the U4/U6 snRNP proteins potentially enables the tight regulation of Brr2 helicase activity at this stage. Release of FBP21 during the B to B<sup>act</sup> transition could play a key role in triggering Brr2 activity during spliceosome activation. Consistent with this idea, FBP21 is the only B-specific protein lost from spliceosomes that are blocked at an





**Figure 5. Organization of B-Specific Proteins in the Human B Complex**

(A) Domain organization of the indicated human B-specific proteins. Domain boundaries are indicated. Protein regions for which high resolution structures are available are indicated by PDB numbers. Structures derived from non-human species are marked with an asterisk.

(B) Overview of the location of major parts of various B-specific proteins.

(C) Location of the N-terminal region (NTR) and the C-terminal WD40 domain of Smu1 (in the unmasked density) at the interface between U2 SF3B3 and Brr2. The short  $\alpha$  helix of RED (aa 209–222) that binds Smu1's NTR is shown.

(D) Smu1's WD40 domain is located at the interface between Brr2's NC and CC domains and contacts the WD40-2 domain of U2 SF3B3. WH, wing helix.

(E) Close up of the fit of the N-terminal helical domain of Prp38 and associated  $\alpha$  helices of Snu23 and MFAP1, and the N-terminal  $\alpha$  helix of FBP21, near the extended U6 ACAGA/5'ss helix. Black asterisks: amino acid positions of the indicated B-specific proteins.

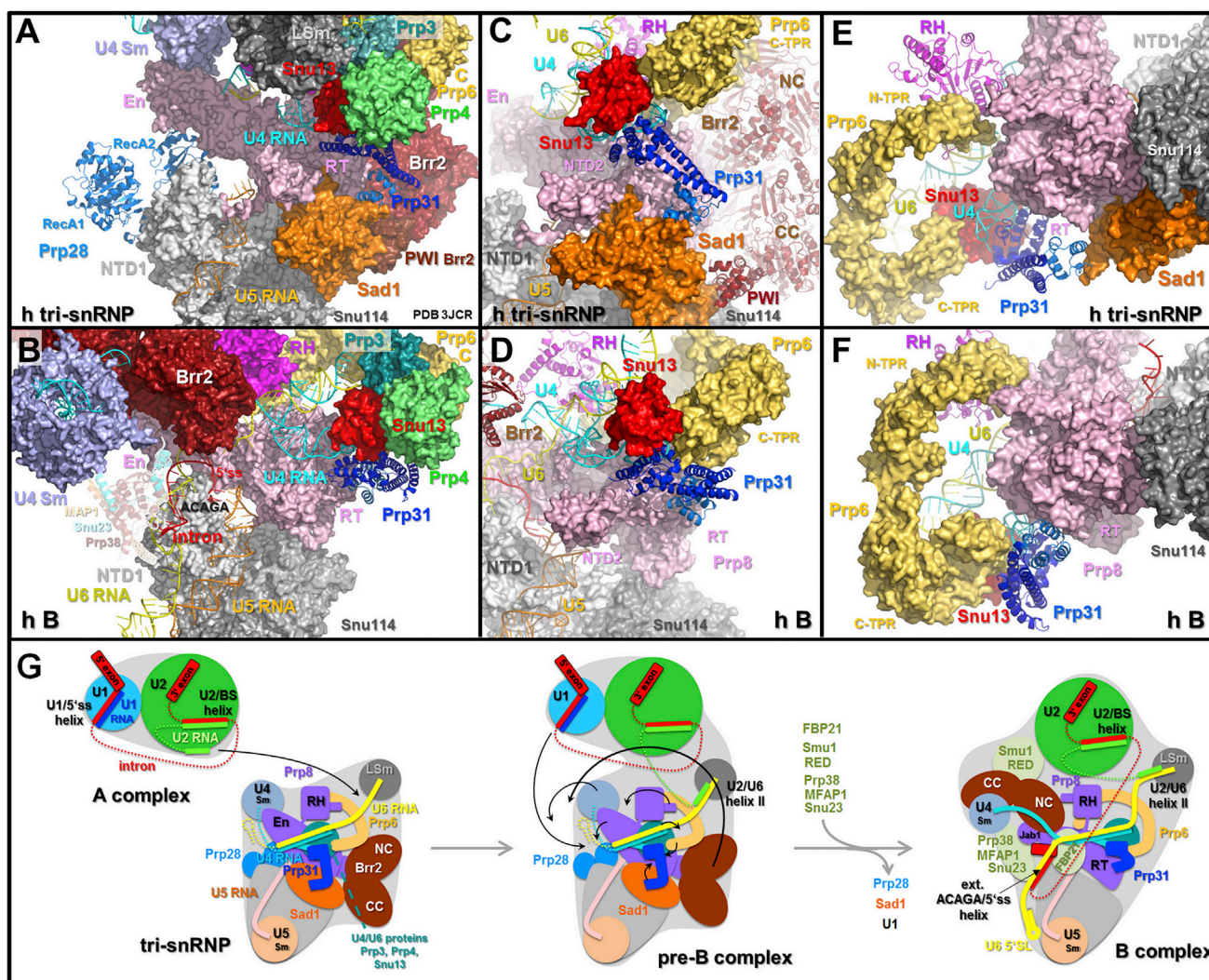
(F) Interaction of the FBP21 N-terminal zinc finger (ZnF) with the U6 ACAGA/5'ss helix and of Snu23's ZnF with the extended region of the U6 ACAGA/5'ss helix. See also Figure S6 and Tables S1 and S2.

early stage of the B to B<sup>act</sup> transition, but in which the U4/U6 duplex has been unwound (Sidarovich et al., 2017).

### Tri-snRNP Rearrangements during Spliceosome Assembly Likely Involve Coordinated Remodeling Events

Comparison of the 3D structures of the B complex and human U4/U6.U5 tri-snRNP reveals that most tri-snRNP proteins undergo extensive repositioning and/or structural rearrangements during B complex formation (Figure 6G). These rearrangements likely occur in a highly coordinated and ordered manner, as the interaction sites of several components in the tri-snRNP and B complex are mutually exclusive. An intriguing question is how these coordinated remodeling events are triggered. Initial docking of the tri-snRNP to the A complex first generates a 37S pre-B complex. In the latter, U1 snRNA is still base paired to the 5'ss, and U2/U6 helix II has formed, but the tri-snRNP is not yet stably bound (Boesler et al., 2016) (Figures 6G and S1). Displacement of U1 from the 5'ss and establishment of the U6 ACAGA/5'ss interaction, which is mediated by the helicase Prp28, leads to a substantial structural change that converts

the 37S pre-B complex into a 50S B complex with stably associated tri-snRNP (Boesler et al., 2015, 2016). Thus many of the rearrangements in tri-snRNP proteins described here likely occur during the pre-B to B transition, and formation of the U6 ACAGA/5'ss helix likely plays a key role in tri-snRNP remodeling. This suggests that Prp28 action may trigger or contribute to tri-snRNP rearrangements, and would likely create new conformational space and enable sampling of different RNP conformations within the spliceosome. The handover of U1 for U6 at the 5'ss likely influences the movement of the RT/En domain toward the Prp8 NTD1. In human tri-snRNPs, Prp28 is located between Prp8's RT/En and NTD1 (Figure 6A). The position of Prp28 in the tri-snRNP, and both the partially closed conformation of Prp8 and the formation/position of the U6 ACAGA/5'ss helix in the B complex, are mutually exclusive (Figures 6A and 6B). Prp28-mediated rearrangements must thus be coordinated with Prp28 displacement from the spliceosome. Interestingly, displacement of Prp28 is likewise a prerequisite for subsequent binding of the Prp38/Snu23/MFAP1 protein complex, whose binding site close to the U6/5'ss helix is also mutually exclusive with that of Prp28 (Figures 5B, 6A, and 6B).



**Figure 6. Coordinated Remodeling of Tri-snRNP Proteins and Displacement of Sad1 during B Complex Formation**

(A and B) Overview of the location of various tri-snRNP proteins and snRNAs in the isolated human tri-snRNP (A) and B complex (B), aligned relative to Prp8 NTD1 and Snu114. Whereas Sad1 bridges Prp8, Brr2 (via Brr2's PWI domain), and Prp31 in the tri-snRNP, it is missing in the B complex.

(C–F) Side (C and D) and top views (E and F) of the tri-snRNP and B complex, showing the remodeling of Prp31's coiled-coil domain upon/after Sad1 displacement (C and E) and creation of new binding sites for Prp6's C-terminal TPRs with the rearranged Prp31 domain in the B complex (D and F). In the tri-snRNP (C and E), Prp31's coiled-coil tip  $\alpha$ -helices (light blue) bind to Sad1, whereas in the B complex (D and F) they interact with Prp8's RT domain in a mutually exclusive manner.

(G) Schematic overview of tri-snRNP rearrangements upon its integration into the spliceosome. Selected proteins and RNAs are shown schematically in the human tri-snRNP (left panel), after initial docking of the tri-snRNP in the pre-B complex (middle panel) and after stable tri-snRNP integration in the pre-catalytic B complex (right panel). Major protein domain movements occurring prior to/during B complex formation are indicated by black arrows in the middle panel. See also Tables S1 and S2.

The Sad1 protein likely also plays a key role in triggering tri-snRNP rearrangements. In the tri-snRNP (Figure 6A), Sad1 contacts the Prp31 coiled-coil domain and the Prp8 NTD2 and RT domains, as well as Snu114 and the Brr2 N-terminal PWI domain, which in turn is associated with Brr2's CC cassette (Agafonov et al., 2016). This suggests that Sad1 not only plays a role in stabilizing the interaction of the U4/U6 and U5 snRNPs, but that it may also help to tether Brr2 in a pre-activation position, i.e., away from the U4/U6 duplex. Strikingly, in the B complex cryo-EM structure we cannot localize any density accommodat-

ing Sad1, consistent with its underrepresentation in our B complex preparation. Thus Sad1 is clearly displaced from its original position in the tri-snRNP, which would allow most of the rearrangements in tri-snRNP proteins observed upon B complex formation. For example, in the B complex, the Prp31 coiled-coil domain is rearranged and is now juxtaposed with Prp3's Nop domain, whereas the  $\alpha$  helices at the coiled-coil tip now directly interact with the Prp8 RT domain (Figures 6C and 6D). This creates new binding sites for Prp6's C-terminal TPRs on Prp31, which would then facilitate the observed shift of the U4/U6





structural rearrangement. Consistent with this, Sad1 is no longer present in the purified yeast tri-snRNP, which could potentially shift the equilibrium between different conformational states toward the B complex structural organization (Huang et al., 2014). This would imply that the tri-snRNP rearrangement that leads to the loading of Brr2 onto its U4/U6 RNA substrate is more highly regulated in higher eukaryotes. Consistent with this idea, Prp4 kinase, which is essential for B complex formation in higher eukaryotes, is absent from *S. cerevisiae*.

Our results indicate that there are significant differences between the activation pathways of human and yeast spliceosomes. The presence of a number of additional proteins in the human B complex, which have no counterparts in *S. cerevisiae* and appear to function first during activation, is a first indication that the transformation of B into an activated B<sup>act</sup> complex is likely to be more complex in higher eukaryotes. Our cryo-EM structure of the human B complex further reveals that at least two of these proteins, namely, FBP21 and Smu1, contact Brr2 and potentially play important roles in regulating its helicase activity. Thus, the initial steps toward the formation of an activated spliceosome, namely, triggering U4/U6 duplex unwinding, appear to be different in the *S. cerevisiae* versus human spliceosome.

After submitting this paper, a cryo-EM structure of the *S. cerevisiae* spliceosomal B complex was reported (Plaschka et al., 2017). A comparison of the molecular architecture of the human and yeast B complexes, reveals that the structure and organization of most proteins common to the yeast and human B complexes are conserved. The location of most human B complex proteins in our model that have homologs in *S. cerevisiae*, for which only short pieces of structural information were available, e.g., MFAP1 (Spp381 in *S. cerevisiae*) or whose position was supported mainly by protein-protein crosslinking, is consistent with the position of their homologs in the yeast B complex structure. A notable difference between the yeast and human B complex structures is the lack of a U6 ACAGA/5'ss interaction in the yeast B complex despite that U1 snRNP is largely absent, which may indicate that the yeast complex is stalled at an earlier stage, but after release of U1.

### Conclusions

Our cryo-EM structure of the human B complex provides important insights into the organization of the spliceosome prior to its activation. The B complex possesses a U2-containing head domain attached via three main bridges (B1-B3) to the main body where the U4/U6.U6 tri-snRNP and B-specific proteins are located. B1, which contains U2/U6 helix II, is likely involved in the initial docking of the tri-snRNP to the spliceosomal A complex. B2 may be important for stabilizing the association of the tri-snRNP within the B complex, whereas B3 allows communication between the U2/BS helix in the head and the extended U6 ACAGA/5'ss helix in the main body of the B complex. One of the most striking observations in this work is how tri-snRNP integration into the B complex leads to extensive rearrangements and/or repositioning of the majority of its proteins. These rearrangements lead, among others, to a partially closed conformation of Prp8 such that its RT/En and NTD1 domains, together with Dim1, form a protein pocket that binds the 5'ss. This is likely fol-

lowed by the movement of the U4/U6 snRNP and Brr2 helicase, such that the latter now contacts its U4/U6 substrate and is poised for the subsequent spliceosome activation step. Thus, these rearrangements generate the functional centers required to prime the B complex for the subsequent activation step. In addition, they create binding sites for the B-specific proteins, ensuring that they are first recruited at this specific stage of spliceosome assembly. Our structure also reveals how several of the human B-specific proteins are organized in the B complex, and suggests they are involved in negatively regulating Brr2, positioning the U6/5'ss helix and/or stabilizing the B complex structure. Finally, our work solidifies the intriguing idea that there is a significant difference between the early events of the activation pathways of human and yeast spliceosomes. Consistent with the large exchange of spliceosomal proteins and major rearrangements in the RNA network that occur during the B to B<sup>act</sup> transition, the general architecture of the B and B<sup>act</sup> complex are dramatically different (Fabrizio et al., 2009). The exchange of more than 40 proteins in the human system will likely occur in multiple steps, with multiple intermediate complexes formed during the transformation of B to an activated B<sup>act</sup> complex, hindering conclusions about the direction or trajectory of most protein rearrangements, as well as about the sequence of remodeling events at the atomic level, during the B to B<sup>act</sup> transition. Future elucidation of the molecular architecture of the human B<sup>act</sup> complex, as well as complexes at intermediate stages of activation, will thus greatly aid in elucidating the RNP remodeling events during spliceosome activation in higher eukaryotes.

### 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
  - In vitro splicing
  - MS2 affinity selection of splicing complexes
  - 2D gel electrophoresis and mass spectrometry
  - ATP sensitivity of purified B complexes
  - Chase of purified B complexes with nuclear extract
  - Crosslinking of the B complex and crosslink identification
  - EM and image processing
  - Model fitting and building
- DATA AND SOFTWARE AVAILABILITY

### SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and one movie and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2017.07.011>.

### AUTHOR CONTRIBUTIONS

D.E.A. and M.N.L. developed the purification strategy and characterized the B complex. O.D. and H.U. analyzed protein-protein crosslinking data. K.B.

D.H., and H.S. prepared grids and collected the EM data. B.K., K.B., H.S., and R.L. analyzed the structure. All authors were involved in data interpretation. The manuscript was written by R.L., B.K., and C.L.W. with input from all authors. R.L., D.E.A., B.K., and H.S. initiated and orchestrated the project.

## ACKNOWLEDGMENTS

We thank T. Conrad for HeLa cell production in a bioreactor; H. Kohansal for preparing HeLa cell nuclear extract; U. Steuerwald, W. Lendeckel, M. Raabe, and U. Pleßmann for excellent technical assistance; and K. Hartmuth for advice and discussions. Majety N. Leelaram was supported by a fellowship from the Alexander von Humboldt Stiftung. This work was supported by the Max Planck Society and the Deutsche Forschungsgemeinschaft (grants SFB 860 to R.L., H.S., and H.U. and LU294/15-1 to R.L.)

Received: May 14, 2017

Revised: June 22, 2017

Accepted: July 11, 2017

Published: August 3, 2017

## 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., Odenwälder, 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 method. *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.
- Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323.
- Boesler, C., Rigo, N., Agafonov, D.E., Kastner, B., Urlaub, H., Will, C.L., and Lührmann, R. (2015). Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site. *RNA* **21**, 1993–2005.
- Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Lührmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* **7**, 11997.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., and Richardson, D.C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
- 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., 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.
- Galej, W.P., Oubridge, C., Newman, A.J., and Nagai, K. (2013). Crystal structure of Prp8 reveals active site cavity of the spliceosome. *Nature* **493**, 638–643.
- Galej, W.P., Nguyen, T.H.D., Newman, A.J., and Nagai, K. (2014). Structural studies of the spliceosome: zooming into the heart of the machine. *Curr. Opin. Struct. Biol.* **25**, 57–66.
- 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.
- Gozani, O., Feld, R., and 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.
- 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.
- Huang, X., Beullens, M., Zhang, J., Zhou, Y., Nicolaescu, E., Lesage, B., Hu, Q., Wu, J., Bollen, M., and Shi, Y. (2009). Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). *J. Biol. Chem.* **284**, 25375–25387.
- Huang, Y.-H., Chung, C.-S., Kao, D.-I., Kao, T.-C., and Cheng, S.-C. (2014). Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. *Mol. Cell. Biol.* **34**, 210–220.
- 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.
- Korneta, I., Magnus, M., and Bujnicki, J.M. (2012). Structural bioinformatics of the human spliceosomal proteome. *Nucleic Acids Res.* **40**, 7046–7065.
- Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc. Natl. Acad. Sci. USA* **95**, 4188–4192.
- Leung, A.K.W., Nagai, K., and Li, J. (2011). Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis. *Nature* **473**, 536–539.
- Liu, S., Li, P., Dybkov, O., Nottrott, S., Hartmuth, K., Lührmann, R., Carlomagno, T., and Wahl, M.C. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science* **316**, 115–120.
- Liu, S., Ghalei, H., Lührmann, R., and Wahl, M.C. (2011). Structural basis for the dual U4 and U4atac snRNA-binding specificity of spliceosomal protein hPrp31. *RNA* **17**, 1655–1663.
- Liu, S., Mozaffari-Jovin, S., Wollenhaupt, J., Santos, K.F., Theuser, M., Dunin-Horkawicz, S., Fabrizio, P., Bujnicki, J.M., Lührmann, R., and Wahl, M.C. (2015). A composite double-/single-stranded RNA-binding region in protein Prp3 supports tri-snRNP stability and splicing. *eLife* **4**, e07320.
- Macias, 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.
- Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.-H., Will, C.L., Urlaub, H., Wahl, M.C., and Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* **26**, 2422–2434.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Lührmann, R., and Wahl, M.C. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* **341**, 80–84.
- Nguyen, T.H.D., Galej, W.P., Bai, X.C., 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.C., 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.
- Papasaiakas, P., and Valcárcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem. Sci.* **41**, 33–45.
- Papasaiakas, P., Tejedor, J.R., Vigevani, L., and Valcárcel, J. (2015). Functional splicing network reveals extensive regulatory potential of the core spliceosomal machinery. *Mol. Cell* **57**, 7–22.



- 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.
- Ragunathan, P.L., and Guthrie, C. (1998). A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* **279**, 857–860.
- 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.
- Reidt, U., Wahl, M.C., Fasshauer, D., Horowitz, D.S., Lührmann, R., and Ficner, R. (2003). Crystal structure of a complex between human spliceosomal cyclophilin H and a U4/U6 snRNP-60K peptide. *J. Mol. Biol.* **331**, 45–56.
- Reuter, K., Nottrott, S., Fabrizio, P., Lührmann, R., and Ficner, R. (1999). Identification, characterization and crystal structure analysis of the human spliceosomal U5 snRNP-specific 15 kD protein. *J. Mol. Biol.* **294**, 515–525.
- Reyes, J.L., Gustafson, E.H., Luo, H.R., Moore, M.J., and Konarska, M.M. (1999). The C-terminal region of hPrp8 interacts with the conserved GU dinucleotide at the 5' splice site. *RNA* **5**, 167–179.
- Santos, K.F., Jovin, S.M., Weber, G., Pena, V., Lührmann, R., and Wahl, M.C. (2012). Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc. Natl. Acad. Sci. USA* **109**, 17418–17423.
- Schellenberg, M.J., Wu, T., Ritchie, D.B., Fica, S., Staley, J.P., Atta, K.A., LaPointe, P., and MacMillan, A.M. (2013). A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat. Struct. Mol. Biol.* **20**, 728–734.
- Scheres, S.H.W. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530.
- Schneider, M., Hsiao, H.-H., Will, C.L., Giet, R., Urlaub, H., and Lührmann, R. (2010). Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. *Nat. Struct. Mol. Biol.* **17**, 216–221.
- Schütze, T., Ulrich, A.K.C., Apelt, L., Will, C.L., Bartlick, N., Seeger, M., Weber, G., Lührmann, R., Stelzl, U., and Wahl, M.C. (2016). Multiple protein-protein interactions converging on the Prp38 protein during activation of the human spliceosome. *RNA* **22**, 265–277.
- Sidarovich, A., Will, C.L., Anokhina, M.M., Ceballos, J., Sievers, S., Agafonov, D.E., Samatov, T., Bao, P., Kastner, B., Urlaub, H., et al. (2017). Identification of a small molecule inhibitor that stalls splicing at an early step of spliceosome activation. *eLife* **6**, e23533.
- Singer, A., Coifman, R.R., Sigworth, F.J., Chester, D.W., and Shkolnisky, Y. (2010). Detecting consistent common lines in cryo-EM by voting. *J. Struct. Biol.* **169**, 312–322.
- Spartz, A.K., Herman, R.K., and Shaw, J.E. (2004). SMU-2 and SMU-1, *Caenorhabditis elegans* homologs of mammalian spliceosome-associated proteins RED and fSAP57, work together to affect splice site choice. *Mol. Cell. Biol.* **24**, 6811–6823.
- Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326.
- Staley, J.P., and Guthrie, C. (1999). An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3**, 55–64.
- Ulrich, A.K.C., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the *Saccharomyces cerevisiae* Spp381 splicing factor. *BMC Evol. Biol.* **17**, 91.
- Ulrich, A.K.C., Schulz, J.F., Kamrad, A., Schütze, T., and Wahl, M.C. (2016a). Structural basis for the functional coupling of the alternative splicing factors Smu1 and RED. *Structure* **24**, 762–773.
- Ulrich, A.K.C., Seeger, M., Schütze, T., Bartlick, N., and Wahl, M.C. (2016b). Scaffolding in the spliceosome via single  $\alpha$  helices. *Structure* **24**, 1972–1983.
- 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. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* **353**, 895–904.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C.L., and Shi, Y. (2016b). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* **351**, 466–475.
- Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* **28**, 2283–2292.
- Xie, J., Beickman, K., Otte, E., and Rymond, B.C. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J.* **17**, 2938–2946.
- 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.J., Zhu, M., Fan, S.B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.X., Chen, H.-F., et al. (2012). Identification of cross-linked peptides from complex samples. *Nat. Methods* **9**, 904–906.
- 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		
EM map of the human B complex, soft mask 2 (4.5 Å)	This paper	EMDB: 3766
EM map of the human B complex, soft mask 1 (5.4 Å)	This paper	EMDB: 3767
EM map of the human B complex, unmasked sharpened	This paper	EMDB: 3768
EM map of the human B complex, unmasked (9.9 Å)	This paper	EMDB: 3769
PDB coordinates of the human B complex	This paper	PDB: 5O9Z
Cryo-EM model of the U4/U6.U5 tri-snRNP of <i>Saccharomyces cerevisiae</i>	Wan et al., 2016b	PDB: 3JCM
Cryo-EM model of the human U4/U6.U5 tri-snRNP	Agafonov et al., 2016	PDB: 3JCR EMDB: 6581
Cryo-EM model of the U4/U6.U5 tri-snRNP of <i>Saccharomyces cerevisiae</i>	Nguyen et al., 2016	PDB: 5GAN
Cryo-EM model of the Bact complex of <i>Saccharomyces cerevisiae</i>	Rauhut et al., 2016	PDB: 5LQW
Cryo-EM model of the human C* complex	Bertram et al., 2017	PDB: 5MQF, EMD: 3545
Crystal structure of the human Prp4-CypH complex	Reidt et al., 2003	PDB: 1MZW
Crystal structure of the human Dim1 protein	Reuter et al., 1999	PDB: 1QGV
Crystal structure of the human Prp31-Snu13-U4 snRNA complex	Liu et al., 2007	PDB: 2OZB
Crystal structure of the human Prp31-Snu13-U4atac 5' stem loop complex	Liu et al., 2011	PDB: 3SIU
Crystal structure of Prp8-Aar2 complex from <i>S. cerevisiae</i>	Galej et al., 2013	PDB: 3ZEF
Crystal structure of the human PRP8 RNase H-like domain	Schellenberg et al., 2013	PDB: 4JK7
Crystal structure of the human Brr2 helicase with the Prp8 Jab1/MPN domain	Mozaffari-Jovin et al., 2013	PDB: 4KIT
Crystal structure of the human U4 snRNP core domain	Leung et al., 2011	PDB: 4WZJ
Crystal structure of the Prp3 ferredoxin-like domain bound to a fragment of U4/U6 di-snRNA of <i>S. cerevisiae</i>	Liu et al., 2015	PDB: 4YHW
Crystal structure of the Smu1-RED complex of <i>Caenorhabditis elegans</i>	Ulrich et al., 2016a	PDB: 5EN6, 5EN7
Crystal structure of the human Prp38-MFAP1 complex	Ulrich et al., 2016b	PDB: 5F5S
Crystal structure of the Snu23-Prp38-MFAP1 complex of <i>Chaetomium thermophilum</i>	Ulrich et al., 2016b	PDB: 5F5U, 5F5V
Experimental Models: Cell Lines		
Human: HeLa S3 cells	Helmholtz Center for Infection Research, Brunswick	N/A
Recombinant DNA		
Plasmid: pMBP-MS2	Macías et al., 2008	Addgene plasmid # 65104
Plasmid: pT7-MINX-M3	Boesler et al., 2016	N/A
Software and Algorithms		
Coot v0.8.3	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>
Gautomatch	Dr. Kai Zhang	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang/">http://www.mrc-lmb.cam.ac.uk/kzhang/</a>
Gctf	Zhang, 2016	<a href="http://www.mrc-lmb.cam.ac.uk/kzhang/">http://www.mrc-lmb.cam.ac.uk/kzhang/</a>

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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	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="http://www.pymol.org">http://www.pymol.org</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>
SpliProt3D	Korneta et al., 2012	<a href="http://iimcb.genesilico.pl/SpliProt3D/home/">http://iimcb.genesilico.pl/SpliProt3D/home/</a>
SWISS-MODEL suite	Guex and Peitsch, 1997	<a href="http://spdbv.vital-it.ch/disclaim.html">http://spdbv.vital-it.ch/disclaim.html</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>

**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, Reinhard Lührmann ([reinhard.luehrmann@mpibpc.mpg.de](mailto:reinhard.luehrmann@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.

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****In vitro splicing**

Uniformly [<sup>32</sup>P]-labeled, m<sup>7</sup>G(5')ppp(5')G-capped MINX pre-mRNA was synthesized in vitro by T7 runoff transcription. HeLa S3 cells were obtained from GBF, Braunschweig (currently Helmholtz Zentrum für Infektionsforschung, Braunschweig) and tested negative for mycoplasma.

To prepare splicing active nuclear extracts, HeLa cells were grown to a density of  $6.5 \times 10^6$  cells/ml and harvested by centrifuging for 10 min at 2000 rpm in a 8 × 2000 mL BIOS rotor (Thermo Scientific). Cells were washed twice with ice cold PBS (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 protease inhibitor cocktail tablets (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 complex assembly as described below without dialysis or freezing.

To isolate B complexes, splicing was performed with 5 nM of <sup>32</sup>P-labeled pre-mRNA and 20% (v/v) HeLa nuclear extract, in buffer containing 0.3 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 50 mM NaCl, 20 mM HEPES-KOH pH 7.9, 2 mM ATP and 20 mM creatine phosphate, and was incubated at 30°C for 2 hr.

**MS2 affinity selection of splicing complexes**

Spliceosomal complexes were isolated by MS2 affinity selection. MINX pre-mRNA containing three MS2 aptamers at its 3' end RNA was incubated with a tenfold molar excess of MBP-MS2 fusion protein and then added to a splicing reaction. After incubating at 30°C for 2 h, centrifuging to remove aggregates, and adding 100 mM NaCl, the reaction was loaded onto a MBP Trap HP column (GE Healthcare). 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, centrifuged at 25,000 rpm for 10 h at 4°C in a Surespin 630 (Thermo Scientific) rotor, and gradient fractions were harvested from the bottom. The distribution of <sup>32</sup>P-labeled MINX pre-mRNA across the gradient was



determined by Cherenkov counting. Fractions were analyzed by denaturing 4%–12% NuPAGE (Life Technologies) followed by autoradiography and SYBR Gold staining to detect RNA. Peak fractions containing B complexes were pooled, concentrated by centrifugation with an Amicon 50 kD cut-off unit, diluted to decrease the sucrose concentration and reloaded on the same gradients with glutaraldehyde as fixative (Kastner et al., 2008). The GraFix gradient contained 0%–0.15% of glutaraldehyde and fractions were quenched with 100 mM aspartate immediately after harvesting. For biochemical sample validation, the same procedure was performed but without fixation in the second gradient. The RNA and protein compositions of purified complexes were determined by denaturing PAGE and two-dimensional (2D) gel electrophoresis.

### 2D gel electrophoresis and mass spectrometry

2D gel electrophoresis of affinity-purified spliceosomal complexes was performed as described previously (Agafonov et al., 2011) using a 7.5% acrylamide mono gel in the second dimension for analysis of proteins larger than 50 kDa, or 15% acrylamide for proteins smaller than 50 kDa. For mass spectrometry, Coomassie-stained protein-spots were cut out of the 1D or 2D gels, and proteins were digested in-gel with trypsin and 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 NCBI non-redundant (nr) database using Mascot as a search engine.

### ATP sensitivity of purified B complexes

Affinity-purified B complexes formed on  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA were incubated with or without 2mM ATP at 30°C for 30 min. The integrity of the complexes was checked by sedimentation in a 5%–20% (w/v) sucrose gradient containing G-150 buffer followed by Bradford assay of gradient fractions to determine the peak. Aliquots of peak fractions were separated by denaturing 4%–12% NuPAGE (Life Technologies) and RNA was visualized by SYBR Gold staining.

### Chase of purified B complexes with nuclear extract

Affinity-purified B complexes formed on  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA were incubated with splicing buffer alone (20 mM HEPES-KOH pH 7.9, 50 mM NaCl, 3 mM  $\text{MgCl}_2$ , 2 mM ATP, 20 mM creatine phosphate) or additionally in the presence of 20% untreated HeLa nuclear extract or extract pre-treated with micrococcal nuclease (NEB) as described previously (Bertram et al., 2017). The splicing reaction was initiated by addition of  $^{32}\text{P}$ -labeled MINX-MS2 pre-mRNA or purified B complexes, and then incubated at 30°C for 0–90 min. Time point aliquots were analyzed by SDS-PAGE, and the  $^{32}\text{P}$ -labeled RNA was visualized with a Typhoon phosphorimager (GE Healthcare).

### Crosslinking of the B complex and crosslink identification

After gradient centrifugation, MS2 affinity-purified spliceosomal complexes were crosslinked with 150  $\mu\text{M}$  BS3 for 30 min at 20°C and further purified by a second gradient centrifugation step. Approximately 25 pmol of B complexes were pelleted by ultracentrifugation and analyzed essentially as described before (Bertram et al., 2017). After digestion with trypsin, peptides were reverse-phase extracted and fractionated on a Superdex Peptide PC3.2/30 column (GE Healthcare). 50  $\mu\text{L}$  fractions corresponding to an elution volume of 1.2–1.8 ml were analyzed on Thermo Scientific Q Exactive HF, Orbitrap Fusion Tribrid or Orbitrap Fusion Lumos Tribrid mass spectrometers. Protein–protein crosslinks were identified by pLink 1.23 search engine (<http://pfind.ict.ac.cn/software/pLink>) and filtered at FDR 1% 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.,  $\text{Score} = -\log_{10}(\text{pLink Score})$ . For model building, a maximum distance of 30 Å between the  $C\alpha$  atoms of the crosslinked lysines was allowed.

### EM and image processing

A negative-stain starting model was built and refined essentially as described previously (Bertram et al., 2017; Singer et al., 2010). 8,168 cryo-images were recorded at  $-193^\circ\text{C}$  in a Titan Krios electron microscope (FEI Company, the Netherlands) on a Falcon III direct electron detector at 120,700x magnification resulting in a pixel size of 1.16 Å at the specimen level. 20 frames were recorded for each micrograph with an average dose of 1.5  $e^-$  per frame per  $\text{Å}^2$ . Motion correction and spatial frequency weighted frame summation was achieved using the MotionCor2 software (Zheng et al., 2017) (<http://www.msg.ucsf.edu/em/software/index.html>). Summed micrograph images were then evaluated based on real space appearance and CTF parameters and only ca 6,000 with good contrast and isotropic Thon rings were used for particle picking and extraction. Using the particle picking software Gautomatch (<http://www.mrc-lmb.cam.ac.uk/kzhang/>) and 40° projections of the negative stain model filtered to 40 Å as a reference, we extracted ~550,000 particle images from the pre-sorted, dose-weighted cryo-micrographs and applied several particle sorting steps at the 2D and 3D level. 2D multivariate statistics and classification were first applied to the non-aligned particle images and subsequently to the aligned particles. In each round, only particles comprising better resolved classes were included in further processing. The remaining ~407,000 particles were then re-extracted using RELION 2.0 and coarsed 2x in the process. 3D classification of these particles in RELION, featuring ten classes, resulted in ~49,000 particles in the overall best defined class. Subsequent rounds of 3D refinement followed by 2D classifications, where particles from the most poorly defined class-sums in RELION were discarded, yielded a 2x coarsed model at 9.9 Å. For the highest resolution structure, the ~44,600 remaining particles were re-extracted at their native pixel size using coordinates refined by previous rounds of classification and consequently utilized for an additional round of

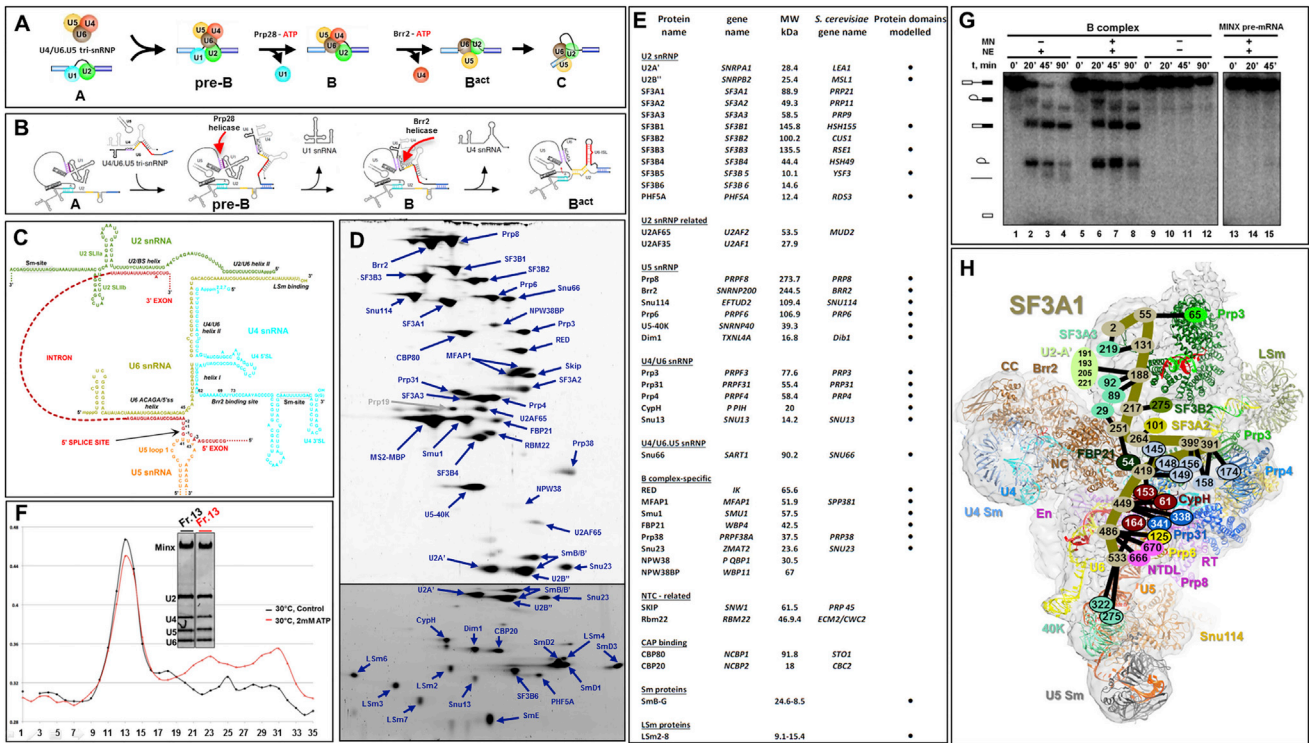
refinement yielding a 9.3 Å resolution unmasked structure. A soft mask (soft mask 1, Figure S2) that included the density harboring Brr2 helicase and the U4 Sm core, with a cut-off of 7 voxel was then used for the refinement and for the determination of resolution. A map with a resolution of 5.4 Å as determined by Fourier shell correlation calculated from two independent datasets with a threshold of 0.143, was obtained. A second soft mask, that excluded the more flexible density region of Brr2 helicase and the U4 Sm core (soft mask 2, Figure S2) was applied during refinement in a similar way as described above and yielded an additional map of this area with a resolution of 4.5 Å as determined by Fourier shell correlation calculated from two independent datasets with a threshold of 0.143. A local resolution plot revealed that there are indeed areas of higher resolution at the RNP core of the B complex that approach 3.5 Å. Some peripheral regions have somewhat lower resolution (Figures 1 and S2). To obtain the most complete B complex structure at the overall highest possible resolution, sorting schemes were adjusted accordingly. Using a starting model that did not include the head of the B complex during 3D classification, the ~407,000 2x coarsened particles (see above) were split into six equally sized groups and 3D classified accordingly, with three classes each. Particles from all classes yielding a (partially) defined structure, regardless of the appearance of the head, were pooled and split again into two subsets of ~120,000 particles each. Both particle sets were then again 3D classified into 6 classes each and inspected visually. Particles from the two best defined classes were then pooled and 3D refined without masking in RELION, yielding a model revealing the best achievable head definition at 9.9 Å resolution according to the FSC 0.143 criterion (which we refer to as the unmasked EM density map).

### Model fitting and building

Available X-ray or homology models of proteins were fit into the EM density using Chimera (Pettersen et al., 2004). Individual models of substructures (e.g., domains or structural motifs) were further fitted as rigid bodies using Coot (Emsley and Cowtan, 2004). The models were adjusted manually to fit into the EM density after visual inspection. Disordered regions were removed and regions that were reorganized or were not present in the initial models (e.g., loops and secondary structure elements) were built using Coot. Homology models of proteins were either obtained using the SWISS-MODEL suite (Guex and Peitsch, 1997), or were directly adapted from the SPlIProt3D database (Korneta et al., 2012). Details of the processing of protein models incorporated into the B structure are described in Table S1. An initial model of the human U5 snRNA was obtained from the C\* complex (PDB: 5MQF, Bertram et al., 2017) and fitted into the 4.5 Å EM density using Coot. MINX pre-mRNA intron bases A56 and G58 could subsequently be modeled as base-pairing with U5 loop 1 bases U41 and U43. This then allowed the placement of all other modeled pre-mRNA bases by tracing in the EM density. U4 snRNA A1-C16 were initially modeled as an idealized double helix base paired with U6 snRNA G59-U74 (helix II) and subsequently refined in Coot. U4 snRNA nucleotides A20-U52 were modeled in a similar manner, based on its crystal structure (PDB: 2OZB, Liu et al., 2007). Missing U4 bases up to U62 were then modeled in Coot by tracing the EM density. The remaining U4 snRNA bases up to A68 were modeled from its 3' terminal helix (which was modeled as an idealized helix) by tracing their path in the EM density. U6 snRNA bases preceding U4/U6 stem II and comprising the helical ACAGA and extended ACAGA box elements, were modeled tracing their path in the EM density using Coot beginning at A30. The U6 5' terminal loop and adjacent bases were adapted from the human C\* reference structure (PDB: 5MQF, Bertram et al., 2017). Once the entire coordinate model was built, all proteins were truncated to poly-alanine and a global minimization real space refinement was conducted against the 4.5 Å or 5.4 Å cryo-EM density, respectively, using the real space refine program from the PHENIX suite (Adams et al., 2010) ([https://www.phenix-online.org/documentation/reference/real\\_space\\_refine.html](https://www.phenix-online.org/documentation/reference/real_space_refine.html)). The RNA model was validated using the MolProbity server (Davis et al., 2007) and exhibited an all atom clash score of less than 12 and no bad bond-lengths or -angles. The scores for individual RNA nucleotides shown in the final B complex model are provided in a separate HTML file. Final visualization was performed with Chimera and PyMOL (<http://www.pymol.org>).

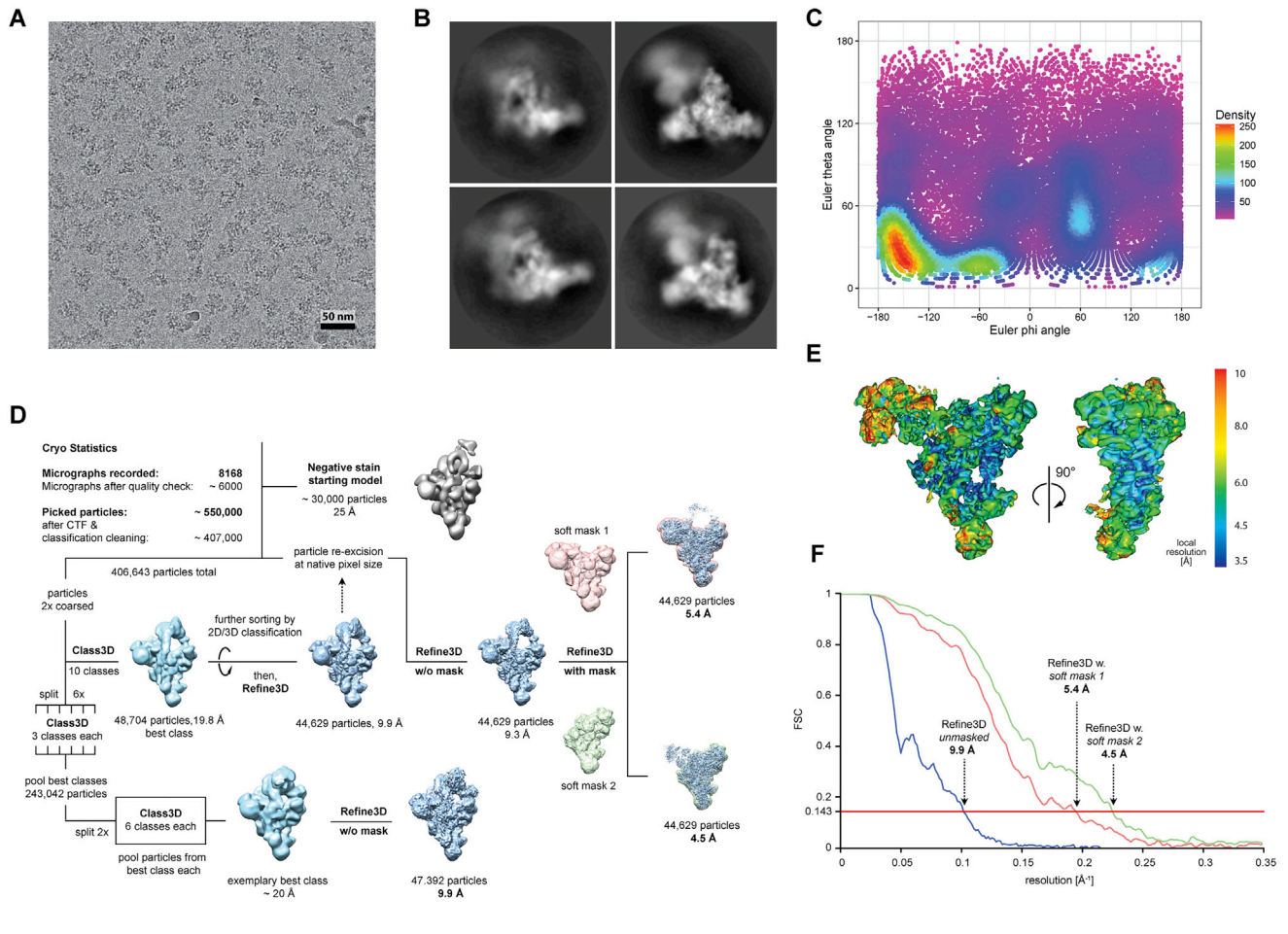
### DATA AND SOFTWARE AVAILABILITY

The cryo-EM maps have been deposited in the Electron Microscopy Data Bank with accession codes EMDB: 3766 (4.5 Å map), EMDB: 3767 (5.4 Å map), and EMDB: 3769 (unmasked 9.9 Å map). A sharpened version of the unmasked map is deposited under EMDB: 3768. The atomic model has been deposited in the Protein Data Bank under accession code PDB: 5O9Z.



**Figure S1. Compositional and Functional Characterization of Affinity-Purified Human B Complexes, Related to Figure 1**

(A) Spliceosome assembly pathway, starting with complex A.  
 (B) RNA-RNA rearrangements occurring during spliceosome activation and catalysis of splicing.  
 (C) Schematic representation of the secondary structure of RNA in the spliceosomal B complex. The complete secondary structure of the U4/U6 duplex is shown. Only selected regions of the pre-mRNA, U2 and U5 snRNAs are shown.  
 (D) Identification of abundant B complex proteins by 2D gel electrophoresis followed by mass spectrometry. Upper and lower panel: proteins > 30 kDa or < 30 kDa, respectively.  
 (E) Summary of proteins in purified, human B complexes. Proteins, or domains thereof, that were modeled into the B complex EM density map are indicated by a bullet point.  
 (F) Purified B complexes are stable in the presence of ATP. Affinity-purified B complexes formed on <sup>32</sup>P-labeled MINX-MS2 pre-mRNA were incubated with (red line) or without 2 mM ATP (black line) at 30°C for 30 min and fractionated by sedimentation on a 5%–20% (w/v) sucrose gradient. RNA in peak fraction #13 was separated by denaturing 4%–12% NuPAGE and visualized by SYBR Gold staining. B complexes incubated with ATP not only exhibit the same sedimentation behavior as the control B complexes, but also the same RNA composition, indicating that Brr2 helicase is negatively regulated in the purified B complexes and does not displace U4 from U6 snRNA in the presence of ATP.  
 (G) Chase of purified B complexes with nuclear extract. Affinity-purified B complexes formed on <sup>32</sup>P-labeled MINX-MS2 pre-mRNA were incubated with untreated HeLa nuclear extract (lanes 1–4) or extract pre-treated with micrococcal nuclease (MN) (lanes 5–8), or with splicing buffer alone (lanes 9–12). The pre-mRNA was efficiently chased into mRNA in the presence of both types of nuclear extract, but not buffer alone, indicating that our purified B complexes are functional and not dead-end complexes. As a control, <sup>32</sup>P-labeled MINX-MS2 pre-mRNA was incubated in the presence of extract pre-treated with MN (lanes 13–15). The pre-mRNA was not spliced with nuclear extract pre-treated with MN, confirming that the vast majority of snRNAs in the MN treated nuclear extract were destroyed. All 2D analyses and in vitro splicing experiments were performed at least twice in two independent experiments.  
 (H) Network of crosslinks between U2 SF3A1 and proteins in the head and the central part of the B complex body. Overview of the positions of selected proteins and RNA in the B complex EM density map (shown is the front view of the B complex; see Figure 1A, left panel), with a schematic diagram showing intermolecular crosslinks between U2 SF3A1 and other indicated proteins, and the likely path of SF3A1. Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Numbers in ovals with black borders indicate the residues in the modeled regions of the proteins, whereas those in ovals without borders are residues within non-modeled regions. The latter are arbitrarily placed close (less than 30 Å) to their crosslinking partners observed in our model. The ovals share the same color as that of the crosslinked protein's name. The crosslinking pattern suggests a path for U2 SF3A1 from the head region along the central axis of the main body close to its foot region, via bridge B3.



**Figure S2. Cryo-EM and Image Processing of the Human B Complex, Related to Figure 1**

(A) Typical cryo-EM raw image of *H. sapiens* B spliceosomes recorded with a Titan Krios (FEI Company) electron microscope at a nominal magnification of 120,700x with a Falcon III direct electron detector resulting in a pixel size of 1.16 Å.

(B) Several representative class averages showing different views of the B complex after reference free 2D classification.

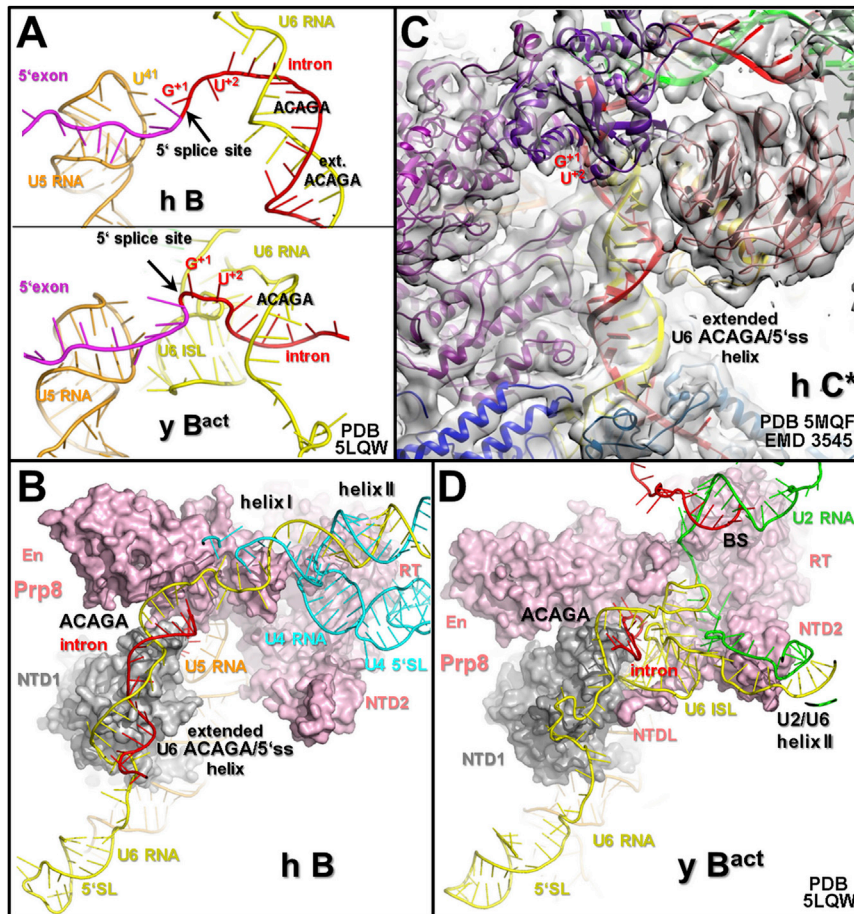
(C) Euler angle distribution of all particle images that contributed to the final high resolution 3D map. The coordinates describe the phi and theta angles. The color and size of the dots reflect the number of particles at any given Euler angle.

(D) Computational sorting scheme. Imaged micrographs were first evaluated according to their real space appearance and the Thon ring quality of local power spectra. Roughly 550,000 particle images were then selected from the remaining micrographs. In a second sorting step, particle images were again discarded based on the quality of Thon rings in classified, local power spectra. After evaluations in Fourier space, particles were subsequently excluded according to multiple rounds of 2D classifications. The remaining ca 407,000 particles were then coarsened 2x to undergo further classification in RELION 2.0. A 3D classification featuring 10 classes was then performed and 48,704 particles were extracted from the dataset. In subsequent rounds of RELION 2D and 3D refinement, only those particles comprising well resolved classes were selected from the dataset, resulting in a 3D volume reconstructed from 44,629 particles at 9.9 Å. To achieve a reconstruction at maximum resolution, only those particles contributing to this final 3D model were re-extracted from their micrographs at the native pixel size of 1.16 Å, using coordinates refined by previous rounds of classification. A further round of 3D refinement in RELION revealed a final, unmasked structure with 9.3 Å overall resolution. To improve details in the more stable areas of the complex, two soft masks (1 & 2) were applied separately, each during one subsequent round of 3D refinement, yielding a final model at 4.5 Å resolution according to the FSC 0.143 criterion. To find the overall best defined structure of the entire complex (including its head region), a modified sorting and classification scheme was applied. Using the 2x coarsened, pre-processed 407,000 single particle images and a 3D starting model lacking the head, the particle dataset was split into 6 equally sized groups and 3D classified into three classes each in RELION. After pooling all particles from the best defined classes, the particle dataset was split again into two equally sized datasets. Subsequent 3D classification of each individual particle dataset, using six classes each, followed by 3D refinement finally yielded the overall best defined structure at 9.9 Å.

(E) Local resolution plot of the complex refined with soft mask 1 reveals a resolution distribution from approximately 3.5 to 10 Å with some less well-defined areas at the periphery of the complex. Higher resolution regions (in blue, up to 3.5 Å resolution) were obtained for the centrally-located core of the spliceosome.

(F) Fourier-shell correlation function of two independently refined half datasets calculated during the 3D Refinement or PostProcessing procedure in RELION indicates a global resolution of 9.9 Å for the unmasked B complex. The same function calculated for the volumes refined with soft mask 1 and 2 reveals a resolution of 5.4 Å and 4.5 Å, respectively.



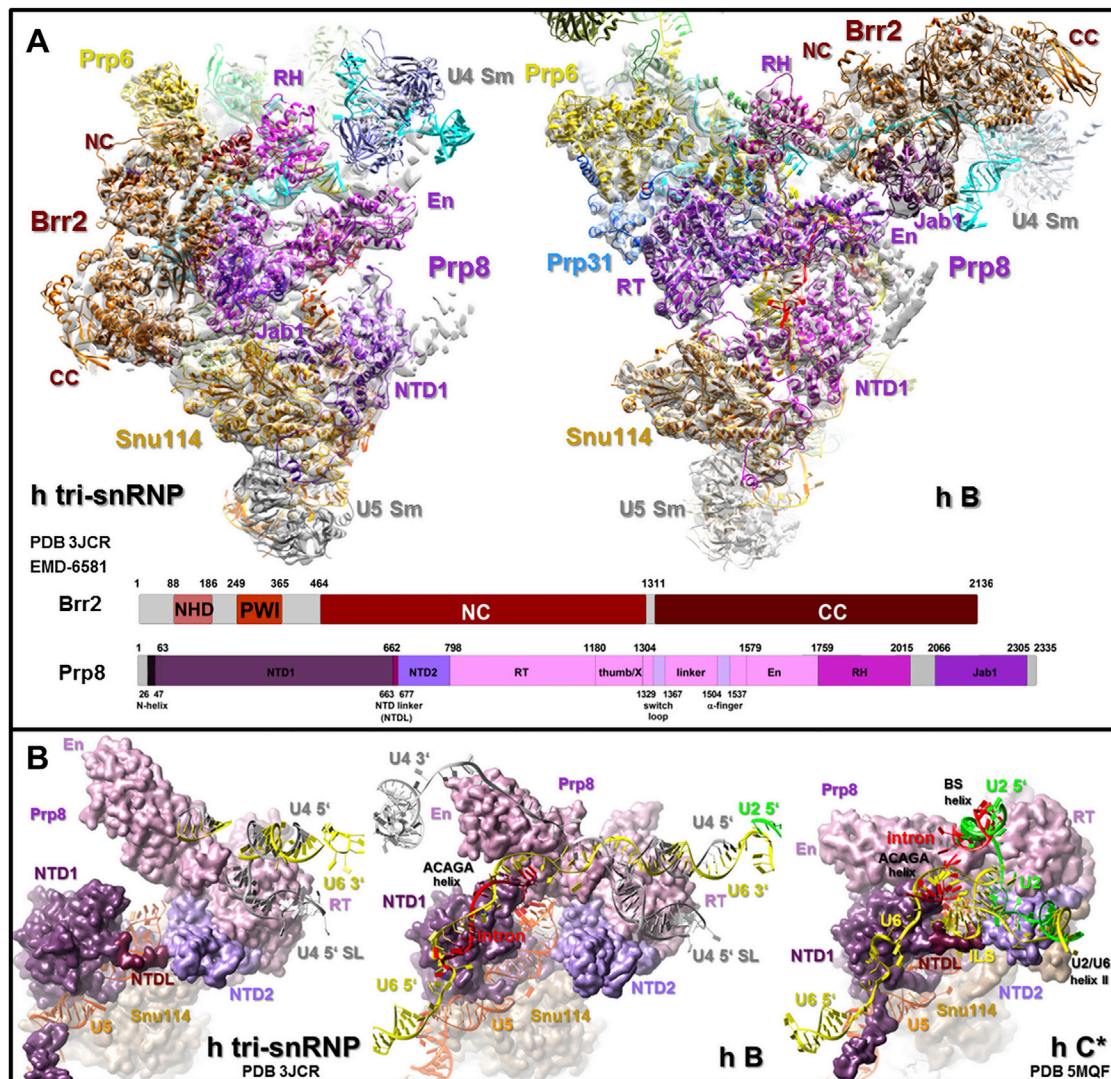


**Figure S3. Arrangements of U6 snRNA and the U6 ACAGA/5'ss Helix in Various Spliceosomal Complexes, Related to Figures 1 and 2**

(A) Substantial differences between the conformations of the first two nucleotides of the 5'ss and the positions of the U6 ACAGA/5'ss helices in the human B complex (upper panel) and the yeast B<sup>act</sup> complex (lower panel).

(B–D) An extended U6 ACAGA/5'ss helix is present in the human B (B) and C\* (C) complexes, but not in the yeast B<sup>act</sup> complex (D). Complexes in (B) and (D) are aligned relative to Snu114 and Prp8's NTD1. In panel C, the extended U6 ACAGA/5'ss helix from the human B complex was fit into the corresponding density of the C\* cryo-EM map. Panels B and D also show that the 5' stem-loop of U6 snRNA adopts already in the B complex a position very similar to its position at later functional stages of the spliceosome.

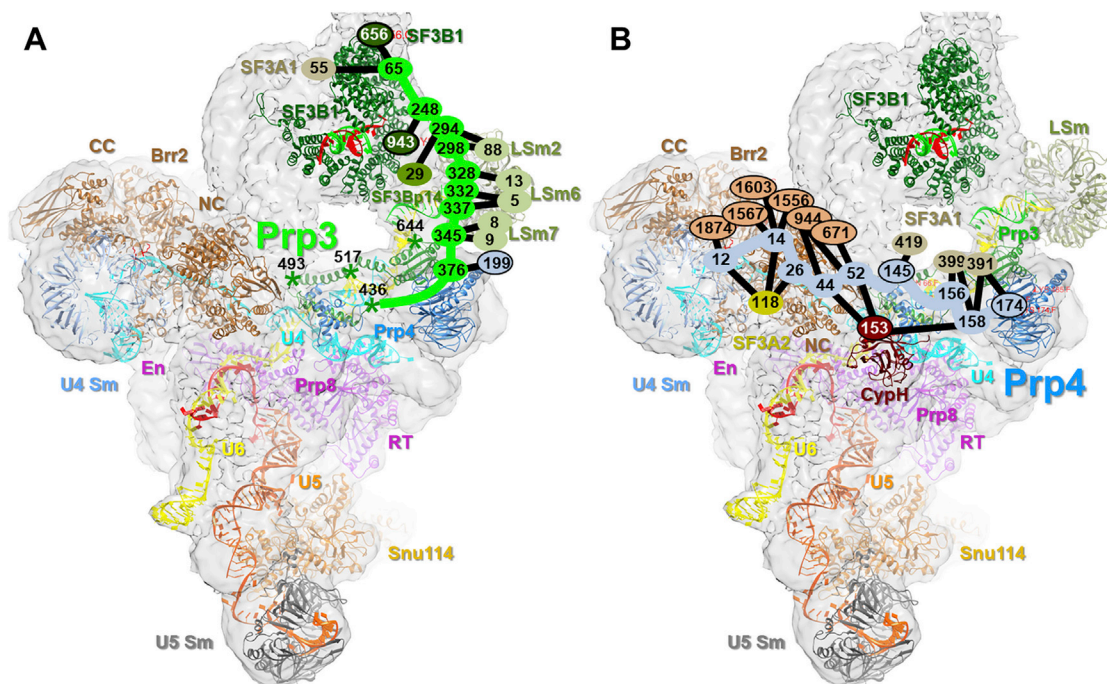




**Figure S4. Differential Arrangements of Selected U5 and U4/U6 Proteins in the Human Tri-snRNP, B and C\* Complex, Related to Figure 2**

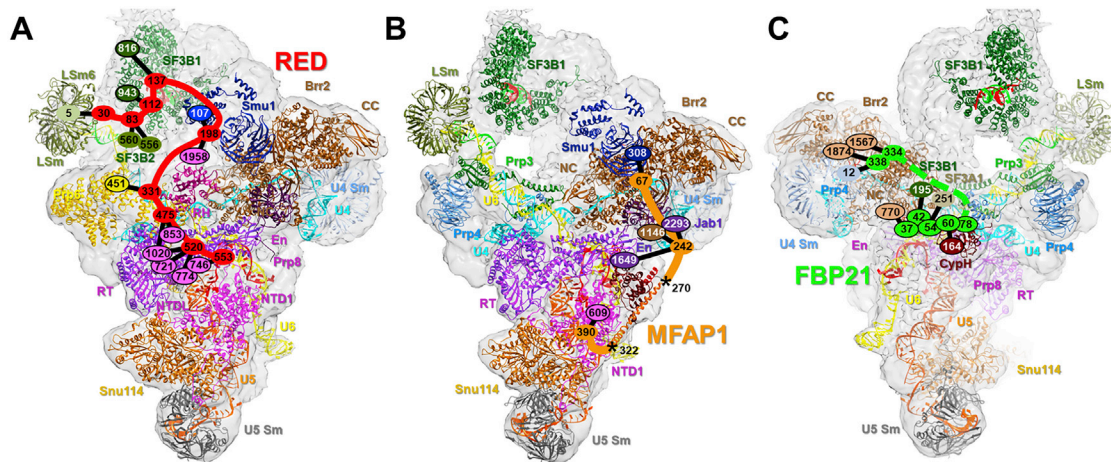
(A) Location of major U5 proteins in the isolated human tri-snRNP and the human B complex. While the U5 protein Snu114 and Prp8's NTD1 domain are structurally organized in a very similar manner in the tri-snRNP (left) and B complex (right), the helicase domain of Brr2 is located at radically different positions and is found at opposite ends of Prp8's RT/En domain in the two complexes. The domain organization of human Prp8 and Brr2 is shown at the bottom of panel A. Amino acids at the domain boundaries are indicated by numbers.

(B) Prp8 adopts significantly different conformations in the human tri-snRNP, B and C\* complexes. (Left panel) Open conformation of Prp8 in the human U4/U6.U5 tri-snRNP, where the En end of the elongated RT/En domain is well-separated from the upper region of the NTD1 domain (see also [Agafonov et al., 2016](#)). (Middle panel) Partially closed conformation of Prp8 in the human B complex. To achieve the B complex conformation, the RT/En domain must move toward the NTD1 domain, such that the En domain just touches the upper region of the NTD1 domain. (Right panel) Closed conformation of Prp8 in the human, catalytically-active C\* complex. To achieve this conformation, the RT/En domain must move even closer to the NTD1 domain, generating a large interface between the Prp8 NTD1 and RT/En domains, which now clamps the 5' exon instead of the 5'ss. In the closed conformation of Prp8, the pocket that accommodates the rearranged catalytic U2/U6 RNA network is generated (see also [Bertram et al., 2017](#)). All complexes are aligned relative to Snu114 and the Prp8 NTD1 domain. Prp8 also adopts a very similar closed conformation in the *S. cerevisiae* B<sup>act</sup>, C, and C\* complexes, as well as in the *S. pombe* intron-lariat spliceosome ([Fica et al., 2017](#); [Galej et al., 2016](#); [Hang et al., 2015](#); [Rauhut et al., 2016](#); [Wan et al., 2016a](#); [Yan et al., 2016, 2017](#)).



**Figure S5. Crosslinks of Prp3 and Prp4 with Other Proteins in the Human B Complex, Related to Figure 3**

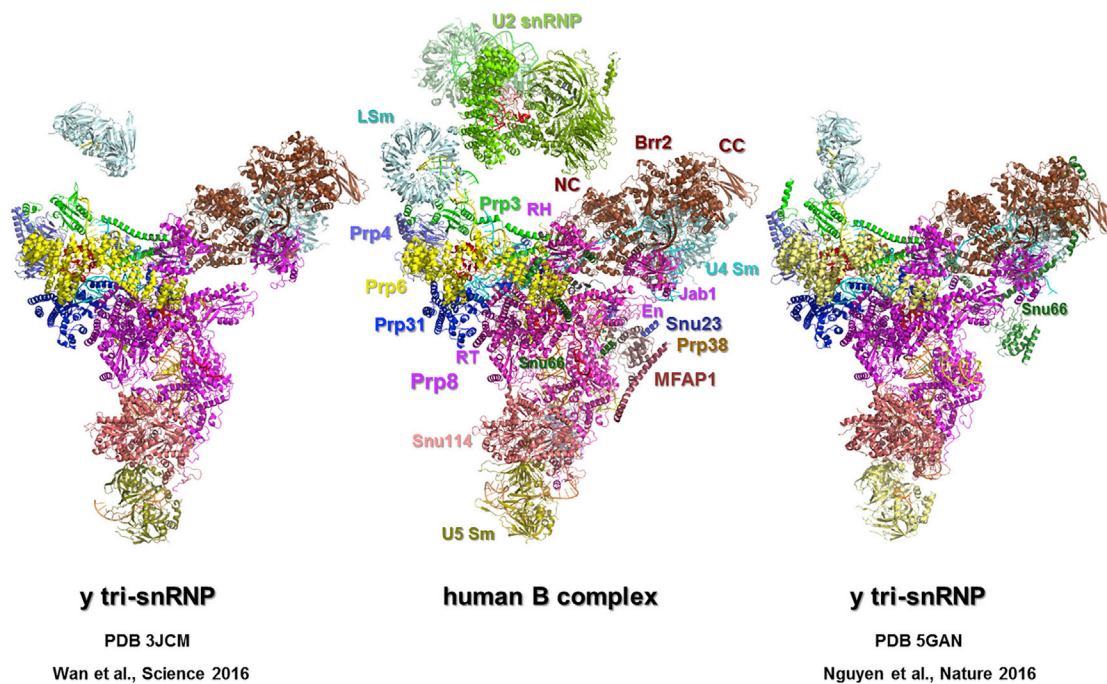
(A and B) Overview of the position of selected proteins and RNA in the B complex EM density map (shown is the front view of the B complex in both panels; see Figure 1A, left panel), with a schematic diagram showing intermolecular crosslinks between Prp3 (A) or Prp4 (B) and other proteins, as indicated. The likely paths of both proteins in the B complex are indicated by a thick line. Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Numbers in ovals with black borders indicate the residues in the modeled regions of the proteins, whereas those in ovals without borders are residues within non-modeled regions. The latter are arbitrarily placed close (less than 30 Å) to their crosslinking partners observed in our model. The ovals share the same color as that of the crosslinked protein's name. Black numbers with asterisks (\*) in (A) represent residues of Prp3  $\alpha$  helices that are modeled in the high resolution B complex structure.



**Figure S6. Crosslinks of RED, MFAP1, and FBP21 with Other Proteins in the Human B Complex, Related to Figure 5**

(A–C) Overview of the position of selected proteins and RNA in the B complex EM density map, with a schematic diagram showing intermolecular crosslinks between RED (panel A), MFAP1 (panel B), FBP21 (panel C) and other indicated proteins, and the likely paths of RED, MFAP1 and FBP21 (thick lines), respectively, in the B complex. The back view of the B complex is shown in panels A and B (see Figure 1A, right panel), whereas the front view is shown in panel C (see Figure 1, left panel). Numbers indicate the positions of crosslinked lysine residues (connected by black lines) in each protein. Numbers in ovals with black borders indicate the residues in the modeled regions of the proteins, whereas those in ovals without borders are residues within non-modeled regions. The latter are arbitrarily placed close (less than 30 Å) to their crosslinking partners observed in our model. The ovals share the same color as that of the crosslinked protein's name.





**Figure S7. Comparison of the Molecular Architecture of the *S. cerevisiae* U4/U6.U5 Tri-snRNP and Human B Complex, Related to Figure 7**

Overview of the structure and organization of proteins in the purified *S. cerevisiae* tri-snRNP as reported by Wan et al. (2016b) (left) or Nguyen et al. (2016) (right), and in the human spliceosomal B complex (middle) as determined by cryo-EM. All complexes are aligned relative to Snu114 and the Prp8 NTD1 domain. The coloring of proteins (labeled in the middle panel) is conserved in all three panels. The structural organization of Prp8, Brr2, Prp6 and the U4/U6 proteins in the yeast tri-snRNP is highly similar to their organization in the human B complex, consistent with the idea that the purified yeast tri-snRNPs are either activated at an early stage, i.e., before they join the spliceosome to form the B complex, or that they might potentially be derived by dissociation of spliceosomal B complexes. In addition, unlike human tri-snRNPs, purified yeast tri-snRNPs contain Prp38 and Snu23, but lack Prp28. Interestingly, one of the cryo-EM structures of a yeast tri-snRNP exhibits density elements, into which the yeast Prp38 NTD in complex with  $\alpha$  helices of Snu23 and Spp381 (the yeast likely homolog of MFAP1) would fit nicely. While this density was tentatively assigned to helical regions of yeast Snu66 by Nguyen et al. (2016), they more likely contain the yeast Prp38 complex, as they are located at the equivalent position as the hPrp38 complex in the human B complex and can be even superimposed. Indeed, in the recently published yeast B complex structure (Plaschka et al., 2017), this density was shown to comprise the yeast Prp38/Snu23/Spp381 protein complex.

## 4 Discussion

The spliceosome-dependent splicing of pre-mRNA is a highly conserved process of exceptional relevance to eukaryotic life. An understanding of the mechanistic details that underlie the procedure, in particularly regarding the human system, is therefore of critical importance. Facilitated by an exceptionally dynamic macromolecular machine, splicing is achieved in a stepwise manner that involves a wealth of structurally and compositionally diverse spliceosome complexes during different stages of the catalytic cycle. The mechanism of catalysis itself, as well as the molecular architecture specifically of the human spliceosome, however, remained poorly understood and elusive for decades following the discovery of the spliceosome. As the structure of any macromolecular machine is tightly coupled to its function, high-resolution models of individual spliceosome assemblies (snapshots) that participate in the splicing cycle are thus likely to contribute towards a better understanding of the mechanistic details and regulation of pre-mRNA splicing.

This study presents the first high-resolution structures of the human spliceosome at the pre-catalytic and catalytically activated stage of the splicing cycle. Utilizing cryo-EM as a tool for molecular imaging, 3D electron density maps of the human B and C\* complexes were solved and refined to overall resolutions of 4.5 Å and 5.9 Å, respectively. In combination with mass spectrometry-coupled crosslinking experiments, the EM densities enabled an unambiguous modelling of the most significant protein and snRNA factors involved in spliceosome assembly and catalysis. Compared to the 3D models of the corresponding complexes in yeast (Fica et al., 2017; Plaschka et al., 2017; Yan et al., 2017), the human spliceosome, particularly in its pre-catalytic state, reveals possible differences in regulation and assembly between both species.

Utilizing the multitude of molecular spliceosome architectures that have been recently unveiled (see text below), it has furthermore become possible to reclassify or validate the interpretation of many biochemical studies of the past. Due to the dynamic nature of the macromolecular machine, numerous important findings that were difficult to interpret without the structural knowledge of today can now be reconciled. The combination of biochemical characterisations and structural relationships is thus a particularly powerful approach in interpreting either results, as elucidated in the following sections.

### 4.1 The Pre-catalytic human spliceosome

The structure of the human B complex presented in this work (Bertram et al., 2017a) allows for an in-depth functional analysis of a pre-catalytic human spliceosome for the first time. Even though its general architecture in humans (Wolf et al., 2009) and yeast (Rigo et al., 2015) as well as its protein composition (Deckert et al., 2006) was proposed some time before, the high-resolution 3D cryo-EM model presented here now allows for the unambiguous specification and evaluation of the location, composition and appearance of functionally important units of the macromolecular machine. The following analysis in particular includes the BSH that is sequestered by components of the U2 snRNP; the putative location of the catalytic centre that is formed in later, catalytically activated complexes; the position of BRR2 and other functionally important protein factors and the location and potential mode of operation of the *B complex-specific* (B-specific) proteins.



### 4.1.1 U2 snRNP components

The major building blocks of the pre-catalytic spliceosome are spatially separated in the B complex, even though their structural integration is in progress. Arranged in a rhombohedral configuration, components of the U2 snRNP and the BS-A nucleotide of the pre-mRNA intron are situated at the upper part of the body (the head), while the tri-snRNP constituents and exon 1 reside at the opposing end, as generally proposed before (Boesler et al., 2015; Wolf et al., 2009). The HEAT repeat containing C-terminal domain of U2-SF3B1/SF3b155 (SF3B1<sup>HEAT</sup>), as also observed in the corresponding yeast spliceosome (Plaschka et al., 2017), tightly sequesters the U2/pre-mRNA BSH. This configuration clearly accounts for the pre-catalytic state of the spliceosome. Although the local resolution of the map at this area is not sufficient to evaluate the atomic details, the SF3B1/BSH model from the human B<sup>act</sup> complex (Haselbach et al., 2018) could be snugly fit into the corresponding B complex density without further adjustments. Thus, besides the coherent large-scale rearrangements of the U2 snRNP components in the B to B<sup>act</sup> complex transition, it appears that relatively little restructuring occurs in this area at the molecular level. In consequence, the mechanistically important BS-A nucleotide appears to be shielded from the catalytic core area of the spliceosome until its presence is actually required to initiate step 1 of the reaction. Supporting this theory, mutations of the U2 snRNP components like for example the SF3B protein family are frequently reported to play a significant role in a variety of malignancies and other diseases. These include for example leukaemia, breast and pancreatic cancers as well as some kinds of melanomas (Yoshida and Ogawa, 2014). Structural studies on the isolated SF3B complex in fact revealed that it is able to switch between an open and closed conformation when sequestering the RNA in its central cavity (Cretu et al., 2016; Finci et al., 2018). The conformation of the wild type SF3B1 protein in the B complex thereby matches the RNA bound, closed conformation that was previously reported in the isolated structures. As further discussed in section 4.2.3, the SF3B assembly and in particular SF3B1 may therefore in general act as a switchable clamp that releases the reactive BS-A nucleotide to the catalytic procedure at the precise moment it is required to participate in the activation of the spliceosome. Mutations that modify or impair this regulatory function may cause a change in splicing kinetics or even entirely aberrant splicing products (Wan and Wu, 2013), which inevitably leads towards an unregulated alteration of the cellular proteome and disease. As an example, specific hotspot mutations of the SF3B1 HEAT domain (e.g. K700E) were repeatedly reported to promote aberrant splicing and cancer (Alsafadi et al., 2016; Wu, 2012). Stunningly, most of the mutated amino acids in fact reside incorporated into the solenoid structures of SF3B1<sup>HEAT</sup>, where they are prone to influence the overall stability of the clamp.

In the past, these correlations could only be interpreted on the basis of spatially ill-resolved biochemical studies (Cazzola et al., 2013; Wu, 2012). The analysis of the corresponding protein factors in the integral spliceosome structures now allows to propose a specific mechanism that explains the phenotypic observations. Further analysis and a potentially better resolved structure of the SF3B and associated proteins in the human B complex may thus aid in proposing even more sophisticated mechanisms of splicing malfunction and disease in the future.

### 4.1.2 Dramatic restructurings after tri-snRNP integration

The spliceosome is significantly restructured during the incorporation of the U4/U6.U5 tri-snRNP into the A complex. A characteristic gap at the centre of the particle then distances both major U2 snRNP and tri-snRNP functional units in the B complex. While the U2 snRNP likely

undergoes little intrinsic restructuring during its incorporation, the tri-snRNP structure (Agafonov et al., 2016) is significantly remodelled during B complex formation. Within the tri-snRNP particle, as well as in all later spliceosomes of the catalytic cycle, PRP8, the neighbouring proteins SNU114, U5-40k, DIM1 and the U5 snRNA represent the core of the spliceosome. PRP8 eventually accommodates the catalytic centre of the complex and thus generally acts as a binding site and platform for many spliceosomal factors and functions. Its dynamic nature was described biochemically (Schellenberg et al., 2013) and many of its pre-catalytic interactions can be elucidated in the B complex structure. Amongst others, the PRP8 Jab1 domain (PRP8<sup>Jab1</sup>) is tightly associated with the N-terminal helicase cassette (NC) of BRR2 (Nguyen et al., 2013). This interaction appears to anchor the mobile helicase BRR2 to the PRP8 core of the spliceosome, allowing BRR2 to abruptly rearrange its position depending on the state of catalytic activation of the spliceosome. In fact, BRR2 dramatically changes its position upon incorporation of the tri-snRNP into the fully assembled pre-catalytic spliceosome. While residing in an RNA-unbound state at the tip of PRP8's RT domain (PRP8<sup>RT</sup>) in the tri-snRNP that is located approximately 10 nm away from its U4 snRNA substrate, BRR2 engages the nucleotides in the correspondingly predicted region of the U4 snRNA (Hahn et al., 2012; Mozaffari-Jovin et al., 2012) in the B complex state. This study thus for the first time captures BRR2 in its substrate-bound state within a functional spliceosome, ready to act in unwinding the pre-catalytic U4/U6 snRNA duplex helix. After the incorporation into the B complex, BRR2 is then situated next to the tip of the PRP8 EN domain (PRP8<sup>EN</sup>), rotated by 180° and located more than 20 nm away from its original position in the human tri-snRNP.

Besides BRR2, other constituents of the pre-catalytic spliceosome are significantly rearranged during B complex formation. While PRP8 does not contact the pre-mRNA in the tri-snRNP particle (Agafonov et al., 2016), it tightly sequesters the first nucleotides of exon 1 upstream of the 5' SS between its N-terminal domain (PRP8<sup>NTD</sup>) and PRP8<sup>RT/EN</sup> in the B complex. Compared to its *“open”* conformation in the tri-snRNP and the more tightly *“closed”* conformation in the catalytically activated spliceosome (Haselbach et al., 2018), PRP8 thus adopts a *“partially-closed”* conformation in the intermediate, pre-catalytic B complex state of the spliceosome. As the catalytic centre has not yet formed there, in particular because the relevant catalytic U6 snRNA moieties are still bound and thus retained in the U4/U6 snRNA duplex helix, the partially-closed conformation of PRP8 represents a structurally feasible intermediate state in the stepwise formation of the catalytically activated spliceosome.

The tight incorporation of the exon 1 pre-mRNA by the U5 snRNP protein PRP8 furthermore indicates a definite transfer of the 5' SS from the U1 snRNA to the core of the spliceosome at this stage of assembly, as predicted biochemically (Konarska, 1998). The spatial configuration of exon 1 is then preserved at the core of the spliceosome throughout all later stages of the splicing cycle that are known to date (Bertram et al., 2017b; Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2017). Additionally, DIM1 clearly contacts the GU dinucleotide of the 5' SS in the B complex. The latter may indicate a potential, previously unknown role of the protein in the selection of the 5' SS.

The observed, functionally significant rearrangements thus contribute towards providing the first structurally-guided understanding of the remarkably dynamic nature of the (pre-catalytic) human spliceosome. The latter thereby in general represents a unique aspect of spliceosome formation and function, as no other large macromolecular machine known to date exhibits this degree of structural dynamics during its catalytic cycle.

### 4.1.3 Functional insights into the role of B-specific proteins

Several proteins associate with the spliceosome exclusively during the pre-catalytic state of the assembly. Accordingly, these factors are termed *B-specific proteins* and their characteristic function within the spliceosomal context is largely unknown. Besides PRP38, SNU23 and MFAP1 (Spp381 in yeast) all other B-specific proteins do not have orthologs in yeast and are thus found to be specific to the metazoan system (Ulrich and Wahl, 2017). Within the B complex structure presented in this work, PRP38, SNU23, MFAP1, SMU1, RED and FBP21 could be identified and localized. The elucidation of their spatial organization thus aids in evaluating the purpose of B-specific proteins during the spliceosomal assembly. PRP38, SNU23 and MFAP1, for example, were identified adjoined in a previously reported bundle (Ulrich et al., 2016b) that contacts PRP8<sup>NTD</sup> and the U6 snRNA/ACAGA helix. The Zn-finger domain of SNU23 thereby characteristically interacts with the 5'SS pre-mRNA at the ACAGA box helix, whereas PRP38 is reported to be essential during pre-catalytic spliceosomal remodelling (Schütze et al., 2016; Xie, 1998). In addition to that of SNU23, the Zn-finger region of the multi domain B-specific protein FBP21 (Huang et al., 2009) likewise contacts the ACAGA box region further downstream of the U6 snRNA. Interestingly, the latter region and the adjacently positioned characteristic U6 snRNA 5' stem loop structure already reside in a configuration that is then conserved throughout all later stages of the catalytically activated spliceosome (Bertram et al., 2017b; Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2017). PRP38, SNU23, MFAP1 and FBP21 may therefore act as transient structural scaffolds that temporarily hold the ACAGA box helix and the 5' SS in its predestined place during catalytic activation until the corresponding RNA elements are eventually fully integrated into the catalytic core of the spliceosome during later stages of activation. Furthermore, MFAP1 is found to contact both the head and body domains of the B complex, potentially participating in the regulation of this remodelling activity. Due to the conserved nature of the proteins' structural configuration in both yeast (Plaschka et al., 2017; Wan et al., 2016b) and metazoan spliceosomes, it appears likely that the presence of the PRP38, SNU23 and MFAP1 bundle is ubiquitously required across species to regulate the formation of the catalytic centre.

Focussing on the role of the B-specific proteins that are absent in lower eukaryotes, SMU1 may play a significant role in regulating the activity of the RNA helicase BRR2 once it is engaged to its U4 snRNA substrate in the B complex. As partially introduced above, BRR2 represents an ATP-binding RNA helicase that is essential for spliceosome activation (Will and Lührmann, 2011). It is comprised of an unusual configuration of two ring-like helicase cassettes (N-terminal, NC, and C-terminal, CC) that are situated side-by-side and act in tandem to stimulate the RNA unwinding activity of the NC only (Santos et al., 2012). Due to its robust attachment to the spliceosome through PRP8<sup>lab1</sup> (section 4.1.2), the activity of BRR2 needs to be tightly regulated in order to prevent premature activation. Multiple pathways have thus been proposed in the literature, ranging from an inhibition by the PRP8 RNase H like domain (PRP8<sup>RNase H</sup>) (Mozaffari-Jovin et al., 2012), PRP8's very C-terminal tail (Mozaffari-Jovin et al., 2013), several regions of FBP21 (Henning et al., 2017) as well as the N-terminal tail of BRR2 itself, comprising a *PWI* and some other plug-like domains that are situated at the interface of both helicase cassettes (Absmeier et al., 2015). Interestingly, in the human tri-snRNP assembly (Agafonov et al., 2016), where the inhibition of BRR2 is essential to maintain the structural integrity of the complex, the *PWI* domain resides at the same interface that was previously reported to down regulate the activity of BRR2 (Absmeier et al., 2015). In the human B complex, however, the *PWI* domain of BRR2 could not be visualized. Hence, it is likely detached from the complex or significantly

destabilized, whereas its NC and CC are clearly structured. Instead, the WD40 domain of the B-specific protein SMU1 binds to the CC of BRR2 close to the NC/CC interface on the one side and to the SF3B3 WD40 core at the head region of the B complex at the other side. In conjunction with its adjacently located *LisH* and *CTLH* domains (Ulrich et al., 2016a), SMU1 is therefore likely to inhibit and regulate BRR2 functionality. Its subsequent destabilisation by the restructuring of the U2 snRNP head of the B complex, which is observed during later stages of catalytic activation, may then lead to the detachment of SMU1 from the critical BRR2 interface and thus ultimately to an activation of the helicase. SMU1 is thus prone to act as a key player in regulating the metazoan spliceosome activation.

#### 4.1.4 Differences in human and yeast spliceosome activation

A comparison of the currently available yeast and human spliceosomal complex structures reveals a potentially more intricate activation pathway for the metazoan spliceosome. Studying the available U4/U6.U5 tri-snRNP structures from both species, it appears that BRR2 resides in diametrically opposite conformations, while most other protein and RNA entities are arranged in a similar manner (Agafonov et al., 2016; Nguyen et al., 2015; Wan et al., 2016b). Whereas BRR2 resides at the tip of PRP8<sup>RT</sup> in the human tri-snRNP, inhibited by SAD1, the helicase is positioned at a roughly equivalent position to that found in the human B complex, lacking SAD1, in the yeast tri-snRNP assemblies. In particular because one of the reported yeast tri-snRNP preparations also contains components of the U2 snRNP and even pre-mRNA (Wan et al., 2016b), it could be argued that the structures that were published as yeast tri-snRNPs may actually represent dissociated B complexes. A comparison to the recently revealed yeast B complex structure (Plaschka et al., 2017) supports this idea, as the tri-snRNP portions of both complexes are largely similar in molecular composition and structure to the isolated tri-snRNP. Further supporting this idea, both purifications of the yeast tri-snRNP structures were reported to dissociate upon ATP treatment (Nguyen et al., 2015; Wan et al., 2016a), whereas the human tri-snRNP purification does not (Agafonov et al., 2016). The observed dissociation of the yeast complexes in a physiological ATP environment is thus likely promoted by the absence of Sad1 (Huang et al., 2014), where the latter protein inhibits any premature activation in the human tri-snRNP counterpart. The yeast tri-snRNP therefore, if not inhibited in any other way *in vivo*, could likely be a dissociation product of the B complex.

On the contrary, spliceosome activation may be entirely different in humans and yeast. Several arguments account for this idea, as in particular the presence of the metazoan exclusive, regulatory B-specific proteins that are reported here could potentially allow for a more complex, or entirely different activation pathway. Interestingly, comparing another suggested tri-snRNP structure and the B complex from yeast (Nguyen et al., 2016; Plaschka et al., 2017), it appears that in neither assembly the most 3' terminal nucleotides of exon 1 are enclosed by PRP8, like it is observed in the human B complex reported in this work (also see section 4.1.2). Whereas the pre-mRNA is bound at the SF3B sequestered BSH in both yeast and human B complexes, Prp8<sup>NTD</sup> and Prp8<sup>RT/En</sup> in yeast likewise reside in a partially closed conformation, however, not enclosing the pre-mRNA but instead parts of the 5' terminal U6 snRNA. Given the otherwise similar, species independent spatial configuration of PRP8 in the B complex, the partially closed configuration of Prp8 would require an additional extensive remodelling step of the yeast B complex in order to eventually arrange the pre-mRNA for catalytic activation, as i.e. observed in the yeast B<sup>act</sup> complex (Rauhut et al., 2016; Yan et al., 2016). The reported yeast B complex structure thus either represents an earlier, general assembly state of the B complex, that

likewise, maybe more transiently also exists in the human system, or the comparison of both structures reveals a significantly different activation pathway regarding the human and yeast splicing machinery.

## 4.2 The catalytically activated human spliceosome

The various structures of catalytically activated spliceosomes that are available to date have greatly contributed towards revealing the biochemical mechanism of pre-mRNA splicing. By supplying the first high-resolution structure of a human spliceosome (Bertram et al., 2017b), this work substantially participated in this process. In combination with the additional models of human spliceosome complexes that were published in the meantime (Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2017), a more detailed analysis of the involved procedures may now be performed.

### 4.2.1 Molecular architecture and structural conservation between species

The structure of the human C\* complex presented here revealed the molecular architecture of a catalytically activated human spliceosome for the first time. Even though the occurrence of this particular assembly state was generally expected within the splicing cycle (Ohrt et al., 2013), the actual existence of the C\* complex as a structurally stable intermediate was poorly understood and hardly characterized. The structure of the post-PRP16- but pre-PRP22 action C\* complex thus generally introduced a new, stable splicing intermediate to the metazoan splicing cycle. The somewhat elongated molecular architecture of the complex thereby consists of three major multi-component domains: The *central domain*, harbouring the catalytic core of the spliceosome; the *top region*, containing mostly U2 snRNP and associated proteins; and an elongated *helical bundle region* that contacts both top and central domain. The areas in between these regions are often populated with proteins that are more transiently associated with the spliceosome, depending on its position in the catalytic cycle.

The central domain of the assembly is approximately 12 x 20 nm in size and includes most of the essential components that are required to form the catalytic centre of the spliceosome. PRP8<sup>NTD</sup>, PRP8<sup>RT/En</sup>, SNU114, the WD40 domain of PLR1, the U5 Sm-core, the U5 snRNA and parts of the U2 and U6 snRNA thereby contribute as the molecular building blocks. The latter arrangement of protein and snRNA domains is found to be particularly stable and conformationally conserved across species in the catalytically activated spliceosome (Fica et al., 2017; Galej et al., 2016; Haselbach et al., 2018; Liu et al., 2017; Rauhut et al., 2016; Wan et al., 2016a; Wilkinson et al., 2017; Yan et al., 2016; 2017; Zhan et al., 2018; Zhang et al., 2017; 2018). The structure of the human C\* complex thus provided the initial evidence for a conserved catalytic core of the spliceosome, which serves as a rigid molecular platform to which other essential factors bind.

Some of these factors are organized in the top region of the human C\* complex, which is populated by the remaining U2 snRNP components that stay stably associated with the spliceosome after its catalytic activation. In particular the U2 Sm-core, large parts of the 3' U2 snRNA including its stem II (Perriman and Ares, 2007) and U2 snRNP proteins A' and B'' are included. Proteins of the SF3 family that are still stably attached in the B<sup>act</sup> complex (Haselbach et al., 2018), however, are lost presumably by the action of RNA helicase PRP2 (King and Beggs, 1990; Schmitt et al., 2018). Compared to the corresponding structures in yeast (Fica et al., 2017;



Yan et al., 2017), the location and composition of the top region is essentially conserved between species; thus, providing further evidence for the high degree of structural conservation of the catalytically activated spliceosome. Contacting this region, the equally well conserved alpha helical *half a TPR* (HAT) domain of the SYF1 protein (Chan and Cheng, 2005; Chen, 2002) contacts the U2 Sm-core and likely guides its remodelling during progression through the catalytic cycle (section 4.2.3). Not conserved and thus absent in yeast is the metazoan specific IBC RNA helicase Aquarius (IBP160), which is docked at the central region of the SYF1 alpha helical solenoids. Aquarius is specifically involved in binding the intron several nucleotides upstream of the ACAGA Box sequence (De et al., 2015), which appears structurally plausible in the human C\* complex structure. However, the intron nucleotides that are involved in the speculated binding to Aquarius cannot be traced in the cryo-EM density due to a drop in local resolution at the corresponding position.

The helical bundle region is attached to the central domain of the C\* complex. It comprises several proteins that initially interact with the spliceosome pre-organized in the PRP19/CDC5L complex (NTC in yeast). As suggested before, based on biochemical evidence (Grote et al., 2010), the long alpha helical elements of PRP19, CDC5L and SPF27 in fact stably interact with each other while being attached to the core of the catalytically activated spliceosome. Interestingly, the central domain of CDC5L, as well as the N-terminal region of Skip (or SNW1) even seem to be intrinsically disordered in solution (Wang et al., 2010). The corresponding amino acids that interact with the catalytically activated spliceosome, however, are distinctively threaded through the main body of the complex, highly structured and stably contacting a wealth of different spliceosomal factors. Given the dynamic nature of the early spliceosome assembly this may be a prime example of how proteins that are intrinsically disordered in solution can locally regain a specific 3D structure to facilitate and modulate the binding of other protein factors.

### 4.2.2 Catalytic centre

The catalytic centre of the spliceosome was biochemically predicted to contain a characteristic three- or four-way snRNA helical junction comprised of nucleotides from the U2 and U6 snRNA (Anokhina et al., 2013). Whether a three-way snRNA junction is solely utilized by the yeast spliceosome (Madhani and Guthrie, 1992) and a four-way helical junction by the mammalian spliceosome instead (Sashital et al., 2004; Sun and Manley, 1995), remained unclear. The structure of the human C\* complex then revealed a universally conserved three-way snRNA helical junction at the core of the complex. As predicted (Anokhina et al., 2013; Madhani and Guthrie, 1992), the U2 and U6 snRNA thereby indeed form *Helix Ia* and *Ib* as well as *Helix II* with one another, which are interconnected by the corresponding nucleotides of the U6 snRNA *internal stem loop* (ISL) structure. Embedded in this intricate 3D RNA network the phosphate backbones of the U6 snRNA nucleotides A53, G54, G72 and U74 (A59, G60, G78 and U80, respectively in yeast) are spatially arranged to putatively bind the two catalytic Mg<sup>2+</sup> ions, as predicted from biochemical evidence (Fica et al., 2013). Although the cryo-EM map was determined at a (local) resolution of ~4.5 Å at that area, the resolution is not sufficient to unambiguously identify metal ions in the density. A comparison to the subsequently published human C\* complex structure at ~3.8 Å (Zhang et al., 2017), however, reveals that the putatively assigned catalytic Mg<sup>2+</sup> ion binding sites and the adjacent RNA moieties were indeed modelled in great accordance to the higher resolved model. Interestingly, the general area that constitutes the active centre in the catalytically activated spliceosome is remarkably conserved between species, as revealed by a comparison of the corresponding yeast structures (Fica et al., 2017; Yan

et al., 2017). The dispute whether the metazoan spliceosome potentially relies on a different catalytic mechanism than that utilized by lower eukaryotes thus appears to be resolved.

### 4.2.3 Large-scale structural rearrangements modulate splicing activity

The spliceosome is significantly remodelled not only during its pre-catalytic state of the assembly (section 4.1.2), but also during the catalytically activated phase of the reaction. Complex macromolecular domains that are comprised of protein, RNA or both are restructured accordingly in order to facilitate splicing. A precise guidance of the involved reactants (i.e. the BS-A or the splice site nucleotides) is thereby of superior importance for the success of the splicing reaction, as the smallest aberration in splice site selection or the misprocessing of RNA in general causes shifts in the open-reading frame, aberrant mRNA products and disease (Solis et al., 2008).

In particular the U2 snRNP components that initially chaperone the pre-catalytic BS area of the pre-mRNA are drastically restructured upon catalytic activation of the spliceosome. Enabling these rearrangements, several hundred kilodaltons of protein and RNA components are repeatedly repositioned and moved across large distances. Tracing for example the dynamic nature of the U2 snRNP components at the B complex stage of assembly, the BSH has formed but is fully embedded in protein, particularly by the alpha helical solenoids of the SF3B1 HEAT domain (SF3B1<sup>HEAT</sup>). In the B complex, the assembly that includes the BS-A is located more than 15 nm away from the future site of catalysis (Bertram et al., 2017a), effectively preventing any kind of pre-mature interaction of the essential step 1 reactant with the pre-mRNA substrate by spatial separation. The functionally important BSH element is then progressively brought in closer to the designated active site of the spliceosome, which thus eventually transitions into the B<sup>act</sup> complex configuration. Here, the access to the BSH is still modulated by the SF3B protein family and the bulged out BS-A nucleotide is tightly sequestered by SF3B1<sup>HEAT</sup> (Haselbach et al., 2018). Even though the catalytic centre of the complex has almost entirely formed and is thus primed for catalysis, the reactive BS-A nucleotide is still kept out of reach and safeguarded, now situated approximately 5 nm away from the active site. A characteristic hyperphosphorylation of SF3B1 is then required to progress with step 1 of the reaction (Wang et al., 1998), even though the necessary structural rearrangements are poorly understood (section 2.4.3.3.). The phosphorylation of SF3B1, however, likely alters the binding characteristics of the protein to the BS-A nucleotide. Hence, the spatial separation and adjustable shielding of the essential step 1 reactant may represent a major regulatory checkpoint during spliceosome activation. The U2 SF3 protein family may therefore act as a significant signal integration hub that particularly coordinates and controls the initiation of splicing.

The potential to precisely rearrange and spatially separate important reactants from each other seems to appear as a subordinate theme in pre-mRNA splicing regulation and enables the spliceosome to effectively modulate its catalytic activity as well as its fidelity. External kinases or helicases (section 4.3.2) are thereby likely to influence the splicing reaction from a distance by interfering with the coordinated spliceosomal rearrangements at characteristic stages of catalysis.

Once step 1 of the reaction is performed and the spliceosomal rearrangements are leading towards a transition into the C complex (Zhan et al., 2018), the U2 snRNP–BSH components are again restructured significantly with respect to their spatial organization in the B<sup>act</sup> complex. At

this stage of the assembly, the SF3B proteins are not only stripped off the functional BSH unit itself, but seem to be released from their stable association with the spliceosomal core entirely. The residual components of the U2 snRNP, such as the U2 snRNA or the Sm-core and directly associated proteins, however, remain. As the SF3B proteins are detached, the BSH is now connected to the intron lariat structure and instead stabilized by various domains of the step 1 associated splicing factors CCDC49, CCDC94 and ISY1 (Zhan et al., 2018). Nevertheless, the formerly reactive BS-A nucleotide is still situated at the catalytic centre of the spliceosome. The entire BSH thus needs to be relocated in order for step 2 of the reaction to occur. Interestingly, the remaining U2 snRNP components again seem to be utilized as a macromolecular handle to perform this task, as initially revealed by the C\* complex structure presented here. In particular the nucleotides that reside towards the 3' end of the U2 snRNA thereby serve as a rod-like element that translates large-scale conformational changes at the exterior of the spliceosome to small-scale restructurings of e.g. the BSH at the core of the complex. At the same time, the significant rearrangements of the involved U2 snRNP components again provide a fair opportunity for regulatory factors to modulate these transitions. The sheer size and accumulated mass of the more peripherally situated mobile components thereby provide a substantially larger area for molecular interactions than that of the relatively small (catalytically active) components at the core of the molecular machine. The modulation of the required restructuring events can thus be more decisive when external factors are able to target the large peripheral domains of the spliceosome. A mechanical coupling between these functional components, like that of the U2 snRNP components and the BSH, then conveys the peripheral signalling to the catalytic centre.

Based on the structural observations, the peripheral U2 Sm-core is in fact significantly rearranged during the B<sup>act</sup>-to-C and the subsequent C-to-C\* complex transition. The BSH likewise matches the corresponding rearrangement vectors of the U2 Sm-core and is thus itself repositioned accordingly (Haselbach et al., 2018; Zhan et al., 2018). As soon as the BSH is guided away from the catalytic centre in the C\* complex, it is stabilized by the RNaseH domain of PRP8 (PRP8<sup>RNaseH</sup>) and the WD40 domain of the step 2-specific protein PRP17 (PRP17<sup>WD40</sup>). Both of the latter have been previously associated with splicing regulation at various steps of the reaction (Jones et al., 1995; Mozaffari-Jovin et al., 2012; Pena et al., 2008). The large-scale rearrangements that eventually relay their trajectories to the much smaller functional entities at the catalytic core of the molecular machine thus seem to be a universal theme in spliceosome regulation. Modulating these rearrangements, factors like kinases or RNA helicases may then take advantage of the extensive target area that is provided by acting on multi subunit complexes like the SF3 protein family or the U2 Sm-core. Once the (large-scale) restructuring process is triggered, binding sites for smaller factors like those belonging to the step-1 or -2 specific protein families seem to emerge rather transiently. A temporary binding of these factors may then further stabilize specific structural states in the splicing cycle. Tracing the large- and small-scale rearrangements of the spliceosome by structural analysis thus enables a thorough analysis of its catalytic mechanism and the specific modes of regulation.

### 4.2.4 Extended ACAGA and BSH helices in the human system

The ACAGA box and BS helices are structural elements that fulfil essential functions in the spliceosome. Their appearance was probed biochemically and predicted accordingly (Will and Lührmann, 2011) and a general structural conservation between species was somewhat expected. As revealed by the structural models of the human spliceosome, however, the essential

RNA double helices formed around the ACAGA box and BS area are significantly elongated in the 5' direction of the U6 snRNA and the 3' direction of the U2 snRNA, respectively, compared to their counterparts in yeast. The functional reason for this extension in the human (and likely the metazoan) system is poorly understood, even though a more abstract comparison between higher and lower eukaryotes may aid in revealing a purpose for this deviation.

Importantly, the consensus sequence of the intron nucleotides that participate in ACAGA box and BSH binding is significantly more degenerate in higher eukaryotes than in lower members of the domain like for example yeast (Burge et al., 1999). A high level of conservation allows the formation of canonical base pairs between nucleotides of the intron and the snRNA, which usually results in the formation of a stable, classical B-form RNA double helix. Within the spliceosome, these helices may not only serve as structurally important entities in the mechanics of splicing (i.e. section 4.2.3) but are likewise also capable of mechanically locking the comprised intron nucleotides into place, hence, ultimately determining their faith in the splicing procedure and ensuring a reliable splice site selection. Due to the lack of canonical base pairing in the metazoan system, however, ACAGA box and BS helices with a comparable stability to that observed in yeast (Fica and Nagai, 2017) cannot be formed by a helix comprised of 7-8 nucleotides only. To compensate this lack of stability, an extension of the helical arrangement by 1-2 turns may effectively restore the functional stability of the helical unit. In fact, all available structures of the catalytically activated human spliceosome reveal a rather non-canonical base pairing in the characteristically distorted, extended RNA helices at the corresponding regions. Surprisingly, the backbone structure of these helices is remarkably conserved, despite the fact that some spliceosomes have been assembled on differing pre-mRNA substrates (Haselbach et al., 2018; Zhan et al., 2018; Zhang et al., 2017) with different, non-conserved nucleotide sequences at the corresponding regions. The observed level of structural conservation may be achieved by a further stabilization of the extended ACAGA box and BS helices through characteristic interactions with proteins, such as a Zn-finger domain of the metazoan specific protein RBM22 or various other proteins (i.e. PRP8<sup>RH</sup> and PRP17<sup>WD40</sup> in the C\* complex), respectively. Supporting this hypothesis by biochemical evidence particularly the interactions of RNA with RBM22 were suggested to promote a stable spliceosome configuration (Rasche et al., 2012).

The degeneracy of metazoan splice sites may be required for alternative splicing (Stepankiw et al., 2015) and has thus likely developed during the evolutionary differentiation from the basic group II self-splicing intron ancestor (section 4.3.1). However, once the splice sites are determined by an altered, more transient interaction with the corresponding initiation factors, they have to be reintroduced to the evolutionarily conserved catalytic mechanism for splicing. A stable incorporation of these sites into the catalytically activated spliceosome is thus a prerequisite for the efficient catalysis of the splicing reaction. As it seems, the metazoan spliceosome employed a mixture of molecular approaches to achieve this task and compensate for the degenerate intron sequences that are nevertheless eventually destined to bind to the corresponding nucleotides of the U6 snRNA ACAGA box and the U2 snRNA BS sequence in order to be spliced correctly.

### 4.3 General aspects of splicing & utilizing cryo-EM as a tool to reveal its mechanism

The splicing of pre-mRNA requires an incredibly complex but well orchestrated succession of remodelling events in the spliceosome. Since it is comprised of hundreds of different molecular building blocks, the assembly pathway has to be well orchestrated and is critical for proper a proper splicing functionality. In order to decipher the underlying mechanisms, cryo-EM on carefully purified spliceosome samples currently represents the most promising method of structural visualisation. Besides the answering of mechanistic questions, however, it becomes increasingly interesting to investigate the evolutionary origins of splicing on the basis of structural conservation. The following section will thus briefly discuss the above-mentioned implications in more detail.

#### 4.3.1 Evolutionary conservation of RNA splicing

The splicing of (m)RNA from precursors does likely represent an ancient process in life, evolutionarily matured over millions of years. Originating in the RNA world, self-splicing introns entirely comprised of RNA once participated in the procedure by catalysing the characteristic transesterification reaction in *cis* or *trans* (Moore et al., 1993). The splicing of pre-mRNA by the spliceosome, however, is catalysed by a large snRNP particle, and its evolutionary origin was initially unclear. Analysing this reaction and its nucleic acid products, it was then revealed that a strikingly similar intron lariat structure was formed in both spliceosomal and Group II self-splicing intron (GII intron) dependent splicing (Padgett et al., 1984; Ruskin et al., 1984). The high conservation of this reaction product lariat structure consequently led to the assumption that the spliceosome may have actually originated from its GII intron ancestor (Cech, 1986; Sharp, 1985). Clear evidence for this theory, however, remained elusive at that time, as a specific structural understanding of the catalytic sites of both splicing machineries was lacking.

An in-depths comparison of the latter is nowadays possible and in fact reveals striking similarities in the composition of the catalytically active centre between the solely RNA comprised GII intron and the spliceosome. The structure of the GII intron thereby possesses six characteristic RNA domains, some of which are specifically conserved in the snRNA and protein structures of the spliceosome (Galej et al., 2018). Beginning at the catalytic centre, Domain V (DV) is actually the most conserved entity in RNA composition and structure between GII introns and the U6 snRNA of the spliceosome (Boulanger et al., 1995). The DV of the GII intron is structurally resembled by the U6 ISL in the spliceosome (e.g. in the C\* complex) and thus also harbours the catalytic triad and the Mg<sup>2+</sup> ions at the active site. The level of conservation regarding the structural fold and partially also the nucleotide sequence is astounding (Robart et al., 2014), making a parallel, independent evolution of both structural entities particularly unlikely. The sequence of the U6 snRNA is furthermore best conserved amongst all spliceosomal snRNAs from various species (Brow and Guthrie, 1988).

DV is directly connected to Domain VI (DVI), which provides the adenosine nucleophile for the first step of the reaction (Lambowitz and Zimmerly, 2011). Its characteristic helical structure thus functionally resembles the BSH of the spliceosome. Comparing the structure of a eukaryotic GII intron (Robart et al., 2014) and that of the catalytically activated spliceosome (e.g., the C\* complex presented in this work), the highly conserved appearance and position of the first exon (exon 1) becomes likewise apparent. Parts of the GII intron Domain I (DI), more



specifically its exon binding sequence (EBS), thereby participate in the binding of the most 3' exon 1 nucleotide in a similar fashion to that observed for the U5 snRNA loop 1 in the spliceosome (Costa et al., 2000). Most of the remaining DI nucleotides, however, provide the general structural scaffold and framework of the GII intron, fully embedding its catalytic centre (Zhao et al., 2015). In evolutionary terms, transitioning from the GII intron to the spliceosome, it thus seems that the size of the DI region has been substantially reduced in the spliceosome, where the interaction of DI nucleotides with exon 1 is now facilitated by the U5 snRNA. Most importantly, however, the scaffolding function of DI nucleotides in GII introns seems to have eventually been handed over to protein moieties (PRP8 in particular) in the spliceosome, leading to the assumption that large parts of the GII intron RNA was replaced by protein during the evolutionary development of the spliceosome. As most components of the modern spliceosome are comprised of proteins, it may be assumed that the general process of replacing RNA with protein is still on going. In particular the observation that more highly developed metazoan spliceosomes contain an even larger number of proteins compared to those of their less developed yeast counterparts supports this assumption. The catalytic core of the assembly, however, is remarkably conserved between species (Fabrizio et al., 2009), and eventually even amongst different splicing machineries in general.

#### **4.3.2 RNA helicases that modulate splicing activity from a distance**

As elucidated in sections 4.1.2 and 4.2.3 the assembly and catalytically active phase of the spliceosome is compellingly modulated by large-scale structural rearrangements. The spliceosome thereby never enters a processive phase, as it is observed for numerous other macromolecular machines such as the RNA polymerases or the ribosome, but is assembled repeatedly on an altering substrate for every round of splicing. A wealth of proteins regulate this functionality, where RNA helicases are found to be of particular importance during the process (Cordin and Beggs, 2013). In particular a family of three DEAD-box, four DEAH-box and one Ski2-like helicase is reported to be essential for all spliceosome functionality in eukaryotes (Cordin et al., 2012). The Ski2-like helicase BRR2 thereby engages its substrate (the U4 snRNA) directly, as elucidated by the B complex structure, and thus likely acts on the spliceosome by exerting a classic unwindase activity. Other essential RNA helicases, however, must act on the remodelling of the spliceosome from a distance, as the structures of the molecular machine that are available to date often reveal the missing engagement of a recruited RNA helicase with its supposed RNA substrate. A prime example of the latter is represented by the DEAH-box helicase PRP16, which resides at the periphery of the C complex (Zhan et al., 2018) (and is absent in the C\* complex) but distanced from any RNA substrate. Instead, it was biochemically proposed that the helicase promotes the dissociation of Cwc25 and Yju2 (CCDC49 and CCDC94, respectively in humans) in an ATP-dependant manner in yeast (Tseng et al., 2011). The latter two factors do in fact interact with the BSH and PRP16 in the post step-1 spliceosome (Zhan et al., 2018) but how a potential RNA helicase activity of PRP16 could remodel these interactions remains unclear.

Another example for an essential RNA helicase that is situated at the periphery of the spliceosome is represented by PRP22, which again belongs to the DEAH family and resides more than 100 Å away from its proposed pre-mRNA target in the C\* complex. PRP22 is reported to be the first helicase that is involved in the disassembly of the spliceosome (Cordin and Beggs, 2013), however, its ATPase activity seems dispensable for the second step of splicing (Company et al., 1991; Schneider et al., 2004). Yeast structures of the P complex then revealed that Prp22 is eventually docked to exon 2 of the pre-mRNA close to the 3' SS at this subsequent stage of

spliceosome assembly (Liu et al., 2017; Wilkinson et al., 2017). Why it is nevertheless recruited to the C\* complex, disengaged from its substrate RNA, and if PRP22 acts similarly in the human P complex can not be determined at this time.

In general, RNA helicases frequently seem to act on the spliceosome from a distance. Whether they exclusively serve as a structural scaffold in remodelling other protein factors or actually act as processive RNA helicases does nevertheless often remain unclear. As their action on the spliceosome is absolutely essential and seemingly not taking place at the catalytic centre, RNA helicases represent a significant class of proteins that modulate the activity of the spliceosome from a distance.

### 4.3.3 Completeness of the reconstructed spliceosome structures

Reconstructed cryo-EM density maps of large macromolecular machines rarely represent the entire appearance of the complex. For instance, in the map of the C\* complex it appears that several peripheral regions of the assembly (e.g. the Aquarius binding site (section 4.2.1) or the PRP19/CDC5L helical bundle) are much more poorly resolved than those regions at the centre of the complex. As the spliceosome possesses a particularly stable catalytic core (section 2.4.3.3), this section is usually the best resolved portion of the structure; whereas factors that are situated in more peripheral regions seem to be less stably associated and thus frequently blur out during averaging and reconstruction. Often, the most dynamic parts of a structure cannot be visualized at all. As a further example, both the B as well as the C\* complex reveal large density elements that are particularly badly resolved and thus usually omitted during an interpretation of the data. Some of these peripheral factors are expected to fulfil regulatory functions in the splicing procedure and sometimes harbour large intrinsically unstructured domains, as for example the 300 kDa SRM300 protein in the C\* complex (Blencowe et al., 2000; Eldridge et al., 1999). Others, however, like the equally large RNA helicase protein BRR2, are usually well structured and stably associated to other known intermediate states of the spliceosome but not clearly identifiable in the density of the C\* complex that is presented here. As a general theme, the estimated mass of all proteins that are biochemically purified during sample preparation and those that can eventually be visualized in the cryo-EM reconstruction differs by 20-40 %. Hence, the spliceosome structures that we analyse today are far from complete in their molecular architecture.

To tackle the difficulties that are caused by the structural inhomogeneity (or “flexibility”) of a macromolecular complex it is thus important to improve the purity and quality of the biochemical preparation itself. Improved sample handling routines may aid in preventing a deterioration of the specimen on the grid or the upstream biochemical pipeline. Furthermore, the physical stability of a specimen can be improved by the mild chemical crosslinking of the macromolecular complexes (Kastner et al., 2008). The latter in particular proved to be a valuable tool in enhancing the structural homogeneity of a typical cryo-EM sample.

### 4.3.4 Non-uniform resolution estimates in cryo-EM

Cryo-EM is currently utilized to elucidate the structures of large macromolecular machines and enzymatic complexes at resolutions that almost match those of crystallographic experiments (Vonck and Mills, 2017). To achieve a particularly high resolution, the biochemical preparation of the specimen, however, must be of excellent quality and great homogeneity (Zhou, 2011). In

particular the latter requirement for an exceptional structural homogeneity of the sample is nevertheless hard to achieve in practice. As introduced in section 2.3.4.1, the most frequently communicated resolution value of a cryo-EM reconstruction is usually estimated as a single value at an arbitrary FSC cut-off level in a “gold-standard” procedure. Even though this estimation routine is broadly accepted and utilized in the field to quantify the resolved detail of a reconstructed 3D EM map, real density maps are rarely structurally homogeneous (section 4.3.3) and thus do not pose a single value for resolution either. The typical FSC based resolution estimation then averages the stable core of the map with those areas that are more flexible (or blurred out entirely). The one, single resolution value obtained through the gold-standard FSC procedure thus generally solely reflects a not further specified “average” resolution, where some parts of the map are often resolved better and others significantly worse. Conclusions that are based on this average resolution value alone should therefore be treated with caution and only serve as a first guideline when assessing the overall quality and information content of a reconstructed 3D density map. Only a careful visual inspection of the map ensures that any kind of resolution-based conclusion may be tested and evaluated appropriately (Penczek, 2010b).

To reduce resolution inhomogeneities in any kind of EM density map through 3D classification and low-pass filtering may be employed (Penczek, 2010b). In this regard, performing extended classification routines may aid in improving the homogeneity of the sample (or particle dataset) *in-silico*. A reasonable filtering of the map then prevents overfitting during the refinement procedure and aids in equalizing the display thresholds of the map during visual inspection and analysis. The success of these measures, however, is hardly quantifiable and should thus again be evaluated with caution by the experimentalist and all others who assess the results. In this study, the maps of the B and C\* complex were deliberately filtered to 4.5 Å and 5.9 Å, respectively, in order to prevent a major distortion of the estimated vs. the actual resolution.

#### **4.3.5 Anisotropic resolution due to preferred particle orientations in Cryo-EM**

The anisotropic distribution of projection angles during a back projection procedure can significantly limit the resolution of a structural model that was obtained by cryo-EM. As introduced in section 2.3, an isotropic recording of 2D sample projections at as many projection angles as possible is important, in particular if the sample molecule is non-symmetric. The latter is usually routinely achieved during standard sample preparation, as particles are ideally adsorbed to the grid in a random fashion during the preparative phase of grid preparation. In certain scenarios, however, the corresponding particles instead interact in a defined (or preferred) way with the substrate that is provided, or the air-water interface of the grid. In these cases, the specimen is often adsorbed to the substrate in a preferred orientation, which may consequently impair the results of the 3D reconstruction process during image processing (Ludtke, 2016). As the spatial information of essential projection angles is missing in these reconstructions, the resolution of the corresponding 3D volume is anisotropic. The magnitude of resolution anisotropy is thereby greatest for (imagined) projection angles that are normal to the angle of the projection in the preferred orientation. Several examples exist in the literature where cryo-EM specimen displayed a severe angular preference in the recorded images (Campbell et al., 2015; Ludtke et al., 2001). The 3D reconstructions in the previously given examples were nevertheless successful, as the high intrinsic symmetry of the particles

eventually allowed to virtually occupy the entire projection sphere. Thus, only a very limited set of angular orientations need to be recorded in these special cases. Unfortunately, symmetry related expansion strategies do not apply to asymmetric particles in the C1 space group.

In this study, the 3D reconstruction of the asymmetric spliceosomal B complex was significantly affected by its preferential adsorption to the grid. Trying to cope with the negative consequences of this effect it was first attempted to record a large dataset and particularly classify the latter in 2D or 3D for occasionally appearing rare views in the actual raw data. As particles in the preferential view were, however, approximately 500x more prominent than others, little isotropic resolution was gained and the 3D refinement and reconstruction of the particle had to be aborted once reaching approximately 20 Å. As no other satisfactory computational method exists to treat the anisotropy in the dataset (Ludtke, 2016) it was consequently attempted to improve the degree of preferential adsorption of the particle by modifying its surface chemistry. In the following, the specimen was quenched with variously charged chemicals like aspartate, lysine or amylamide after crosslinking, grids were deliberately glow discharged in air atmosphere and the sample buffer treated with detergents like *Lauryl Maltose Neopentyl Glycol* (LMNG), as in parts described elsewhere (Glaeser et al., 2016; Zhang et al., 2010). Furthermore, recording a dataset at substantial tilt angles is suggested as a solution to address a preferred specimen orientation on the grid (Tan et al., 2017). The latter was likewise attempted here at a tilt angle of both 30° and 40°. However, the experiment did not yield feasible improvements in resolution as tilting a cryo-EM specimen to high angles inevitably induces other artefacts to the data such as an increased sample drift or charging effects. Instead, new datasets were recorded for every chemical modification listed above and the data could eventually be refined to a final resolution of 4.5 Å (Bertram et al., 2017a). The angular preference was thus improved compared to the initial attempts, but was not entirely eliminated. The preferential view is still prominent in the final reconstruction and thus limits the maximum resolution that is isotropically attainable. In the future, even more chemical modifications or an entirely different purification protocol could be utilized in order to try eliminating the preferential orientation of the particle entirely.

### 4.4 Perspectives

The spliceosome structures presented in this work have greatly improved the understanding of human pre-mRNA splicing. Significant discoveries include the mapping of the dramatic restructuring events that are required to assemble and activate the human pre-catalytic spliceosome. Furthermore, the role of the U2 snRNP components and the B specific protein family could be elucidated in unprecedented detail, ultimately revealing an intricate network of regulatory and functional interactions among a wealth of different splicing factors. A direct comparison of the yeast and human pre-catalytic spliceosome structures then revealed a previously unexpected, significant difference in the general initiation of the splicing procedure between species, even though further studies (ideally in the pre-catalytic yeast system) must be conducted to make such comparisons more robust.

Concerning the catalytically activated human spliceosome, the C\* complex revealed the generally conserved catalytic core of the spliceosome for the first time. In contrast to the pre-catalytic configurations, the core of the catalytically activated spliceosome appears to be entirely conserved between higher and lower eukaryotes. The structure further elucidated the catalytic configuration of the corresponding U2 and U6 snRNA nucleotides that finally position the two catalytic Mg<sup>2+</sup> ions. The 5' SS as well as the BS nucleotides are thereby held in place by

characteristically extended RNA double helices – a concept of potential evolutionary adaptation in higher eukaryotes that was generally unknown before and thus first described here in the human system.

Even though both models of the B and C\* complexes revealed intriguingly new insights into the mechanics of splicing, neither assembly eventually represented a complete representation of the spliceosome with all of its peripheral factors. As in particular the peripheral factors are poorly resolved, it will be necessary to improve the protocols that are used for biochemical purification, sample preparation and image processing to enhance the overall completeness of the structures in the future.

Besides improving the level of structural completeness, many more, potentially unknown functional states of the spliceosome likely exist. The biochemical and microscopic search for novel snapshots of this molecular machine will therefore continue. The discoveries of the past have largely revealed the mechanism of how the splicing reaction is actually catalysed once the splice sites were identified on the pre-mRNA substrate. How exactly, however, the pre-spliceosomal components are dedicated towards a specific splice site, and how this process is regulated, is unknown. In particular when considering the alternative splicing events that are essential to higher eukaryotes it will become a crucial task improve the understanding of the underlying splice site selection procedures in the future. Further structural models of the pre-catalytic spliceosome such as the metazoan A complex or even earlier assemblies thus likely aid in answering these questions.

The unparalleled large-scale rearrangements that productively drive the catalytic activity of the molecular machine prevail in both pre-catalytic and catalytically activated states of the spliceosome. How these rearrangements mechanistically modulate the splicing reaction, however, is only partially understood. In particular the question of how specific factors like helicases at the periphery of the assembly can influence the splicing reaction, or even proofread its fidelity, at the heart of the complex is far from fully answered. Learning more about the spliceosome's dynamic mode of operation may therefore eventually lead to a generally better understanding of the assembly, regulation and function of large macromolecular machines in the cell. A precisely targeted interference with these procedures can then be utilized to specifically manipulate the enzymatic productivity of such machines, opening the potential to cure diseases or other malfunctioning.

The research of splicing mechanics has thus just arrived at an intermediate level of maturity. Future structural studies of whole spliceosome complexes – in combination with biochemical optimization of the underlying purification procedures – will therefore contribute to a progressively deeper understanding of RNA splicing in particular and the function of macromolecular machines in general.



# References

- Absmeier, E., Wollenhaupt, J., Mozaffari-Jovin, S., Becke, C., Lee, C.-T., Preussner, M., Heyd, F., Urlaub, H., Lührmann, R., Santos, K.F., et al. (2015). The large N-terminal region of the Brr2 RNA helicase guides productive spliceosome activation. *Genes Dev.* *29*, 2576–2587.
- Achsel, T., Brahm, H., Kastner, B., Bachi, A., Wilm, M., and Lührmann, R. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J* *18*, 5789–5802.
- 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., Agafonov, D.E., Deckert, J., Deckert, J., Wolf, E., Wolf, E., Odenwalder, P., Odenwalder, P., Bessonov, S., Bessonov, S., et al. (2011). Semiquantitative Proteomic Analysis of the Human Spliceosome via a Novel Two-Dimensional Gel Electrophoresis Method. *Mol Cell Biol* *31*, 2667–2682.
- Agafonov, D.E., Kastner, B., Dybkov, O., Dybkov, O., Hofele, R.V., Hofele, R.V., Liu, W.T., Liu, W.T., Urlaub, H., Lührmann, R., et al. (2016). Molecular architecture of the human U4/U6.U5 tri-snRNP. *Science* *351*, 1416–1420.
- 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*, 10615.
- Anokhina, M., Bessonov, S., Miao, Z., Westhof, E., Hartmuth, K., and Lührmann, R. (2013). RNA structure analysis of human spliceosomes reveals a compact 3D arrangement of snRNAs at the catalytic core. *EMBO J.* *32*, 2804–2818.
- Arenas, J.E., and Abelson, J.N. (1997). Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc. Natl. Acad. Sci. U.S.A.* *94*, 11798–11802.
- Baker, M.L., Baker, M.R., Hryc, C.F., and DiMaio, F. (2010). Chapter One - Analyses of Subnanometer Resolution Cryo-EM Density Maps. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 1–29.
- Baker, M.L., Baker, M.R., Hryc, C.F., Ju, T., and Chiu, W. (2012). Gorgon and pathwalking: Macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* *97*, 655–668.
- Bammes, B.E., Jakana, J., Schmid, M.F., and Chiu, W. (2010). Radiation damage effects at four specimen temperatures from 4 to 100K. *J. Struct. Biol.* *169*, 331–341.
- Bandziulis, R.J., Swanson, M.S., and Dreyfuss, G. (1989). RNA-binding proteins as developmental regulators. *Genes Dev.* *3*, 431–437.
- Barford, D., and Hopfner, K.-P. (2016). Editorial overview: Macromolecular machines and assemblies. *Curr. Opin. Struct. Biol.* *37*, vi–viii.
- Behrens, S.E., and Lührmann, R. (1991). Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev.* *5*, 1439–1452.
- Berman, A.L., Kolker, E., and Trifonov, E.N. (1994). Underlying order in protein sequence organization. *Proc. Natl. Acad. Sci. U.S.A.* *91*, 4044–4047.
- Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017a). Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* *170*, 701–713.

- Bertram, K., Agafonov, D.E., Liu, W.-T., Dybkov, O., Will, C.L., Hartmuth, K., Urlaub, H., Kastner, B., Stark, H., and Lührmann, R. (2017b). Cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* *542*, 318–323.
- Blatch, G.L., and Lässle, M. (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* *21*, 932–939.
- Blencowe, B.J., Baurén, G., Eldridge, A.G., Issner, R., Nickerson, J.A., Rosonina, E., and Sharp, P.A. (2000). The SRm160/300 splicing coactivator subunits. *Rna* *6*, 111–120.
- Boesler, C., Rigo, N., Agafonov, D.E., Kastner, B., Urlaub, H., Will, C.L., and Lührmann, R. (2015). Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site. *Rna* *21*, 1993–2005.
- Boesler, C., Rigo, N., Anokhina, M.M., Tauchert, M.J., Agafonov, D.E., Kastner, B., Urlaub, H., Ficner, R., Will, C.L., and Lührmann, R. (2016). A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity. *Nat. Commun.* *7*, 11997.
- Boulanger, S.C., Belcher, S.M., Schmidt, U., Dib-Hajj, S.D., Schmidt, T., and Perlman, P.S. (1995). Studies of point mutants define three essential paired nucleotides in the domain 5 substructure of a group II intron. *Mol. Cell. Biol.* *15*, 4479–4488.
- Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B., and Jacob, M. (1982). U2 RNA shares a structural domain with U1, U4, and U5 RNAs. *EMBO J.* *1*, 1259–1265.
- Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H., and Lührmann, R. (1984). Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* *3*, 1357–1363.
- Brow, D.A., and Guthrie, C. (1988). Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* *334*, 213–218.
- Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P., Murshudov, G., IUCr (2015). Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.* *71*, 136–153.
- Brüggeller, P., and Mayer, E. (1980). Complete vitrification in pure liquid water and dilute aqueous solutions. *Nature* *288*, 569–571.
- Burge, C.B., Tuschl, T., and Sharp, P.A. (1999). 20 Splicing of Precursors to mRNAs by the Spliceosomes. Cold Spring Harbor Monograph Archive; Volume 37 (1999): the RNA World, 2nd Ed.: the Nature of Modern RNA Suggests a Prebiotic RNA World.
- Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S., and Carragher, B. (2015). 2.8 Å resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *Elife* *4*, e06380.
- Carlson, M.A., Haddad, B.G., Weis, A.J., Blackwood, C.S., Shelton, C.D., Wuerth, M.E., Walter, J.D., and Spiegel, P.C. (2017). Ribosomal protein L7/L12 is required for GTPase translation factors EF-G, RF3, and IF2 to bind in their GTP state to 70S ribosomes. *FEBS J.* *284*, 1631–1643.
- Cazzola, M., Rossi, M., and Malcovati, L. (2013). Biologic and clinical significance of somatic mutations of SF3B1 in myeloid and lymphoid neoplasms. *Blood* *121*, 260–269.
- Cech, T.R. (1986). The generality of self-splicing RNA: Relationship to nuclear mRNA splicing. *Cell* *44*, 207–210.
- Chan, S.-P., and Cheng, S.-C. (2005). The Prp19-associated Complex Is Required for Specifying Interactions of U5 and U6 with Pre-mRNA during Spliceosome Activation. *J. Biol. Chem.* *280*, 31190–31199.

- Chen, C.H. (2002). Functional and physical interactions between components of the Prp19p-associated complex. *Nucleic Acids Res.* *30*, 1029–1037.
- Company, M., Arenas, J., and Abelson, J. (1991). Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature* *349*, 487–493.
- Cong, Y., and Ludtke, S.J. (2010). Chapter Eight - Single Particle Analysis at High Resolution. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 211–235.
- Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* *25*, 888–896.
- Cordin, O., and Beggs, J.D. (2013). RNA helicases in splicing. *RNA Biol.* *10*, 83–95.
- Cordin, O., Hahn, D., and Beggs, J.D. (2012). Structure, function and regulation of spliceosomal RNA helicases. *Curr. Opin. Cell Biol.* *24*, 431–438.
- Cornish-Bowden, A. (2014). Current IUBMB recommendations on enzyme nomenclature and kinetics. *Perspect. Sci.* *1*, 74–87.
- Costa, M., Michel, F., and Westhof, E. (2000). A three-dimensional perspective on exon binding by a group II self-splicing intron. *EMBO J* *19*, 5007–5018.
- Cretu, C., Schmitzová, J., Ponce-Salvatierra, 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.
- Crowther, R.A., and Klug, A. (1975). Structural Analysis of Macromolecular Assemblies by Image Reconstruction from Electron Micrographs. *Annu. Rev. Biochem.* *44*, 161–182.
- Crowther, R.A., and Sleytr, U.B. (1977). An analysis of the fine structure of the surface layers from two strains of Clostridia, including correction for distorted images. *J. Ultrastruct. Res.* *58*, 41–49.
- Crowther, R.A., AMOS, L.A., Finch, J.T., de Rosier, D.J., and Klug, A. (1970). Three Dimensional Reconstructions of Spherical Viruses by Fourier Synthesis from Electron Micrographs. *Nature* *226*, 421–425.
- Dahlberg, J.E., and Lund, E. (1991). How does III x II make U6? *Science* *254*, 1462–1463.
- Dainty, J.C., Shaw, R., and Cutrona, L.J. (1976). Image Science: Principles, Analysis and Evaluation of Photographic-type Imaging Processes. *Physics Today* *29*, 71–72.
- Davies, J.F., Hostomska, Z., Hostomsky, Z., Jordan, S.R., and Matthews, D.A. (1991). Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. *Science* *252*, 88–95.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* *35*, W375–W383.
- de Rosier, D.J., and Klug, A. (1968). Reconstruction of Three Dimensional Structures from Electron Micrographs. *Nature* *217*, 130–134.
- De, I., Bessonov, S., Hofele, R., Santos, dos, 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.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Lührmann, R. (2006). Protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* *26*, 5528–5543.

- DiMaio, F., and Chiu, W. (2016). Chapter Ten - Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps. In *Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, R.A. Crowther, ed. (Academic Press), 255–276.
- Dobro, M.J., Melanson, L.A., Jensen, G.J., and McDowell, A.W. (2010). Chapter Three - Plunge Freezing for Electron Cryomicroscopy. In *Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 63–82.
- Draper, D.E. (2004). A guide to ions and RNA structure. *Rna* *10*, 335–343.
- Draper, D.E., Grilley, D., and Soto, A.M. (2005). Ions and RNA folding. *Annu. Rev. Biophys. Biomol. Struct.* *34*, 221–243.
- Dubochet, J., Adrian, M., Lepault, J., and McDowell, A.W. (1985). Emerging techniques: Cryo-electron microscopy of vitrified biological specimens. *Trends Biochem. Sci.* *10*, 143–146.
- Dubochet, J., Adrian, M., Chang, J.-J., Homo, J.-C., Lepault, J., McDowell, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. *Q. Rev. Biophys.* *21*, 129–228–228.
- Dybkov, O., Stützer, A., Bertram, K., Kastner, B., Stark, H., Lührmann, R., and Urlaub, H. (2018). Protein-Cross-Linking zur Aufklärung von komplexen Strukturen. *BIOspektrum* *24*, 278–282.
- Dziembowski, A., Ventura, A.-P., Rutz, B., Caspary, F., Faux, C., Halgand, F., Laprèvote, O., and Séraphin, B. (2004). Proteomic analysis identifies a new complex required for nuclear pre-mRNA retention and splicing. *EMBO J.* *23*, 4847–4856.
- Eldridge, A., ISSNER, R., Li, Y., Reifenberg, E., Sharp, P.A., and BLENCOWE, B.J. (1999). Role of the SRm160/300 splicing coactivator in exon-enhancer function. *Biochem. Cell Biol.* *77*, 393.
- Elmlund, D., and Elmlund, H. (2015). Cryogenic Electron Microscopy and Single-Particle Analysis. *Annu. Rev. Biochem.* *84*, 499–517.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst. D* *60*, 2126–2132.
- Erickson, H.P., and Klug, A. (1970). The Fourier Transform of an Electron Micrograph: Effects of Defocussing and Aberrations, and Implications for the Use of Underfocus Contrast Enhancement. *Berichte Der Bunsengesellschaft Für Physikalische Chemie* *74*, 1129–1137.
- 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.
- Fedor, M.J., and Williamson, J.R. (2005). The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* *6*, 399–412.
- Ferré-D'Amaré, A.R., Zhou, K., and Doudna, J.A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* *395*, 567–574.
- Fica, S.M., and Nagai, K. (2017). Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine. *Nat. Struct. Mol. Biol.* *24*, 791–799.
- 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.
- 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 catalyses nuclear pre-mRNA splicing. *Nature* *503*, 229–234.

- 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.
- 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.
- Fischer, U., Sumpter, V., Sekine, M., Satoh, T., and Lührmann, R. (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J* *12*, 573–583.
- Fourmann, J.-B., Dybkov, O., Agafonov, D.E., Tauchert, M.J., Urlaub, H., Ficner, R., Fabrizio, P., and Lührmann, R. (2016). The target of the DEAH-box NTP triphosphatase Prp43 in *Saccharomyces cerevisiae* spliceosomes is the U2 snRNP-intron interaction. *Elife* *5*, e15564.
- Fourmann, J.-B., Schmitzová, J., Christian, H., Urlaub, H., Ficner, R., Boon, K.-L., Fabrizio, P., and Lührmann, R. (2013). Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system. *Genes Dev.* *27*, 413–428.
- Fourmann, J.-B., Tauchert, M.J., Ficner, R., Fabrizio, P., and Lührmann, R. (2017). Regulation of Prp43-mediated disassembly of spliceosomes by its cofactors Ntr1 and Ntr2. *Nucleic Acids Res.* *45*, 4068–4080.
- Frank, J. (1975). Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy* *1*, 159–162.
- Frank, J. (1981). chapter 16 Three-Dimensional Reconstruction of Single Molecules. In *Three-Dimensional Ultrastructure in Biology*, J.N. Turner, ed. (Academic Press), 325–344.
- Galej, W.P., Toor, N., Newman, A.J., and Nagai, K. (2018). Molecular Mechanism and Evolution of Nuclear Pre-mRNA and Group II Intron Splicing: Insights from Cryo-Electron Microscopy Structures. *Chem. Rev.* *118*, 4156–4176.
- 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.
- Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. *Nucleic Acids Res.* *36*, 2257–2267.
- Garces, D.H., Rhodes, W.T., and Peña, N.M. (2011). Projection-slice theorem: a compact notation. *J. Opt. Soc. Am. A* *28*, 766–769.
- Gilbert, W. (1978). Why genes in pieces? *Nature* *271*, 501.
- Glaeser, R.M., Han, B.-G., Csencsits, R., Killilea, A., Pulk, A., and Cate, J.H.D. (2016). Factors that Influence the Formation and Stability of Thin, Cryo-EM Specimens. *Biophys. J.* *110*, 749–755.
- Golden, B.L., Gooding, A.R., Podell, E.R., and Cech, T.R. (1998). A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* *282*, 259–264.
- Gotoh, O. (2018). Modeling one thousand intron length distributions with fitld. *Bioinformatics* *34*, 3258–3264.
- Gozani, O., Feld, R., and 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.
- Grigorieff, N. (2013). Direct detection pays off for electron cryo-microscopy. *Elife* *2*, 32.



- Grote, M., Wolf, E., Will, C.L., Lemm, I., Agafonov, D.E., Schomburg, A., Fischle, W., Urlaub, H., and Lührmann, R. (2010). Molecular architecture of the human Prp19/CDC5L complex. *Mol. Cell. Biol.* *30*, 2105–2119.
- Guthrie, C., and Patterson, B. (1988). Spliceosomal SNRNAs. *Annu. Rev. Genet.* *22*, 387–419.
- Hahn, D., Kudla, G., Tollervey, D., and Beggs, J.D. (2012). Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning. *Genes Dev.* *26*, 2408–2421.
- Harauz, G., and van Heel, M. (1986). Exact filters for general geometry three dimensional reconstruction. *Optik* *78*, 146–156.
- Haselbach, D., Komarov, I., Agafonov, D.E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2018). Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell* *172*, 454–464.e11.
- Hashimoto, C., and Steitz, J.A. (1984). U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* *12*, 3283–3293.
- Heel, M., and Stöffler-Meilicke, M. (1985). Characteristic views of *E. coli* and *B. stearothermophilus* 30S ribosomal subunits in the electron microscope. *EMBO J* *4*, 2389–2395.
- Henderson, R., Baldwin, J.M., Ceska, T.A., Zemlin, F., Beckmann, E., and Downing, K.H. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.* *213*, 899–929.
- Henderson, R. (1995). The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* *28*, 171–193.
- Henderson, R., Sali, A., Baker, M.L., Carragher, B., Devkota, B., Downing, K.H., Egelman, E.H., Feng, Z., Frank, J., Grigorieff, N., et al. (2012). Outcome of the First Electron Microscopy Validation Task Force Meeting. *Structure/Folding and Design* *20*, 205–214.
- Henning, L.M., Santos, K.F., Sticht, J., Jehle, S., Lee, C.-T., Wittwer, M., Urlaub, H., Stelzl, U., Wahl, M.C., and Freund, C. (2017). A new role for FBP21 as regulator of Brr2 helicase activity. *Nucleic Acids Res.* *45*, 7922–7937.
- Hiller, D.A., Singh, V., Zhong, M., and Strobel, S.A. (2011). A two-step chemical mechanism for ribosome-catalysed peptide bond formation. *Nature* *476*, 236–239.
- Hodnett, J.L., and Busch, H. (1968). Isolation and Characterization of Uridylic Acid-rich 7 S Ribonucleic Acid of Rat Liver Nuclei. *J. Biol. Chem.* *243*, 6334–6342.
- Horowitz, D.S. (2011). The mechanism of the second step of pre-mRNA splicing. *Wiley Interdiscip. Rev. RNA* *3*, 331–350.
- Hosokawa, F., Tomita, T., Naruse, M., Honda, T., Hartel, P., and Haider, M. (2003). A spherical aberration-corrected 200 kV TEM. *J Electron Microsc. (Tokyo)* *52*, 3–10.
- Huang, X., Beullens, M., Zhang, J., Zhou, Y., Nicolaescu, E., Lesage, B., Hu, Q., Wu, J., Bollen, M., and Shi, Y. (2009). Structure and function of the two tandem WW domains of the pre-mRNA splicing factor FBP21 (formin-binding protein 21). *J. Biol. Chem.* *284*, 25375–25387.
- Huang, Y.-H., Chung, C.-S., Kao, D.-I., Kao, T.-C., and Cheng, S.-C. (2014). Sad1 counteracts Brr2-mediated dissociation of U4/U6.U5 in tri-snRNP homeostasis. *Mol. Cell. Biol.* *34*, 210–220.
- Irimia, M., and Roy, S.W. (2014). Origin of Spliceosomal Introns and Alternative Splicing. *Cold Spring Harb. Perspect. Biol.* *6*, a016071–a016071.
- Ishizuka, K. (1994). Coma-free alignment of a high-resolution electron microscope with three-fold astigmatism. *Ultramicroscopy* *55*, 407–418.

- Jaeger, K.-E., and Eggert, T. (2004). Enantioselective biocatalysis optimized by directed evolution. *Curr. Opin. Biotechnol.* *15*, 305–313.
- Jensen, G.J. (2010). *Cryo-EM, Part C: Analyses, Interpretation, and Case Studies* (Elsevier Science).
- Jones, M.H., Frank, D.N., and Guthrie, C. (1995). Characterization and functional ordering of Slu7p and Prp17p during the second step of pre-mRNA splicing in yeast. *Proc. Natl. Acad. Sci. USA* *92*, 9687–9691.
- Jones, S., Stewart, M., Michie, A., Swindells, M.B., Orengo, C., and Thornton, J.M. (1998). Domain assignment for protein structures using a consensus approach: characterization and analysis. *Protein Sci.* *7*, 233–242.
- Joyce, G.F. (1989). RNA evolution and the origins of life. *Nature* *338*, 217–224.
- Kambach, C., Walket, S., and Nagai, K. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr. Opin. Struct. Biol.* *9*, 222–230.
- Karijovich, J., and Yu, Y.-T. (2010). Spliceosomal snRNA modifications and their function. *RNA Biol.* *7*, 192–204.
- 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. Meth.* *5*, 53–55.
- Kimoto, K. (2014). Practical aspects of monochromators developed for transmission electron microscopy. *Microscopy (Oxf)* *63*, 337–344.
- King, D.S., and Beggs, J.D. (1990). Interactions of PRP2 protein with pre-mRNA splicing complexes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* *18*, 6559–6564.
- Kobe, B., Gleichmann, T., Horne, J., Jennings, I.G., Scotney, P.D., and Teh, T. (1999). Turn up the HEAT. *Structure/Folding and Design* *7*, R91–R97.
- Konarska, M.M. (1998). Recognition of the 5' splice site by the spliceosome. *Acta Biochim. Pol.* *45*, 869–881.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* *31*, 147–157.
- Laggerbauer, B., Achsel, T., and Lührmann, R. (1998). The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplexes in vitro. *Proc. Natl. Acad. Sci. USA* *95*, 4188–4192.
- Lambowitz, A.M., and Zimmerly, S. (2011). Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb. Perspect. Biol.* *3*, a003616–a003616.
- Lardelli, R.M., Thompson, J.X., Yates, J.R., and Stevens, S.W. (2010). Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing. *Rna* *16*, 516–528.
- Lenz, F.A. (1971). Transfer of Image Information in the Electron Microscope. In *Electron Microscopy in Material Science*, U. VALDRÈ, ed. (Academic Press), 541–569.
- Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980). Are snRNPs involved in splicing? *Nature* *283*, 220–224.
- Lerner, M.R., and Steitz, J.A. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U.S.A.* *76*, 5495–5499.
- Leung, E.K.Y., Suslov, N., Tuttle, N., Sengupta, R., and Piccirilli, J.A. (2011). The mechanism of peptidyl transfer catalysis by the ribosome. *Annu. Rev. Biochem.* *80*, 527–555.

- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Meth.* *10*, 584–590.
- Liang, W.-W., and Cheng, S.-C. (2015). A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence. *Genes Dev.* *29*, 81–93.
- Liu, H.-L., and Cheng, S.-C. (2012). The Interaction of Prp2 with a Defined Region of the Intron Is Required for the First Splicing Reaction. *Mol. Cell. Biol.* *32*, 5056–5066.
- 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.
- Londei, P., Teichner, A., Cammarano, P., De Rosa, M., and Gambacorta, A. (1983). Particle weights and protein composition of the ribosomal subunits of the extremely thermoacidophilic archaeobacterium *Caldariella acidophila*. *Biochem. J.* *209*, 461–470.
- Longo, D.M., and Uhlenbeck, O.C. (1993). Self-cleaving catalytic RNA. *FASEB J.* *7*, 25–30.
- Ludtke, S.J. (2016). Chapter Seven - Single-Particle Refinement and Variability Analysis in EMAN2.1. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, R.A. Crowther, ed. (Academic Press), 159–189.
- Ludtke, S.J., Jakana, J., Song, J.-L., Chuang, D.T., and Chiu, W. (2001). A 11.5 Å single particle reconstruction of GroEL using EMAN. *J. Mol. Biol.* *314*, 253–262.
- Madhani, H.D. (2013). snRNA Catalysts in the Spliceosome's Ancient Core. *Cell* *155*, 1213–1215.
- Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* *71*, 803–817.
- Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J.* *23*, 2381–2391.
- Marabini, R., and Carazo, J.M. (1996). On a new computationally fast image invariant based on bispectral projections. *Pattern Recognit. Lett.* *17*, 959–967.
- Marcia, M., and Pyle, A.M. (2012). Visualizing Group II Intron Catalysis through the Stages of Splicing. *Cell* *151*, 497–507.
- Marquez, S.M., Chen, J.L., Evans, D., and Pace, N.R. (2006). Structure and Function of Eukaryotic Ribonuclease P RNA. *Mol. Cell* *24*, 445–456.
- Mattaj, I.W. (1986). Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* *46*, 905–911.
- Mattick, J.S. (1994). Introns: evolution and function. *Curr. Opin. Genet. Dev.* *4*, 823–831.
- McLeod, R.A., Kowal, J., Ringler, P., and Stahlberg, H. (2017). Robust image alignment for cryogenic transmission electron microscopy. *J. Struct. Biol.* *197*, 279–293.
- McMullan, G., Faruqi, A.R., and Henderson, R. (2016). Chapter One - Direct Electron Detectors. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, R.A. Crowther, ed. (Academic Press), 1–17.
- McMullan, G., Faruqi, A.R., Clare, D., and Henderson, R. (2014). Comparison of optimal performance at 300keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* *147*, 156–163.

- Moore, M.J., Query, C.C., and Sharp, P.A. (1993). 13 Splicing of Precursors to mRNA by the Spliceosome. Cold Spring Harbor Monograph Archive; Volume 24 (1993): the RNA World.
- Mozaffari-Jovin, S., Santos, K.F., Hsiao, H.H., Will, C.L., Urlaub, H., Wahl, M.C., and Lührmann, R. (2012). The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA. *Genes Dev.* *26*, 2422–2434.
- Mozaffari-Jovin, S., Wandersleben, T., Santos, K.F., Will, C.L., Lührmann, R., and Wahl, M.C. (2013). Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8. *Science* *341*, 80–84.
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D* *67*, 355–367.
- Müller, M.M., Allison, J.R., Hongdilokkul, N., Gaillon, L., Kast, P., van Gunsteren, W.F., Marlière, P., and Hilvert, D. (2013). Directed evolution of a model primordial enzyme provides insights into the development of the genetic code. *PLoS Genet.* *9*, e1003187.
- Narlikar, G.J., and Herschlag, D. (1997). Mechanistic aspects of enzymatic catalysis: lessons from comparison of RNA and protein enzymes. *Annu. Rev. Biochem.* *66*, 19–59.
- Neer, E.J., Schmidt, C.J., Nambudripad, R., and Smith, T.F. (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* *371*, 297–300.
- Nesbitt, S., Hegg, L.A., and Fedor, M.J. (1997). An unusual pH-independent and metal-ion-independent mechanism for hairpin ribozyme catalysis. *Chem. Biol.* *4*, 619–630.
- Nguyen, T.H.D., Galej, W.P., Bai, X.-C., 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.
- Nguyen, T.H.D., Galej, W.P., Bai, X.-C., 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., Li, J., Galej, W.P., Oshikane, H., Newman, A.J., and Nagai, K. (2013). Structural Basis of Brr2-Prp8 Interactions and Implications for U5 snRNP Biogenesis and the Spliceosome Active Site. *Structure/Folding and Design* *21*, 910–919.
- Nguyen, T.H.D., Tam, J., Wu, R.A., Greber, B.J., Toso, D., Nogales, E., and Collins, K. (2018). Cryo-EM structure of substrate-bound human telomerase holoenzyme. *Nature* *557*, 190–195.
- Noble, S.M., and Guthrie, C. (1996). Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics* *143*, 67–80.
- Nogales, E., and Scheres, S.H.W. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol. Cell* *58*, 677–689.
- Ohr, T., Odenwälder, P., Dannenberg, J., Prior, M., Warkocki, Z., Schmitzová, J., Karaduman, R., Gregor, I., Enderlein, J., Fabrizio, P., et al. (2013). Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system. *Rna* *19*, 902–915.
- Ohr, T., Prior, M., Dannenberg, J., Odenwälder, P., Dybkov, O., Rasche, N., Schmitzová, J., Gregor, I., Fabrizio, P., Enderlein, J., et al. (2012). Prp2-mediated protein rearrangements at the catalytic core of the spliceosome as revealed by dcFCCS. *Rna* *18*, 1244–1256.
- Oxley, M.P., Lupini, A.R., and Pennycook, S.J. (2017). Ultra-high resolution electron microscopy. *Rep. Prog. Phys.* *80*, 026101.
- Pabo, C.O., Peisach, E., and Grant, R.A. (2001). Design and selection of novel Cys2His2 zinc finger proteins. *Annu. Rev. Biochem.* *70*, 313–340.

- Padgett, R., Konarska, M., Grabowski, P., Hardy, S., and Sharp, P. (1984). Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* *225*, 898–903.
- Pena, V., Rozov, A., Fabrizio, P., Lührmann, R., and Wahl, M.C. (2008). Structure and function of an RNase H domain at the heart of the spliceosome. *EMBO J.* *27*, 2929–2940.
- Penczek, P.A. (2010a). Chapter Two - Image Restoration in Cryo-Electron Microscopy. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 35–72.
- Penczek, P.A. (2010b). Chapter Three - Resolution Measures in Molecular Electron Microscopy. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 73–100.
- Penczek, P.A. (2010c). Chapter One - Fundamentals of Three-Dimensional Reconstruction from Projections. In *Tools for Model Building and Optimization Into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps*, G.J. Jensen, ed. (Academic Press), 1–33.
- Penczek, P.A., Grassucci, R.A., and Frank, J. (1994). The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy* *53*, 251–270.
- Penzo, M., and Montanaro, L. (2018). Turning Uridines around: Role of rRNA Pseudouridylation in Ribosome Biogenesis and Ribosomal Function. *Biomolecules* *8*, 38.
- Perriman, R.J., and Ares, M. (2007). Rearrangement of competing U2 RNA helices within the spliceosome promotes multiple steps in splicing. *Genes Dev.* *21*, 811–820.
- 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.
- Ponomarenko, E.A., Poverennaya, E.V., Ilgisonis, E.V., Pyatnitskiy, M.A., Kopylov, A.T., Zgoda, V.G., Lisitsa, A.V., and Archakov, A.I. (2016). The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* *2016*, 7436849–6.
- Popp, D., Koh, F., Scipion, C.P.M., Ghoshdastider, U., Narita, A., Holmes, K.C., and Robinson, R.C. (2018). Advances in Structural Biology and the Application to Biological Filament Systems. *Bioessays* *40*, 1700213.
- Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1986). A NEW 3-D RECONSTRUCTION SCHEME APPLIED TO THE 50S RIBOSOMAL SUBUNIT OF E. COLI. *J. Microsc.* *141*, RP1–RP2.
- Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of *Escherichia coli*. *J. Microsc.* *146*, 113–136.
- Rasche, N., Dybkov, O., Schmitzová, J., Akyildiz, B., Fabrizio, P., and Lührmann, R. (2012). Cwc2 and its human homologue RBM22 promote an active conformation of the spliceosome catalytic centre. *EMBO J.* *31*, 1591–1604.
- 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.
- Reddy, R., and Busch, H. (1983). Small Nuclear RNAs and RNA Processing. In *Progress in Nucleic Acid Research and Molecular Biology*, W.E. Cohn, and K. Moldave, eds. (Academic Press), 127–162.



- Reimer, L. (2013). *Transmission Electron Microscopy: Physics of Image Formation and Microanalysis* (Berlin, Heidelberg: Springer Berlin Heidelberg).
- Renaud, J.-P., Chari, A., Ciferri, C., Liu, W.-T., Rémy, H.-W., Stark, H., and Wiesmann, C. (2018). Cryo-EM in drug discovery: achievements, limitations and prospects. *Nat. Rev. Drug. Discov.* *17*, 471–492.
- Rhode, B.M., Hartmuth, K., Urlaub, H., and Lührmann, R. (2003). Analysis of site-specific protein–RNA cross-links in isolated RNP complexes, combining affinity selection and mass spectrometry. *Rna* *9*, 1542–1551.
- Richardson, J.S. (1981). The Anatomy and Taxonomy of Protein Structure. In *Advances in Protein Chemistry*, C.B. Anfinsen, J.T. Edsall, and F.M. Richards, eds. (Academic Press), 167–339.
- Rigo, N., Sun, C., Fabrizio, P., Kastner, B., and Lührmann, R. (2015). Protein localisation by electron microscopy reveals the architecture of the yeast spliceosomal B complex. *EMBO J* *34*, 3059–3073.
- Robart, A.R., Chan, R.T., Peters, J.K., Rajashankar, K.R., and Toor, N. (2014). Crystal structure of a eukaryotic group II intron lariat. *Nature* *514*, 193–197.
- Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E.V. (2012). Origin and evolution of spliceosomal introns. *Biol. Direct.* *7*, 11.
- Root-Bernstein, M., and Root-Bernstein, R. (2015). The ribosome as a missing link in the evolution of life. *J. Theor. Biol.* *367*, 130–158.
- Root-Bernstein, R., and Root-Bernstein, M. (2016). The ribosome as a missing link in prebiotic evolution II: Ribosomes encode ribosomal proteins that bind to common regions of their own mRNAs and rRNAs. *J. Theor. Biol.* *397*, 115–127.
- Rosenthal, P.B., and Henderson, R. (2003). Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* *333*, 721–745.
- Rossmann, M.G., Morais, M.C., Leiman, P.G., and Zhang, W. (2005). Combining X-Ray Crystallography and Electron Microscopy. *Structure/Folding and Design* *13*, 355–362.
- Roy, B., Haupt, L.M., and Griffiths, L.R. (2013). Review: Alternative Splicing (AS) of Genes As An Approach for Generating Protein Complexity. *Curr. Genomics* *14*, 182–194.
- Rubinstein, J.L., Brubaker, M.A., Zhao, J., and Benlekber, S. (2015). Description and comparison of algorithms for correcting anisotropic magnification in cryo-EM images. *J. Struct. Biol.* *192*, 209–215.
- Ruskin, B., Krainer, A.R., Maniatis, T., and Green, M.R. (1984). Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* *38*, 317–331.
- Saibil, H.R. (2012). Electron Microscopy of Macromolecular Machines. In *Macromolecular Crystallography*, (Dordrecht: Springer, Dordrecht), 115–123.
- Saldanha, R., Mohr, G., Belfort, M., and Lambowitz, A.M. (1993). Group I and group II introns. *FASEB J.* *7*, 15–24.
- Santos, K.F., Jovin, S.M., Weber, G., Pena, V., Lührmann, R., and Wahl, M.C. (2012). Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc. Natl. Acad. Sci. U.S.A.* *109*, 17418–17423.
- Sashital, D.G., Cornilescu, G., and Butcher, S.E. (2004). U2–U6 RNA folding reveals a group II intron-like domain and a four-helix junction. *Nat. Struct. Mol. Biol.* *11*, 1237–1242.
- Saxton, W.O., and Baumeister, W. (1982). The correlation averaging of a regularly arranged bacterial cell envelope protein. *J. Microsc.* *127*, 127–138.

- Saxton, W.O., and Frank, J. (1976). Motif detection in quantum noise-limited electron micrographs by cross-correlation. *Ultramicroscopy* *2*, 219–227.
- Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T., and Lührmann, R. (2004). RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J.* *23*, 3000–3009.
- Schatz, M., and van Heel, M. (1990). Invariant classification of molecular views in electron micrographs. *Ultramicroscopy* *32*, 255–264.
- Schatz, M., and van Heel, M. (1992). Invariant recognition of molecular projections in vitreous ice preparations. *Ultramicroscopy* *45*, 15–22.
- Schellenberg, M.J., Wu, T., Ritchie, D.B., Fica, S., Staley, J.P., Atta, K.A., LaPointe, P., and MacMillan, A.M. (2013). A conformational switch in PRP8 mediates metal ion coordination that promotes pre-mRNA exon ligation. *Nat. Struct. Mol. Biol.* *20*, 728–734.
- Scheres, S.H.W. (2012a). A Bayesian View on Cryo-EM Structure Determination. *J. Mol. Biol.* *415*, 406–418.
- Scheres, S.H.W. (2012b). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* *180*, 519–530.
- Scherzer, O. (1936). Die schwache elektrische Einzellinse geringster sphärischer Aberration. *Z. Physik* *101*, 23–26.
- Schmidt, C., and Urlaub, H. (2017). Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Curr. Opin. Struct. Biol.* *46*, 157–168.
- Schmitt, A., Hamann, F., Neumann, P., and Ficner, R. (2018). Crystal structure of the spliceosomal DEAH-box ATPase Prp2. *Acta Crystallogr. D Struct. Biol.* *74*, 643–654.
- Schneider, S., Campodonico, E., and Schwer, B. (2004). Motifs IV and V in the DEAH Box Splicing Factor Prp22 Are Important for RNA Unwinding, and Helicase-defective Prp22 Mutants Are Suppressed by Prp8. *J. Biol. Chem.* *279*, 8617–8626.
- Schütze, T., Ulrich, A.K.C., Apelt, L., Will, C.L., Bartlick, N., Seeger, M., Weber, G., Lührmann, R., Stelzl, U., and Wahl, M.C. (2016). Multiple protein–protein interactions converging on the Prp38 protein during activation of the human spliceosome. *Rna* *22*, 265–277.
- Scott, W.G., and Klug, A. (1996). Ribozymes: structure and mechanism in RNA catalysis. *Trends Biochem. Sci.* *21*, 220–224.
- Seipelt, R.L., Zheng, B., Asuru, A., and Raymond, B.C. (1999). U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway. *Nucleic Acids Res.* *27*, 587–595.
- Ségault, V., Will, C.L., Sproat, B.S., and Lührmann, R. (1995). In vitro reconstitution of mammalian U2 and U5 snRNPs active in splicing: Sm proteins are functionally interchangeable and are essential for the formation of functional U2 and U5 snRNPs. *EMBO J.* *14*, 4010–4021.
- Sharp, P.A. (1985). On the origin of RNA splicing and introns. *Cell* *42*, 397–400.
- Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* *34*, 3955–3967.
- Siebert, X., and Navaza, J. (2009). UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions. *Acta Cryst. D* *65*, 651–658.
- Sigworth, F.J. (1998). A Maximum-Likelihood Approach to Single-Particle Image Refinement. *J. Struct. Biol.* *122*, 328–339.

- Solis, A.S., Shariat, N., and Patton, J.G. (2008). Splicing fidelity, enhancers, and disease. *Front. Biosci.* *13*, 1926–1942.
- Sontheimer, E.J., and Steitz, J.A. (1993). The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* *262*, 1989–1996.
- Staley, J.P., and Guthrie, C. (1999). An RNA Switch at the 5' Splice Site Requires ATP and the DEAD Box Protein Prp28p. *Mol. Cell* *3*, 55–64.
- Staněk, D., Rader, S.D., Klingauf, M., and Neugebauer, K.M. (2003). Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.* *160*, 505–516.
- Stark, H., Dube, P., Lührmann, R., and Kastner, B. (2001). Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* *409*, 539–542.
- Stark, H., Zemlin, F., and Boettcher, C. (1996). Electron radiation damage to protein crystals of bacteriorhodopsin at different temperatures. *Ultramicroscopy* *63*, 75–79.
- Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* *90*, 6498–6502.
- Stepankiw, N., Raghavan, M., Fogarty, E.A., Grimson, A., and Pleiss, J.A. (2015). Widespread alternative and aberrant splicing revealed by lariat sequencing. *Nucleic Acids Res.* *43*, 8488–8501.
- Sun, J.S., and Manley, J.L. (1995). A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* *9*, 843–854.
- Svoboda, P., and Di Cara, A. (2006). Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.* *63*, 901–908.
- Swanson, L., and Schwind, G. (2009). Review of ZrO/W Schottky Cathode. In *Handbook of Charged Particle Optics*, Second Edition, (CRC Press), 1–28.
- Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B., and Lyumkis, D. (2017). Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Meth.* *482*, 73–796.
- Terwilliger, T.C., Read, R.J., Adams, P.D., Brunger, A.T., Afonine, P.V., Grosse-Kunstleve, R.W., and Hung, L.-W. (2012). Improved crystallographic models through iterated local density-guided model deformation and reciprocal-space refinement. *Acta Crystallogr. D Biol. Crystallogr.* *68*, 861–870.
- THON, F. (1971). Phase Contrast Electron Microscopy. In *Electron Microscopy in Material Science*, U. VALDRÈ, ed. (Academic Press), 571–625.
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* *22*, 1616–1625.
- Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. *Science* *320*, 77–82.
- Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein Structure Fitting and Refinement Guided by Cryo-EM Density. *Structure/Folding and Design* *16*, 295–307.
- Tseng, C.-K., Liu, H.-L., and Cheng, S.-C. (2011). DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps. *Rna* *17*, 145–154.
- Tsuno, K. (2011). Monochromators in electron microscopy. The Eighth International Conference on Charged Particle Optics *645*, 12–19.

- Turunen, J.J., Niemelä, E.H., Verma, B., and Frilander, M.J. (2012). The significant other: splicing by the minor spliceosome. *Wiley Interdiscip. Rev. RNA* *4*, 61–76.
- Ulrich, A.K.C., and Wahl, M.C. (2017). Human MFAP1 is a cryptic ortholog of the *Saccharomyces cerevisiae* Spp381 splicing factor. *BMC Evol. Biol.* *17*, 364.
- Ulrich, A.K.C., Schulz, J.F., Kamprad, A., Schütze, T., and Wahl, M.C. (2016a). Structural Basis for the Functional Coupling of the Alternative Splicing Factors Smu1 and RED. *Structure* *24*, 762–773.
- Ulrich, A.K.C., Seeger, M., Schütze, T., Bartlick, N., and Wahl, M.C. (2016b). Scaffolding in the Spliceosome via Single  $\alpha$  Helices. *Structure* *24*, 1972–1983.
- Unno, M., Mizushima, T., Morimoto, Y., Tomisugi, Y., Tanaka, K., Yasuoka, N., and Tsukihara, T. (2002). The structure of the mammalian 20S proteasome at 2.75 Å resolution. *Structure/Folding and Design* *10*, 609–618.
- Unwin, P.N., and Henderson, R. (1975). Molecular structure determination by electron microscopy of unstained crystalline specimens. *J. Mol. Biol.* *94*, 425–440.
- Urlaub, H., Raker, V.A., Kostka, S., and Lührmann, R. (2001). Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* *20*, 187–196.
- Valadkhan, S., and Manley, J.L. (2001). Splicing-related catalysis by protein-free snRNAs. *Nature* *413*, 701–707.
- van Heel, M. (1987). Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* *21*, 111–123.
- van Heel, M., Schatz, M., and Orlova, E. (1992). Correlation functions revisited. *Ultramicroscopy* *46*, 307–316.
- Vonck, J., and Mills, D.J. (2017). Advances in high-resolution cryo-EM of oligomeric enzymes. *Cryo Electron Microscopy: Exciting Advances in CryoEM Herald a New Era in Structural Biology • Biophysical Methods: Behind the Scenes of the Cryo-EM Revolution* *46*, 48–54.
- Vulovic, M., Vulovi, M., Brandt, P.L., Ravelli, R.B.G., Koster, A.J., Ravelli, R., van Vliet, L.J., and Rieger, B. (2010). Estimation of defocus and astigmatism in transmission electron microscopy. (*IEEE*), 1121–1124.
- Wade, R.H. (1992). A brief look at imaging and contrast transfer. *Ultramicroscopy* *46*, 145–156.
- Wahl, M.C., and Lührmann, R. (2015). SnapShot: Spliceosome Dynamics I. *Cell* *161*, 1474–1474.e1.
- 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. (2016a). Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution. *Science* *353*, 895–904.
- Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C.C.L., and Shi, Y. (2016b). The 3.8 Å structure of the U4/U6.U5 tri-snRNP: Insights into spliceosome assembly and catalysis. *Science* *351*, 466–475.
- Wan, Y., and Wu, C.J. (2013). SF3B1 mutations in chronic lymphocytic leukemia. *Blood* *121*, 4627–4634.
- Wang, C., Chua, K., Seghezzi, W., Lees, E., Gozani, O., and Reed, R. (1998). Phosphorylation of spliceosomal protein SAP 155 coupled with splicing catalysis. *Genes Dev.* *12*, 1409–1414.
- Wang, R.Y.-R., Kudryashev, M., Li, X., Egelman, E.H., Basler, M., Cheng, Y., Baker, D., and DiMaio, F. (2015). De novo protein structure determination from near-atomic-resolution cryo-EM maps. *Nat. Meth.* *12*, 335–338.

- Wang, X., Zhang, S., Zhang, J., Huang, X., Xu, C., Wang, W., Liu, Z., Wu, J., and Shi, Y. (2010). A Large Intrinsically Disordered Region in SKIP and Its Disorder-Order Transition Induced by PPIL1 Binding Revealed by NMR. *J. Biol. Chem.* *285*, 4951–4963.
- White, H.B., III (1976). Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* *7*, 101–104.
- Wilkinson, M.E., Fica, S.M., Galej, W.P., Norman, C.M., Newman, A.J., and Nagai, K. (2017). Postcatalytic spliceosome structure reveals mechanism of 3′-splice site selection. *Science* *358*, 1283–1288.
- Will, C.L., and Lührmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell. Biol.* *13*, 290–301.
- Will, C.L., and Lührmann, R. (2006). 13 Spliceosome Structure and Function. Cold Spring Harbor Monograph Archive; Volume 43 (2006): the RNA World, 3rd Ed.
- Will, C.L., and Lührmann, R. (2011). Spliceosome Structure and Function. Cold Spring Harb. Perspect. Biol. *3*, a003707.
- Wolf, E., Kastner, B., Deckert, J., Merz, C., Stark, H., and Lührmann, R. (2009). Exon, intron and splice site locations in the spliceosomal B complex. *EMBO J.* *28*, 2283–2292.
- Woolford, D., Ericksson, G., Rothnagel, R., Muller, D., Landsberg, M.J., Pantelic, R.S., McDowall, A., Pailthorpe, B., Young, P.R., Hankamer, B., et al. (2007). SwarmPS: Rapid, semi-automated single particle selection software. *J. Struct. Biol.* *157*, 174–188.
- Wu, C.J. (2012). Understanding the Role of Mutations in SF3B1 and Splicing in Chronic Lymphocytic Leukemia. *Blood* *120*, SCI–15–SCI–15.
- Xie, J. (1998). Progression through the spliceosome cycle requires Prp38p function for U4/U6 snRNA dissociation. *EMBO J.* *17*, 2938–2946.
- Xu, D., and Nussinov, R. (1998). Favorable domain size in proteins. *Fold. Des.* *3*, 11–17.
- 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., Hang, J., Wan, R., Huang, M., Wong, C.C.L., and Shi, Y. (2015). Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* *349*, 1182–1191.
- 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.
- Yarus, M. (2011). Getting Past the RNA World: The Initial Darwinian Ancestor. Cold Spring Harb. Perspect. Biol. *3*.
- Yarus, M. (2018). Eighty routes to a ribonucleotide world; dispersion and stringency in the decisive selection. *Rna* *24*, rna.066761.118–rna.066761.1055.
- Yoshida, K., and Ogawa, S. (2014). Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* *5*, 445–459.
- Yoshihisa, T. (2014). Handling tRNA introns, archaeal way and eukaryotic way. *Front. Genet.* *5*, 213.
- Zemlin, F., Weiss, K., Schiske, P., Kunath, W., and Herrmann, K.H. (1978). Coma-free alignment of high resolution electron microscopes with the aid of optical diffractograms. *Ultramicroscopy* *3*, 49–60.
- Zhan, X., Yan, C., Zhang, X., Lei, J., and Shi, Y. (2018). Structure of a human catalytic step I spliceosome. *Science* *359*, 537–545.



- Zhang, J., Baker, M.L., Schröder, G.F., Douglas, N.R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C.J., Levitt, M., Ludtke, S.J., et al. (2010). Mechanism of folding chamber closure in a group II chaperonin. *Nature* *463*, 379–383.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *Journal of Structural Biology* *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.
- Zhang, X., Yan, C., Zhan, X., Li, L., Lei, J., and Shi, Y. (2018). Structure of the human activated spliceosome in three conformational states. *Cell Res.* *28*, 307–322.
- Zhao, C., Rajashankar, K.R., Marcia, M., and Pyle, A.M. (2015). Crystal structure of group II intron domain 1 reveals a template for RNA assembly. *Nat. Chem. Biol.* *11*, 967–972.
- 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. Meth.* *14*, 331–332.
- Zhou, Z.H. (2011). Chapter 1 - Atomic resolution cryo electron microscopy of macromolecular complexes. In *Recent Advances in Electron Cryomicroscopy, Part B*, S.J. Ludtke, Venkataram Prasad, B.V., eds. (Academic Press), 1–35.
- Zhou, Z.H., Hardt, S., Wang, B., Sherman, M.B., Jakana, J., and Chiu, W. (1996). CTF Determination of Images of Ice-Embedded Single Particles Using a Graphics Interface. *Journal of Structural Biology* *116*, 216–222.
- Zou, X., Hovmöller, S., and Oleynikov, P. (2011). Phase contrast, contrast transfer function (CTF) and high-resolution electron microscopy (HRTEM). In *Electron Crystallography Electron Microscopy and Electron Diffraction*, (Oxford University Press), 131–155.

